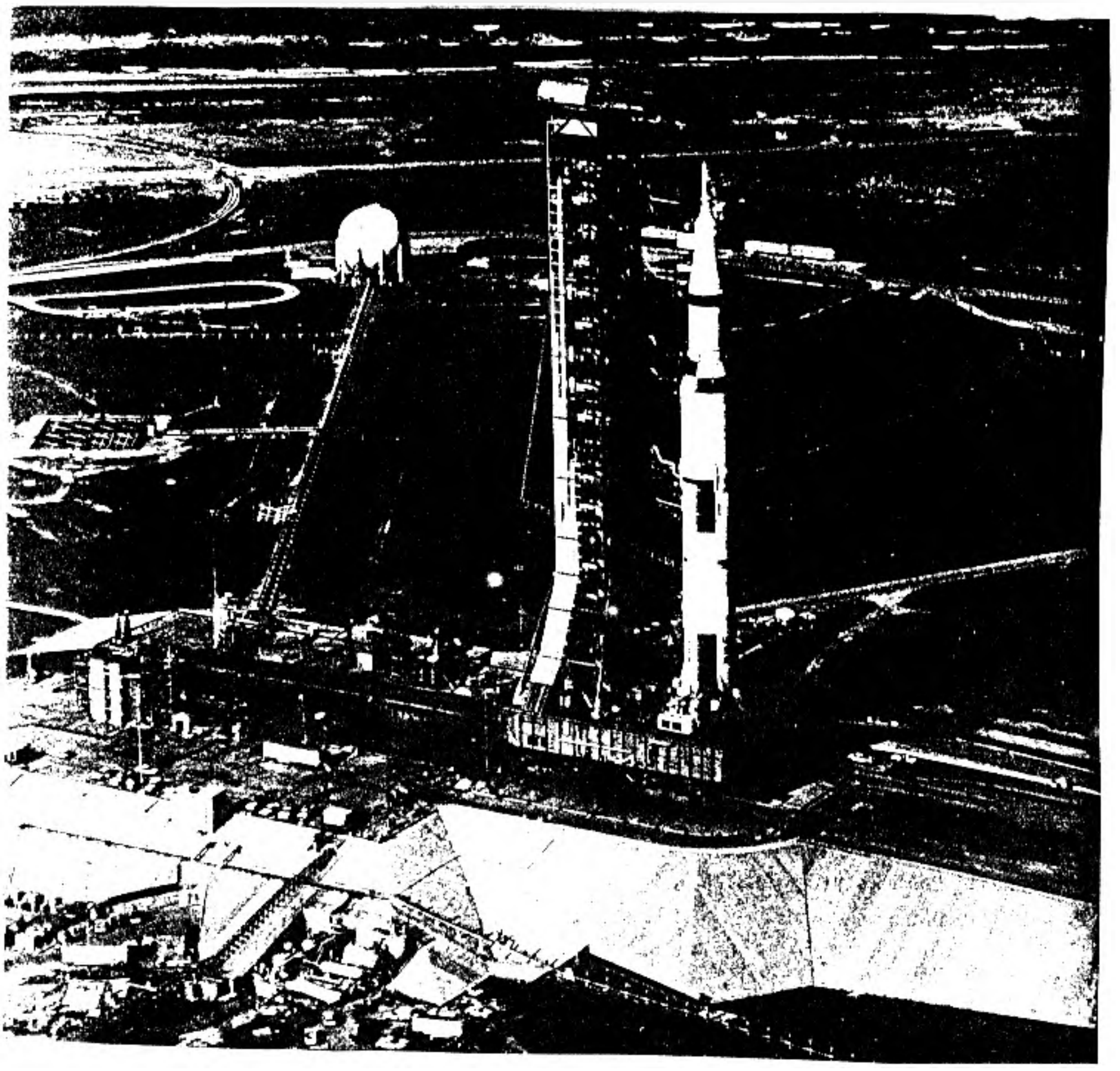
GUIDED MISSILES

Fundamentals

DEPARTMENT OF THE AIR FORCE

AFM 52-31



DEPARTMENT OF THE AIR FORCE
Headquarters US Air Force
Washington, D.C. 20330

AF MANUAL 52-31
1 July 1972

Technical Training

GUIDED MISSILES FUNDAMENTALS

This manual provides personnel in the missile electronic maintenance career field and missile operations utilization field with reference information on the fundamentals of guided missiles. It also provides a basic coverage of the subject for students in these and closely associated career and utilization fields. A prerequisite to the understanding of these fundamentals is a knowledge of electronic principles.

CONTENTS

Chapter		*Page*
1	The Story of Guided Missiles	1-1
2	Missile Aerodynamics	2-1
3	Propulsion of Guided Missiles	3-1
4	Physics Involved in Guided Missile Design	4-1
5	Control Systems Components	5-1
6	Guidance Systems Components	6-1
7	Guided Missile Control Systems	7-1
8	Trajectory Considerations	8-1
9	Guidance Systems	9-1
10	Guided Missile Operations	10-1
11	Principles of Lasers	11-1
12	Instrumentation	12-1
Attachments		
1	Department of Defense Guided Missiles	A1-1
2	Index	A2-1

Supersedes AFM 52-31, 15 January 1964. (For summary of revised, deleted, or added material, see signature page.)

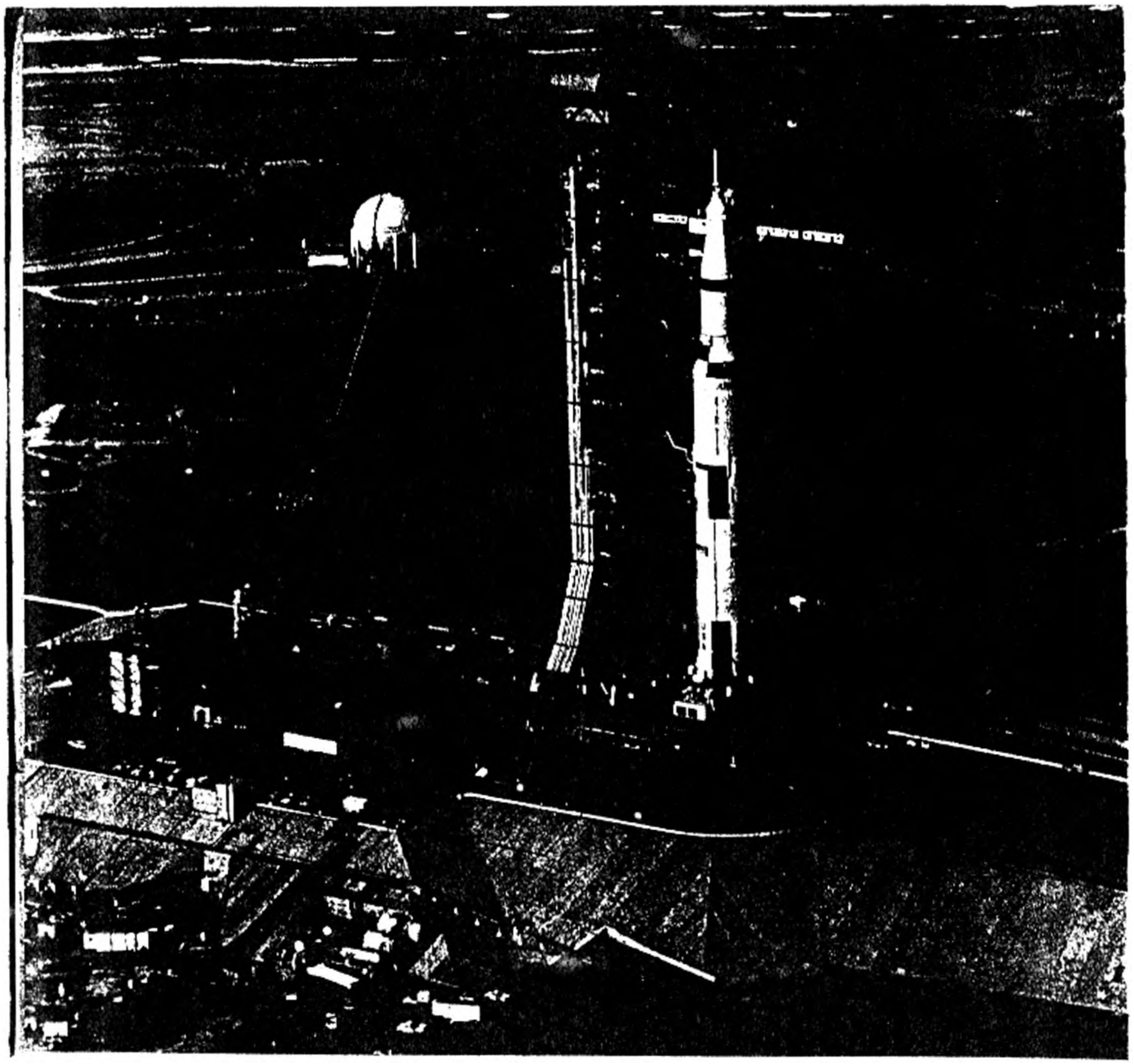
OPR: DPPTB (Prepared by Hq ATC/TTSS)

DISTRIBUTION: F

Published by Air Training Command in accordance with AFM 5-1.

☆ U.S. GOVERNMENT PRINTING OFFICE:1972 O—488-481

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $4
Stock Number 0270-0321



Missile transporter with the Apollo 11 Saturn V space vehicle, and mobile launcher, nears the top of the incline at Cape Kennedy, Florida. The 12.5 million pound load, consisting of the 12 million pound launcher and 6-million pound unfueled space vehicle, was positioned atop pedestals some seven hours after rollout began. The transporter covered the 3.5-mile distance from the Vehicle Assembly Building to the launch pad at an average speed of less than one mile per hour. The space vehicle sent Apollo 11 astronauts Neil A. Armstrong, Michael Collins, and Edwin E. Aldrin, Jr. on the nation's first manned lunar landing mission.

Apollo 11, shown on the cover, riding a pillar of flame, rises to clear its mobile launcher at Kennedy Space Center, Florida. Liftoff occurred at 9:32 am EDT, initiating man's first lunar landing mission.

CHAPTER 1

The Story of Guided Missiles

INTRODUCTION

Never before in history has mankind been faced with so much knowledge to be learned in such a short period of time as technological advances continue to be made in the field of guided missiles and space vehicles. Thirty years ago it did not seem possible, but today the skies are filled with missiles, satellites, space probes, and new types of supersonic aircraft. The world seems to get smaller and smaller as probes continue to outer space, revealing secrets of the unknown.

More than any single device, the guided missile has shaped the aerospace forces of the world today. Representative of a nation's aerospace power, these forces include the entire aeronautical and astronautical capacity of a nation. In the United States, for example, aerospace forces are composed of the following systems:

- Aircraft systems
- Missile systems
- Space vehicle systems

Continuous research and development programs have made it possible to include or integrate guided missiles in each of the systems. Combat aircraft, for example, are fitted with airborne weapons which can be launched against enemy aircraft, his ground forces, or strategic targets deep inside his nation. Ground-based missile systems have various range capabilities from a few miles to several thousand miles. These ground-based missiles are ballistic or nonballistic types, depending on their mission requirements. Short-range guided missiles are usually very mobile so that they may be transported easily and quickly to locations where they are most needed. The U. S. Army, for example, uses several missiles which can be carried and launched by the individual soldier.

Very long-range guided missiles require large fuel supplies and extremely complex guidance and control systems. These missiles are usually stored within specific and specially designed areas. Later generation ballistic missiles are designed for underground hardened-site storage to be launched as retaliatory measures in the event of attack by missiles from an unfriendly nation. Certain of these larger ballistic missiles have been integrated with the space vehicle systems. In these cases, the guided missile has been used for the booster and sustainer stages to carry vehicles into outer space. The guided missile possesses many, if not all, of the desirable characteristics which are predominant in aerospace forces. These are as follows:

- Range
- Mobility
- Flexibility
- Speed
- Penetration
- Firepower delivery

The mobility of a weapon system is of great importance. Certain guided missiles are very mobile and ideal for selective functions needed to resolve a limited war. The same holds true for special operations conducted during a cold war to counter activities harmful to this country and its allies. The use of mobile guided missiles is especially attractive since only a short notice is needed to transport them to a new location where they can be used immediately. When mobility is combined with flexibility, then the missile becomes a very versatile weapon system.

The flexibility of some guided missiles is evident by their integration with certain manned and unmanned weapon systems employed by land, sea, and air forces. An excellent example of flexibility is shown by the *Polaris* missile and its mobile underseas launching vehicle—the nuclear submarine.

The guided missile can be employed as a diversionary weapon to confuse ground and air forces. For instance, the *Quail*, which is carried and air-launched by bomber aircraft, can be equipped with electronic gear that emits electromagnetic radiations. These radiations are capable of diverting enemy attacking aircraft and their missiles, as well as those missiles that are launched from bases on the ground surface.

The surface-to-air missiles can be made an integral part of hostile aircraft and early warning networks. Each such radio and radar network, as, for example, BMEWS (Ballistic Missile Early Warning System) is a separate weapon system. The flexibility of certain guided missiles is demonstrated by their usage in more than one branch of the armed forces. Examples of these weapons are given in following chapters.

The penetrative ability of a guided missile is largely attributed to three things. First is the peculiar medium, aerospace, in which a certain type of ballistic missile system best operates. The advantage here is apparent during strategic air operations since nations have no aerial shores or definite, marked boundaries. Natural geographic features such as high mountains or large oceans are no longer considered as barriers to air operations in which manned and unmanned weapon systems are important components. This means that the entire military, political, economic, and psychosocial structure of a nation can be brought under aerial attack. More important is the influence that a powerful aerospace force exerts upon politico-military decisions in all major nations during peacetime.

Second is the wide choice of altitudes and speeds that are open to guided missiles. At the present time, there is very little that can be done to divert or destroy ballistic missiles which are capable of traveling over intercontinental distances and at hypersonic speeds. Technological developments in the future may bring about beam-directed energy weapons which can be used to intercept such missiles and their decoys. The ABM (Anti-ballistic Missile) is designed to provide limited protection in this area.

Third, the penetration ability of guided missiles is enhanced by employing a variety of tactics in manned and unmanned operations. The use of guided missiles no longer requires air superiority over enemy terrain as was necessary for conventional aircraft to drop their gravity bombs on selected targets. Aerospace forces provide the means to penetrate enemy defenses and deliver tremendous firepower at any point on earth or beyond its sensible atmosphere. As an integral part of these forces, the guided missile is capable of a high degree of accuracy, fast reaction, and high launch rates.

The intent of this manual is to present the fundamentals of the missiles field—to inform you of the aerodynamics, propulsion, instrumentation, control, and guidance systems involved in guided missiles. You will also learn of the capabilities of missiles and how they can be used.

Earlier, it was mentioned that both the aeronautical and astronautical capacity of a nation must be considered in evaluating its aerospace power. Our nation has been actively engaged in various space probes of which the Venus fly-by, made by the *Mariner II* spacecraft, is an example. The distance traveled, before scientific observations of the surface of Venus could be made, was more than 180,200,000 miles. This tremendous range or spacial distance was covered over a period of several months. The launch vehicle, which placed the *Mariner II* in orbit, was a modified *Atlas* missile combined with an *Agena B* vehicle. The *Atlas* ballistic missile was formerly an integral part of our overall defense.

Later developments in the space program saw the launching of the *Mercury* and *Gemini* programs, carrying man into outer space. These space vehicles used the *Atlas* and *Titan* missiles as boosters.

The most noteworthy achievements of 1969 were the *Apollo* manned space flights. The first flight, the *Apollo 10*, made 31 orbital sweeps of the moon. The three astronauts traveled 750,000 miles on this eight-day mission, and made two low passes over the moon's Sea of Tranquillity,

one at 9.4 miles and the second at a height of 12 miles. This flight could only have been accomplished by the very finest of technical accuracy and engineering achievement. The booster, which operated perfectly in putting the *Apollo 10* in orbit, was the *Saturn* missile.

The ultimate goal was reached in July 16, 1969 when the *Apollo 11* astronauts Neil A. Armstrong, Michael Collins, and Edwin E. Aldrin Jr. were launched to the moon by a *Saturn V* missile from Cape Kennedy, Florida. Armstrong and Aldrin made world history by landing on the moon July 20, 1969. The "moon walk" was televised back to earth and viewed by all countries.

After take-off from the moon July 21, the two astronauts rejoined Collins in the Command Module circling the moon. They splashed down in the Pacific Ocean and recovery was made on July 24, 1969. The world was astounded at the perfection of this entire operation, which lead to other planned space explorations.

BRIEF HISTORY OF GUIDED MISSILES

Just as World War II was not fought along the same lines as World War I, victory in a future war may not be realized by using World War II aircraft practices. The solution to guided missile problems of the future will stem from past experiences. If we are to profit from the lessons learned from the missiles of the past, we must analyze and apply these lessons in the light of the current situation.

The idea of using guided missiles came during World War I. The use of the airplane as a military weapon gave rise to the idea of using remote-controlled aircraft to bomb targets. The leaders in the field were Orville Wright, who flew the first airplane, E. A. Sperry of the Sperry Gyroscope Company, and Charles F. Kettering of General Motors Corporation. These men devised and tested our first missile, the *Bug*, a small version of the aircraft used in those days. While the first missile did not get into combat, a most important result of these early tests was the recommendation that future experimentation should be done with radio-controlled aircraft to correct the missile while in flight.

In 1924, funds were allocated to develop a missile using radio control. Many moderately successful flights were made during the 1920s with this control but by 1932 the project was closed because of lack of funds.

About 1935, two brothers named Good, amateur model airplane builders, built and flew a model plane that was remotely controlled by radio signals transmitted from the ground. These flights were the first completely radio-controlled flights on record.

Radio-controlled target planes were the first airborne remote-controlled aircraft used by the Army and Navy.

By December of 1941, just before entry into World War II, remote-controlled aircraft were developed to the point where they were seriously considered for use as a weapon of warfare by General Arnold, then Chief of Staff of the Army Air Corps.

So far only missiles powered by internal combustion engines and propellers have been discussed. Work was also being done to develop missiles using reaction-type engines (including rocket engines), which contain within themselves all the elements needed for power, and jet engines, which depend on the surrounding atmosphere as a source of oxygen. Both the jet and rocket engines may become obsolete insofar as missile powerplants are concerned when a nuclear powerplant for aircraft is developed.

Development of American Rocketry

Dr. Robert H. Goddard, at one time a physics professor at Clark University, Worcester, Massachusetts, was largely responsible for the interest in rockets back in the twenties. When Dr. Goddard first started his experiments, no related technical information was available. He started a new science, industry, and field of engineering. Through his scientific experiments, he pointed the way to the development of rockets as they are known today. The Smithsonian Institute agreed to finance his experiments in 1920. From these experiments he wrote a paper titled *A Method of Reaching Extreme Altitudes*, in which he outlined a space rocket of the step (multistage) principle, theoretically capable of reaching the moon.

Dr. Goddard discovered that with a properly shaped, smooth, tapered nozzle he could increase the ejection velocity 8 times with the same weight of fuel. This would not only drive a rocket 8 times

faster, but 64 times farther, according to his theory. Early in his experiments he found that solid propellant rockets would not give him the high power or the duration of power needed for a dependable supersonic motor capable of extreme altitudes. On 16 March 1926, after many trials, Dr. Goddard successfully fired, for the first time in history, a liquid propellant rocket into the air. It attained an altitude of 184 feet and a speed of 60 mph. This seems small as compared to present day heights and speeds of missile flights, but instead of trying to achieve speed or altitude at that time, Dr. Goddard was trying to develop a dependable rocket motor.

Dr. Goddard later was the first to fire a rocket that reached a speed faster than the speed of sound. He was first to develop a gyroscopic steering apparatus for rockets, first to use vanes in the jet stream for rocket stabilization during the initial phase of a rocket flight, and first to patent the idea of step rockets. After proving on paper and in actual test that a rocket can travel in a vacuum, he developed the mathematical theory of rocket propulsion and rocket flight, including basic designs for long-range rockets. All of this information was available before World War II, but evidently its immediate use did not seem applicable. Near the end of World War II, intense work started on rocket-powered guided missiles, using the experiments and developments of Dr. Goddard and the American Rocket Society.

The American Rocket Society developed rockets and rocket motors after its organization in 1930. Its first motor was based mainly on German designs obtained from the German Rocket Society in 1931. The American Rocket Society was first to build a sectional rocket motor that could test rockets of different sizes, thus cutting down the cost of a new motor for each type tested.

In 1941 some members of the American Rocket Society formed a company known as Reaction Motors, Inc. It was organized to develop and manufacture rocket motors for both military and civilian use.

Development of German Rockets

The first flight of a liquid propellant rocket in Europe occurred in Germany on 14 March 1931, five years after Dr. Goddard made his first successful rocket test. A German scientist named Winkler was in charge. Winkler lost his life a short time later during one of his experiments.

Germany by this time had begun to sense the future importance of liquid propellant rockets in warfare. In 1932 Captain Walter Dornberger (later a general) of the German army obtained the necessary approval to develop liquid propellant rockets for war purposes. By 1936, Germany decided to make research and development of guided missiles a major project. Germany spent $40,000,000 on a project, known as the Peenemunde Project, for establishing a large rocket research and development laboratory. Hitler put the members of the German Rocket Society to work there, closing to the rest of the world German developments on rockets until after the war. Unlike Germany, the U.S.A. during this time paid little attention to the development of jet and rocket propulsion for any specific purpose.

Evolution of Jet Engines

The rocket was just one type of jet propulsion powerplant that was being proposed and worked on in this century. As early as 1913, Rene Lorin, a French engineer, proposed and first patented the idea for a ramjet powerplant. Lorin's patent was followed by a Hungarian patent for a similar device in 1928 and another French patent in 1933. None of the proposed ideas resulted in a workable engine. The failures occurred not because the fundamentals of operating such a device were not known but because technical information on high-speed fluid flow was unavailable. A period of 32 years separated Lorin's original idea and the first free-flight testing of a ramjet-powered vehicle capable of developing thrust in excess of drag. This test occurred in June 1945 when the Applied Physics Laboratory of Johns Hopkins University successfully flew the first ramjet-powered aircraft.

The forerunner of the present day turbojet was a mechnical type. In 1927, the Italian Air Ministry began investigating the possibility of propelling an aircraft by placing a conventional propeller inside the mouth of a venturi-shaped fuselage. This so-called "ducted propeller" installation was a form of mechanical jet propulsion. Tests with this "ducted propeller" installation demonstrated that it possessed excellent maneuverability and

stability characteristics, although its overall performance was only mediocre. In 1932, an Italian by the name of Campini designed and later flew an aircraft propelled by a thermaljet engine. His jet-powered aircraft was not a turbojet because it depended upon a conventional reciprocating engine instead of gas turbine for compressor power.

Evaluation of Campini's engine had hardly been made when improved jet engines began to appear in various countries. A young British engineer and Royal Air Force officer, Frank Whittle, had filed a patent for a thermaljet engine as early as 1930. Whittle's design eliminated the reciprocating engine as the power source for driving the compressor. Instead, the mixture of air and gasses was used after combustion to drive a gas turbine. The turbine drove the compressor. On 7 April 1941, a Gloster "Pioneer" aircraft, powered by Whittle's engine, became airborne during taxiing tests and flew about 150 yards at an altitude of about six feet. On 15 May 1941, this same aircraft made the first official takeoff for a turbojet-powered aircraft and flew for 17 minutes.

After these successful flights, the Army Air Corps sent a special group of men to England to study the engine. Further development became a British-American project. At this time, only 10 hours of jet engine operation had been accumulated in 15 flights.

Turbojet development in the U.S. was turned over to the General Electric Company because of its experience in the development and production of turbosuperchargers. Today, turbojet engines are built and developed by nearly all aircraft engine companies. At the present time, the turbojet engine is the only air-jet engine capable of getting an aircraft or missile to take off under its own power without the assistance of a booster.

Another air-jet engine is the pulsejet. This type of engine was patented by a German engineer in 1930, but a good, workable pulsejet engine was not perfected until World War II. It became famous during the war as the powerplant of the German *Buzz Bomb* or V-1. This engine was capable of propelling the V-1 at 450 miles per hour.

In 1940, General Motors Corporation was given a contract to build and develop jet-powered, controllable bombs which in the final version were to be command-controlled with television. These bombs were tested extensively in 1941 and 1942. The testing led to the development of new types of jet-powered bombs.

Enemy Guided Missiles of World War II

The Japanese were far behind the Germans in missile development during World War II. The *Baka*, used by the Japanese during the war, was not a guided missile in the true sense of the word. It was a rocket-propelled, piloted glide bomb designed for use against shipping targets. The *Baka* was known as a suicide bomb. Although it achieved a certain degree of success, it had poor maneuverability, a characteristic which resulted in many *Bakas* being shot down by antiaircraft fire.

The Japanese also tried an air-launched, radio-controlled, rocket-assisted glide bomb. This missile had to be dropped from a low altitude, and the control plane had to get within two and one-half miles of the target. The procedure made the control plane an easy target for antiaircraft fire. The project was dropped before the end of the war.

The German developments in the field of guided missiles during World War II were the most advanced. Their most widely known missiles were the V-1 and V-2 surface-to-surface missiles. As early as the spring of 1942, the original V-1 had been developed and flight tested at Peenemunde. In 1943, Germany was working on 48 different antiaircraft missiles. These were later consolidated into 12 projects for immediate development into useful weapons. Toward the end of the war, efforts were being directed toward the successful production of an antiaircraft missile capable of intercepting allied bombers.

The V-1 (a robot bomb) was a pilotless, pulsejet, midwing monoplane, lacking ailerons but using conventional airframe and tail construction. All guidance and control was accomplished internally by gyro stabilization and preset compass guidance. It was launched from a ramp 150 feet long and 16 feet above the ground at the highest end. A speed of approximately 200 mph had to be reached before the V-1 propulsion unit could maintain the missile in flight. The missile carried a warhead weighing 1988 pounds.

The V-1 (vengeance weapon 1) was not accurate, and it was susceptible to destruction by

antiaircraft fire and aircraft. However, the interruptions it caused in the functioning of a vital war center such as London, together with the amount of physical damage it did, made the V-1 effective in lowering morale.

The V-2 (vengeance weapon 2) was the first long-range, rocket-propelled missile to be put into combat. Concentrated efforts began in 1941. The V-2 was put into mass production, and the first V-2 landed in England in September 1944.

The V-2 was a supersonic missile, vertically launched and automatically tilted to a 41° to 47° angle a short time after launching. The maximum range was about 200 miles, and the top speed was about 3300 mph. The V-2 was a large missile, having a length of 46 feet 11 inches and a diameter of 4 feet 5 inches. Its total weight at takeoff was over 14 tons, including a 1650-pound warhead.

Active countermeasures against the V-2 were impossible. Except for its initial programmed turn, it operated as a free projectile at extremely high velocity.

Several other German missiles were also highly developed during World War II and were in various stages of test.

One of these, the *Rheinbote*, was also a surface-to-surface missile. This rocket was a 3-stage device with booster-assisted takeoff. Its range was about 135 miles, with the third stage reaching over 3200 mph in about 25 seconds after launching. Overall length of the rocket was about 37 feet; but after having dropped a rearward section at the end of both the first and second stages, it had a length of only 13 feet. The 13-foot section of the third stage carried an 88-pound high-explosive warhead.

A surface-to-air missile, the *Wasserfall*, was a radio-controlled supersonic rocket, similar to the V-2 in general principles of operation. Fully loaded it had a weight of slightly less than 4 tons. Its length was 25 feet. Designed for intercepting aircraft, this missile had specifications which called for maximum altitude of 65,000 feet, speed of 560 mph, and range of 30 miles. Its 200-pound warhead could be detonated by radio after the missile had been command-controlled to its target by radio signals. It also was to use an infrared proximity fuze and homing device for control on final approach to the target and for detonating the warhead at the most advantageous point in the approach. Propulsion was to be obtained from a liquid propellant powerplant, with nitrogen-pressurized tanks.

Another surface-to-air missile, the *Schmetterling* (HS-117), was still in the development stage at the close of the war. All metal in construction, it was 13 feet long and had a wing span of 6½ feet. Effective range against low-altitude targets was 10 miles. It traveled at 540 mph at altitudes up to 35,000 feet. A proximity fuze would set off its 55-pound warhead. Propulsion was obtained from a liquid propellant rocket motor with additional help from two booster rockets during takeoff. Launching was to be accomplished from a platform which could be inclined and rotated toward the target.

A third German surface-to-air missile was the *Enzian*, designed to carry payloads of explosives up to 1000 pounds. It was to be used against heavy-bomber formations. The *Enzian* was about 12 feet long. It had a wing span of approximately 14 feet and weighed a little over 2 tons fully loaded. Propelled by a liquid propellant rocket, it was assisted during takeoff by four solid propellant rocket boosters. Launching, as in the case of the *Schmetterling*, was from a rotatable launcher, with range elevation possible. Its range was about 16 miles, speed 560 mph, and maximum altitude 48,000 feet. Guidance was by command control.

A German air-to-air missile, the X-4, was designed to be launched from fighter aircraft. Propelled by a liquid propellant rocket, it was stabilized by four fins placed symmetrically. Length was about 6½ feet and span about 2½ feet. Its range was slightly over 1½ miles, and its speed was 560 mph at an altitude of 21,000 feet. Guidance was accomplished by electrical impulses transmitted through a pair of fine wires from the fighter aircraft. The wires unrolled from two coils mounted on the tips of two opposite fins of the missile. This missile was claimed to have been flown, but it was never used in combat.

United States Guided Missiles of World War II

A project for developing missiles in the U.S.A. during World War II was started in 1941. In that year the Army Air Corps asked the National Defense Research Committee to undertake a project for the development of a vertical, controllable

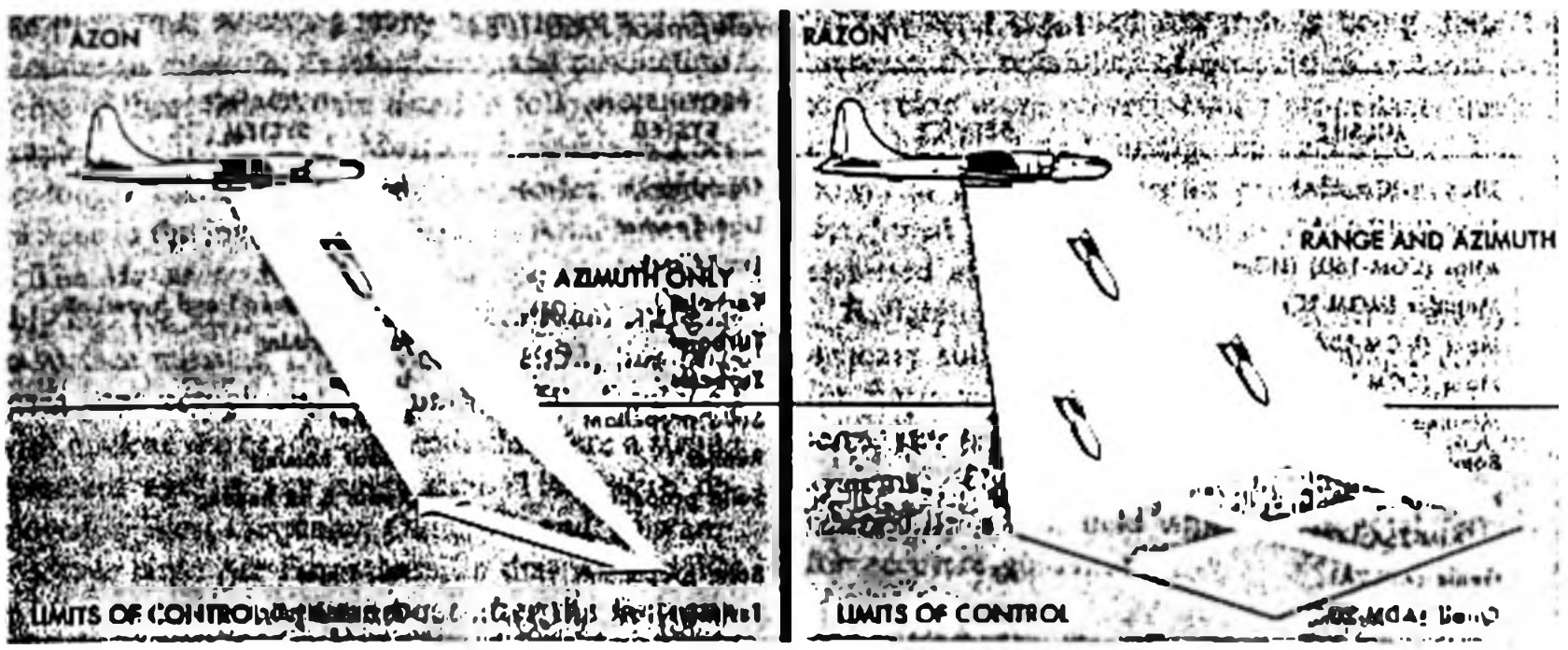


Figure 1-1. Limits of Control of the Azon and Razon

bomb. The committee initiated a glide-bomb program which resulted in standardization of a preset glide bomb attached to a 2000-pound demolition bomb. The *Azon*, a vertical bomb controlled in azimuth only, went on the production line in 1943. *Project Razon*, a bomb controlled in both azimuth and range, was started in 1942 but not completed until the end of the war. The limits of control of both *Azon* and *Razon* bombs are shown in figure 1-1. A medium-angle glide bomb called the *ROC* and a 12,000-pound bomb known as the *Tarzon*, both controllable in azimuth and range, were also under development at this time. The two bombs did not reach the combat stage during World War II. The *Tarzon* project was dropped in 1946 and picked up again in 1948. The *Tarzon* and *Razon* bombs were used successfully in the Korean action.

In 1943, a project was initiated for development of a glide torpedo using standard Navy torpedoes. In the final days of the war, these glide torpedoes were used on several missions in the Pacific theater.

In 1944, a glide-bomb mission was made against Cologne, Germany. Most of the bombs reached the target area. In this same year, remote television-control equipment was developed and installed in bombing aircraft. These aircraft were used to control television-sighted, explosive-laden bombers unfit for further service. These radio-controlled bombers saw some service over Germany under the *Weary Willie* project. Light and radar target-seeking devices were developed for use with glide bombs and were tested until 1945.

Our first jet-propelled missile was actually a flying-wing, jet-powered, and radio-controlled bomb. The second version of a jet-propelled bomb was a copy of the German V-1 with a few improvements. Another model consisted of a combination of the two mentioned above, using the flying wing together with the pulsejet engine of the V-1. This project became obsolete in 1946. By 1946 the *Tiamat*, a guided aircraft rocket, had been completed to the extent that full-size versions were being tested.

The Navy had a number of guided missile projects under development by the end of the war. One of these, the *Gargoyle*, was an air-launched, powered, radio-controlled glide bomb using a flare for visual tracking. The Navy also had a glide bomb called the *Glomb*. It was guided to a target by radio control, monitored by television. The *Loon*, a modification of the German V-1, was to be used from ship-to-shore and to test guided missile components. Another missile, known as *Gorgon IIC*, used a ramjet engine with radar tracking and radio control.

During the war, these weapons were developed under pressure for immediate use. At the end of the war in 1945, nearly all previous development on guided missiles, controllable bombs, and guided aircraft rockets was considered obsolete. New military characteristics and specifications were drawn up with future weapon possibilities in mind.

Table 1-1. Missile Development 1950-1964

MISSILE	SERVICE	PROPULSION SYSTEM	GUIDANCE SYSTEM
Titan I (HGM-25A)	AF	Liquid rocket	Radio-inertial
Titan II (LGM-25C)	AF	Liquid rocket	Inertial
Atlas (CGM-16D) (HGM-16F)	AF	Liquid rocket	Radio-inertial
Matador (MGM-1C)	AF	Turbojet	Radar command and hyperbolic
Mace (MGM-13A)	AF	Turbojet	Map-matching
Mace (CGM-13B)	AF	Turbojet	Inertial
Minuteman (LGM-30A, B, F)	AF	Solid propellant	Inertial
Bomarc (CIM-10A & 10B)	AF	Ramjet	Radar homing
Falcon (AIM-4A, C, E, F) (AIM-26A, 47A)	AF	Solid propellant	Radar & IR homing
Genie (AIR-2A)	AF	Solid propellant	Free-flight
Quail (ADM-20C)	AF	Turbojet	Gyro-autopilot
Hound Dog (AGM-28)	AF	Turbojet	Inertial
Davy Crockett	Army	Solid propellant	Free-flight
ENTAC (MGM-32A)	Army	Solid propellant	Wire-guided
Honest John (MGR-1)	Army	Solid propellant	Free-flight
Little John (MGR-3A)	Army	Solid propellant	Free-flight
Pershing (MGM-31A)	Army	Solid propellant	Inertial
Hawk (MIM-23A)	Army	Solid propellant	Radar-homing
Sergeant (MGM-29A)	Army	Solid propellant	Inertial
Shillelagh (MGM-51A)	Army	Solid propellant	Command
Nike-Hercules (MIM-14B)	Army	Solid propellant	Command-tracking radar
Polaris (UGM-27)	Navy	Solid propellant	Inertial
Regulus (RGM-6)	Navy	Turbojet	Inertial
Subroc (UUM-44A)	Navy	Solid rocket	Inertial
Talos (RIM-8E)	Navy	Ramjet	Beam-rider — Homing
Tartar (RIM-24B)	Navy	Solid propellant	Beam-rider
Terrier (RIM-2E)	Navy	Solid propellant	Beam-rider — Homing
Shrike (AGM-45A)	Navy	Solid propellant	Radar homing
Sidewinder 1-C (AIM-9D)	Navy, AF	Solid propellant	IR homing
Sparrow III-6B (AIM-7E)	Navy, AF	Solid propellant	Homing
Bullpup (AGM-12B)	Navy, AF	Solid propellant	Radio command
Bullpup (AGM-12C)	Navy, AF	Liquid propellant	Radio command

GUIDED MISSILE DEVELOPMENT AND CLASSIFICATION

A high respect for the far-reaching potentials of guided missiles was realized in the closing phases of World War II. Immediately following the war's end, the Army Air Corps began a crash program of missile development. The program was later drastically curtailed because of peacetime budgeting and the need for conservation of the limited numbers of technically qualified personnel and limited supplies of material. Contracts were cancelled for interim-type guided missiles being developed until such time that longer range, supersonic missiles could be realized.

Research and Development

Attention was again focused on guided missile programs in 1950 when this nation became involved in the Korean war, and possible involvement with other countries. With a higher budget for defense, all three services began extensive mis-

sile programs, involving many companies and universities in research, development, and production. Some of these missiles are listed in following paragraphs and in table 1-1. Most of these missiles are no longer carried in current inventories. Current missiles in the inventory are listed in Attachment 1.

The *Matador* and its successor, the *Mace*, probably had the longest tenure as operational surface-to-surface missiles. Operational in 1952, this program was not phased out until 1969. Equipped with nuclear warheads, these missiles were a strong deterrent against any aggression. They were deployed in West Germany, Okinawa, and Taiwan.

Other large Air Force missiles that were highly publicized but were phased out after the first production were the SM-62 *Snark*, the GAM-63 *Rascal*, and the GAM-87 *Skybolt*.

The *Snark* was a long-range strategic weapon with turbojet propulsion. The *Rascal* was designed as an air-to-surface missile for strategic use within a 90-mile range. The *Rascal* was too dangerous a load to carry on bombers because of the liquid fuel system needed for its rocket engines. The *Skybolt* held the most promise during this period to fill the need for an IRBM (intermediate range ballistic missile). The *Skybolt* featured a stellar-supervised (star-tracker) inertial guidance system that was highly accurate over a long range.

All of the missile programs that have come and gone have served as a basis for the constantly improving research and development programs for the current missiles.

The research program is a continuing process, not only for the production missiles, but also for the many individual system components. The program of component research is based on realizing major aims and overcoming problems which are inherent in the development of a missile system. Research has resulted in the development of dependable solid rocket motors that provide reliable high altitude, supersonic operation. This development has made it possible to store a completely fueled missile in a silo for extended time periods with minimum reaction time. It is not necessary to store the fuel separately and load it prior to launching the missiles. The complex fuel storage and transfer systems associated with liquid propellants are not required. Gas turbine and other jet engines are rapidly becoming obsolete for missile propulsion because of altitude and speed limitations. Continued research is being done to increase the reliability of propulsion systems and to develop more powerful solid propellants having safer handling and storage characteristics. Large powerful motors with solid propellants are being developed that can be reliably shut down and restarted at will.

Reliability of the guidance systems is always a primary subject for research. The major effort is for improvement of components of inertial systems, star trackers, and radar and infrared homing systems. The introduction of lasers has now opened up a new field which holds great promise for accurate guidance.

Guided Missile Classification

In much the same manner as aircraft, missiles are typed by their general characteristics grouping. Such a grouping may show in what manner a missile is used; but it will not identify a particular missile. This general classification makes use of three items: launch environment, target environment (or mission), and type of vehicle.

Launch environment may be air, ground, underground, or underwater. Thus the letters are "A" for air, "G" for ground, "L" for underground or silo launched, and "U" for underwater. The second letter is used to designate the target environment or mission. This letter may be "I" for interceptor, "G" for surface target, or "Q" for drone. The third letter designates the type vehicle as "M" for missile, or "R" for rocket. Some examples of this general classification are as follows:

AIR—air-interceptor rocket
AGM—air-to-surface missile
LGM—silo launched surface-to-surface missile
UGM—underwater-to-surface missile

The tables 1-2 through 1-5 give more complete designations. The uniform system of designations may also be found in AF Pamphlet 82-1, Army Pamphlet 700-6 and NAVAIR Instruction 13100, for the Air Force, Army, and Navy respectively. In addition to the general designator for missile identification, additional items of information may be included as follows:

1. Status prefix
2. Launch environment
3. Primary mission

Table 1-2. Status Prefix Symbols

LETTER	TITLE	DESCRIPTION
J	Special Test (Temporary)	This letter is reserved for vehicles on special test programs by authorized organizations. Some vehicles, particularly those on bailment contract, may have a special configuration to accommodate this test. Upon completing the test, these vehicles will either be returned to their original configuration or be returned to standard operational configuration.
N	Special Test (Permanent)	This letter is used for vehicles that undergo special testing programs by authorized organizations. As above, these include vehicles on bailment contract. Vehicle configurations are so drastically changed that they cannot practically or economically be returned to the original or standard operational configurations.
X	Experimental	This letter is limited to a vehicle in its developmental or experimental stage where the basic type and series have been designated. This vehicle has not yet been established as a standard vehicle for service use.
Y	Prototype	This letter is seen on vehicles which have been procured in limited quantities in order to develop the potentialities of the model.
Z	Planning	This letter is used to identify a vehicle during its planning and predevelopment stage.

4. Vehicle type
5. Vehicle design number
6. Vehicle series
7. Manufacturer's code
8. Serial number

Items 1 through 5 may be stenciled on the missile itself. Items 6 and 7 are important particularly to missile maintenance personnel. The letters "X" and "Y," which are status prefix symbols, are used to identify experimental and prototype vehicles, respectively. Other prefix symbols and when they are used are shown in table 1-2.

Missile designators, when the occasion warrants, will have a status prefix symbol but not necessarily a launch environment symbol. A typical designator is shown below for an early *Minuteman* missile. Note that it contains eight items of essential information.

J—Status prefix symbol
L—Launch environment symbol
G—Mission symbol
M—Vehicle type symbol
30—Design number
B—Series symbol
BO—Manufacturer's code
00003—Serial number

Table 1-3 shows the various methods of protecting, storing, and launching a military rocket or guided missile. Rocket systems employed for line-of-sight fire against ground targets are not included. Some typical examples of missile designators are given in this table.

Note that several missiles are designed for similar tasks; only the method of launching differs. This similarity is noted by the second symbol with the missile designator. These tasks or missions are given in table 1-4 along with their characteristic identifying letter and description.

In the last example given in table 1-4, a missile bearing this designator would be a weather probe.

Table 1-3. Launch Environment Symbols

1st LETTER	TITLE	DESCRIPTION
A	Air	Launched from aircraft while in flight. Example: AGM-45A (Shrike)
B	Multiple	Capable of being launched from more than one environment. Example: BQM-34A (Firebee)
C	Coffin	Horizontally stored in a protective enclosure and launched from the ground. Example: CGM-13B (Mace)
F	Individual	Carried by one man Example: XPIM (Redeye)
H	Silo Stored	Vertically stored below ground level and launched from the ground. Example: HGM-25A (Titan)
L	Silo Launched	Vertically stored and launched from below ground level. Example: LGM-30G (Minuteman III)
M	Mobile	Launched from a ground vehicle or movable platform. Example: MIM-23A K (Hawk)
P	Soft Pad	Partially or nonprotected in storage and launched from the ground. Example: PGM-17A (Thor)
R	Ship	Launched from a surface vessel such as a ship, barge, etc. Example: RIM-66A (Sea Mauler)
U	Underwater	Launch from a submarine or other underwater device. Example: UGM-27C (Polaris)

Table 1-4. Mission Symbols

2nd LETTER	TITLE	DESCRIPTION
D	Decoy	Vehicles designed or modified to confuse, deceive, or divert enemy defenses by simulating an attack vehicle. Example: ADM-20A *(Quail)*
E	Special Electronic	Vehicles designed or modified with electronic equipment for communications, countermeasures, electronic radiation sounding, or other electronic recording or relay missions. Example: XPEM-43B *(Redeye)*
G	Surface Attack	Vehicles designed to destroy enemy land or sea targets. (See examples in Table 1-3.)
I	Intercept-Aerial	Vehicles designed to intercept aerial targets in defensive or offensive roles. Example: AIM-9E *(Sidewinder)*
Q	Drone	Vehicles designed for target, reconnaissance, or surveillance purposes. Example: BQM-34A *(Firebee)*
T	Training	Vehicles designed or permanently modified for training purposes. Example: ATM-12B *(Bullpup)*
U	Underwater Attack	Vehicles designed to destroy enemy submarines or other underwater targets. Example: UUM-44A *(SUBROC)*
W	Weather	Vehicles designed to observe, record, or relay data pertaining to meteorological phenomena. Example: PWN-3A

Table 1-5. Vehicle Type Symbols

3rd LETTER	TITLE	DESCRIPTION
M	Guided Missile	As the third letter in a missile designator, it identifies an unmanned, self-propelled vehicle. Such a vehicle is designed to move in a trajectory which may be entirely or partially above the earth's surface. While in motion this vehicle can be controlled remotely or by homing systems, or by inertial and/or programmed guidance from within. The term "guided missile" does not include space vehicles, space boosters, or naval torpedoes, but it does include target and reconnaissance drones. (See table 1-4.)
N	Probe	The letter "N" is used to indicate nonorbital instrumented vehicles which are not involved in space missions. These vehicles are used to penetrate the space environment and transmit or report back information.
R	Rocket	This identifies a self-propelled vehicle without installed or remote control guidance mechanisms. Once launched, the trajectory or flight path of such a vehicle cannot be changed. Example: AIR-2B (Super Genie)

Such a missile is not intended to go into orbit about the earth as, for example, the weather satellite *Tiros*. Table 1-5 shows the types of vehicles which have a combat-related mission.

The last two items of a missile designator are the design number and series symbol. Each vehicle type of the same basic design is identified by the same design number. Where more than one design is present for a single vehicle type, consecutive design numbers are assigned. When major modifications are present in a vehicle type, then each modification is indicated by a sequential letter; e.g., A, B, etc.

There is another way in which aerospace and other vehicles are identified. This is by the weapon system designator which consists of the letters "WS" followed by the appropriate numbers to pinpoint the desired weapon.

Actually the weapon system designator, as it is used in the Air Force, is intended to identify all related facilities, equipment, material, services, and personnel required solely for the operation of a particular aerospace vehicle and its major elements. For example, the *Mace* is the main aerospace vehicle of Weapon System—309A. The major elements of this weapon system are the missiles, the launch facility, aerospace ground equipment, facilities, and personnel qualified in the loading, handling, testing, maintenance, and launching of the missile. As an instrument of combat, a weapon system becomes a self-sufficient unit of striking power in its intended operational environment. Here are some typical weapon systems, with the major aerospace vehicles by their popular names.

WS-107A-2	*Titan*
WS-131B	*Hound Dog*
WS-133A	*Minuteman*
WS-200A	*Bomarc*
WS-208A	*Falcon*

WS-221A	*Sidewinder*
WS-309A	*Mace*
WS-315A-1	*Thor*
WS-321A	*Bullpup*
WS-429L	*Firebee*

The last entry, *Firebee*, is a drone vehicle which is specially designed and developed for target missions. Other drones can be made of conventional aircraft which are fitted with a guidance system for remote control. Their official vehicle designator is then prefixed by the letter "Q," as for example QB-47 or QF-80. Missiles can also be used for target missions, as for example the Navy's *Regulus* vehicle which is an aerodynamic missile.

The term "aerodynamic missile" brings up still another method by which aerospace vehicles are classified. In this method, all missiles are either in the ballistic missile classification or in the aerodynamic missile classification.

Essentially the difference between the ballistic missile and the aerodynamic missile lies in the fact that the former does not rely upon aerodynamic surfaces to produce lift and consequently follows a ballistic trajectory when thrust is terminated. Aerodynamic missiles have a winged configuration; examples of such vehicles are the *Hound Dog* and *Mace*. The term "ballistic" leads to yet another method of identifying and describing some aerospace vehicles.

In common usage today are the following abbreviations which use the term "ballistic missile" in the sense that the type of missile and its capacity are indicated.

ICBM—Intercontinental ballistic missile
IRBM—Intermediate range ballistic missile
AICBM—Anti-intercontinental ballistic missile
FBM—Fleet ballistic missile
MMRBM—Mobile mid-range ballistic missile
ALBM—Air-launched ballistic missile

The range has much to do with using this kind of missile designator, which, like the point-to-point designator, is used with the vehicle's popular name. Ballistic missiles capable of attaining very long range (over 5000 nautical miles) or intercontinental range, are given the ICBM designator. Currently, these are the *Minuteman* and *Titan* missiles. Figure 1-2 shows the first test launch of the *Titan* missile. The *Thor*, *Jupiter*, and *Polaris* missiles are all classified as IRBMs.

All of these IRBMs are capable of reaching the upper limit of intermediate or mid-range which is 1500 nautical miles. Lower-limit of this range is 300 nautical miles. Missiles with less than a 300-mile range, such as those in the U. S. Army weapons inventory (*Pershing*, *Sergeant*, *Hawk*, etc.), are considered short-range weapons.

The last item on the list, ALBM, has nothing to do with the range capability of a particular ballistic missile. Rather, this abbreviation deals with the concept of launching a ballistic missile from a moving aerial platform which would be difficult for an enemy to track or destroy. This is similar to the *Polaris* missile which can be launched from continually roving submarines. The latest Air Force missile that could be designated as an ALBM is the AGM-69A SRAM which is air-launched to a ground target.

Missiles are sometimes referred to according to their airspeed relative to the speed of sound and their type of propulsion system. Generally, the highest rate of airspeed that can be reached safely and still assure correct operation is considered as that missile's classification. At best, this is only a general means of classification since the airspeed is related to the speed of sound (or Mach 1) which varies with respect to ambient air temperature. There are only four possible groups which are considered in classifying a particular missile by this means:

1. Subsonic—airspeeds less than Mach 1
2. Sonic—airspeeds equal to Mach 1
3. Supersonic—airspeeds ranging between Mach 1 and Mach 5
4. Hypersonic—airspeeds exceeding Mach 5

Almost all AIM and SAM missiles can be placed in the supersonic classification since modern military aircraft are capable of attaining Mach 1 speed. Some NASA (National Aeronautics and Space Administration) vehicles are used as satellite launch vehicles. To overcome the earth's gravitational influence, these vehicles must attain a velocity of from 4.9 miles per second (18,000 mph) to 7 miles per second (25,000 mph) depending on the purpose of the payload. This speed would place satellite launch vehicles in the hypersonic classification. At the present time, very few



Figure 1-2. First Test Launch of Titan Missile

missiles are found in the subsonic or sonic classification. Typical among these slower vehicles are the *Regulus* and *Mace*—which are also known as "air breathers."

The term "air breathers" pertains to the kind of propulsion unit upon which a missile must depend as its main source of power. In this case, the form of powerplant must use oxygen from the atmosphere to support fuel combustion inside its jet engine. In many instances, an air breather may obtain its initial launching speed from a solid propellant rocket motor known as a booster. Once the proper cruising speed for which the vehicle was designed is reached, its main or sustainer engine takes over for the remainder of the flight. This engine uses liquid fuel (not to be confused

with liquid propellant that is used in some missile rocket engines). Basically, the propulsion system method of missile classification is very simple and currently includes the following three categories:

1. Air breathers
2. Liquid propellant (chemical rockets)
3. Solid propellant

Each of the last two missile classifications could be further subdivided into single and multistage groupings which would include missiles with one or more propulsion units at each stage. Each stage, after the propellant is consumed, drops away and the next stage ignites to propel the missile on its flight. As a result of research and development programs conducted by the Government and by private industry, two more categories may be added to the three given previously—electric and nuclear rockets.

Classification as to the type of guidance system contained in the missile is quite often used to differentiate between aerospace vehicles. The two basic categories would be unguided missiles, such as free-flight rockets, and guided missiles which would contain one or more of the following systems:

1. Command guidance (wire, radio, or radar)
2. Inertial guidance
3. Position-fixing guidance
4. Seeker or homing guidance

The simplest command guidance system uses fine wires over which the proper navigation signals can be relayed from the carrier aircraft to the airborne weapon. Both the U.S. Army and Navy have used this kind of system to guide certain of their missiles.

Some guided missiles may contain combinations of the above systems. One such missile, the *Bomarc*, had a command guidance system which controlled the weapon from the ground to the approximate altitude and general area of a hostile aircraft, whereupon the *Bomarc's* own homing system then took over.

The main disadvantage of the command system is that it may be subject to enemy electronic counter-measures. This is not the case with a dead-reckoning or inertial guidance system. Extreme accuracy is possible with this type of guidance system since only the precise locations of the launching site and the target are needed. All other factors, even those which may vary (wind, speed, etc.), are compensated for as the missile streaks toward its fixed target. Most ICBMs and IRBMs use some form of an inertial guidance system.

Again a combined inertial and position-fixing guidance system may be used. The latter may occasionally refer to a map, chart, or star to check the missile trajectory. Infrared seekers and radar homing devices are employed in guidance systems for many AIMs such as the *Falcon, Sidewinder,* and *Sparrow* missiles mentioned previously.

Newspapers and trade journals make constant use of a missile's popular name and the abbreviations ICBM or IRBM, as the case may be, as well as the point-to-point designator, either in its abbreviated form or in its completely spelled out form. To talk shop effectively and easily and to keep abreast of the continual changes in the missile field, you should become familiar with all of the methods of classifying aerospace vehicles.

SUMMARY

As unguided missiles, the use of rockets had been known as far back as A.D. 1232 when they were employed to repel the Mongol besiegers of the city of Pein-King (Peiping). Since that time and in one form or another, rockets have been used as weapons and machines of war, for amusement through their colorful aerial bursts, as life-saving equipment, and for communications or signals. The lack of suitable guidance and control systems may have accounted for the rocket's slow improvement over the years. Strangely enough, it was the airplane rather than the rocket that stimulated the development of a guided missile as it is known today.

In this country, Kettering's *Bug* is considered to be the first guided missile. Lacking sufficient range, the *Bug* never became an operational weapon. In the time between the two World Wars, rocket and guided missile developments were greatly overshadowed by the feats and records accomplished by aircraft.

The apparent success of the German V-1 missile against such large land targets as the cities of London and Antwerp had much to do with stimulating development of guided missiles in Japan and the United States. Since several countermeasures could be applied against the V-1 missile, the de-

velopment of the V-2 was stepped up. The fact that a V-2 could impact on its target unannounced and with tremendous speed pointed out the importance of guided missiles as a new weapon of war. Success of the V-2 missile came too late to affect the outcome of the war in favor of Germany. After World War II, several nations pushed development of guided missiles, both as military weapons and as instruments of science to probe outer space.

In this country, the rapid and numerous developments of various types of aerospace vehicles were guided by the government which stimulated research by private industry. Larger ballistic missiles can travel over intercontinental distances with extreme accuracy, while other missiles with short ranges possess a high degree of mobility and flexibility. Still other missiles are capable of placing instrumented satellites into orbits to approach other planets for scientific studies of their surfaces.

With so many missiles available for such a variety of missions, many methods of classifying and designating each missile are available. The most popular designator is, of course, the popular name of the missile. Another designator, the point-to-point method, is frequently used.

Guided missile research and developments cover many fields as, for example, electricity, electronics, chemistry, metallurgy, and aerodynamics. The next chapter presents aerodynamics and what effects it has upon guided missiles.

CHAPTER 2

Missile Aerodynamics

The detailed study of air in motion and the mathematical analysis of the various forces present are involved and lengthy. Such detailed study is necessary only to the aeronautical engineer and designer. This presentation is confined to those basic theories and principles necessary for you to understand the effects of aerodynamic forces on missile structures.

PHYSICS OF FLIGHT

In a study of aerodynamics, you must understand several basic laws of physics. Once you understand these laws, applying them to the particular problems of missile aerodynamics becomes relatively simple.

Missile aerodynamics are generally the same for subsonic or supersonic flight. In fact, generally speaking, aerodynamics are the same for any craft that is intended to fly. Because of the rapid advancement in speeds and altitudes, new concepts and problems continually arise, such as the problem of the shock-wave which occurred when sonic (speed of sound) speeds were reached. Other problems which have been encountered are those concerning pressures and temperatures at extremely high altitudes.

First, we consider phenomena common to subsonic and supersonic speeds.

Relativity of Motion

The first things to consider in the study of aerodynamics are the forces which act on a body that is moving through the air. If you watch an airplane flying overhead, it appears that the air is standing still and the airplane is moving. In this case it seems that the opposing force on the body is due to the velocity of the body moving through the air as illustrated by the ball in figure 2-1.

If you are riding in the airplane, it seems that the airplane is standing still and that the air is moving. This is illustrated by the ball in figure 2-2.

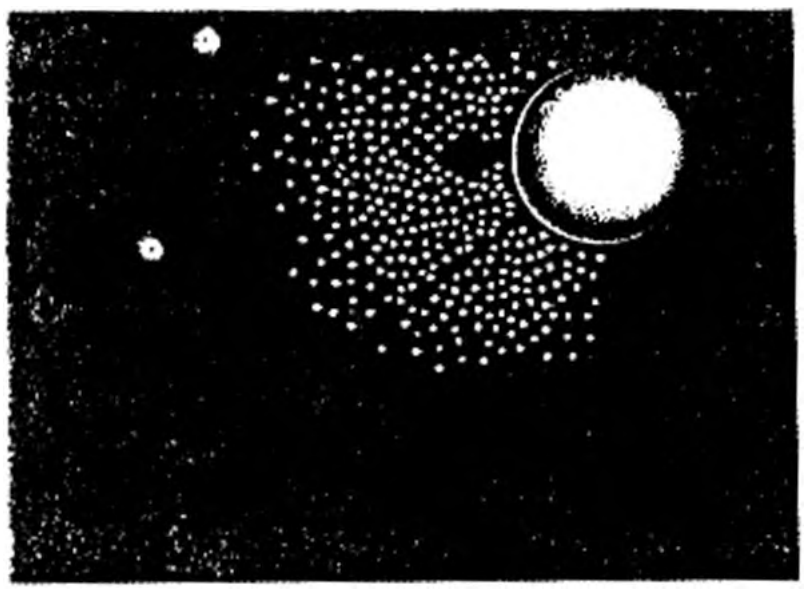


Figure 2-1. Observation from Stationary Point

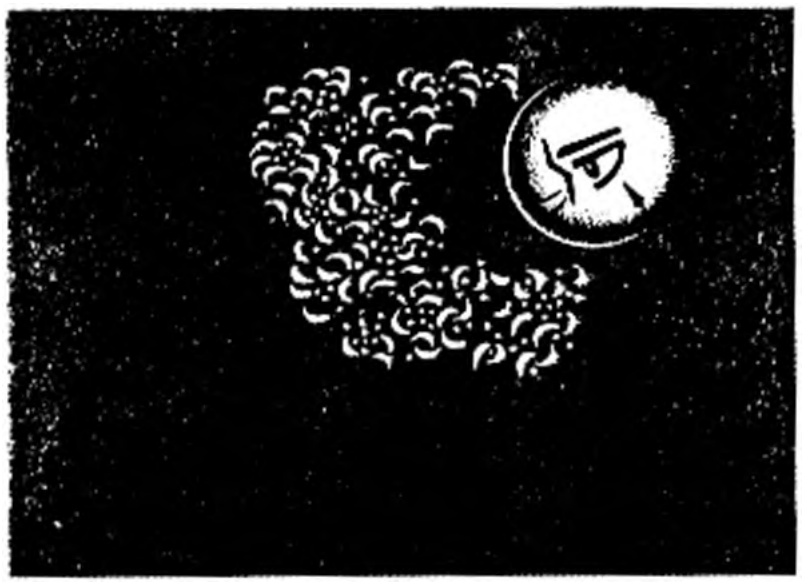


Figure 2-2. Observations from Moving Point

The preceding paragraphs explain the concept of *relativity of motion*. In either case the forces exerted on the body by the air are the same. This principle may be stated as: *The force exerted by the air on a body does not depend on the absolute velocity of either one but only on the relative velocities between them.*

In the study of aerodynamics, this principle is put into practice by the use of wind tunnels in which the aircraft is held stationary and high wind velocities are made to pass over it. This simplifies the study of the forces acting on the airframe, and test aircraft and missiles can be virtually "flown" while they are actually standing still.

Forces Acting on a Missile

Numerous forces act on all parts of a missile due to *air resistance*, *gravity* and other factors. Consider again a body moving through the air at a certain speed. There are four forces present. One is the force the body exerts on the air in moving forward. In opposition to this is the force the air delivers to the body. The other two forces are—the force of gravity which is pulling the body toward the center of the earth and the force the body exerts in the opposite direction to keep itself aloft.

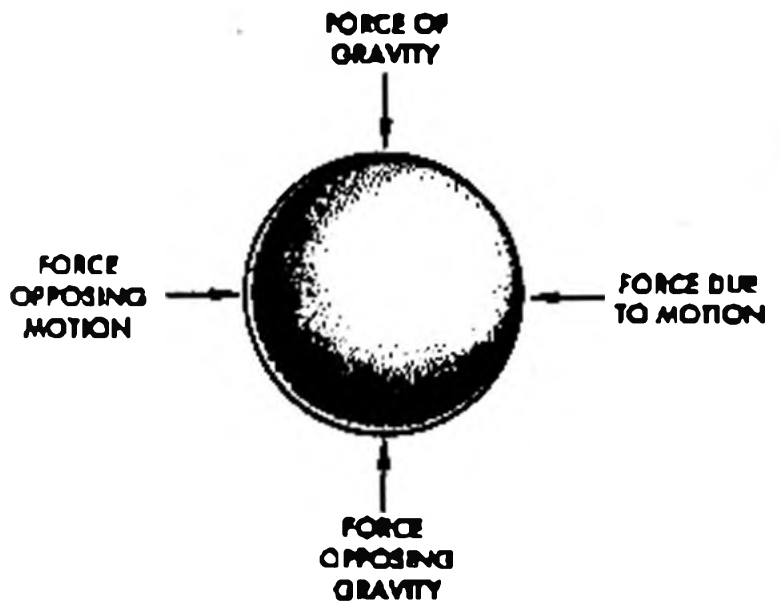


Figure 2-3. Forces Acting on Body Moving Through Air

In figure 2-3, the body will not move if all the opposing forces are equal; instead the body will remain in a state of equilibrium. If any of the forces acting on the body are not equal, a movement will result in the same direction as the greater force. For example, the ball in figure 2-4 is a representation of the forces acting on a body. The lengths of the vector arrows indicate the magnitudes of the forces, and the arrowheads point in the direction of the forces. Force "B" is greater than opposing force "A". Force "D" is greater than the opposing force "C". As indicated, the body will move in the direction indicated by the large arrow.

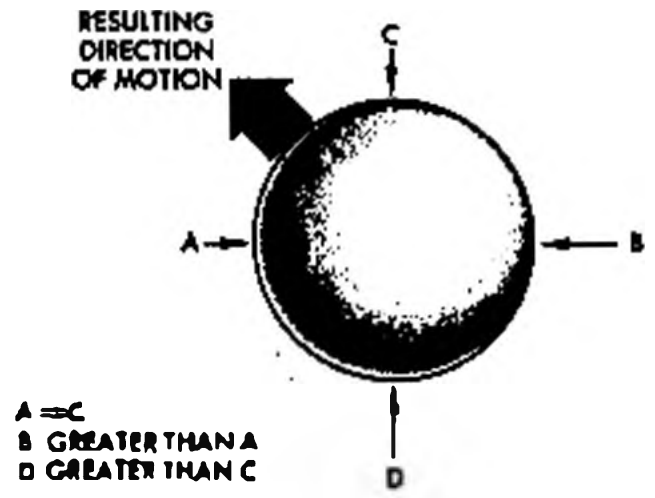


Figure 2-4. Unequal Forces Acting on Body

Figure 2-4 illustrates what is known as *vector representation* of the forces acting on the body. Any number of forces may be shown by a vector representation and can be resolved, or simplified, into resultant forces to determine what effect may result. Force diagrams are essential in designing structures that are to be subjected to a number of different forces.

In the previous paragraph, as well as later in this manual, reference is made to vector representation and resultant motion or force. A brief review of vectors is presented in the following paragraphs.

A vector is defined as a graphic representation of quantity having two characteristics, direction and magnitude. Figure 2-5 illustrates vectors and their uses.

Suppose you wish to indicate the effect of two or more forces acting at angles to each other. Figure 2-5B illustrates this condition. If a parallelogram is constructed as shown by the dotted lines and a line is drawn connecting the original point of the vectors to the opposite corner of the parallelogram, the resulting diagonal will represent, both in direction and magnitude, a single force which

would produce the same effect as the two original forces. This diagonal is the *resultant*.

The science of vector analysis can become quite complicated, but the above information is adequate for our application to the forces explained in this manual.

NEWTON'S LAWS OF MOTION. Now examine three basic laws of physics known as Newton's laws of motion.

The first law states: "A body in a state of rest remains at rest, and a body in motion tends to remain in uniform motion unless acted upon by some outside force." This law tells us that whenever there are unequal forces upon a body, the body must move. After the body is in motion, it will stay in motion as long as there is no force present to change that motion. For example, if you push against a book lying on a table, you find that there is a certain amount of force required to overcome friction to set the book in motion. If you could eliminate all of the restraining forces acting on the book once it is in motion, the book would continue to move uniformly until acted upon by some outside force. These restraining forces are the main concern in the study of aerodynamics.

Newton's second law states: "The rate of change in the momentum of a body is proportional to the force acting on the body and is in the direction of the force."

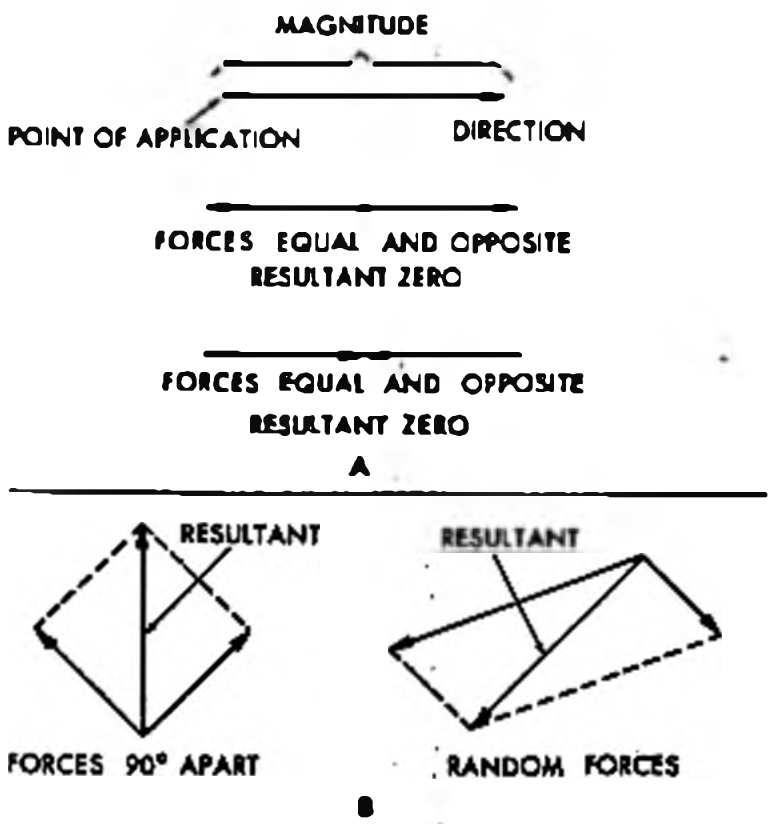


Figure 2-5. Vectors

His third law states: "To every action there is an equal and opposite reaction." If a force is applied, there must be a reaction opposite to and equal to the applied force.

DEFINITION OF FORCE. What is meant by a "force applied"? Consider a pound of air moving in some direction. The pound of air is capable of producing a force. The force of that pound of air is directly proportional and equal to the weight and the velocity change. This can be stated simply as: Force = Mass × Acceleration, or $F = ma$.

Since air has mass, once air is put into motion it is capable of applying a force.

Whenever a force is applied through some distance, work is done. In simple form: Work = Force × Distance, or $W = Fd$.

Any mass in motion is capable of applying a force to another body and thereby doing work. Whenever the motion of a mass is changed, there is a change in its momentum.

EDDIES. When air moves over a surface, a certain number of particles of the air tend to rebound or swirl past the surface as shown in figure 2-6.

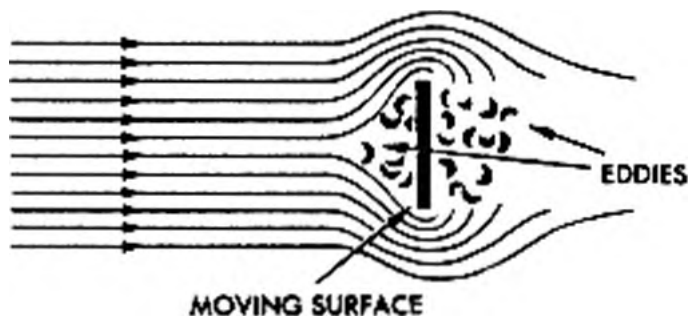


Figure 2-6. Eddies from Airflow over Flat Surface at Right Angle to Airflow

The turbulence of swirling particles is known as eddies. The number of eddies depends on such factors as smoothness, size, and shape of the surface. The number of eddies also varies with the angle the surface makes with the air, with relative speeds, and with the density of the air.

LIFT AND DRAG. Because of the friction between the mass of air and the inclined surface as illustrated in figure 2-7, there is a force which tends to move the surface parallel to itself. There is also a force perpendicular to the surface, caused by the deflecting airstream. These combined forces give a resultant force backward in the direction



Figure 2-7. Forces Acting on Flat Surface in Airstream

of the moving air mass. Breaking this resultant force down, we have one force with its vector perpendicular to the moving mass of air called lift and another force parallel to the mass of air called drag. The angle that the mass of air makes with the surface is called the angle of attack. The figure shows the forces which make up the resultant force on the surface at a particular angle of attack.

Bernoulli's theorem. Scientist Daniel Bernoulli discovered that total energy in any system remains constant. In other words, if one element of any energy system is increased, another decreases to counterbalance it.

Before applying Bernoulli's theorem to a wing section, consider what happens to a baseball when a pitcher throws a curve. By snapping the wrist, the pitcher puts a spin on the ball. After leaving the pitcher's fingers, the ball has two motions as shown in figure 2-8. The ball is turning on an axis perpendicular to the line of motion but parallel to the ground, and it is moving forward.

These two motions of the ball cause different velocities of air to rush around the ball. Now, on the side of the ball that is turning into the rush of air, the relative velocity of the air is slower because the surface spin of the ball is in the same direction as the ball was pitched so it partly counteracts the speed of the air rushing past it. On the side of the ball which is turning away from the rush of air, the velocity of the spinning air and the velocity of air flowing past the ball are in the same direction. The relative velocity of the air is greater on the side of the ball which is turning away from the rushing air.

Keeping in mind that Bernoulli's theorem states that total energy in a system remains constant, it becomes evident that if the speed of air over a surface is increased, the pressure exerted by the air on the surface must decrease, thus keeping the total energy constant. Conversely, decreased speed of air over a surface must result in increased pressure on the surface.

Returning to the pitched baseball, you now can see that unbalanced pressure is exerted on the

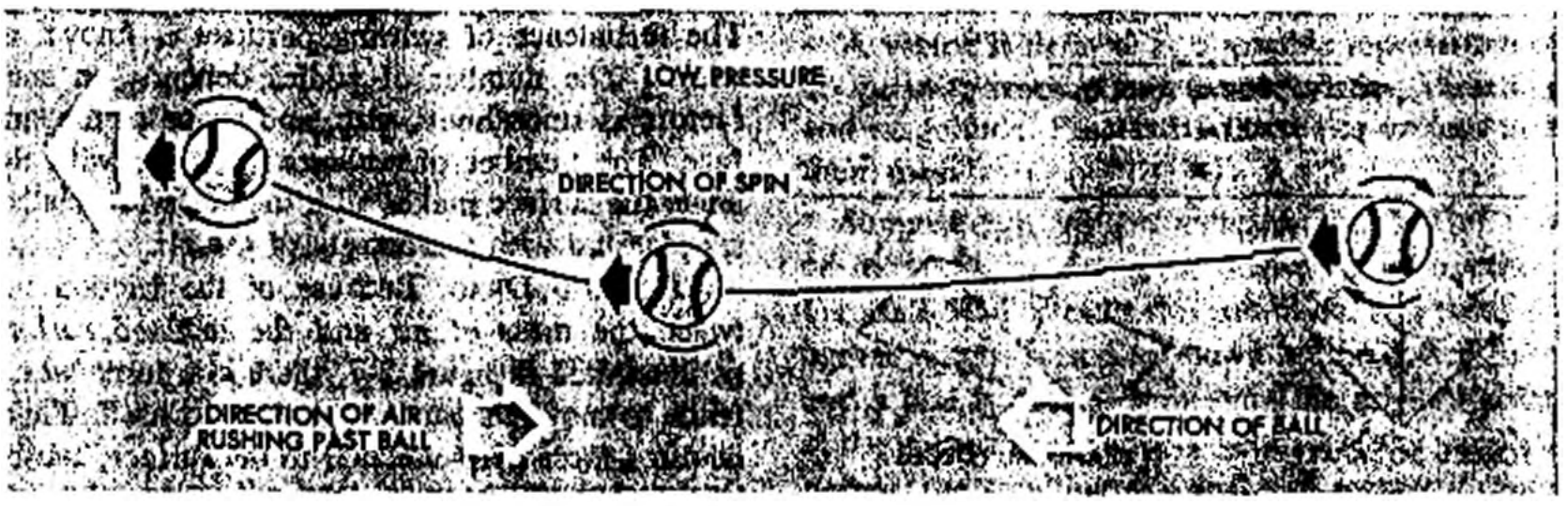


Figure 2-8. Bernoulli's Theorem Applied to Pitched Baseball

spinning ball. On the side of the ball turning away from the rushing air, the pressure is decreased because the speed of the air is increased, while on the other side the pressure is increased because the counteracting air has decreased the total speed of the air.

The building up of the lift force on an airfoil differs only in application from the pitching of a curve ball. Notice that the top surface of the wing section shown in figure 2-9 has greater curvature than the lower surface.

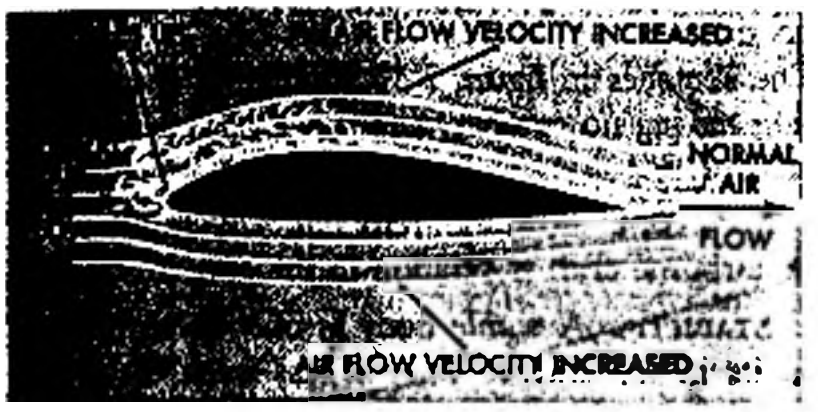


Figure 2-9. Airflow Over Wing Section

The difference in curvature of the upper and lower surfaces of the wing builds up the lift force. Air flowing over the top surface of the wing must reach the trailing edge of the wing in the same amount of time as the air flowing under the wing. To do this, air passing over the top surface has to move at a greater speed than air passing below the wing because of the greater distance the air must travel via the top surface. The increased speed means a corresponding decrease of pressure on the surface. A pressure differential is created between the upper and lower surfaces of the wing, giving it lift.

You now can see that lift force is created by a change of momentum in the air mass. In this presentation, lift is approached from either the standpoint of increased pressure on the under surface of an airfoil or from the standpoint of increased speed of the airstream passing over the top surface.

Always keep in mind that whenever an object changes its direction or rate of motion, unequal forces are acting on it.

Boundary layer. Boundary layer refers to a condition occurring as a result of friction between an airfoil surface and the air moving past it. This clinging of air to an airfoil, such as a wing, is a serious problem in aircraft design. Lift depends upon circulation around an airfoil. When circulation is restricted, lift is reduced. Boundary layer drag can be reduced by polishing the airfoil surface and by eliminating such protrusions as rivets.

Now that you are familiar with the basic laws of physics which are involved in flight, you are ready to consider other factors contributing to aerodynamics.

PROBLEMS OF AERODYNAMIC FORCES

A presentation of the problems of aerodynamic forces involves several flight terms which you need to understand. A few of these terms are explained in the following paragraphs.

Basic Flight Terminology

The following explanations of flight terms are intended to give the basic meanings of the terms; they are not an engineer's definitions.

AIRFOIL. An airfoil may be described as any structure around which air flows in a manner that is useful in flight. The most important airfoils on an aircraft are the wings. Other airfoils of an aircraft consist of tail surfaces and the fuselage.

DRAG. The resistance of an object to the flow of air around it is called drag. Drag is due in part to the adhering of air to a surface (boundary layer) and in part to the piling up of air in front of the object. One of the main objectives in aerodynamics is to reduce this resistance yet maintain a high amount of lift and stability.

STREAMLINES. The paths of air particles as they flow past an object are called streamlines. Note figure 2-10.

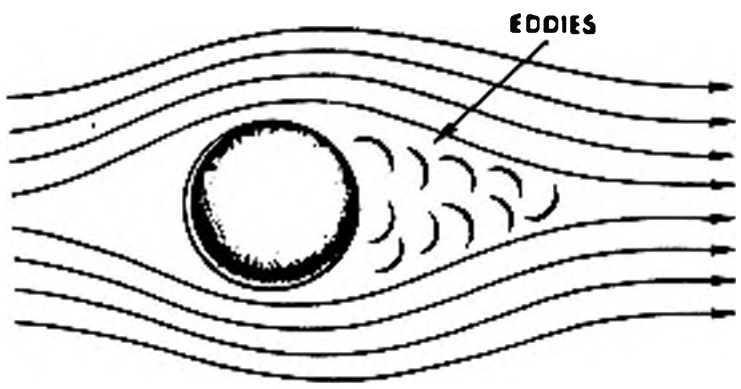


Figure 2-10. Streamline Flowing Past Spherical Body

The streamlines travel both above and beneath the spherical body. In the case of a wing, the streamlines must travel a greater distance and move with greater speed over the top of the wing. The increase in speed over the top of the wing results in decreased pressure on the upper curved surface of the wing. When a winged aerospace vehicle is traveling through the atmosphere, the decreased pressure on the upper surface of the wing is less than atmospheric pressure.

WING SPAN. Wing span is the measured distance from the tip of one wing to the tip of the other.

CHORD. The distance from front to back of a wing is called the chord as illustrated in figure 2-11. The chord usually varies in length along

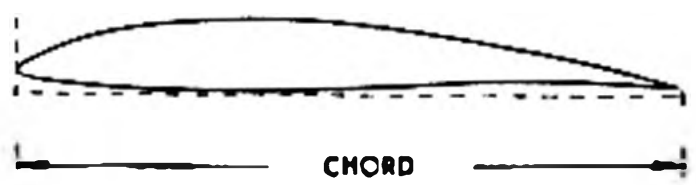


Figure 2-11. Measurement of Chord on Wing Section

the wing, so that we usually speak of the average chord. The average chord is considered to be the ratio of the wing to the wing span. For example, if a wing has an area of 48 square feet and a span of 16 feet, the average chord would be determined by dividing the area (48 sq ft) by the span (16 ft). The average chord would be 3 feet.

CAMBER. Camber refers to the rise of a curve of an airfoil section. Camber is usually expressed as the ratio of the departure of the curve from a straight line joining the extremities of the curve to the length of this straight line. Upper camber refers to the upper surface, lower camber to the lower surface, and mean camber to the mean line of the section. Camber is positive when the mean camber line curves upward, negative when the mean camber line curves downward, and there is no camber if the mean camber line is straight. The sketches in figure 2-12 indicate a chord section having no camber, and one having a positive camber.

ATTITUDE. Attitude refers to the position of an airborne craft in relation to the ground.

STABILITY. A *stable* body is one which returns to its initial attitude after it has been disturbed by some outside force. If outside forces disturb a stable aircraft from its normal flight, the aircraft tends to return eventually to its original attitude.

Sometimes a body, if disturbed from its original attitude, assumes a new attitude and neither returns to its origin nor moves any farther from it. Such a body is *neutrally stable*. If the attitude of a neutrally stable aircraft is changed by an outside force or change in controls, the aircraft does not tend to return to the original attitude. It re-

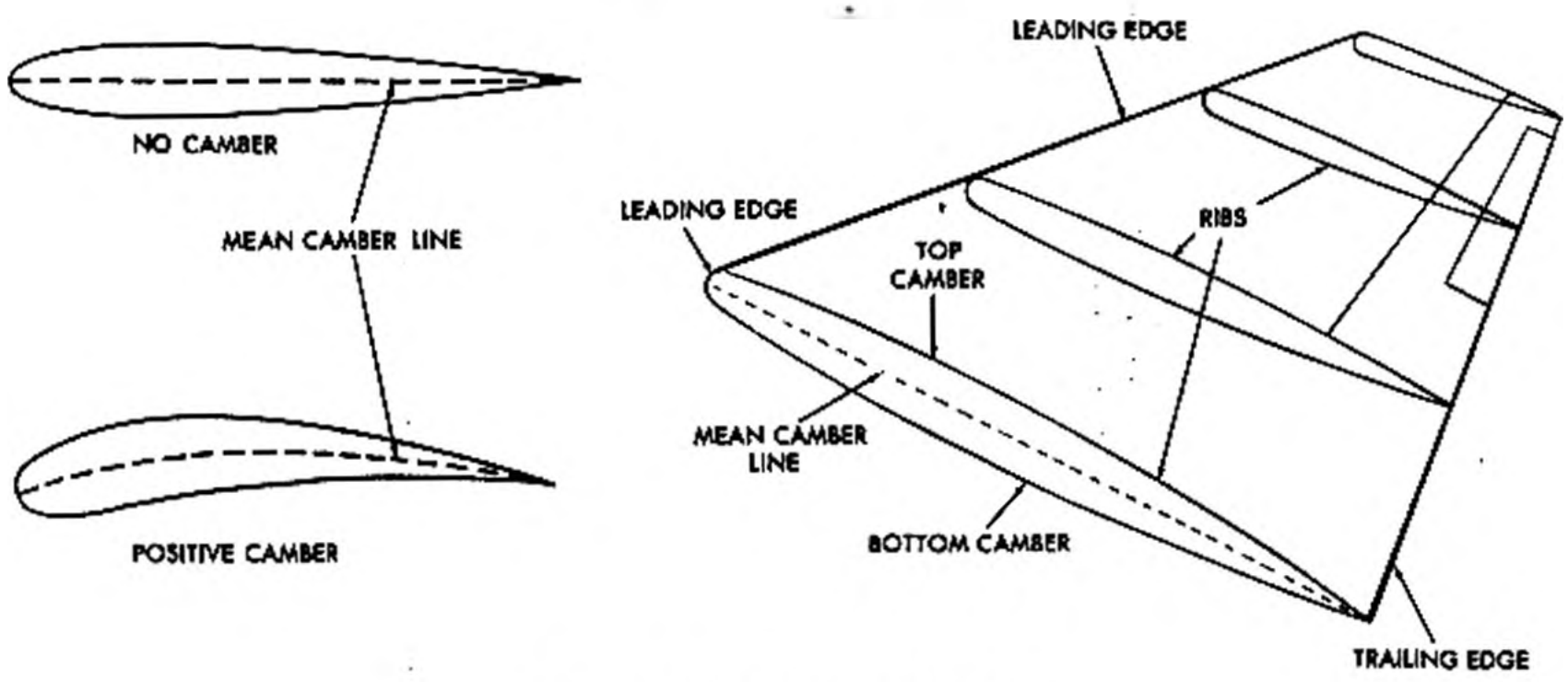


Figure 2-12. Camber of Airfoils

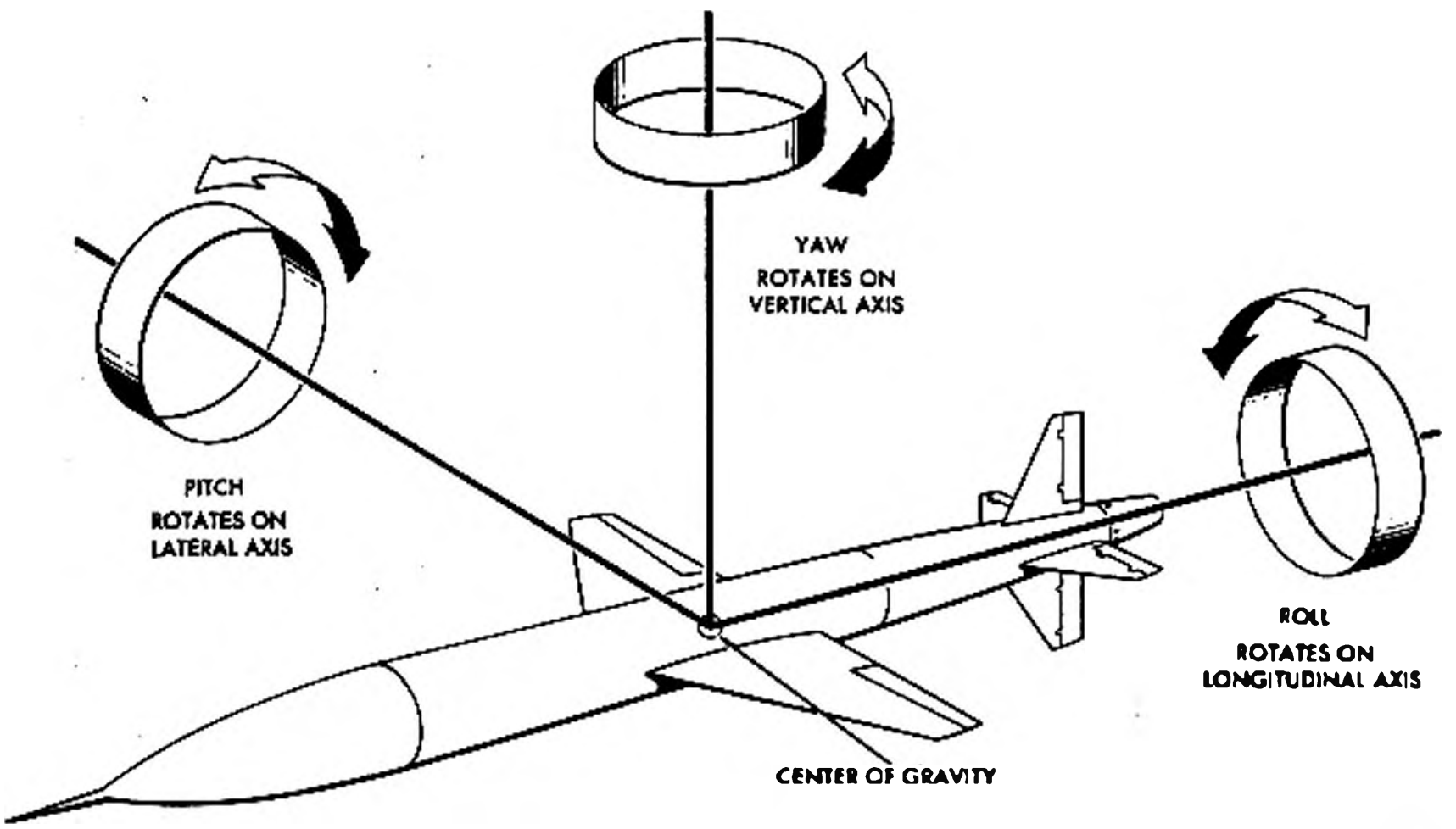


Figure 2-13. Flight Attitude of Missiles

mains, instead, in the new attitude until other forces influence it.

A third type of stability is *negative stability or instability*. In a case of instability, a body displaced from its original attitude tends to move farther away. For example, if an unstable aircraft is put into a climb, it tends to climb more and more steeply until it stalls.

Axis. A missile moves about three axes as shown in figure 2-13. The vertical line through the center of gravity is considered as the *yaw axis*, the longitudinal line through the center of the fuselage is called the *roll axis*, and the line that is perpependicular to the line through the fuselage and parallel to the wings is called the *pitch axis*. Whenever there is any displacement of the missile about any of these axes, the outcome may be any of the following actions.

- It may oscillate about the axis and return to its initial position.
- It may remain in the new attitude.
- It may increase its displacement and get out of control.
- It may return to its original position readily without any oscillation.

The last possibility, which indicates a stable body, is desired. We discuss later how this problem of stability is met.

DIHEDRAL AND CATHEDRAL. A dihedral angle is the angle formed by a reference line through the wing surface and lateral axis of the aircraft. This angle lies in a plane perpendicular to the longitudinal axis. Cathedral, often referred to as negative dihedral, is the angle that the wings make in a downward direction toward the wingtips from the fuselage. Cathedral is often used on vehicles designed for supersonic speed. Study figure 2-14.

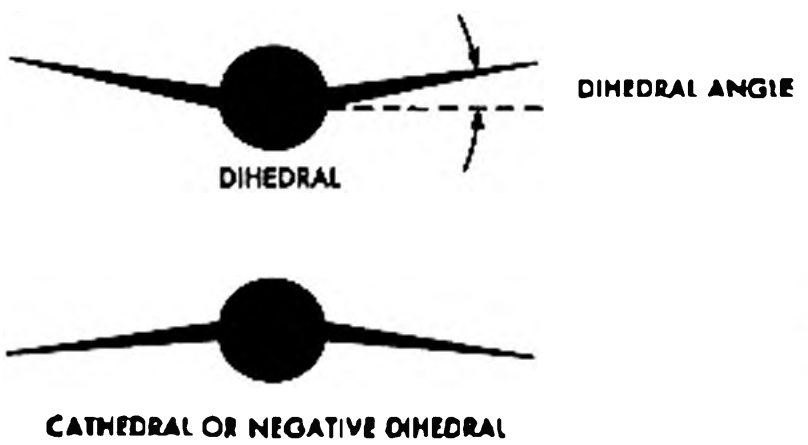


Figure 2-14. Dihedral and Cathedral

Problems Pertaining to the Force of Air

Now that you are familiar with some aerodynamic terminology and have a general idea of the forces and actions that are taking place on an airfoil, consider some of the more specific problems of air forces. Since a wing section is the most important of the lifting surfaces, much of the following presentation centers around this section.

RELATIONSHIP OF MODEL MISSILE TO FULL-SIZE MISSILE. Although a model made to perfect scale flies in a wind tunnel under nearly the same conditions as a full-size missile in the air, the airflows related to each are not the same. This difference is present because the airmass force and the frictional forces do not follow the same laws. The airmass force follows what is called the square law while the friction force follows the direct law. The mass force is proportional to:

(Length) × (Length) × (Velocity) × (Velocity).

The friction forces are directly proportional to:

(Length) × (Velocity).

A true relationship hardly ever exists in model tests because the forces on a model generally are much less than the forces on a full-size missile. Because of such discrepancies between model tests and full-size operation, full-scale test missiles must be built to get a true indication from flight test. Differences between the effects of air forces on models and on full-size missiles become more prominent at supersonic speeds.

CENTER OF PRESSURE. On each infinitesimal part of a wing surface there is a small force present. This force is different in magnitude and direction from the force acting on any other area forward or rearward from this point. It is possible to add mathematically all of these small forces. The sum of all the tiny forces over this surface is called the resultant. This resultant has magnitude, direction, and location. The point of intersection of the line of direction with the chord is called the *center of pressure.*

In actual flight there are different airspeeds for different angles of attack. But if the velocity of the airstream is maintained constant for test purposes while the angle of attack is changed, the results on a nonsymmetrical wing are as shown in the sketches in figure 2-15. The sketches show a wing section at various angles of attack and the effect of the different angles of attack on the resultant force and the position of the center of pressure.

The burble point (the stalling angle) referred to in figure 2-15, is a point at which airflow over

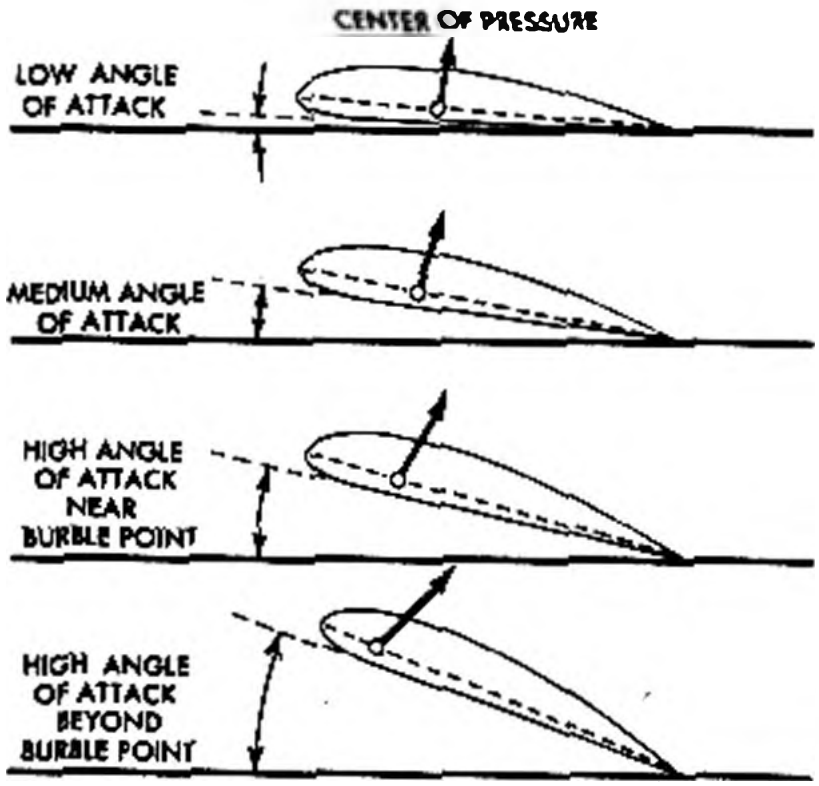


Figure 2-15. Changes of Center of Pressure as Result of Varied Angles of Attack

the upper surface becomes rough, causing an uneven distribution of pressure.

At small angles of attack, the resultant is comparatively small. Its direction is upward and back from the vertical, and its center of pressure is well back from the leading edge. Note that the center of pressure changes with the angle of attack, and the resultant has an upward and backward direction. At a positive angle of attack of 3° or 4°, the resultant has a nearly vertical direction. Either increasing or decreasing the angle causes the direction of the resultant to be farther from the vertical.

To determine the amount of force that is applied to a surface, you must first determine the mass rate of air that is working on that surface. To find the mass rate of air, you must know the velocity of the air and the area that it is working on. The mass rate of air is determined as follows: Mass rate = (Cross section) × (Velocity) × (Density), where density, by simple definition, is the quantity per unit volume.

Suppose a mass of air is moving in a horizontal direction and is deflected slightly downward at a certain velocity in feet per second. The lift that the mass of air exerts in pounds is equal to the product of the mass multiplied by the change in velocity:

Lift = (Cross section) × (Velocity) × (Density) × (Deflection velocity).

This is a simplified formula to compute lift. This formula is basic; you will see later that there are other factors that must be taken into consideration in the computation of lift.

As you learned before, the resultant force on a wing under a particular condition can be described as two components in two directions, the chosen directions being perpendicular and parallel, respectively, to the relative wind. These components, you will recall, are called lift and drag.

DETERMINATION OF LIFT. Lift force depends on the contour of a wing, angle of attack, air density, area of the wing, and the square of the airspeed. The common equation for lift is given as:

$$L = C_L \frac{\rho}{2} SV^2,$$

where L is the lift in pounds, C_L is the lift coefficient which depends on the wing contour and the angle of attack, ρ is the air density in slugs per cubic foot, S is the area of the wing in square feet, and V is the velocity (airspeed) in feet per second. (A slug is defined as the mass which would receive an acceleration of 1 foot per second when it is acted upon by an unbalanced force of 1 pound.)

The coefficient of lift (C_L) and the coefficient of drag (C_D) are determined by wind-tunnel tests and are plotted as a characteristic curve for the particular airfoil under consideration. Figure 2-16 shows a typical set of curves on a given airfoil.

To maintain level flight, total lift must equal the weight it is supporting. As the angle of attack increases, the lift coefficient increases to a certain maximum value. The maximum value of the lift coefficient is the point where the air no longer flows evenly over the wing surface but tends to break away. The breaking away (the burble point) is called the *stalling angle*. After the stalling angle is reached, the lifting force is rapidly lost, as is the airspeed.

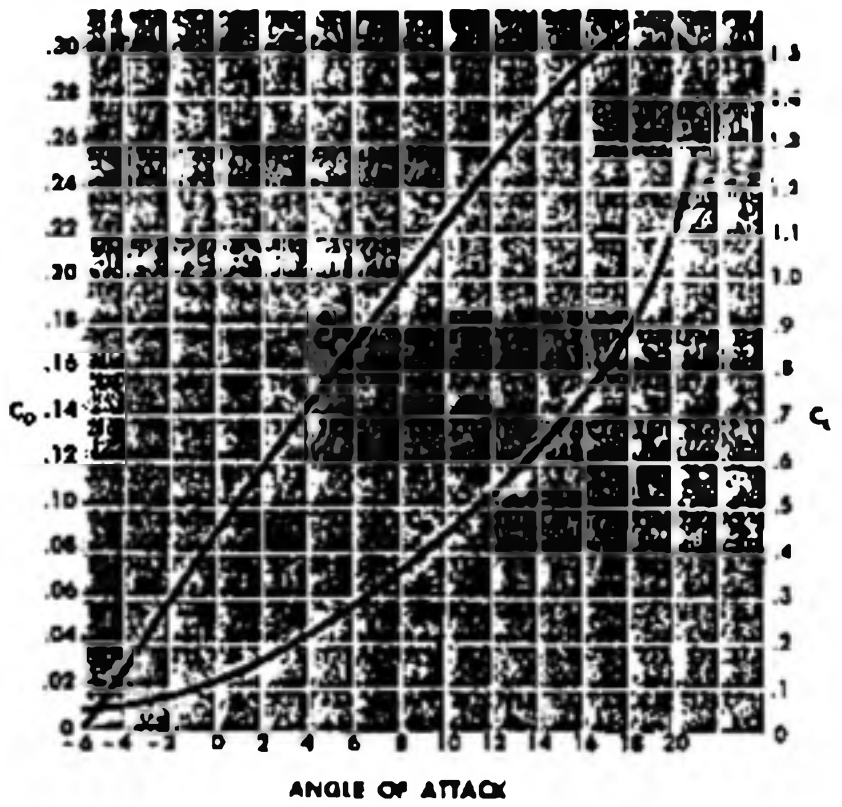


Figure 2-16. Coefficients of Lift (C_L) and Drag (C_D) Curves of Airfoil

The least possible airspeed occurs when flying at the angle of the maximum-lift coefficient. Another name for this angle of maximum-lift coefficient is the angle of minimum speed. When weight is added to an aircraft, wing loading is increased and the airspeed must be increased to fly at the same angle of attack as could be flown before the weight was added.

"Clipping the wings" is sometimes done to increase the speed of an aircraft. This decrease in wing area means an increased wing loading, and a higher velocity is required for any angle of attack as compared with the velocity before the wing was reduced.

DETERMINATION OF DRAG. Drag is the resistance of air to forward motion. The drag component of the resultant force on a wing is the component parallel to the direction of motion. It is this force that resists the forward motion. In horizontal flight, drag is the force that must be overcome by the thrust (the force that is pushing the vehicle forward). The drag force (in pounds) is the product of a drag coefficient, obtained from characteristic curves of airfoils, times half the air density times wing area in square feet times the square of the velocity. The drag force in formula is

$$D = C_D \frac{\rho}{2} S V^2,$$

where C_D is the coefficient of drag, ρ is air density, S is the wing area, and V is the velocity (in feet per second). For small angles of attack, the drag coefficient changes very little with angle of attack. As the angle of attack increases, the drag coefficient increases. The drag coefficient is usually quite small in comparison with the lift coefficient.

In addition to the wings on a missile, the fuselage, tail airfoils, and other surfaces resist motion. The resistance of the parts of a missile which do not contribute to lift is called *parasite drag*. That part of the drag of the airfoils which contributes to lift is called *induced drag*. Both parasite and induced drag vary as the square of the velocity. The function of the driving force is to furnish a forward-acting thrust to balance the opposing drag.

A force producing an acceleration is an unbalanced force; it is not counteracted by an equal force acting in the opposite direction. Such a force acting on a body produces an acceleration in the direction of the force. If a body is in motion at a certain speed and the forward-acting force of thrust is just equal to the opposing force of the total drag, then there is no acceleration; the velocity remains the same.

If the thrust is decreased in magnitude so that drag is greater than thrust, an unbalanced backward force results. This causes a deceleration of forward velocity. Decreasing the forward velocity causes a decrease in drag. When the drag force has decreased in magnitude until it is equal to thrust, there is no further deceleration, and the body continues moving at the new, slower velocity.

Wing drag varies with air density and as the square of the velocity of the missile. In level flight, a missile flies faster at altitude than at the same angle of attack at sea level. This occurs because air density is less at higher altitudes; thus the drag on the wing is less and the missile must travel faster for lift resulting from the decrease in air density. The effect of decrease in density on the wing is exactly neutralized by the increase in velocity squared. For a given aircraft, the wing drag depends on the angle of attack and is independent of altitude. The known drag for one altitude is essentially the same at any other altitude, provided the angle of attack remains unchanged.

To maintain the forward movement of the wing through the air, a force equal to the drag must be constantly exerted. This force multiplied by the velocity is equal to the power which must be expended in maintaining forward motion. If the force in pounds is multiplied by the velocity in feet per second, the product is power in foot pounds per second. By definition, horsepower (hp) is 550 foot pounds per second. The horsepower needed to move a wing through the air is: $hp = DV/550$, where D is the drag in pounds and V is the velocity in feet per second.

ASPECT RATIO. In discussing the forces on a wing, the area of the wing has been mentioned frequently. An area may be in the form of a square or a rectangle with different ratios of length of sides. There is a difference in the forces on a wing depending on the shape of the wing, even though the area remains the same.

There is a relationship between the span and the chord of the wing. This relationship is called the aspect ratio. In a rectangular-shaped wing, it is the ratio of span to chord. If the wing is not a simple rectangle but is tapered or has elliptic tips, the span is still the extreme distance from tip to tip but the chord varies from one position to another along the span. We can either find an average chord, or define the *aspect ratio* as the ratio of span squared to the area. In formula this ratio reads: Aspect ratio $= b^2/S$, where b is the span of the wing and S is the area.

We must know the aspect ratio of a wing to fully determine the extent of the forces acting on the wing.

WINGTIP VORTEX. As air flows about a wing, the pressure of the air immediately above the upper surface is less than the air pressure immediately below the lower surface. With the air at higher pressure below the wing, air will spill by the wingtips to the upper surface as shown in figure 2-17.

This flow of air from the lower surface combines with the normal flow of air, causing a swirl of air at the wingtips. The swirl is called a *wingtip vortex*. At each wingtip the action of the vortex is to throw the air inward and downward. The downflow caused by the wingtip vortices has a greater effect in disturbing the ordinary flow about a wing with a shorter span than a wing with a longer span. For two wings of the same area, the

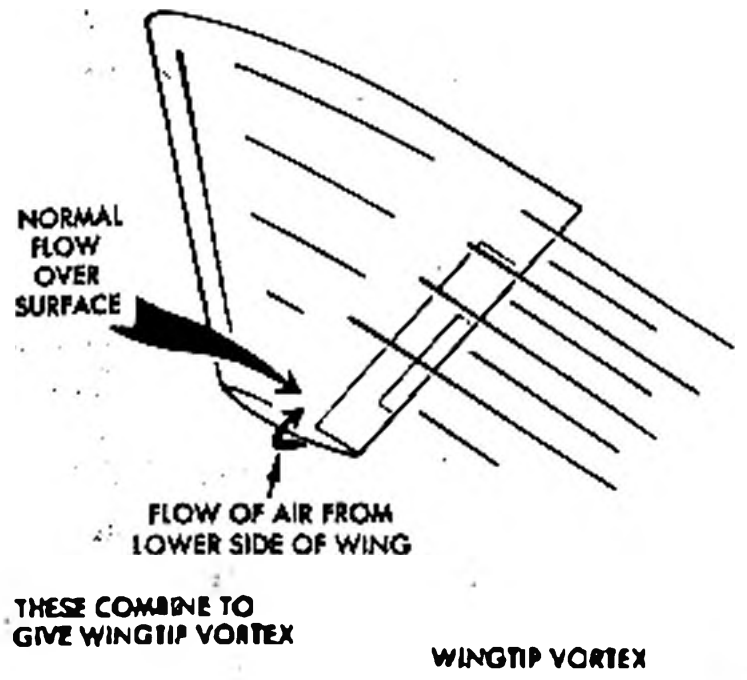


Figure 2-17. How Wingtip Vortices are Formed

wing with the smaller aspect ratio has a shorter span than a wing with a larger aspect ratio. Wings with large aspect ratios are less affected by wingtip vortices than wings with small aspect ratios.

In analyzing the effect of the wingtip vortices, consider the drag of the wing as being made up of two parts, called profile drag and induced drag. *Profile drag* is essentially the skin friction of air on the wing and is practically independent of the angle of attack and the aspect ratio. *Induced drag* is related to the downflow of the wingtip vortices. The magnitude of induced drag depends on both the aspect ratio and the angle of attack.

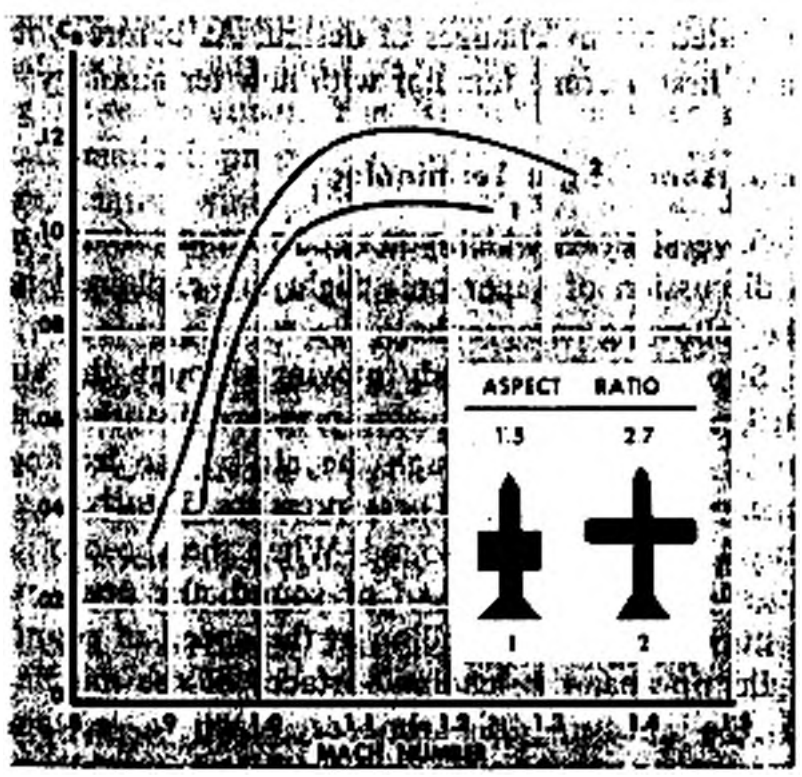


Figure 2-18. Effect of Aspect Ratio on Wing

Since the total drag of a wing is the sum of its profile drag and its induced drag and since the induced drag changes with various aspect ratios, the total drag changes with various aspect ratios. With higher aspect ratios, the effect of the wingtip vortices is to give the same lift coefficient as smaller angles of attack. The net result of changes in aspect ratio is shown in figure 2-18. For a given aspect ratio at any one angle of attack, the lift coefficient is greater and the drag coefficient is less than for a wing of smaller aspect ratio.

The figure shows that from a theoretical standpoint the biggest possible aspect ratio is most desirable, but for practical reasons of structural strength the aspect ratio is seldom greater than 8.5 for aircraft.

CONTROL ABOUT THE THREE AXES. Aircraft must be constructed in such a manner that they will fly a course without continual correction in direction. Such stability is made possible by devices that control an aircraft about its three axes.

Stability about the vertical axis. Stability about the vertical axis is commonly provided by a vertical fin. If an aircraft tends to turn to the right (yaw to the right), the pressure on the left side of the fin is increased. This increased pressure resists the rotation and forces the tail to the opposite direction.

In conventional aircraft, the fin may be divided and have a movable part called the rudder that is used for directional control. In addition to the rudder, there may be trim tabs that can be set for a particular direction of flight relative to prevailing air currents. The tabs compensate for the unbalanced air forces on the rudder when the wind direction is not directly from the front or rear of the flight path.

Along with the fin, the vertical sides of the fuselage act as stabilizing surfaces. The same action takes place here as on the fin.

Another means of obtaining yaw stability is by sweepback of wings. *Sweepback* is the angle that the leading edge of a wing makes with the longitudinal axis of the missile measured from the nose. If a missile yaws to the right, the leading edge of the left sweepback wing becomes more perpendicular to the relative wind while the right wing

becomes less perpendicular to the relative wind. This condition places more drag on the left wing and less on the other. The unbalanced drag tends to equalize, forcing the missile back to its original attitude.

Stability about the longitudinal axis. Stability about the longitudinal axis is provided in a missile by dihedral and the position of the wing.

Dihedral produces stability by causing a change of lift on the wing surfaces. As a missile starts to roll, it will sideslip slightly and thus create a relative wind component. This component increases the lift on the lower wing and decreases the lift on the higher wing. The lift increases on the lower wing because the angle of attack becomes greater on that wing.

The positioning of the wings at the time an aircraft is constructed is another means of obtaining stability about the longitudinal axis. An aircraft has greater stability if the center of lift on the wings is placed above the aircraft center of gravity than if placed below the center of gravity. Aircraft generally are classified according to four wing types: low-wing, mid-wing, high-wing, and parasol-wing.

Stability about the lateral axis. Stability around the lateral axis is called pitch stability. Pitch stabilizing is accomplished by a horizontal surface at the tail of the aircraft. This surface is known as the stabilizer. The stabilizer may be considered as consisting of two sections: the stationary part as the stabilizer and the movable part as the elevator. Stability with reference to pitch is accomplished by the increasing forces present on the tail surface when the aircraft changes its angle of attack. For example, if a missile tends to nose downward, the force on the upper surface of the stabilizer would increase, forcing the tail down and bringing the missile back to its original attitude.

Elevator tabs are used to correct for any existing factors causing pitch, just as the trim tabs are used in the case of the rudder.

Other types of control surfaces and stabilizing devices are taken up later in this chapter.

AERODYNAMICS OF SUPERSONIC FLIGHT

You now have a basic understanding of the forces that are constantly working on an airframe. Such information helps you understand the problems that arise in supersonic flight. The major problems in high-speed aerodynamics are listed as follows:

- Drag reduction.
- Lift effectiveness.
- Aerodynamic control.
- Stability: static, dynamic, automatic.
- Propulsion aerodynamics.
- Aerodynamic loads.
- Atmospheric conditions.
- Techniques for studying high-speed aerodynamics.

Up to the end of World War II, improvements in aircraft performance and speed came slowly. Progress was made primarily in the development of larger engines and minor aerodynamic changes in former design. As jet powerplants came into prominence, the speeds of new aircraft increased greatly. The old design procedures, formerly satisfactory, were changed for several reasons. For one thing, the flow of air at high speeds follows a different set of laws. Near-sonic speeds means dealing with compressibility and Mach-number effects. At these speeds, the problem of heating must be dealt with. High speeds also require changes in wind-tunnel design, scale effects, types of airfoil sections, problems of stability, and other factors that up to this time had little significance.

The next pages introduce you to some of these problems and show how the problems are compensated for by changes of design. As before, you must first become familiar with new terminology.

Supersonic Flight Terminology

Several terms which you need to understand for a discussion of supersonic flight are explained in the following paragraphs.

SHOCK WAVE. A body moving through the air strikes air molecules causing pressure disturbances to be propagated through the air by the leading surfaces of the body. These pressure disturbances travel at the speed of sound. When the speed of a missile reaches the speed of sound, the pressure disturbances also travelling at the speed of sound build up near the leading surfaces of the missile. If the pressure disturbances reach a certain strength, a shock wave is established just ahead of the leading surfaces. As air passes through a

shock wave, the static pressure rises sharply. Consequently, the forces and pressure distribution on the missile are altered. Strong changes in trim are necessary. The flow of air over the wings separates, much as in a low-speed stall. Tail surfaces may get a good buffeting. Wing drag rises. And deflection of control surfaces may cause new shock waves, which interact so that the controls may become ineffective at certain speeds.

MACH NUMBER. Since the speed of sound plays a big role in determining shock waves and airflow characteristics at high speeds, the ratio of flight speed to the speed of sound is significant. This ratio is called Mach number in honor of an Austrian scientist, Ernst Mach (pronounced "mock"), who first pointed out its importance in 1887. If a missile travels at a speed twice the speed of sound, the missile has a flight speed of Mach 2.0. If a missile has a speed half that of sound, it has a flight speed of Mach 0.5. The speed of sound varies with altitude, decreasing from 760 miles per hour at sea level to about 675 miles per hour at 30,000 feet.

REYNOLDS NUMBER. When scale models are used to predict the aerodynamic characteristics of full-size vehicles, two flow similarities must be satisfied. First, the flows must be geometrically similar. This requirement means that the scale model must be an exact replica of the full-size vehicle.

Secondly, the flows must be dynamically similar. Among the conditions for dynamic similarity is that the Reynolds numbers for the two corresponding flows be equal. The Reynolds number is a mathematical proportion which includes model size, relative windspeed, air viscosity, and air density. A change in Reynolds number can change the lift and drag on a body in that flow.

HEAT BARRIER. As air is compressed by the ram effect at high speeds, a temperature rise takes place. The energy of the moving body transformed into a temperature is known as the heat barrier. The temperature rise is directly proportional to the square of the supersonic velocity. As the speed of sound is doubled, the ram temperature is increased four times. The standard temperature at sea level is considered to be 59° Fahrenheit. This temperature decreases with altitude up to about 46,500 feet. At sea level ram temperature is about 88° F at 760 miles per hour, 29° hotter than the standard sea level temperature. Ram temperature is about 260° F at 1300 miles per hour, and about 1000° F at 2600 miles per hour. As flying speeds increase, materials which have high strength at high temperatures are used.

SPEED CLASSIFICATION. So far, mention has been made of two general terms when relating speeds to sound: subsonic and supersonic. Speed categories can be classified more specifically as subsonic, transonic, supersonic, and hypersonic. Speeds are called *subsonic* when a body moves through air at a speed less than the speed at which the pressure wave is building up. That is, air in front of the body has a forewarning of the approaching body and is able to get out of the way.

When a point is reached at which the speed of the body is the same as that of the pressure wave, waves build up into a single pulse of greater magnitude. This condition exists when the body moves at the speed of sound; the body is then at *sonic* speed. This speed is where *transonic* speeds also occur; that is, when a body is moving at sonic speed, there are points at which air velocities about the body are above the speed of sound and there are points at which air velocities are below the speed of sound. Considerable turbulence and buffeting of missiles take place in this region; therefore, it is desirable to pass through the transonic speeds rapidly to prevent the unwanted disturbances.

Supersonic speed is present when the airflow at all points about the body is greater than the speed of sound. In supersonic flow, little turbulence is present.

The fourth classification of speed is hypersonic, sometimes called ultrasonic flow. As a body moves through the air at high speeds, a short amount of time is necessary for the molecules of air to adjust themselves to the presence of the body and to readjust themselves after the body has passed. This period of adjustment and readjustment is termed the relaxation time. If a body is moving at a speed greater than the relaxation time, it is in a new velocity range which is called *hypersonic*.

MACH ANGLE. To explain what is meant by Mach angle, picture a boat on a lake. If the boat is at rest and is rocked by a wind, ripples will radiate from the boat in all directions. The speed at which these ripples travel is called the propagation rate. Assume that these ripples were propo-

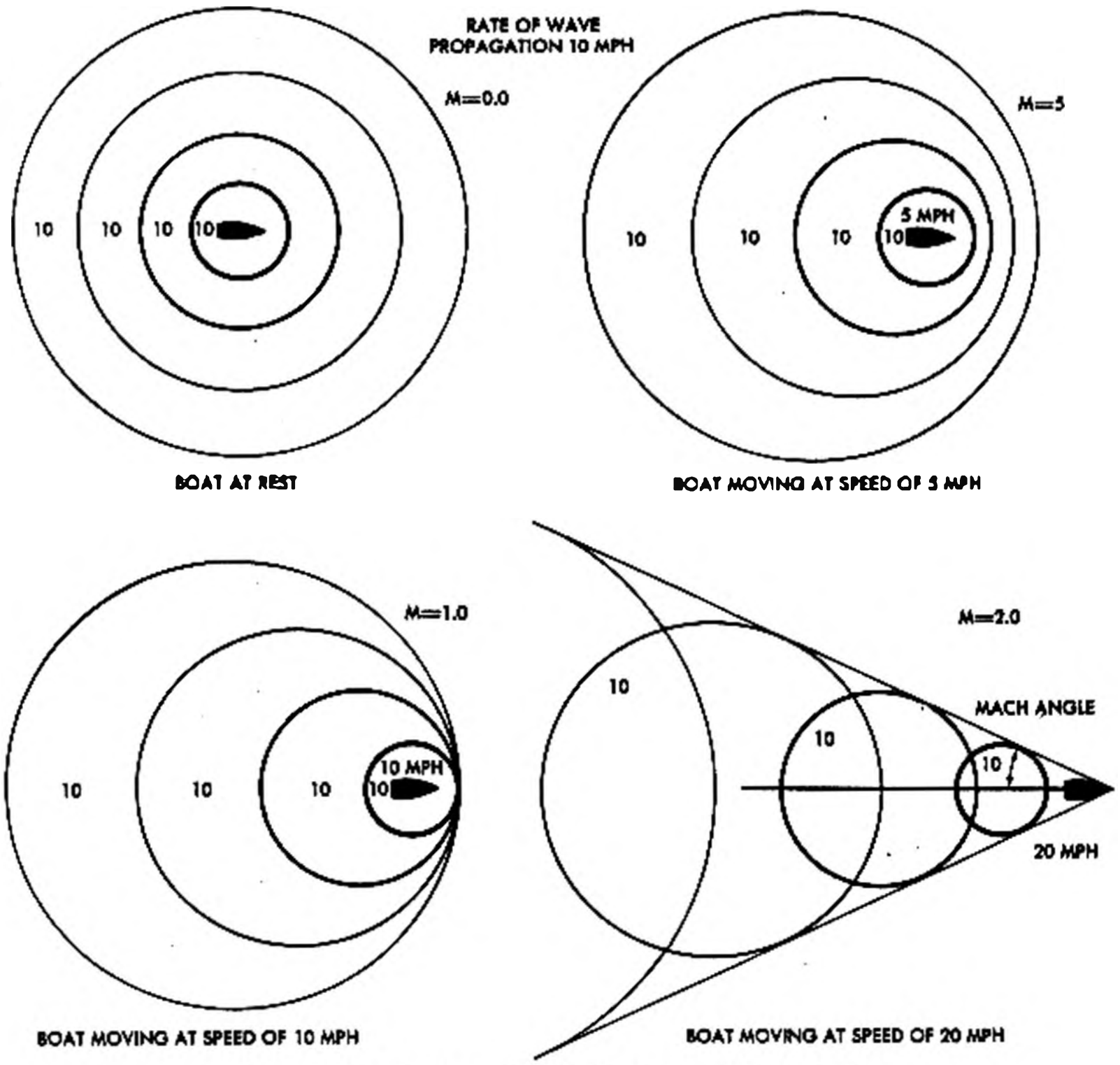


Figure 2-19. Analogy of Mach Numbers and Mach Angle

gated at a speed of 10 mph. The ripples would be concentric, and each would remain inside the previous one.

Now if the boat moves forward with a speed of 5 mph, the ripples will still radiate at 10 mph. Each circle will stay inside of the previous one, but they will not be concentric. The ratio between the boat speed and the propagated wave speed is 5/10. If the propagation rate of the ripples from the boat is thought of as being the speed of sound, the above relationship of 5/10 corresponds to Mach 0.5. If the boat were to increase its speed to 10 mph, then the relationship corresponding to the speed of sound would be Mach 1.0. Now assume that the speed of the boat is increased to 20 mph. This relationship corresponds to Mach 2.0. The ripples still form at 10 mph, but the center of disturbance is moving twice as fast. The wave pattern now becomes a wedge on the surface of the water. In the air, with 3-dimensional flow, the pattern would be a cone. The semivertex angle is the Mach angle. This angle is illustrated in figure 2-19.

NORMAL SHOCK WAVE. Pictured in figure 2-20 is a normal shock wave. The airfoil section has a speed (the free-stream velocity) of 675 miles per hour. The figure indicates the speed of sound to be 761 miles per hour. Dividing 675 by 761 gives

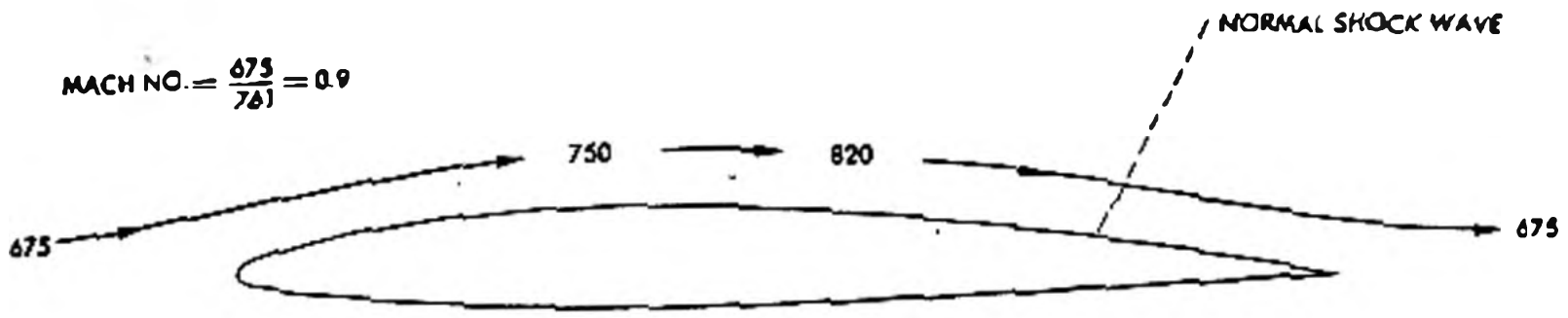


Figure 2-20. Normal Shock Wave

a Mach number of 0.9. As the air gains speed over the airfoil, it increases to 750 mph and then to 820 mph, which is supersonic. The air then begins to slow down to its original speed of 675 mph. Now if you will recall, Bernoulli's theorem says that if the velocity of air changes, its pressure changes. Since the velocity has changed rather suddenly from 820 mph to 675 mph, you can expect a large pressure difference to exist. The boundry between the two areas of pressure difference is where the normal shock wave occurs. Normal shock waves absorb a large amount of energy which is dissipated in the form of heat.

OBLIQUE SHOCK WAVE. In supersonic flight, an additional type of wave called an oblique shock wave occurs. It has this name because it meets the free stream at an oblique angle, an angle greater than 90°. An oblique shock wave is shown in figure 2-21. The component of flow perpendicular to the wave must go from supersonic to subsonic velocity just as in the case of the normal shock wave. If you assume these values to be 900 mph to 600 mph, the free-stream velocity may be 1200 mph and the component parallel to the shock wave may be 794 mph. The resultant velocity (say 950 mph) behind the wave is equal to the vector resultant of the 600 and 794-mph components. This resultant velocity may or may not be supersonic. The original velocity of the free stream and the oblique angle of the shock wave are the two factors that determine the final velocity.

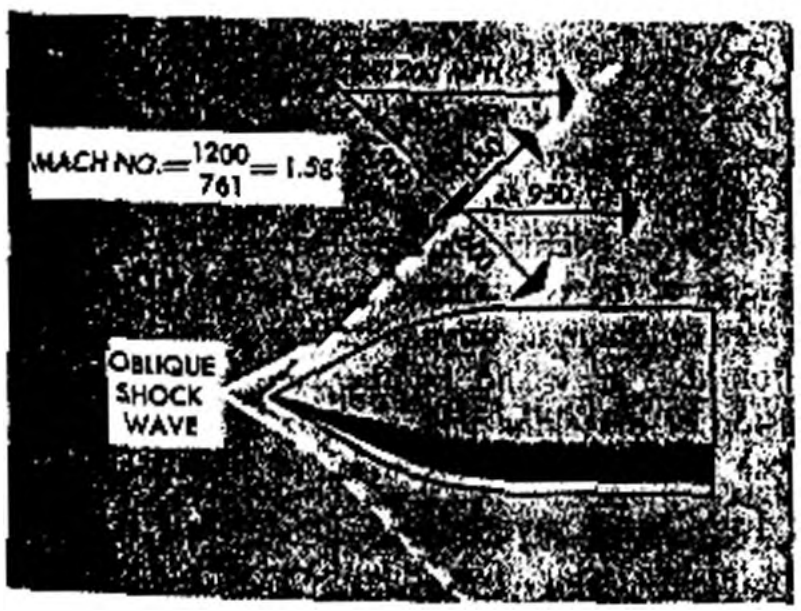


Figure 2-21. Oblique Shock Wave

Figure 2-22 illustrates the effects of speed on air. Note the changes in airflow and shock waves as the wing accelerates from subsonic to supersonic velocity.

Now that you have been introduced to the meaning of several terms that pertain to high-speed flight, you are ready to take up the problems of supersonic flight.

Guided Missile Configuration

The configuration of a missile is a primary concern of missile designers, because a missile's configuration largely determines the extent of drag and lift acting on the missile. These two forces in turn largely determine the overall efficiency of the missile.

DRAG REDUCTION. The achievement of low drag in supersonic flight is of great importance. With low drag configuration, smaller missile powerplants can be used and less fuel capacity is necessary.

The drag of a missile is made up of fuselage drag, wing and fin drag, and drag developed by the interference of the various drags. The effects of thickness distribution, Reynolds number, surface imperfection, and Mach number all influence the drag. Wing drags also are greatly influenced by thickness ratio, sweepback, aspect ratio, and section of airfoil.

Contrary to subsonic experience, the total drag of a supersonic missile is not necessarily the sum of the separately measured drags of its elements.

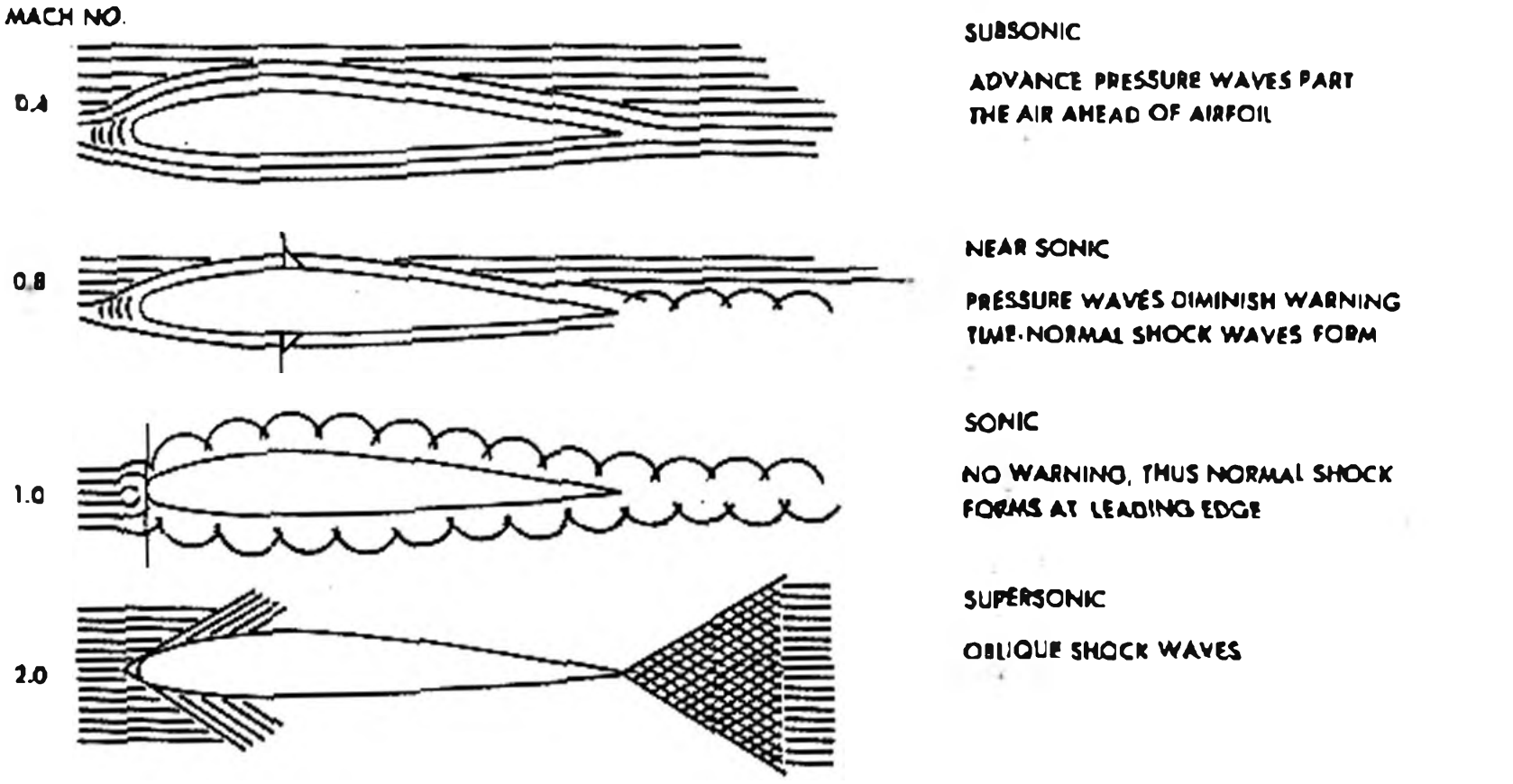


Figure 2-22. Effects of Speed on Air

For example, the drag of a wing seems to be determined by the body shape on which it is mounted.

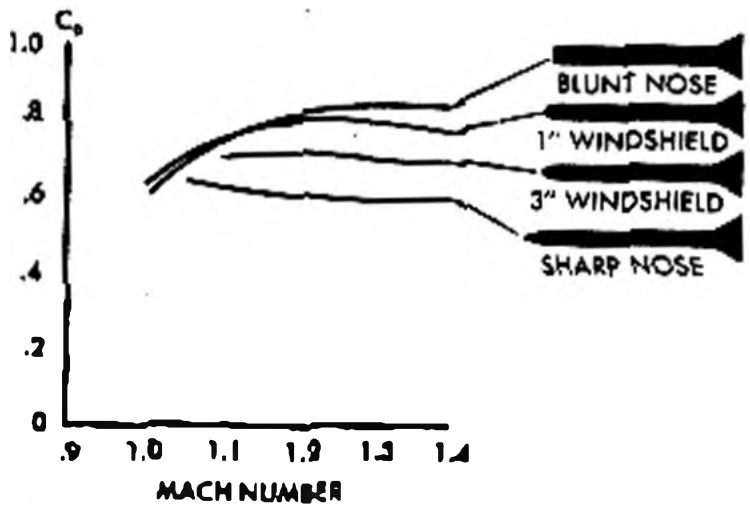


Figure 2-23. Effects of Conical Windshields on Body Drag

A simple illustration of an interference effect in supersonic airflow is shown in the graph in figure 2-23. The curves were obtained with free-flight tests of rocket-powered models.

The drag of a blunt-nosed body is shown by the upper curve. Adding the conical point, which in itself has a positive drag, greatly reduced the drag of the combination of conical point and body. Apart from overcoming interference effects, windshields of the type shown may be useful in cases where optimum aerodynamic shapes must be compromised, as would be the case when forward vision of a seeking unit requires a hemispherical nose shape to avoid distortion of signals.

LIFT EFFECTIVENESS. A lift force is required for maneuvering. For a long-range winged missile, the lift must provide the required support at minimum drag. Lift must also vary smoothly with

angle of attack if control response is to be satisfactory. The lift behavior of three different wing designs is shown on the graph in figure 2-24.

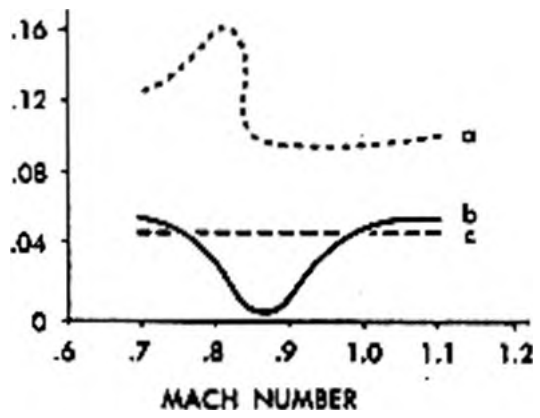


Figure 2-24. Lift Effectiveness

The upper curve (a) shows the behavior of a high aspect ratio, 10 percent-thick wing with no sweepback. In this case, an abrupt reduction of lift effectiveness occurs slightly above Mach 0.8. However, definite positive lift effectiveness is retained through the speed of sound into the supersonic range. The second curve (b) shows typical behavior of a relatively thick wing at high speeds. In this case, a complete loss of lift effectiveness occurs at high subsonic Mach numbers. The third curve (c) shows the effect of sweepback in delaying and minimizing the critical changes produced in the transonic range. The sweptback wing of (c) has the same section as the high aspect ratio wing of (a). With the sweptback wing, there is very little variation in the lift behavior over the speed range tested.

AIRFLOW AT DIFFERENT SPEEDS. The conditions of flight associated with subsonic flow are well known. Supersonic conditions appear to be orderly and, consequently, of such a nature as to be readily analyzed mathematically. But in the transonic speed range, major design problems arise.

The drawings in figure 2-25 show the airflow pattern, along with lift and drag changes, as a circular arc wing section moves from subsonic to supersonic speed.

Figure 2-25, top figure, shows an airfoil in high subsonic flow. The free-stream Mach number is 0.8. This condition is well known and is the kind encountered every day. The lift is much greater than the drag. The second figure indicates a con-

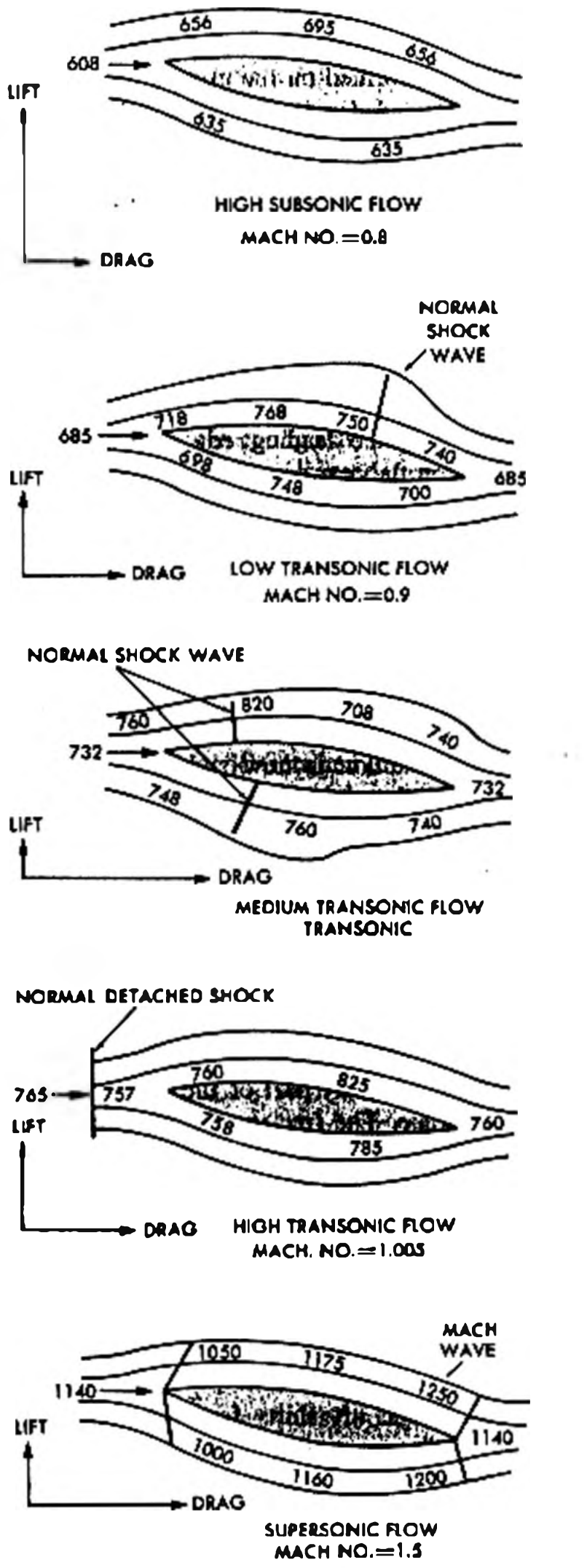


Figure 2-25. Airflow Patterns at Different Speeds

dition of early transonic flow. The free-stream Mach number has increased to 0.9, and a normal shock wave has formed on the upper surface. Lift has decreased, and drag has increased. The third figure is in the middle of the transonic range. The free stream Mach number has now increased to 0.95, and a normal shock wave has formed on both the upper and lower surfaces. Lift has decreased even more, and the drag has increased sharply.

The fourth figure shows late transonic flow. The free-stream Mach number is over 1.0. A weak normal shock wave has just started to form a few inches ahead of the leading edge. The normal shock waves on the rear upper and lower surfaces have virtually disappeared, since the flow is now sonic over that portion of the section. Lift is about the same, and the drag has decreased a small amount. The bottom figure shows fully developed supersonic flight conditions. The free-stream Mach number is now 1.5. The normal shock wave has moved back and become a Mach wave, and the usual Mach wave has formed on the trailing edge. The airflow condition is now stable. The lift is equal to the weight of the missile. The drag is high, but this is caused by the high speed and not by energy losses in the shock waves as was the case in transonic flow.

WING DESIGN. Airflow over an ideal wing would be subsonic until a velocity of Mach 1 is reached, and then it immediately would become supersonic. In other words, the transonic range would be eliminated. Actually the transonic region begins when the flow over any part of the wing first becomes supersonic. The free-stream Mach number at which transonic flow begins is called the critical Mach number for the wing.

Figure 2-26 illustrates the evolution of a high-speed airfoil section. Pictured first is a typical low-subsonic airfoil with a critical Mach number of 0.7 at sea level. Sonic speed occurs first near the leading edge. When the normal shock wave forms, it decreases the lift over most of the upper surface. The center of the lift shifts forward, and drag increases rapidly.

The next sketch is a typical high sub-sonic airfoil with a critical Mach 0.85. The airflow does not reach its maximum value until it reaches a point far back on the wing. When the normal shock wave does form, the loss in lift and the increase in drag are much less than with the first airfoil.

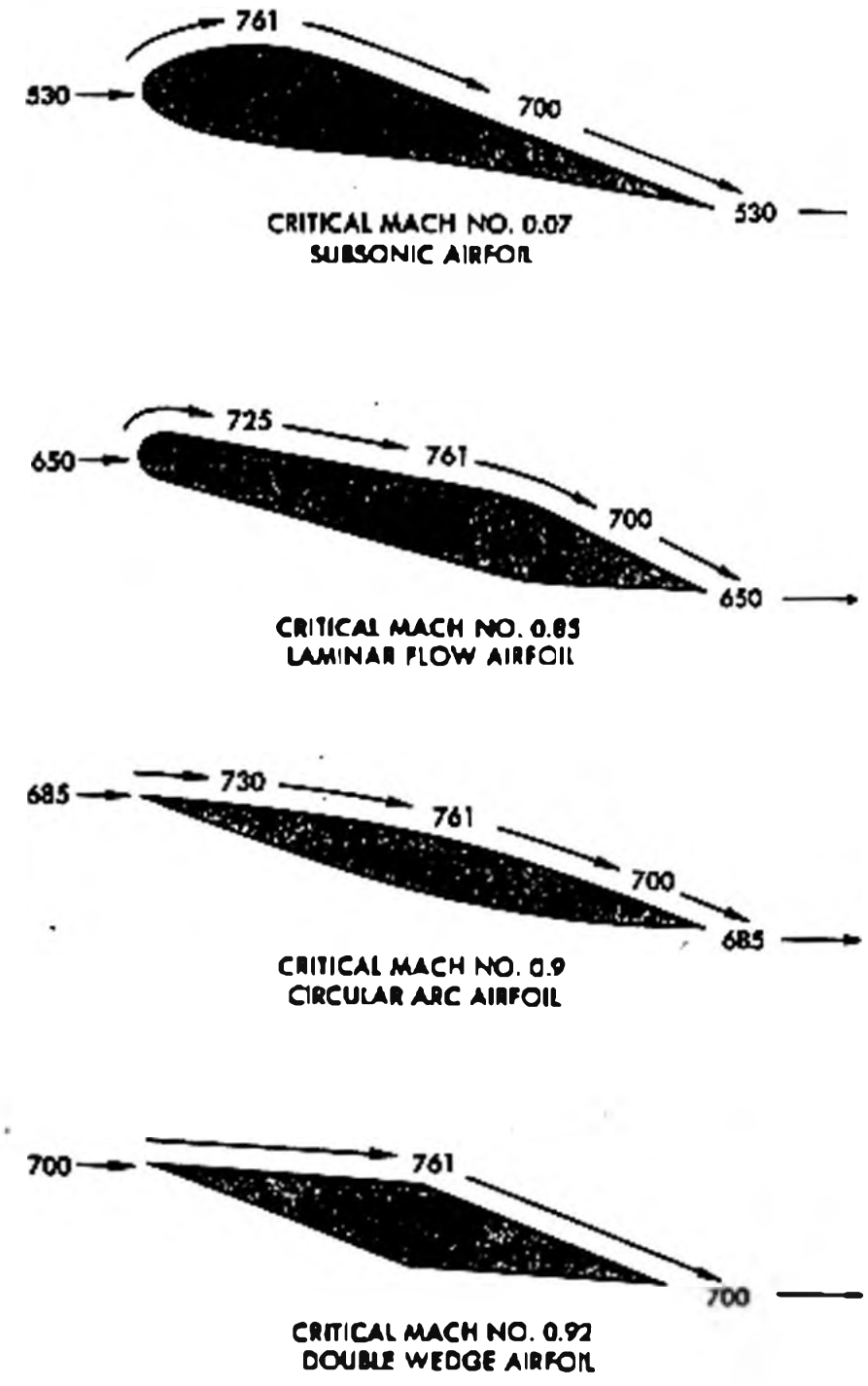


Figure 2-26. Evolution of High-Speed Airfoil Sections

The third sketch is a low-supersonic-type airfoil. The critical Mach number may be as high as 0.9. This is known as a circular-arc airfoil. The acceleration of air over the surface is gradual. Because of its symmetrical shape, the center of pressure on the airfoil changes very little when the lift decreases. This characteristic greatly simplifies the control problem.

The fourth type is known as a wedge airfoil. It also is a supersonic section having a critical Mach number of 0.9 or better. It can be shown mathe-

matically that the ideal supersonic airfoil is a flat plate; however, such an airfoil would not have sufficient strength to support the load. This wedge then is a modified flat plate designed to give more strength than a flat plate.

The low-speed characteristics of the wedge and circular-arc airfoil are poor. These four typical airfoil sections should be sufficient to give an idea of the trends in design; thin symmetrical sections are desired.

ARRANGEMENT OF AIRFOILS. Arrangement of airfoils on a missile are governed by many factors,

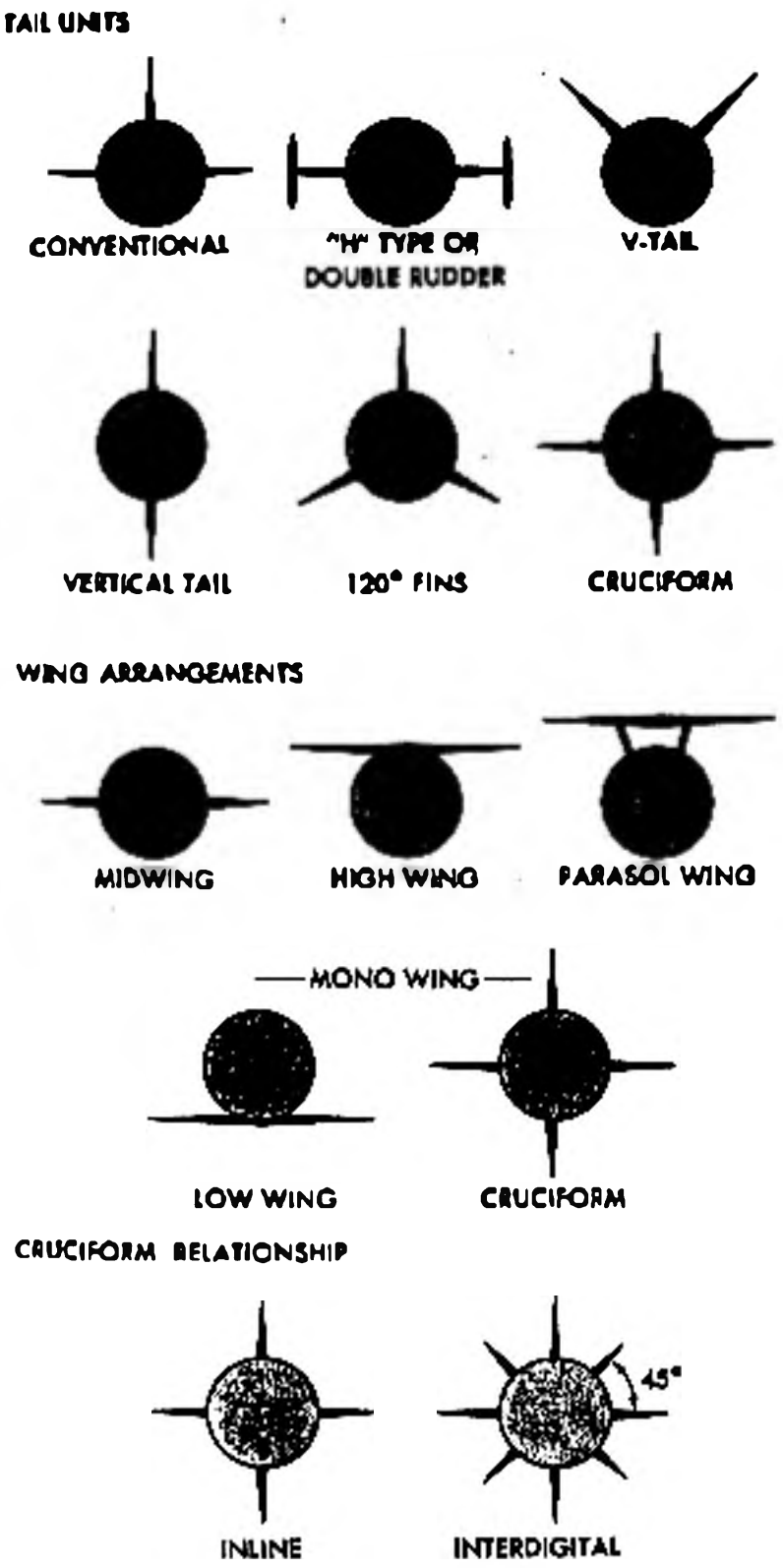


Figure 2-27. Common Arrangements of Airfoils

such as speed, range, launching period, and whether or not recovery of the missile is desired. Various combinations of the types shown in figure 2-27 may be used. The sketches illustrate only the most common type of arrangements. Figure 2-28 shows some of these arrangements on missiles. Note also, the various plan forms that are illustrated.

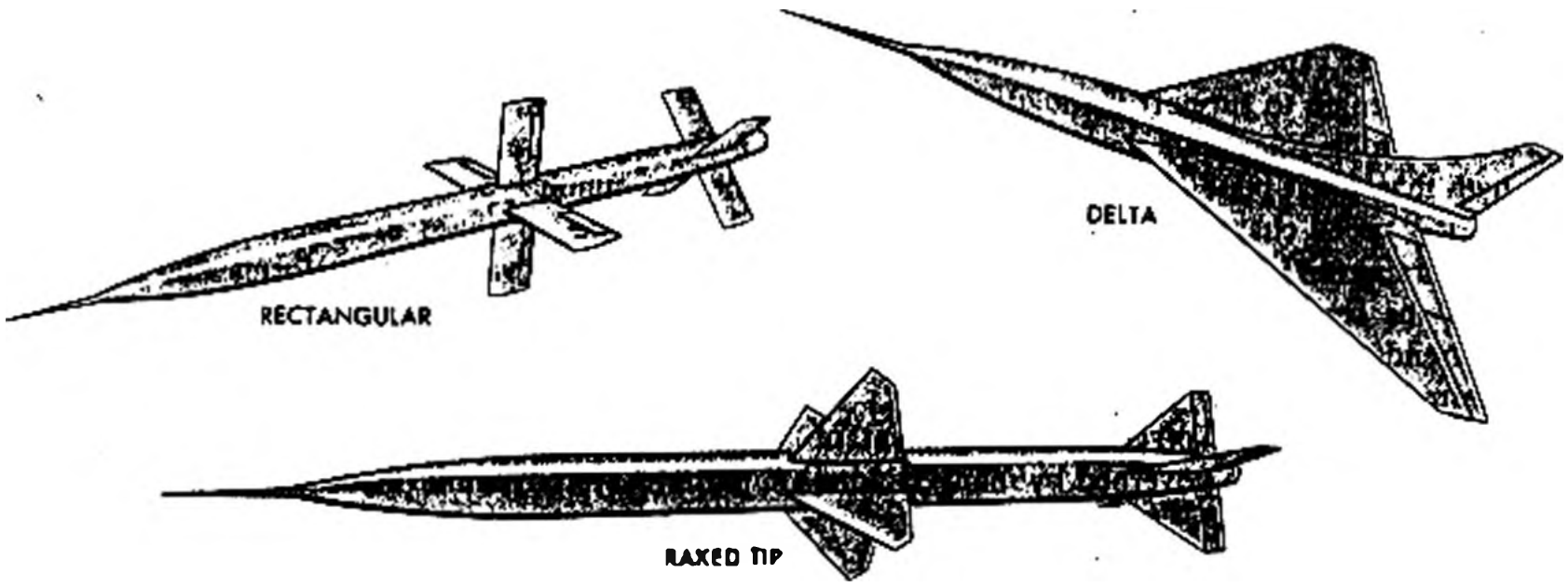
AERODYNAMIC LOADS. One of the problems of missile designing is aerodynamic loading. Such loading results from any change in the airflow pattern. Aerodynamic loads are in general compensated for by control-surface positioning, sweepback, and fuselage configuration. Auxiliary equipment such as antennas, airspeed measuring devices, and various telemetering devices all add to the aerodynamic loads. All of these obstacles to an even airflow are being eliminated as new devices are made available to accomplish the same purpose.

In the transonic range, still more data is needed to fully determine the general behavior of lift and drag on a missile and the basic laws governing this behavior. Systematic tests are still needed over the supersonic range to determine fully the general effects of various wing sections, airflow interference, and the combined-surfaces interference. The design of missiles for supersonic flight has come a long way, and we can well expect continued improvement of guided missile design.

Aerodynamic Control

Aerodynamic control is the connecting link between the guidance system and the missile flightpath. This point is where much regard is given to a smooth and exact operation of the control surfaces of the missile. The control must be powerful enough to produce the necessary change of direction. At the same time, the control must have the best design configuration for the intended speed. Methods must be found for balancing controls and varying the control surfaces to meet the variations of lift and drag at different Mach numbers. The operating mechanism (or servo systems) must be made to operate readily to prevent instability in the missile.

EXTERNAL CONTROL SURFACES. Fixed guide fins are the simplest type of control system for stability. The flight of a common arrow is an example of this type of stability, since the feathered



Figures 2-28. Airfoil Plan Forms

fins at the rear of an arrow provide for a stable line of flight. This same principle of obtaining stability is used in missile design. Fixed guide fins are used in one way or another on most missiles. These fins are generally referred to as stabilizers of a specific type, such as horizontal stabilizer or vertical stabilizer.

In addition to fixed control surfaces, there are movable control surfaces referred to as flaps or movable vanes.

Keep in mind that a control surface is not effective until the airflow across the surface has attained sufficient speed to develop a force. As the speed is increased, the reaction to the control surfaces becomes more acute and often results in overcontrol.

The disadvantage of using guide fins without movable control lies in the lack of the precise control necessary to follow a given course. To accurately control a missile, two general types of

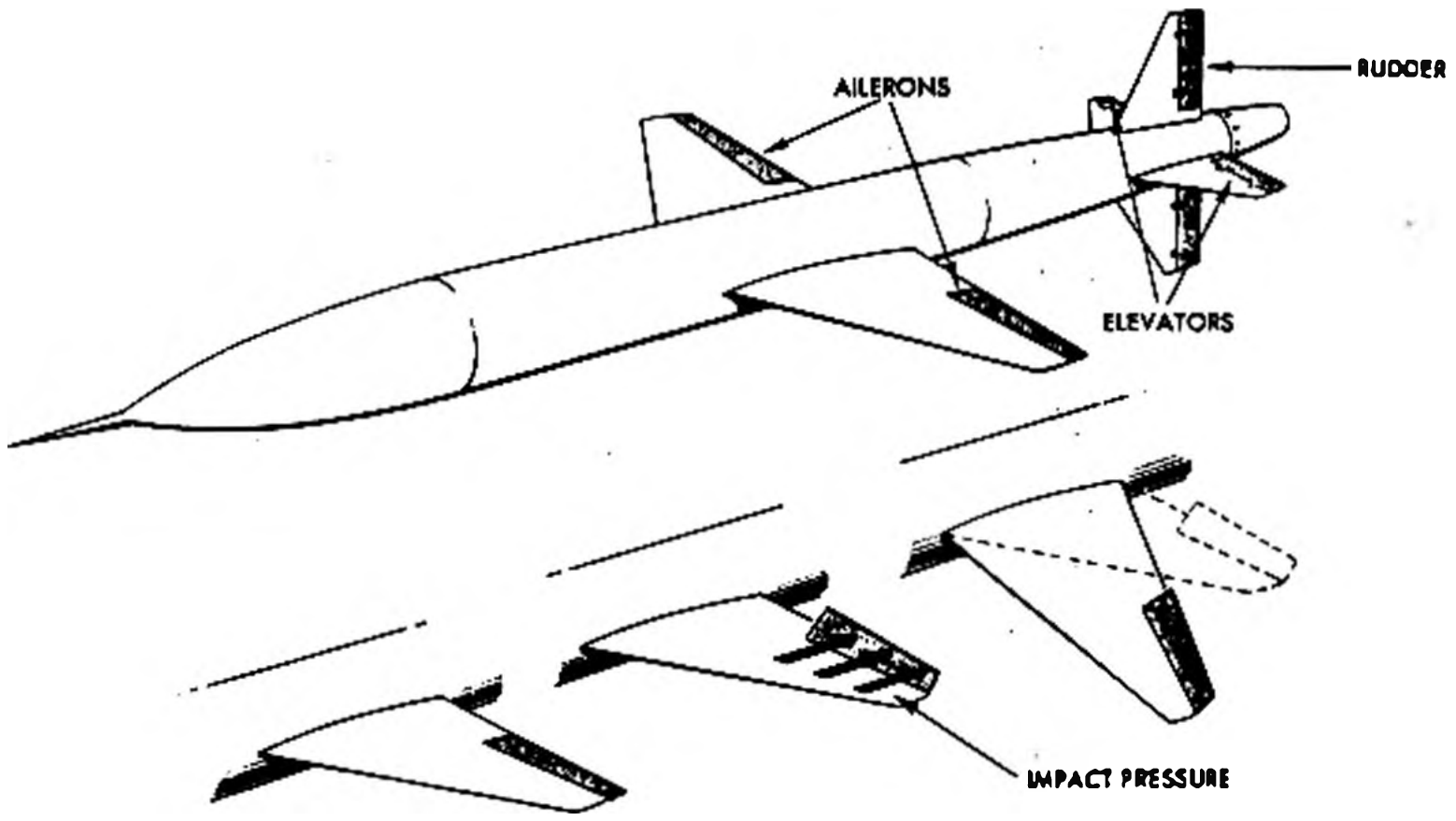


Figure 2-29. Primary Control Surfaces

control surfaces are used. These types are primary and secondary controls. Primary control can be looked upon as the main controlling factor of the missile's path. This group alone could, under certain conditions, give satisfactory results. If there were no unstabilizing conditions present, primary control could function satisfactorily unaided. However, by the use of secondary controls of various combinations which are discussed later, a missile can be controlled much more accurately and efficiently.

Primary controls. Ailerons, elevators, and rudders are considered primary controls. A conventional aileron is attached to the outer trailing edge of the wings or main lifting surfaces. This conventional arrangement appears in figure 2-29. When one aileron is lowered, the opposite one is raised, thus controlling roll. They are coupled to the governing control system. Various control systems are taken up in later chapters.

Elevators are attached to the horizontal stabilizer on the tail. They are used for pitch control and are raised and lowered together.

Finally, a rudder is used to maintain directional control. It is attached to the vertical stabilizer and gives yaw control.

Figure 2-30 shows that the rudder, ailerons, and elevators control movement about the yaw, roll, and pitch axes, respectively.

The previously mentioned control surfaces attain control by presenting a surface to the existing airflow at an angle which causes a force to exist. This force pushing against the control surface moves the wing or tail to which the control surface is attached in a direction opposite to the control-surface movement. Control surfaces along with the direction of movement resulting from the positioning of these surfaces are shown in figure 2-31.

The wing, horizontal stabilizer, and vertical stabilizer can be considered a part of the primary control surfaces since they are the leading factors in determining the missile's path. But for presentation of controls, we will regard the elements that tend to change the direction of travel as control surfaces.

Secondary control surfaces. In the secondary group of controls are tabs, spoilers, and slots.

Tabs can be divided into three types: fixed, trim, and booster. Their reaction in the relative airstream, however, is the same in all three cases. They always affect the control of a missile indirectly. They do not in themselves determine the direction of motion of the missile. For example, suppose you want to trim in pitch. To raise the nose of the missile, the tab must be moved down. The primary control surface, in this case the elevator, is hinged to the horizontal stabilizer. The tab, in turn is hinged to the elevator. A small movement of the tab on the trailing edge of the elevator causes a small force to be exerted on the primary control surface. The result of this force is a small movement of the elevator in the opposite direction; therefore, if the tab is moved down, the elevator moves up. Since the missile responds only to the primary control action, the tail is lowered, thus raising the nose of the missile. Note figure 2-31. This function is exactly the same for directional trim (yaw) and for lateral trim (roll).

A fixed tab consists of a piece of metal attached permanently to the trailing edge of the primary control. The metal is bent uniformly in the required direction to trim the missile. The trimming is done in anticipation of a certain set of unbalanced conditions around the center of gravity at a given airspeed.

A trim tab is more complex in construction than a fixed tab, and is controllable. Trim can be changed as attitude, speed, or altitude vary.

A booster tab, sometimes known as a servo tab, is widely used where large primary control surfaces are used. The force and stress on the control system are greatly reduced by using the booster tab. This type of tab is directly connected to the control system. When the system is actuated, the tab moves first. The tab movement, in turn, moves the primary control in the normal manner. The effect of the movement of the tab on the primary surface is considered instantaneous. The location of tabs on a missile is shown in figure 2-32.

To summarize, a fixed tab is preset for a given condition of stability. A trim tab is controllable, and its setting can be changed at will over a wide range of conditions. A booster tab is used to assist in moving control surfaces of large areas. Since the performance and design of these controls are well understood, they probably will be used on missiles in one form or another for sometime to come. Even so, in the case of supersonic missiles, new systems of aerodynamic control have been developed.

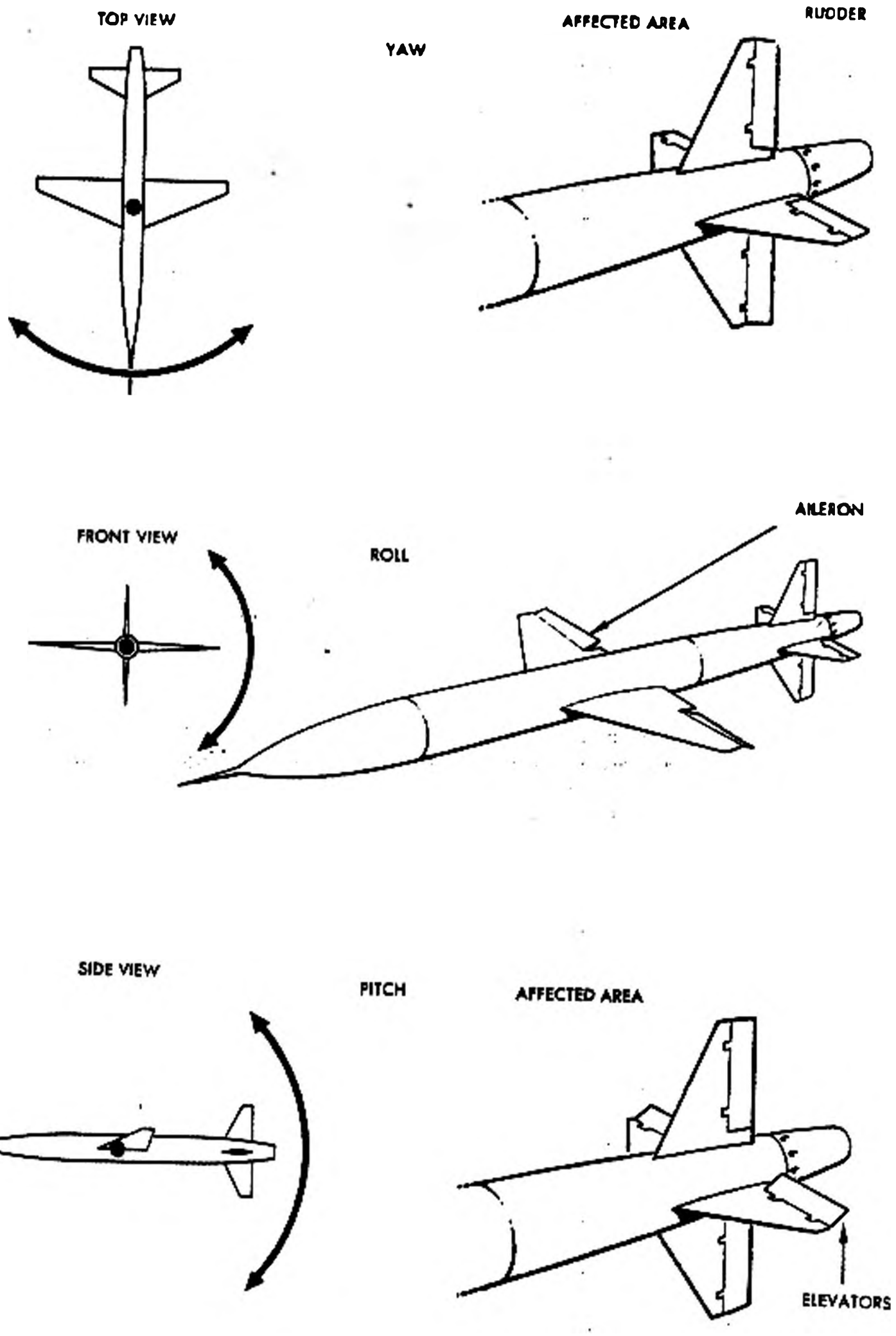


Figure 2-30. Functions of Primary Control Surfaces

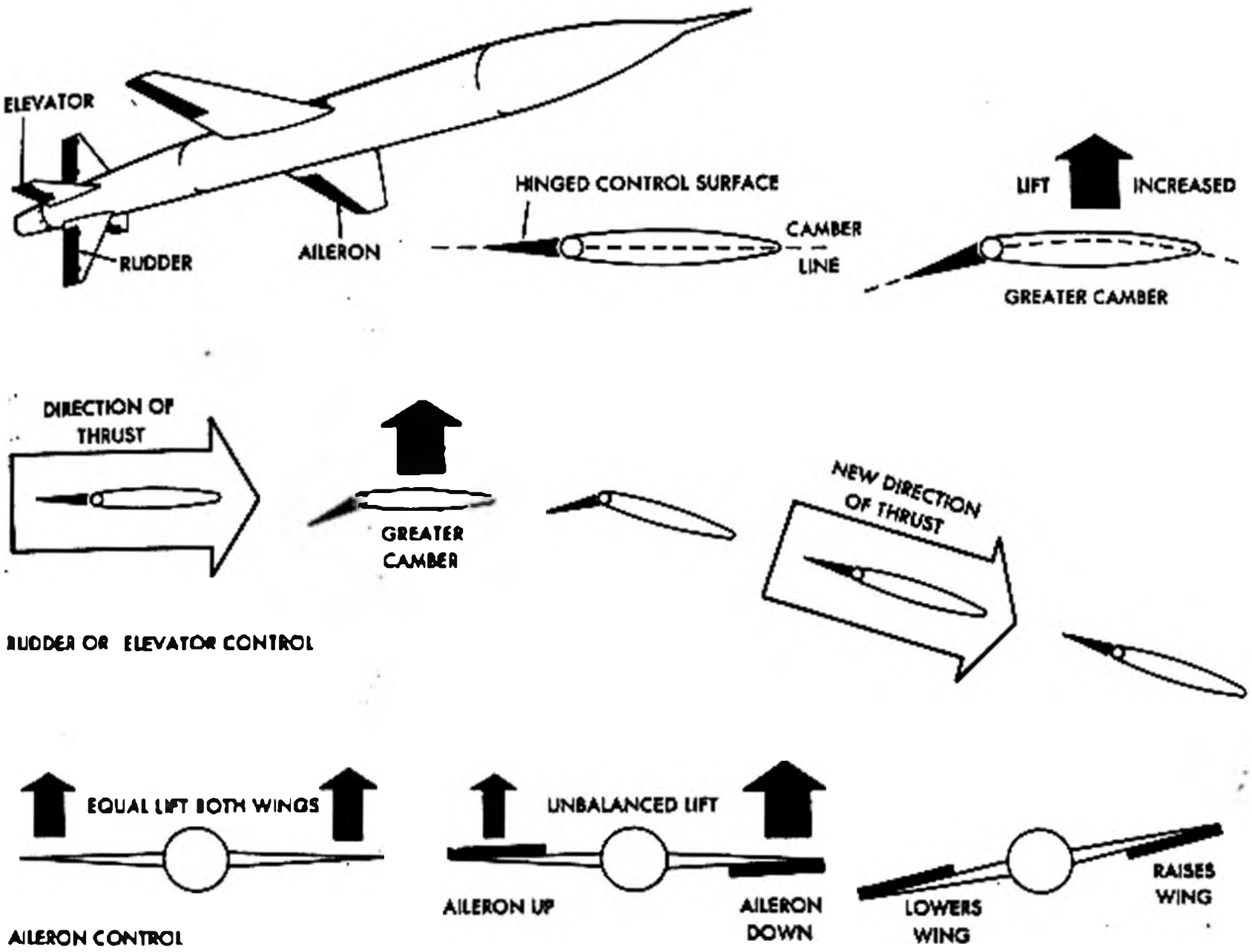


Figure 2-31. Control of Flight

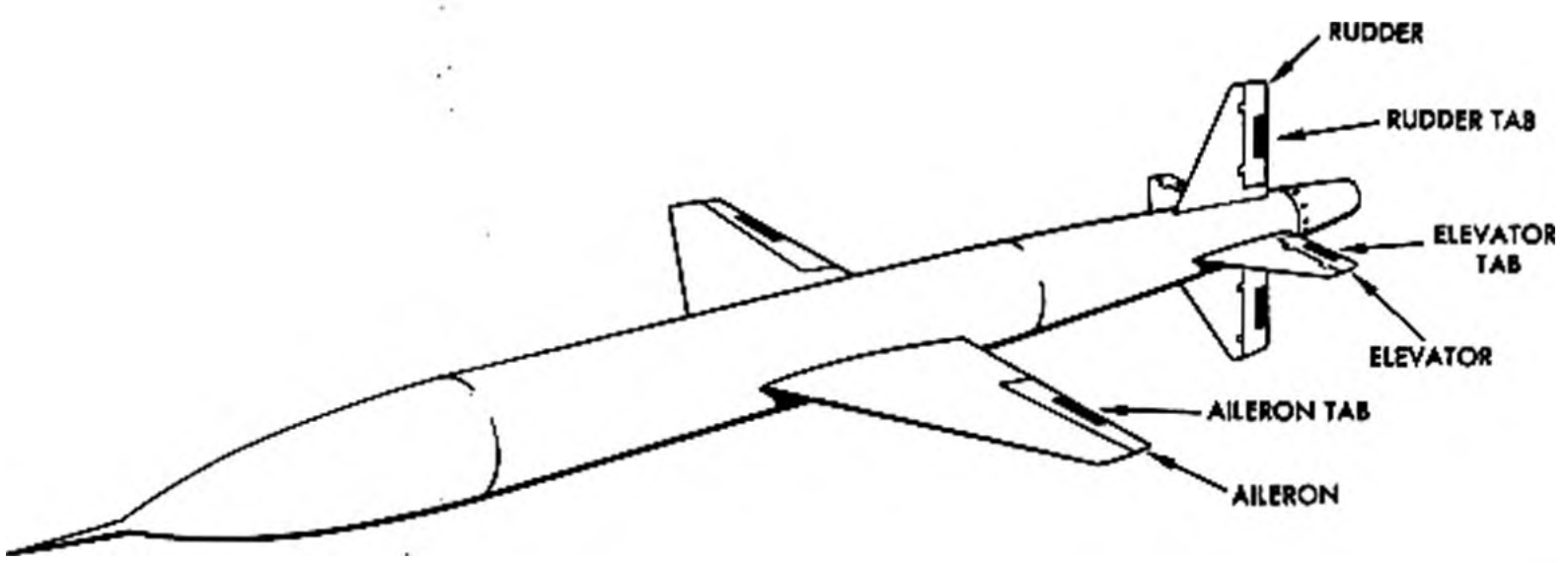


Figure 2-32. Location of Tabs

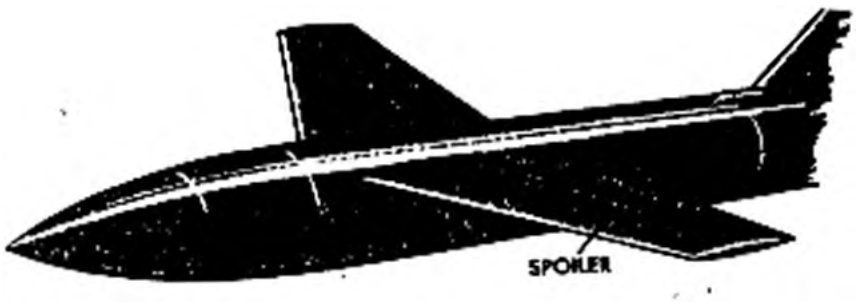


Figure 2-33. Location of Spoilers on Wing Surface

As aircraft speeds were increased closer to the speed of sound, the performance of surface controls became more and more critical. A surface control is nothing more than a hinged flap, or airfoil, attached to a larger airfoil which may be, as previously stated, a wing or a stabilizer. Engineers soon found that, at speeds over 300-mph range, peculiar vibrations called flutter occured on controls, especially in the case of ailerons. These vibrations became more severe as speed increased until in some cases they reached such proportions that surface controls disintegrated and sometimes caused failure of the entire wing or tail structure. This has been overcome by increasing the rigidity and strength of the primary controls and by balancing these surface controls statically and dynamically.

A method of overcoming bending on a wing is by the use of spoilers. They may be of one solid unit or a series of units that tend to interrupt the negative lift on a wing. Spoilers are recessed into the upper camber of the wings.

The spoiler in figure 2-33 consists of a solid hinged flap. When the spoiler is not used, the flow of air over the wing is smooth and uninterrupted; thus the full lifting power of the wing is realized. Assume that a gust of air has caused the left wing to drop. The control system instantly calls for the spoiler on the right wing to extend. As the spoiler extends, the negative lift pattern on the right wing is "spoiled," or reduced a considerable amount, by the turbulence created by the spoiler. The wings then tend to return to the original position. A spoiler may cause more drag than the conventional aileron; therefore, when the spoiler is extended, a yawing moment may be initiated strong enough to cause the rudder to be actuated simultaneously with roll control.

A slot is basically a high-lift device and is located along the leading edge of the wing. In the region of normal angle of attack, the slot is ineffective. But as shown in figure 2-34, when a missile reaches high angles of attack, the slot can be opened to allow air to spill through. A slot performs a function exactly opposite to that of a spoiler. It is a valuable device for increasing lateral stability, for landing high-speed aircraft, and for preventing stalls in slow aircraft.

One type of slot operates automatically with an increased pressure difference resulting from high angles of attack. A disadvantage of this mechanism is that it must be kept locked until needed because maneuvers at high speed calling for large angles of attack would automatically open it, causing serious reduction in performance or even structural failure of the slot.

Dual-purpose controls. Up to this point the controls presented have been of a conventional type. As missile speeds increased, new control surfaces were developed. New types of primary controls are elevons, ruddervators and ailevators.

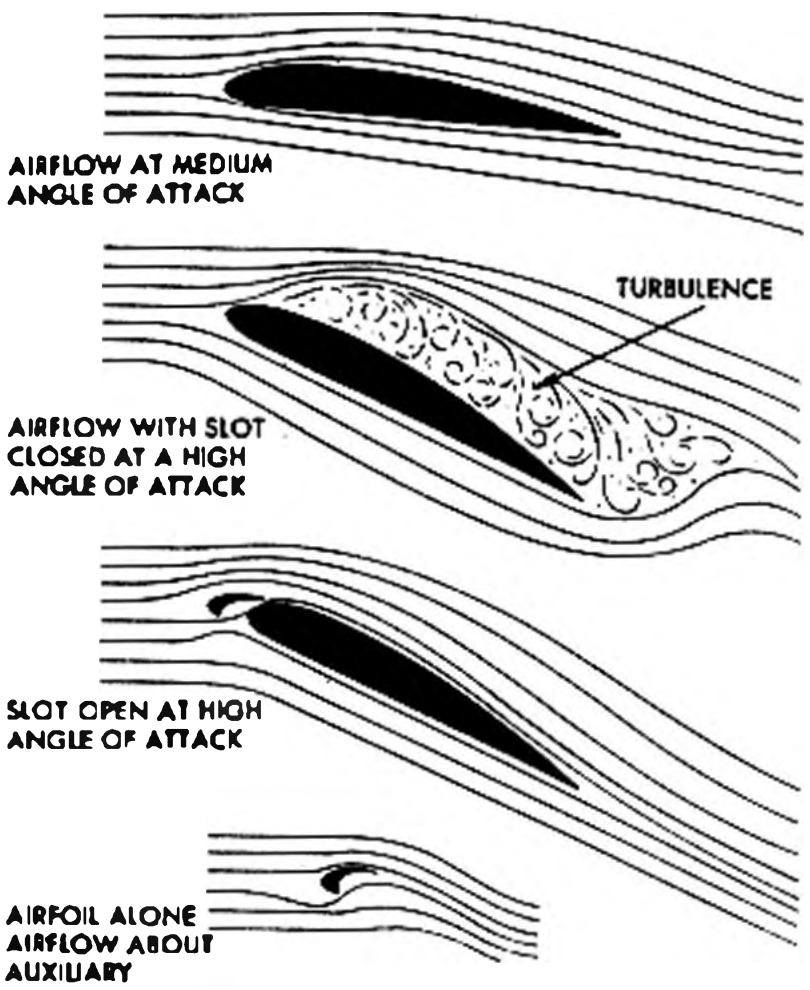


Figure 2-34. Airflow with Slot



Figure 2-35. Dual Purpose Control Surfaces

As the names indicate, they consist of control surfaces that accomplish two purposes. For example, an elevon takes the place of an elevator and an aileron, giving control of pitch and roll. An ailevator is the same as an elevon. Ruddervators are used for yaw and pitch control. Observe figure 2-35.

Variable incidence control. Use is sometimes made of a variable incidence control to overcome the problem of flutter and the need for structural strength of control surfaces and yet have a control that is sensitive and effective at various speed ranges. This type of control varies the positioning of an entire airfoil section rather than just part of it. This method eliminates the disturbing factors that are set up between two surfaces, such as the surfaces of a slot and a wing, which make an angle with each other. At high speeds, this type of control surface is especially well adapted because little control movement is required to displace the missile a considerable distance. The variable incidence control can be used on the wing, horizontal stabilizer, or vertical stabilizer.

Canard structure. A canard structure consists of a stabilizing plane and elevator control which is located ahead of the center of gravity, or wing and lifting surfaces. During the early development of the airplane, it was found to be easier from the standpoint of construction to install the tail plane at the aft end of the fuselage. This trend has been followed throughout the years. The canard adaptation of missiles is now being used by some manufacturers. The forward wing may consist of a stabilizing plane with a control surface attached to the trailing edge, or it may consist of a pivoted wing in which case the entire stabilizing plane has

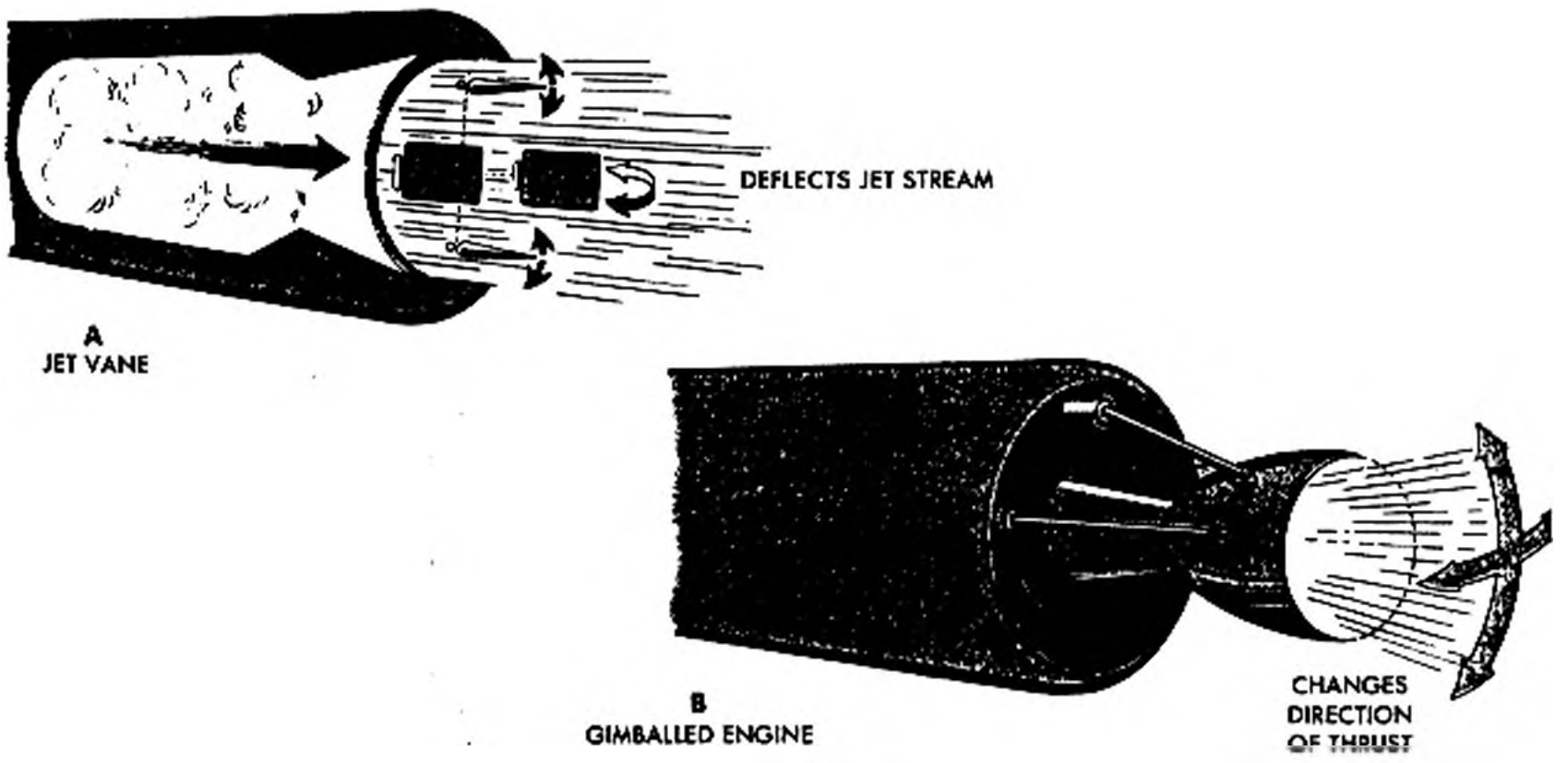


Figure 2-36. Steering Methods

a movable angle of incidence (variable incidence control), either positive or negative. The Nike-Zeus employs a canard type structure.

CONTROL AT STARTING SPEEDS. Until it attains a speed at which the airfoil sections have an aerodynamic stabilizing effect, the missile must have a means of control other than the external control surfaces. This means may be supplied by the use of exhaust vanes or jet control.

Jet vanes. Jet vanes are surfaces which are installed directly in the exhaust path of the jet engine as shown in figure 2-36A. When the positions of the vanes are changed, the exhaust is deflected causing the thrust to be directed in opposition to the jet vanes. The velocities of the exhaust are sufficient to give adequate control until the time the missile has attained enough speed for the external controls to take over. Because of the tremendous heat in the exhaust, the life of jet vanes is generally short.

Jet control. Various systems of jet control have been developed for supersonic flight.

- One means of jet control is changing the position of the jet engine so as to give the desired direction of motion. This method uses the gimbaled engine as shown in figure 2-36B.
- Liquid injection thrust vector control (LITVC) is the method used for control of the second stage of the LGM-30F *Minuteman*. Figure 2-37 is two views of the LGM-30F second stage

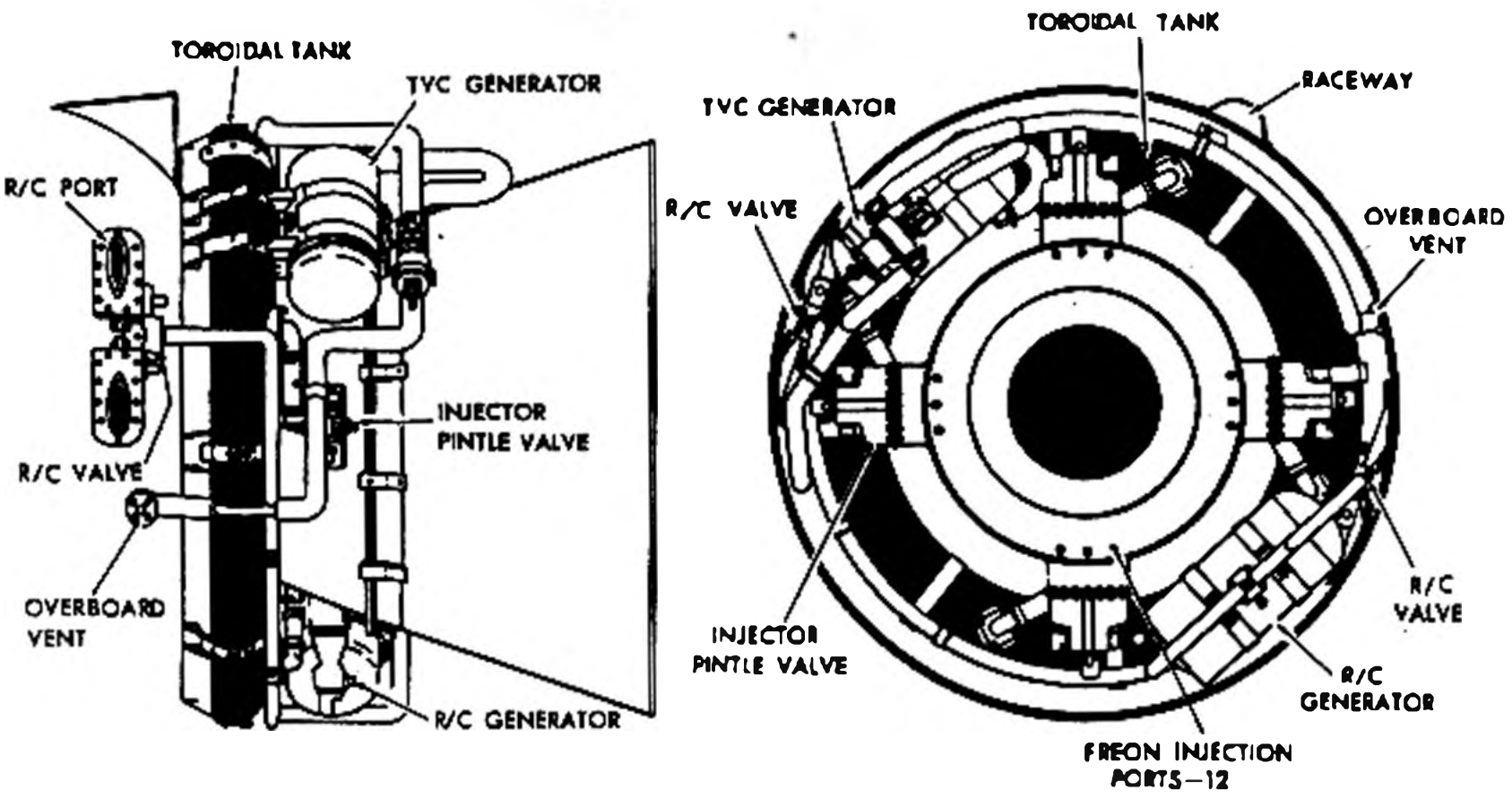


Figure 2-37. Liquid Injection Thrust Vector Control System

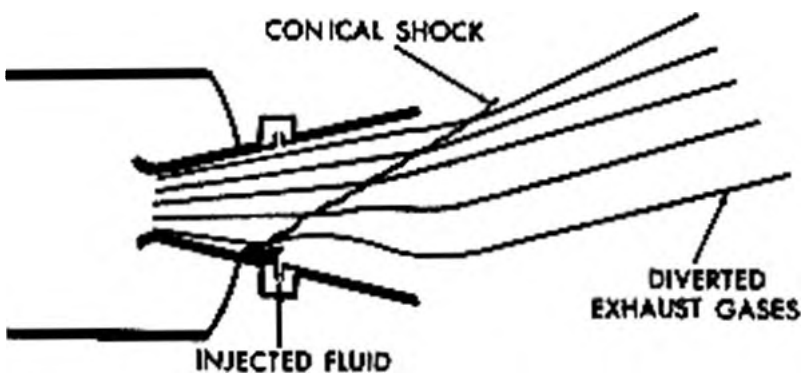


Figure 2-38. Freon Injection

nozzle section. Thrust vector control is accomplished by injecting Freon under pressure through the ports located down stream of the nozzle throat. The Freon is injected at any of four locations, 90° apart, through pintle valves. Injection of the Freon causes a shock wave, illustrated in figure 2-38, which alters the main thrust vector. The deflection of the thrust vector causes a change in the missile's pitch or yaw attitude.

- Another system of jet control is by placing roll control jets at various points about the missile

body. Control is accomplished by using one or another of these jets as necessary to give different directions of thrust. The second stage of the LGM-30F accomplished roll control through the use of two roll control (R/C) valves also illustrated in figure 2-37. Each valve has two opposing exhaust ports. The roll control generator provides gas pressure that is exhausted equally through all four ports during normal flight. If a roll correction is required, the appropriate R/C valves are closed to unbalance the exhausting gas, producing a thrust in the proper direction.

SPECIFIC CONTROL APPLICATIONS TO GUIDED MISSILES. The standard method for control of aircraft around the vertical axis is the rudder located in the tail assembly. Likewise, the control around the lateral axis is performed by the elevators located in the tail assembly. The control of roll around the longitudinal axis is performed by the ailerons attached at the trailing edge of the outer wing panel as near the wing tip as possible. Early cruise type missiles used these conventional type controls. Increased missile speeds made necessary different types of controls. For example, in one configuration the tail assembly is fixed. Directional control is performed by flaps located at the trailing edge of vertical stabilizer surfaces as shown in figure 2-39. When the flaps are moved from neutral position, the chord of the vertical wing is effectively changed to a given angle of attack, depending upon angular displacement of the flaps. Changing the angle of attack creates a differential in pressures on the wing, thus affecting lift. This results in the missile turning about its vertical axis, and a change of direction takes place.

Correction for roll is performed by ailerons located in the wing tips of the vertical wings. Channel guides hold the ailerons at a fixed angle of incidence with respect to the wing panel chord. When the aileron is extended, the positive angle of attack exerts an aerodynamic lift which tends to correct an error in roll.

The change in attitude of the missile around its lateral axis is accomplished by means of flaps on its horizontal wings. When the flaps are lowered, an increase in lift tends to lift the nose of the missile, and conversely, the raising of the flaps lowers the nose.

With the advent of the large ballistic missiles that travel at hypersonic speeds and reach altitudes of hundreds of miles new methods of control were necessary. Aerodynamic controls are ineffective at these extreme speeds and altitudes. Most ballistic missiles now use some form of thrust vector control such as thrust chamber gimbaling, LITV control, or combinations of these as used on the *Minuteman*.

Stability of Missiles

Since stability of missiles has a direct effect on the behavior of the controls, a high degree of stability must be maintained. Increased missile speeds cause definite stability changes by center-of-pressure shifts. A pressure shift causes variation in the flows that are acting on the surfaces of a missile. Even in pure supersonic flow, variations of the Mach number cause center-of-pressure shifts. Referring to figure 2-40A, note that the missile is in level flight. It is longitudinally stable about its lateral axis through the center of gravity.

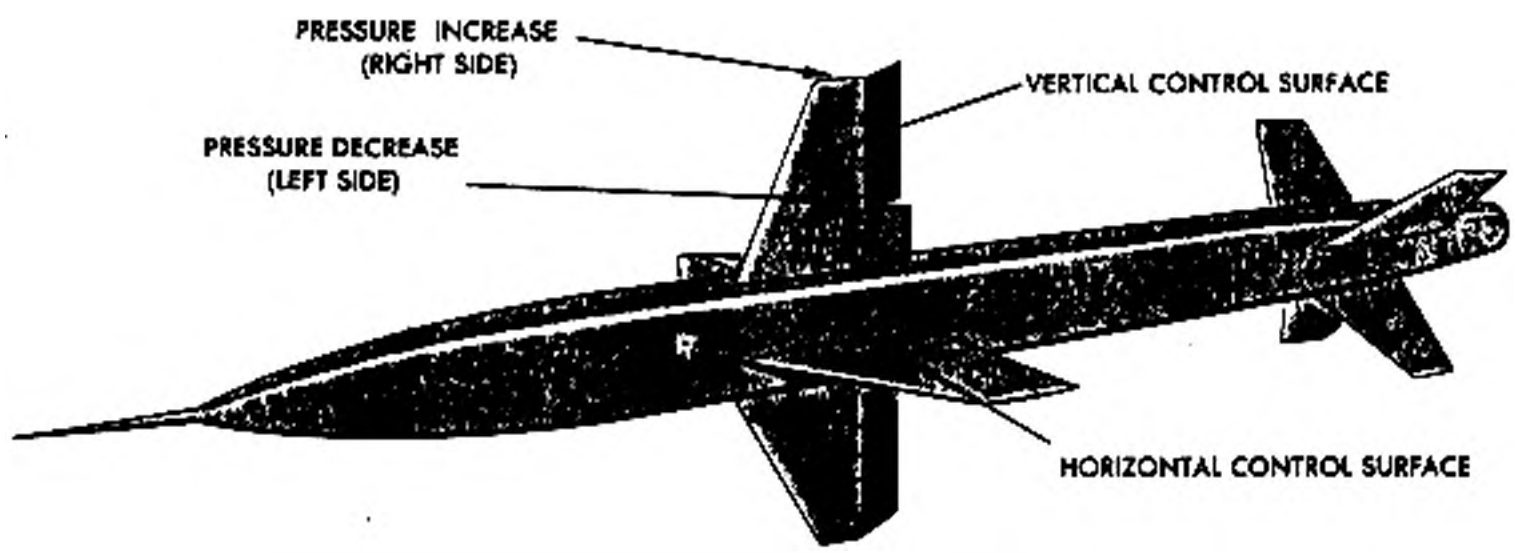


Figure 2-39. Directional Control Using Flaps on Vertical Control Surfaces

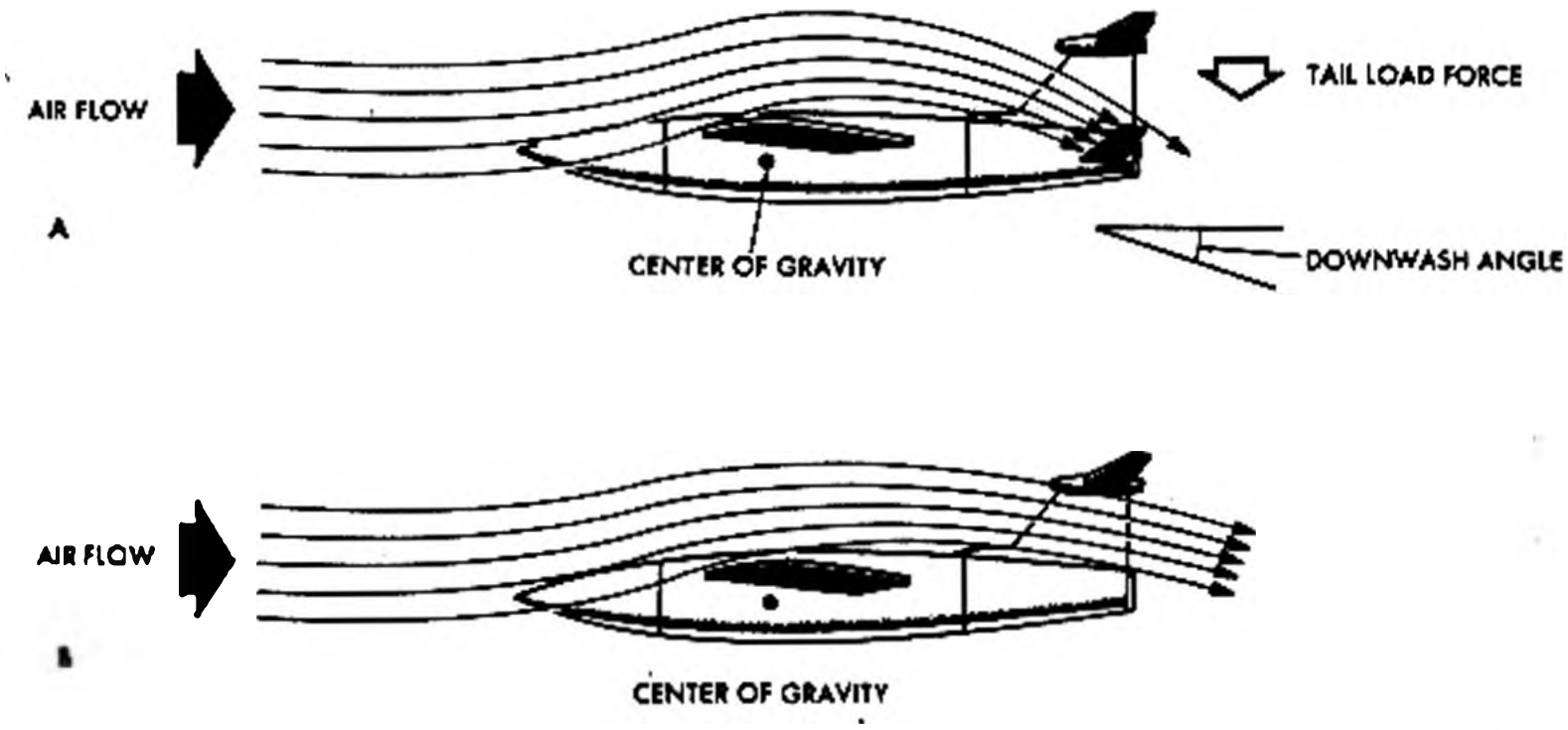


Figure 2-40. Downwash

Airflow over the wing is deflected downward toward the elevator. This angle of deflection is called the downwash angle. When lift decreases as a result of reduced speed, this downwash angle decreases, resulting in pressure changes. There will be points where unstable conditions are set up as a result of such pressure shifts resulting from different airflow patterns. Once an uncontrollable condition is set up, it must be compensated for by the control surfaces or by quickly changing the speed to a point of stability.

As you learned earlier, unstable conditions are dominant at transonic speeds. Most missiles have some dive control and roll recovery devices to overcome unstable conditions. An example of such a device is an arrangement in which horizontal tail surfaces are placed high on the fin to help overcome downwash conditions as illustrated in figure 2-40B.

Ailerons and rudders come in for their share of trouble. The unstable flow over the wings causes the ailerons to oscillate, creating a condition known as "buzz." A similar condition called "snaking" may exist about the directional axis. A part of these troubles may be compensated for by variable incidence control surfaces.

Propulsion Aerodynamics

Propulsion aerodynamics includes such items as the design of low-loss air intakes for ramjets; design of subsonic diffusers, nozzles, and tail pipes; and the effects of jet exhaust on drag and flow about the body of a missile. Aerodynamic-mechanical problems of fuel control, regulation of the proper amount of air intake, and the stability of the internal flow and combustion processes must also be taken into consideration.

BALLISTIC MISSILE AERODYNAMICS

The design of any airframe involves a continual struggle and compromise in satisfying requirements of both weight and strength. For a long-range ballistic missile, which must be power-accelerated all the way to thrust cutoff, excess weight is a critical problem because it results in a drastic reduction of maximum range. Yet at the same time, the airframe must have sufficient strength and rigidity to support the other components and the propellants without failing under normal static and flight loads. Included in these loads are the stresses caused by vibrations of the engines, propellant sloshing, maneuvers, and aerodynamic heating.

A significant reduction in the mass that must be accelerated through the entire distance to thrust cutoff, and also in the propellant weight at takeoff, is obtained by using a multistage design in which one or more engine sections, with or without associated propellant tanks, are jettisoned during the powered flight. Advance design techniques also permit reduction in airframe weight.

The development of new materials may offer opportunities for weight reduction. The designer must consider not only the physical properties of the materials (such as strength-to-weight ratio and corrosion resistance), but also such factors as availability and cost of the materials and their adaptability to standard fabrication techniques.

One effect needing thorough investigation is the shift of the center of gravity of the vehicle as the propellant is consumed. This shift relative to the center of pressure is a determining factor in missile stability. In the ICBM and IRBM, consideration must also be given to the rapid change in the relative positions of the centers of gravity and pressure when a stage with its rocket engines is jettisoned. Begin your considerations by studying the environment that must be withstood by the airframe.

Airframe Environment

Various conditions of temperature, atmospheric pressure, humidity, and degree of water resistance that must be withstood by the airframe are specified by the military specifications. A missile may be stored for long periods of time in the arctic region where the temperature is many degrees below freezing; it may be stored in the tropics where the sun can raise the temperature of a metal body above the boiling point of water. A missile may be transported many ways to the theater of operation; it will be handled roughly and subjected to many shocks in the process. The military services, with many years of experience and background in the handling of weapon equipment, are best able to specify the environmental conditions to which the missile will be subjected prior to its use. This experience is translated into specifications which outline the environmental tests for all service equipment.

The airframe during the launching phase is subjected to severe accelerations. If a booster is employed, there may be severe transient forces imposed upon the airframe at the time of booster separation. The missile maneuvers may exert many gravities of acceleration. The propelling motor may have resonances and thus initiate vibrations within the missile structure. Under certain conditions, missile airfoils may flutter and induce vibrations. The internal structure of the missile may have different resonances in different sections. Such resonances may develop severe internal vibrations without appreciably affecting the behavior of the airframe in flight.

All of these conditions and others must be considered in establishing the environment of the missile airframe. The only adequate means of obtaining knowledge of this environment is by making statistical tests during missile flights. Such tests are usually conducted by the contractor as part of the design of the airframe. The data developed are unique to the specific missile and cannot be applied generally (as in the case of environment which is governed by the theater of operations).

One of the major decisions to be made, during the early preliminary design phase after the general configuration is fixed, relates to the type of basic airframe structure to be used and the manner in which the basic loads will be reacted. What is basic airframe structure?

Basic Airframe Structure

The aerodynamic analysis establishes allowable center of gravity (cg) shift and airloads and pressure distribution from which flight regime, shear, and bending moments, may be derived by conventional means. With this data, the designer is then able to develop an airframe design and to establish the properties for each major structural section. With the structural design roughed out, materials and allowable stresses are selected with consideration to type of load, weight, temperature, redundancy, cost (including evaluation of critical materials), and environment. In the allocation of space to the various systems, the usual practice is to locate the engine and fuel tanks first and to package the accessory equipment in such a manner as to eliminate need for longitudinal or roll balance if at all possible. In making a decision on the type of structure, an evaluation is necessary on the use of longerons and nonload carrying doors or covers versus a pure or semimonocoque structure with the attendant problem of local cutouts or load carrying access doors.

There is no way to predict the lightest structure for a missile; in fact, acceptance of less than the optimum weight design is usually necessary to obtain other desirable features. Using different types of basic structure for different sections of the missile to obtain certain design or packaging advantages is frequently good practice. For example,

if a fuel tank or a solid rocket sustainer engine (usually a thin shell structure with relatively heavy end closures) must be located between two package compartments, advantage might be taken of the inherent stiffness of such a structure to stiffen the entire missile.

Monocoque designs based on the inner skin carrying the bending loads (as in the case of the ramjet-propelled missiles), as well as on the outer skin performing a like function (as in the case of rocket-propelled missiles), have been used. There is no easy way of determining the optimum airframe construction. Advantages of designing to cause the inner skin to carry the load, as compared to the use of an inner skeleton or framework (if there is no central duct), are related to accessibility for packaging and servicing. These generally offset the decreased stiffness due to the smaller moment of inertia resulting from the smaller radius of the "active" material.

Loading the inner skin of a ducted body (which is of some importance for certain configurations such as the cruciform) gives more uniform bending characteristics in the principal guidance planes. Carrying all bending loads through the outer skin is most efficient structurally but has the attendant disadvantages of requiring semipermanent load-carrying doors and carefully designed cutouts, stringers, doublers, and other space consuming elements.

Cellular construction, if feasible for packaging, gives a stiff efficient structure and may be used profitably for both primary and secondary components.

A typical ICBM has a total flight time of about 30 minutes or less. It remains only 4 or 5 minutes in the atmosphere. Consider just the first 2 to 2½ minutes of flight as the missile leaves the launching pad. A typical ballistic missile trajectory consists of a vertical rise, followed by a transition to a gravity turn (a turn in which the missile axis is maintained parallel to the velocity), and then a constant attitude which is held until burnout. During these stages of flight, serious aerodynamic forces are encountered, such as drag, stability, heating, etc.

Aerodynamic Forces

A ballistic missile in flight has two main aerodynamic forces acting upon it, lift and drag. These forces are expressed by the following equations:

$$L = C_L \frac{1}{2} \rho V^2 S$$

$$D = C_D \frac{1}{2} \rho V^2 S$$

In these equations C_L is the lift coefficient, C_D is the drag coefficient, S is a reference area (usually the area of a cross-section of the missile), and $1/2\ \rho\ V^2$ is called the dynamic pressure. Lift force is needed in the flight of aircraft; it is an unfortunate byproduct in the flight of ballistic missiles. The location on the missile at which the lift force acts is called the center of pressure (cp). Lift and drag are obtained by integrating the pressure acting over the entire missile and resolving the resulting force into two components, one normal and the other parallel to the velocity. Frequently, the integrated pressures are resolved into forces normal and parallel to the axis of the missile, rather than normal and parallel to the velocity vector. We call these normal (N) and the axial (A) forces.

For the types of ballistic missiles which we know today, the cp is usually located forward of the missile's center of gravity (cg). Since the missile is a free body in space, any disturbing force will tend to rotate it about its cg. The lift force causes instability. For example, whenever there is an angle between the velocity vector and the axis of the missile (this is called angle of attack), the lift force which occurs tends to rotate the missile about its cg in such a direction as to increase further the angle of attack. This is known as "static instability."

Lift occurs whenever there is an angle of attack; however, drag is always present. Drag slows down the missile and therefore reduces range. The lift or normal force imposes bending moments on the missile airframe and must be taken into consideration in the structural design.

The following information gives some insight into the physical concepts of aerodynamic phenomena to show how the aerodynamicist obtains his data, and to indicate some practical results.

Fundamentals of Aerodynamics

The nature of the flow field about a body moving through the air, the pressure distribution, and

particularly the coefficients C_L and C_D are functions of certain dimensionless parameters. Among the most important of them are the following:

1. The Mach number, M. $M = V/a$, where a is the speed of sound—this is a measure of the compressibility of the air; density changes become important as M approaches 1.

2. The Reynolds number, R. $R = \rho VL/\mu$, where L is some characteristic length of the body and μ is the coefficient of dynamic viscosity—the Reynolds number expresses the ratio of inertia forces to viscous forces. When R is low, viscous effects are large, but at high R, inertia forces predominate. Note also that R is a scaling parameter since it depends on length.

3. The ratio of specific heat of air at constant pressure to the specific heat of air at constant volume, γ. $\gamma = c_p/c_v$.

This parameter becomes important when compressibility effects are large. The aerodynamicist has at his disposal a set of general equations for the analysis of airflow about a body. These are called the Navier-Stokes equations. They are extremely general and amount to a statement of Newton's laws for a viscous compressible fluid, together with equations specifying conservation of mass and energy. In their general form they are too complicated to permit practical solutions, and several simplifications are imposed on them. Among these are the following:

1. *Steady flow*—this assumes that all physical quantities are either stationary or very slowly varying in time.

2. *Nonviscous flow*—it has been found that viscosity is relatively unimportant except for a thin layer of air next to the body (the boundary layer). This permits all the rest of the flow field to be handled by the nonviscous equations and is an excellent approximation at high Reynold numbers and still moderately good at relatively low R. The nonviscous flow assumption results in considerable simplification of Navier-Stokes equations. The flow is now called *potential flow*. The simplified equations are still nonlinear and difficult to solve.

3. *Small disturbance flow*—this assumes that the physical properties of the flow (such as pressure, velocity, and density) are represented as a steady component, plus a very small variation thereof. The equations are written in terms of the small variations only, with the result that linear equations only are obtained, for which solutions can be written. The linearized equations can handle a great number of supersonic and subsonic problems with a surprising degree of accuracy. They are most applicable to slender, pointed bodies and, therefore, are not too useful for the thick blunt-nosed configuration of our current ballistic missiles.

4. *Newtonian flow*—at high supersonic Mach numbers and also at large angles of attack, the Newtonian flow concept has physical validity. This assumes that a flow encountering a solid surface loses its normal component of momentum and moves along the surface with its tangential component unchanged. When applied to blunt-nosed bodies at moderate supersonic Mach numbers, the Newtonian approach gives surprisingly accurate pressure distributions.

Small disturbances spread in all directions, moving at the velocity of sound. If the body creating the disturbance is moving at subsonic speed (M less than 1), the disturbances outrace the body, and the entire field of flow is affected. If the body is moving at a supersonic speed (M greater than 1), that body outruns the disturbance. All the disturbances are contained within a cone whose vertex is at the disturbance.

Once a disturbance is created by a part of a body, for example, by a fin on a missile, its effect can never be felt by any upstream part of the missile if the speed is supersonic.

A large disturbance such as would be set up by a blunt nose of a body traveling at supersonic speeds is propagated as a discontinuous wave with speed greater than the velocity of sound. This discontinuous wave, which is called a shock wave, stands in front of the body and travels at the same speed as the body rather than at the speed of sound. Across the shock wave there is a discontinuous jump in pressure, temperature, density, and velocity. The flow across the shock is no longer potential, and viscosity plays an important role. The shock wave is a dissipative mechanism since the stagnation pressure of the flow behind the shock is always less than the stagnation pressure of the flow in front. By stagnation pressure we mean the total pressure which would be obtained if the flow were decelerated isentropically (with-

out a loss or gain in heat) to zero velocity. The static temperature is higher behind the shock than in front, but the stagnation temperature is the same since the law of the conservation of energy still holds across the shock wave.

Base and Friction Drag

There are two important sources of drag: base drag and friction drag. At low-subsonic speeds, these two items comprise practically the entire drag of the vehicle, for pressure drag is virtually negligible. Pressure drag starts to become important at transonic speeds where compressibility first becomes important; at supersonic speeds pressure or wave drag predominates.

The origin of base drag can be described roughly as follows. The flow separates at the base of the missile and expands around the corner to form a turbulent wake. Adjacent to the base, a wake of turbulent eddies is left, the pressure level of which is below the static pressure of the main stream. This acts in effect like a suction against the base and constitutes the source of base drag. Calculating the pressure in the base region is difficult, and experimental data are used almost exclusively. The complete flow mechanism in the base region is not fully understood, particularly the effect of the high-velocity flow out of the main rocket nozzles on the base flow patterns. The lowest base pressures (and hence the largest contribution to drag) occur at low-supersonic Mach numbers, and the base pressure gradually approaches free stream ambient pressure as the Mach number increases.

Another source of drag is friction. As stated earlier, the effects of viscosity appear to be confined to a thin region of air immediately in contact with the body, as for example the boundary layer. Just outside the boundary layer, the air is moving at high velocity. At the surface of the body, the velocity of the air is zero. The occurrence of a velocity gradient across the boundary layer gives rise to a shear stress which is the source of friction drag.

Friction between the air in the boundary layer and the surface of the missile gives rise to another important effect, namely, aerodynamic heating. This phenomenon is not significant at low speeds; however, it becomes very important at high speeds even if the free-stream air temperature is at or lower than the temperature of the missile surface. The reason for this is that the energy of motion of the air is large compared to the thermal energy. In the boundary layer, the action of viscosity causes a large fraction of the kinetic energy of motion to dissipate as heat.

Which jobs does the aerodynamicist do with respect to heating? He determines the nature of the boundary layer. He calculates or estimates the value of the Stanton number throughout the trajectory. He determines the flow properties (ρ, V, T, and M) at the edge of the boundary layer all over the body. In general, these are not the same as free-stream properties of the flat plate example. The aerodynamicist must calculate all the properties of the flow field about the body and not merely the pressure, as we indicated earlier. The calculations to determine the time-history of the skin temperature are too complicated to be performed by hand and are generally solved on a computer. The following section shows how the aerodynamicist furnishes the input data for these calculations.

Techniques for Determining Aerodynamic Effects

The physical laws governing the flow of gases are well known for most cases. Exact solutions to the equations are seldom practical in engineering problems. Approximate theoretical treatment (where theory is used as a guide) is a more valuable approach and forms the basis of most analytical design work. The use of large computing machines helps in obtaining exact solutions to certain equations; but even here, the complexity of the equations, the difficulty of programming them on the machines, and the length of computing time needed to obtain solutions make this approach impractical for solving individual problems. The use of machines becomes practical when many calculations of the same type have to be made. In a few areas, the physical laws are not fully understood, as for example, in the case of turbulent flow. In such cases, the theory to date just correlates experimental data but does not explain the real mechanism of the physical phenomenon.

Making use of available theory as best he can, the aerodynamicist calculates the flow field about

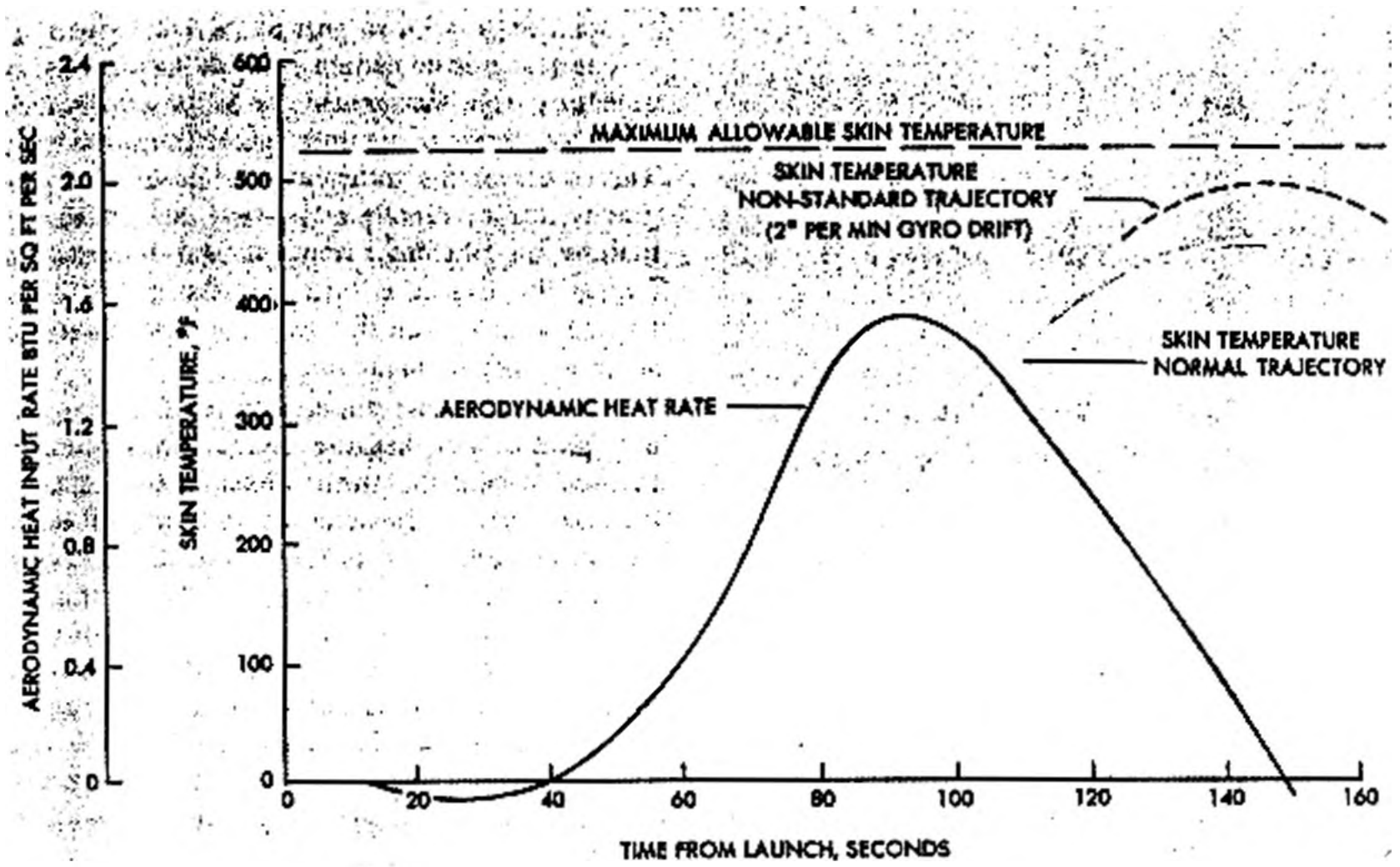


Figure 2-41. Heating of Guidance Compartment of Typical Ballistic Missile

a missile through the trajectory. He determines the pressure distribution over the body for loads and stability data. He determines the density, temperature, and velocity distributions for aerodynamic heating calculations. Since Mach number in a typical trajectory varies from zero at launch to values as high as 15 by the time the missile leaves the atmosphere, the aerodynamic analyses are generally performed at a series of discrete Mach numbers. Continuous behavior between these discrete points is assumed. The Reynolds number also varies considerably throughout the trajectory, and is of considerable significance in determining the aerodynamic heating. For this problem, the aerodynamicist has to consider the variations of both M and R throughout the trajectory.

Missile Aerodynamic Heating

Figure 2-41 illustrates the boundary layer heat input and the skin temperature of a typical ballistic missile. The particular location studied is on the guidance compartment of the vehicle. The skin material is an aluminum alloy. The results are for a nornial trajectory. The effect of a nonstandard trajectory (such as would be obtained, for example, if gyro drift occurred) is also illustrated. You can see that the skin material, its thickness, the trajectory, and trajectory deviations must all be determined so as to keep the maximum skin temperature within allowable limits.

The structure design and environment of the ballistic missile are now presented.

Structure Design and Environment

One of the most important objects in the structural design of a ballistic missile is the reduction of weight to a minimum consistent with the desired performance and reliability of the missile. To do this, the loads and environment to which the structure is subjected must be accurately known. This section reviews the characteristics of a ballistic trajectory which influence structural

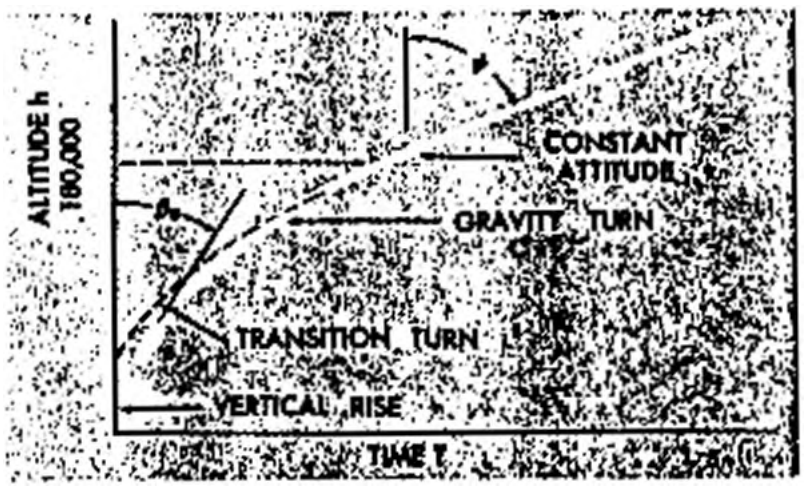


Figure 2-42. Typical Ballistic Missile Trajectory

loads and then presents the various types of loads which result.

First, a few standard terms used in the description of structural requirements must be defined. These terms emphasize the importance attached to weight in the missile design. The "limit load" of a structure is the maximum load that the structure is required to sustain without permanent deformation. The "ultimate load" is the maximum load which the structure is required to sustain without failure, as for example, rupture or collapse. The ratios of these loads to the maximum expected loads are the limit and ultimate factors of safety, respectively.

These factors allow for such things as size tolerances, variation in material properties, and variations in the loading conditions from those expected.

Trajectory Characteristics Which Cause Loads

A typical ballistic trajectory, as shown in figure 2-42 starts off with a short vertical rise during which the missile is rolled to the desired azimuth direction. Then, for the mathematical trajectories, there is an instantaneous tip over into a "zero lift" or "gravity" turn. In this turn, the missile axis is maintained parallel to the velocity vector so that, ideally, there is zero aerodynamic lift and the forces on the missile are minimized. After leaving the atmosphere (approximately 40 miles), a "constant attitude" trajectory is followed until burnout. This trajectory results in maximum range and is so named since the thrust vector is maintained at a constant angle with respect to the launch vertical.

The real missile cannot possibly change instantaneously from the vertical rise to the gravity turn. Actually, there is an intervening transition turn. The object of the transition turn is to attain the gravity turn as rapidly as possible, subject to limitations on pitching, velocity, and acceleration. During this transition turn, the missile axis is not aligned with its velocity vector and is, therefore, subject to aerodynamic lift.

During flight, the missile acceleration increases at a gradual rate because of the consumption of the fuel, while the thrust remains the same or increases slightly. In determining structural loads, we are not interested in the acceleration which is the derivative of the velocity, but the total acceleration including that due to gravity. Expressed in terms of the number of Gs, this is called the load factor, n, and is equal to the thrust minus the drag divided by the weight, $n = T - D/W$. The force required to support any object in the missile is then equal to n multiplied by the weight of the object.

One of the major disturbances to the missile trajectory is wind. The comparison of the wind profile with altitude is shown in figure 2-43. This is nominally a 99-percent profile; that is, the velocities shown are not expected to be exceeded more than 1 percent of the time. The maximum value of the wind, 300 ft/sec, occurs at 35,000 ft, the same altitude at which the maximum dynamic pressure occurs. Now consider the types of loads acting on the missile.

AXIAL LOADS. To appreciate the general character and distribution of axial missile loads, con-

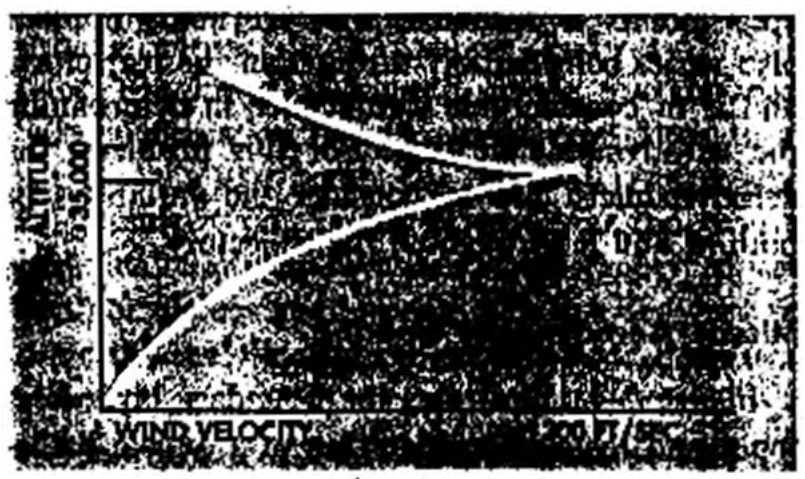


Figure 2-43. Wind Velocity

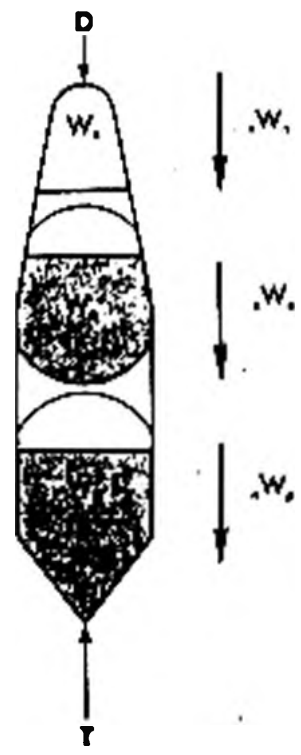


Figure 2-44. Axial Missile Loads

sider figure 2-44, a schematic sketch of a typical ballistic stage indicating the principal weight items and axial loads acting on the structure. The principal weights are the nose cone (or later stage for a multistage missile) and the propellants. Compared to these, the effect of the distributed weight of the structure itself is small.

Starting from the front end, you can see that the structure supporting the nose cone must resist its inertia reaction, nW_n, and its drag, D. The drag is proportional to the dynamic pressure, while the inertia reaction is proportional to the load factor, n. For a blunt high-drag nose cone, the maximum axial load will occur at the maximum dynamic pressure; while for a heavy low-drag nose cone (or later stage), the maximum load will occur at burnout.

Next, consider the section behind the forward tank (in this case, the oxidizer tank). In addition to the load carried by the nose support structure, this section must support the inertia reaction of the oxidizer. Since the oxidizer's weight decreases at approximately the rate the load factor increases, the inertia reaction remains constant. The total load on this section increases gradually with time.

For the section between the rear tank and the engine, the axial load almost equals the thrust minus the inertia reaction of the engine; therefore, the axial load increases slightly during flight due to the increase of thrust with altitude.

BENDING LOADS. Bending loads are primarily caused by maneuvers and by aerodynamic lift forces resulting from angles between missile axis and relative velocity of the air (angles of attack).

Ideally, the only maneuvers and angles of attack occur during the transition turn. The problem is to devise a turning program which is completed before the dynamic pressure becomes high but which, at the same time, minimizes pitching accelerations and velocities.

During the gravity turn, small angles of attack result from such things as gyro drift in the turning program and nonstandard weights and thrusts. The largest angles are caused by winds. As the result of a transverse wind, the velocity of the missile with respect to the air is no longer parallel to the missile axis, and aerodynamic lift is developed. The resultant of the lift acts on the forward part of the missile tending to pitch the missile up and increase the angle. The control system then swivels the engine to prevent this. As a result of these forces (the lift at the front end, the engine side force at the rear, and the missile's own inertia), the missile is subjected to moments which tend to distort it as shown in figure 2-45.

On the concave side the stresses due to the bending moment add to those caused by the axial loads, while on the convex side they subtract. Since the bending moments are primarily due to the aerodynamic forces resulting from the winds, they have their peak value at the time of maxi-

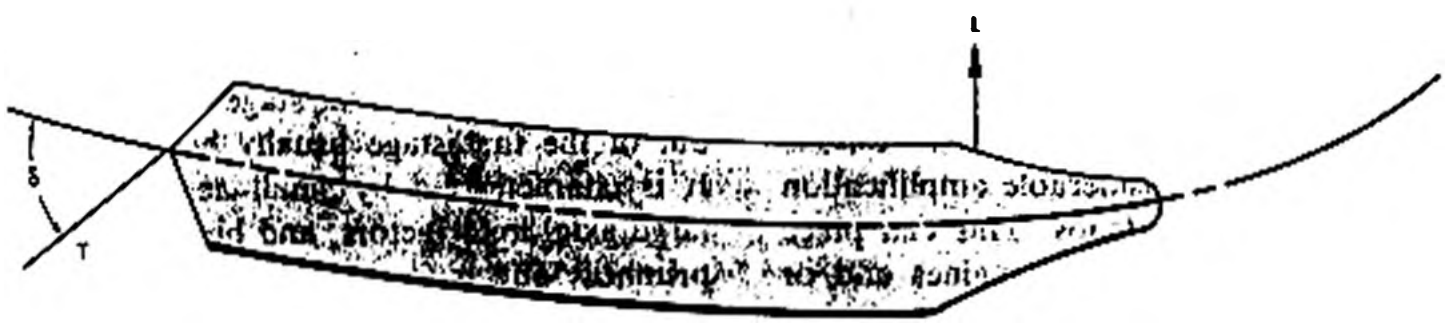


Figure 2-45. Bending Loads

mum dynamic pressure, that is where the winds also have their maximum value.

THERMAL LOADS. Although the effects of heating are not strictly loads in the sense that the previously discussed axial and bending loads are, they have to be similarly studied in designing the structure. High temperatures affect the structure in two ways:

1. They affect the properties of the structural materials and thus reduce the loads which the structure can carry.

2. They cause internal stresses due to differential expansion of structures. This differential expansion results either from two parts of different materials being heated equally or parts of the same material being subjected to temperature gradients.

Of the two effects, the reduction in material properties is generally more important since it requires either the change to more heat-resistant and less efficient materials, or the use of additional material or insulation to reduce the temperature. The second effect can generally be minimized by detail design of the structure.

One method of reducing the heating problem is to fly a more nearly vertical or lofted trajectory. This is undesirable since it reduces the performance of the missile; however, it may be necessary for an existing missile for which some change, such as increased thrust, has resulted in a higher heating rate.

INTERNAL PRESSURE. Internal pressure in the tanks has to be treated in two ways:

1. As a load which determines the wall thickness of the tank.

2. As a part of the structure which helps resist axial and bending loads.

DYNAMIC LOADS. Up to now we have considered steady loads or loads which varied slowly compared to the principal frequencies of the missile. Now, we consider some of the important suddenly applied or rapidly oscillatory loads.

At launch. The sudden increase in thrust at engine firings or the sudden release of the missile on launching can cause considerable amplification of the normal steady state loads. This can place limits on the starting time of the engines and/or require the missile to be released gradually or prior to the full buildup of engine thrust.

Vibration. The greatest source of vibration is the nonsteady thrust of the engine. Other causes are oscillatory aerodynamic forces and bending of the missile. Vibration primarily affects the missile's electrical and mechanical equipment and their mountings, and it may cause their structural panels to fail due to fatigue.

Propellant sloshing. Another cause of oscillatory loads is the sloshing of the propellants in the tanks. By itself, the effect on structural loads is not great, but by coupling back through the control system, it can be very great.

This brings up one of the major problems in missile design, the coupling between the control system and the missile. This problem is primarily in the controls area, but there is a strong interaction with the structure and the structural loads. The control system tries to guide the missile in a predetermined attitude program. The attitude is measured by a gyro mounted on the missile. Because of the bending of the missile, the gyro does not measure the true attitude. The control system may make a false correction, leading to instabilities. The problem is to find a position for the gyro and a circuit design for the control system which will make the system stable. This is complicated by the fact that because of the large changes of weight the bending characteristics of the missile change considerably during flight.

Flight Loads

You can see that there are two primary critical times of flight for the structure of a vertically rising ballistic missile.

The first occurs at maximum dynamic pressure. It is characterized by large aerodynamic forces, bending moments, relatively low-axial load factors, and low temperatures. It primarily affects rearward sections where the variation of axial load with load factor is small. Being primarily an aerodynamic condition, it is sensitive to changes in the pitch program but is not particularly sensitive to changes in weight or thrust.

The second occurs at stage burnout, the burnout of the first stage usually being most critical. It is characterized by small aerodynamic forces, large axial load factors, and high temperatures. It primarily affects forward sections where the loads increase rapidly with axial acceleration and the temperatures are high. This condition is sensitive

to changes in weights and thrust, but is only secondarily affected by the pitch program's influence on heating.

Secondary but sometimes important critical conditions occur at launch and during the transition turn.

General design philosophy holds that the ground loads should not determine the design of any major structural element of the missile. Because of the importance of missile weight, only flight load should be considered in the design of missile structure. This means that for handling, transporting, and erecting, special equipment is developed which will not subject the missile to loads which are greater than those it incurs in flight. Except for regions primarily affected by drag, the axial load increases continuously during the time of flight of a stage and the rate of increase is much greater for sections at the front end than for those at the rear.

Techniques for Studying High-Speed Aerodynamics

Techniques for studying high-speed aerodynamics are one of the major problems of supersonic aerodynamics. Developing missiles for supersonic speeds requires special procedures and special equipment. New wind-tunnel designs have helped to solve this problem.

WIND TUNNELS. While wind tunnels for some time have been used with great success in the subsonic field, they have had to be improved and enlarged to meet the needs of missile study.

As soon as a missile reaches Mach 1, the effects of shock waves must be considered. In setting up such velocities as this in wind tunnels, engineers found that shock waves are reflected off the tunnel walls. The diameter of the tunnel compared to that of the model must be very large to approach the conditions of actual flight.

New techniques in wind-tunnel design have eliminated the reflection of the waves, or "choking" effect. The problem of supersonic and hypersonic tunnel design is not in many ways as great as that of subsonic tunnel design but a bottleneck in progress developed because of transonic speeds.

Wind-tunnel choking, scale effect, or Reynolds number, which up to this time have been well understood, take new aspects in the transonic region. Results of scale-model tests can no longer be truly reliable; consequently, development of larger supersonic wind tunnels is necessary. Wind tunnels have been constructed which accommodate complete missiles of the smaller sizes. Most of the larger missiles are tested by stages, using full-size stages. Figure 2-46 shows the basic layout of a simple wind tunnel.

REPORTS OF WIND TUNNEL TESTS. The following are parts of reports made on wind-tunnel tests:

> . . . Eight data runs were made at Mach number 2.23 to measure the turbulent-boundary-layer shear stress along a flat plate, and to obtain total pressure surveys through the boundary layer. The plate was tested at angles of attack of 0, -1.25, and -3.00

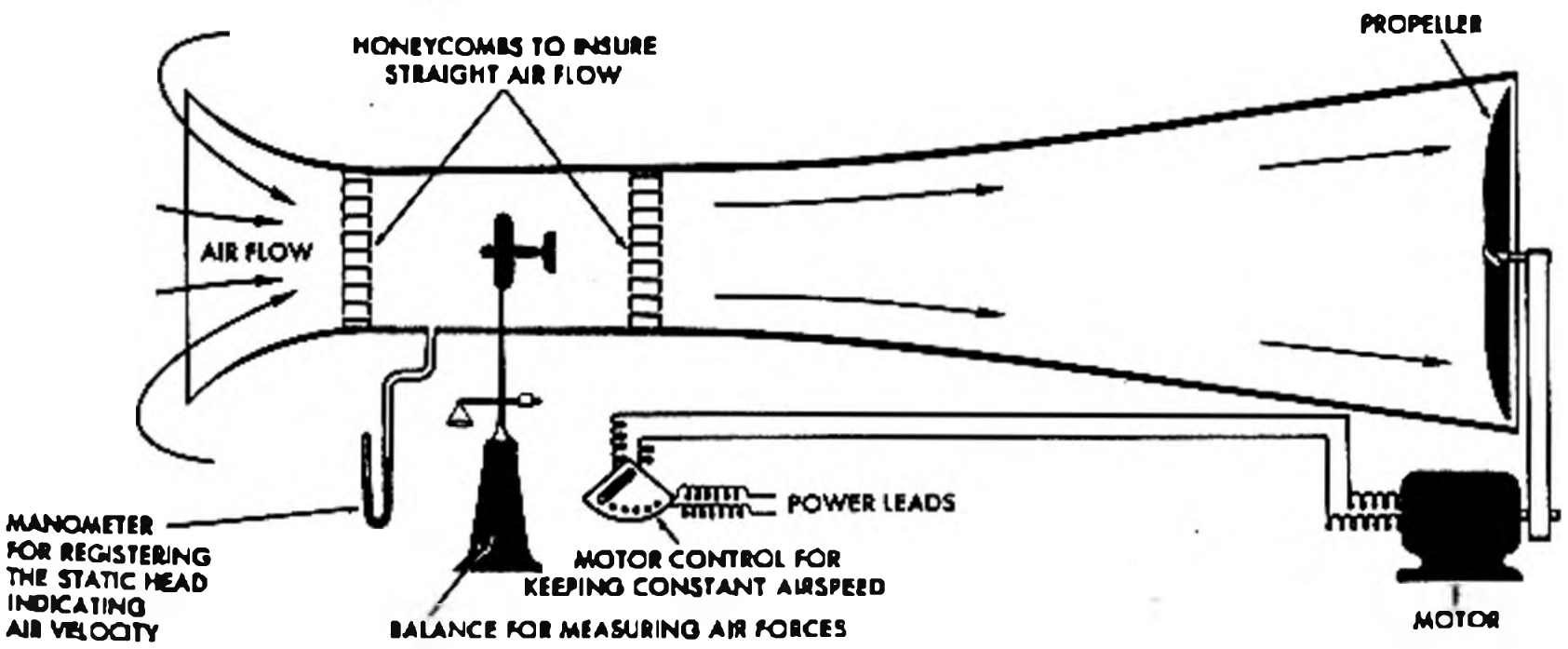


Figure 2-46. Wind Tunnel Construction

degrees to vary the effective free-stream Mach number . . .

. . . Thirty-five data runs were made at Mach number 1.73 to obtain the stability, control, and drag characteristics of composite configurations of the 1/82.5 scale model. The model was mounted with -24 roll-indexing balance to record six-component data between -8 and +12 degrees indicated angle of attack at 0 and 90 degrees roll attitude. Eleven surface deflections of 0, -6, and -10 degrees were tested . . .

These samples give you some idea of what is included in the thousands of tests that must be made in the development of a missile.

Progress Reports on Guided Missiles

During the development of any missile, periodic reports are published giving new developments and the results from tests made on current models. Parts of such reports are given on this and the next page. The examples are given so that you may realize the complexity of the problems that confront designers in the development of new missiles.

The following exemplifies a typical progress report on a missile, showing what the related influencing factors are.

. . . The tactical requirements of this guided missile made it necessary that the tactical version of the missile be capable of carrying a warhead of 2000 pounds and have a range of approximately 1000 miles. Also desired was that the tactical version would be in production in the shortest time possible. This made advisable the use of standard and proven design features. To use a reliable and proven powerplant, a low supersonic speed was chosen that allowed the use of a conventional turbojet engine.

The cruising speed was chosen as Mach 0.85. This speed helped in making the vulnerability less. An added reduction in vulnerability was realized by selecting a high cruising altitude of 60,000 feet. High wing loading made it impractical to choose a cruising altitude above 60,000 feet. Since this missile was to attain supersonic speed during the terminal dive, it had to exhibit good stability and control characteristics in the transonic and supersonic regions. These high speed requirements also called for clean design. To have a clean design, it was decided to have a symmetrical configuration throughout the missile. This symmetrical design in the transonic region of speed is desirable because one advantage for transonic speeds occurs at the zero-lift dive condition.

A sharp sweepback angle and low aspect ratio were also decided upon to improve the Mach speed characteristics. The airfoil sections were made thin and symmetrical; this also aided the high-speed performance. Having a sweptback wing made possible the elimination of the horizontal tail surface since longitudinal control could be obtained by the trailing edges of the wing. Ailevators provided the necessary lateral and longitudinal control. Elimination of a horizontal tail surface greatly simplified the design and eliminated air-flow problems that would otherwise be present. It also eliminated other problems that would have been met in designing separate control surfaces to have the same critical Mach number. The ailevators were placed near the fuselage which reduced the length of the connections to the actuators within the fuselage.

The design of the vertical tail section was determined by the same considerations as those for the wing sections. Good directional stability was attained by proper location of the single vertical tail section. The use of a double vertical tail had been considered, but since it was desirable to provide a landing gear for recovery, the double tail was abandoned. The cost of flight-testing the missile was so great it was worthwhile to provide this landing gear.

Control response was investigated at the cruising speed to be sure that the missile had the required maneuverability. It was decided to use banked turns since the skid-turn performance was inadequate. This change caused some further complications within the autopilot and controls system . . .

The following information might be typical of another progress report of a considerably smaller missile than the one reported above.

. . . In this particular missile, the size and shape were the main consideration; therefore, the main attention was given to the number, size, and location of control surfaces. Designing a missile according to specified size and shape gives rise to particular aerodynamic problems. The factors of greatest importance are those concerning drag, stability, and maneuverability.

It was desirable that this missile exhibit good stability at low speeds and in adverse wind conditions. A highly sweptback, delta wing was chosen to give the desired lift and stability at low speed and at the desired upper limit of speed of Mach 3. Since lifting surfaces suffer a loss in lift at higher speeds, the chosen wing showed good performance under all conditions.

To limit the overall length of the missile, it was decided to eliminate a tail section and use four symmetrically placed wing surfaces. Modifications of the control system were made to function with the planned configuration.

To attain the desired maneuverability of this missile, having a small lifting surface, it was decided to use adjustable jet vanes in addition to the wing controls. These jet vanes were located in the motor exhaust and coupled directly to the control surfaces. The jet vanes provided good control at low speeds before the wing surfaces became aerodynamically

effective. These jet vanes were graphite coated to withstand the high temperatures of the motor exhaust . . .

From the examples above, you can readily see that considerations given to missile design are many and varied. No one problem stands alone when a new missile is being developed.

TRENDS IN CONSTRUCTION OF MISSILES

Up to recent times, fuselage construction of missiles followed the usual pattern of semimonocoque or full monocoque design. A semimonocoque type usually has four main longitudinal members with secondary members attached to them. In full monocoque, the construction depends entirely on the skin being attached to the secondary members, thus forming a shell-like structure.

One metal used to cover structures is titanium. This metal is about 60 percent heavier than aluminum but only about half the weight of steel. Its alloys are several times stronger than aluminum and rival the best steel alloys. Titanium is used for skin-stressed structures as well as nonstructural parts. New methods of forming and machining titanium have resulted in its increased use for missile and aircraft structures.

Another important characteristic of titanium is its resistance to high temperature. The problem of surface temperatures has been mentioned before in regard to the structural problems. Apart from this, the temperature rise presents a serious problem of premature explosions of warheads, resulting in combustion in the fuel lines and damage to operating equipment. Titanium can be used in the 300° to 800° temperature range encountered in the transonic and supersonic speed ranges.

Materials other than metals have been developed as structural materials and are being used extensively. One of these is known commercially as Spiralloy. This material is composed of continuous glass fibers helically wound on a mandrel and bound with epoxy resin to form very strong components such as rocket motor cases. The third-stage motor of the *Minuteman III* is one example of the use of this material. One of the latest materials in use is a graphite composite or graphite yarn wound on a mandrel and impregnated with epoxy.

There is a continuing search for stronger and lighter structural materials and methods of fabricating missile structures. In the future, many new materials will undoubtedly be found, tested, and improved to provide lighter and stronger vehicles capable of carrying larger payloads.

SUMMARY

This chapter has presented a general picture of the problems involved in the field of aerodynamics. As you can see, it would be beyond the scope of any one text to cover every phase of aerodynamics in detail.

In your work with a missile organization, you don't need a complete knowledge of aerodynamics. It is to your advantage to have an understanding of the basic concepts of aerodynamics as it has been presented here.

CHAPTER 3

Propulsion of Guided Missiles

Guided missiles must move with a high velocity to lessen the probability of interception and destruction by enemy countermeasures. Missiles must also be able to intercept and destroy high-velocity enemy missiles and manned aircraft in flight.

The reciprocating engine-propeller combination was a satisfactory propulsion device for aircraft until the start of World War II when necessity demanded the development of aircraft capable of higher subsonic and even supersonic speeds. As speeds increased, the combination proved unsatisfactory because with a constant power input the propellor thrust diminished rapidly after a certain speed was reached. This condition necessitated extremely large engines to produce sufficient horsepower to give further increase in speed. In addition, when approaching the speed of sound, shock-wave formations on the propeller drastically reduced the thrust for an engine of given horsepower.

Lack of sufficient power was one of the main reasons why supersonic flight was not realized at a much earlier date. Aerodynamicists have calculated that 35,000 hp would be required to propel an F-51 at a speed of Mach 1.5 at 50,000 feet. On the other hand, an aircraft the same size and weight as the F-51, but incorporating all the latest aerodynamic refinements, including jet propulsion, would require only 8000 hp under the same conditions.

BASIC FORMULAS AND LAWS UNDERLYING JET PROPULSION

Jet propulsion is defined as a means of locomotion brought about by a change in the momentum of matter ejected from within the propelled body.

Because of this definition of jet propulsion, jet-propulsion engines are sometimes referred to as reaction-type engines. This label is not sufficiently specific, since any body moving in a fluid works on the reaction principle if it is self-propelled. The action of a conventional propeller consists of increasing the momentum of the air, and the propeller thrust is the resultant reaction. The ordinary propeller-driven aircraft is not a form of jet propulsion because the working fluid is not ejected from within the vehicle. If the propeller were ducted and the air allowed to pass through the vehicle, then the vehicle would have mechanical jet propulsion. None of the jet-propulsion engines in missiles use the mechanical method.

To understand the principles involved in jet-propulsion systems, you must consider a basic mathematical analysis of how the thrust of a jet engine is developed. Remember that thrust is any force tending to produce motion of a body.

Newton's second and third laws of motion are the underlying principles on which jet propulsion is based. Newton's second law of motion states that an unbalanced force acting upon a body causes the body to accelerate in the direction of the force. The acceleration produced is directly proportional to the unbalanced force and inversely proportional to the mass of the body. The formula for force is

$$F = Ma, \qquad (1)$$

where F is the unbalanced force in pounds, M is the mass of the body in slugs, and a is the acceleration produced in feet per pound. Mass is a fun-

damental property of matter which causes it to resist a change in velocity. A body at rest is acted on by the force of gravity which is usually referred to as the acceleration of gravity. This acceleration or "g" as it is usually called, amounts to approximately 32.2 ft/sec/sec at the surface of the earth. The weight of a body is, therefore, the force exerted on that mass that causes it to be accelerated at 32.2 ft/sec/sec. Mass, weight, and acceleration can be mathematically related by the following expression:

$$M = \frac{W}{g} \tag{2}$$

Where M = mass in slugs, W = weight in pounds, and g = the acceleration of gravity. Although the value of g decreases with increasing altitude, causing a corresponding decrease in weight, this effect can safely be ignored in atmospheric and low orbital flight.

Now, by substituting equation 2 in equation 1, Newton's second law may be represented by

$$F = \frac{W}{g} a \tag{3}$$

In the application of this equation to jet propulsion, F is the unbalanced force which accelerates the working fluid through the exhaust nozzle, and a is the acceleration in ft/sec². In accordance with Newton's third law of motion, the forward thrust (T) of a jet-propulsion unit is equal and opposite to this unbalanced force (F).

By definition, acceleration a is the rate of change of velocity v and may be expressed as

$$a = \frac{v_2 - v_1}{t}, \tag{4}$$

where v_1 is the initial velocity of some mass, v_2 is the final velocity of that same mass, and t is the time required to change the velocity from v_1 to v_2. The original equation, F = Ma, may be rewritten as

$$F = \frac{Mv_2 - Mv_1}{t} \tag{5}$$

Since by definition, Mv is called momentum, we can state that the thrust of a jet engine is equal to the rate of change of momentum of the working fluid. The above equation is frequently rewritten as

$$F = m (v_2 - v_1). \tag{6}$$

where m represents M/t and is called the mass rate of flow of the working fluid in slugs per second.

Now let W equal the total weight in pounds of working fluid that flows through the unit during the thrust-producing portion of operation, and let t equal the total time required. Then, W/t will equal the weight rate of flow in pounds per second. Letting w = W/t, the equation for thrust becomes

$$T = \frac{w}{g} (v_2 - v_1). \tag{7}$$

where: T = thrust in pounds
w = weight rate of flow of working fluid in pounds per second
v_1 = intake or initial velocity of working fluid
v_2 = exhaust or final velocity of working fluid
g = acceleration of gravity (taken as 32.2 ft/sec²)

While the above equation is actually the method for calculating the force exerted upon the

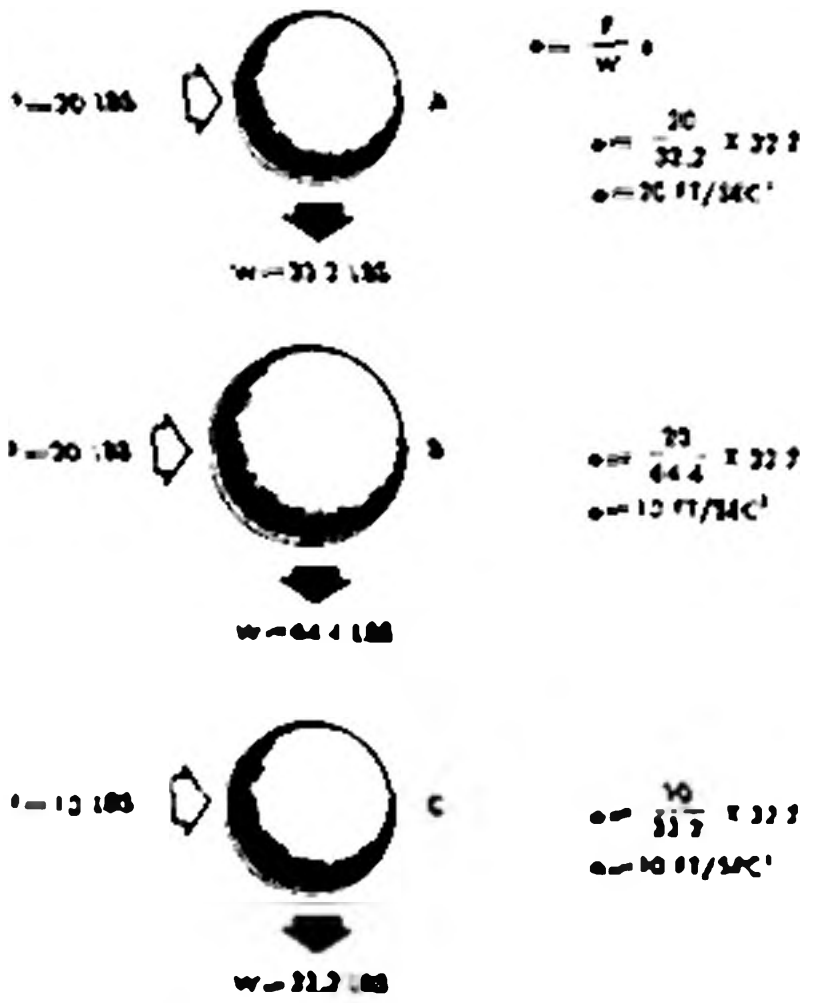


Figure 3-1. Examples of Newton's Second Law of Motion

working fluid. It also gives the value of the thrust forcing the jet engine forward.

Examples of Newton's second law are shown in figure 3-1. The first figure shows a body weighing 32.2 lbs subjected to a force of 20 lbs. resulting in an acceleration of 20 ft/sec². The next figure shows a body weighing twice as much as the previous one but acted upon by a like force of 20 lbs. Note that the acceleration imparted to the body in this case is 10 ft/sec². The third figure again shows a body weighing 32.2 lbs, but the applied force is decreased to 10 lbs. Note that the acceleration has also decreased as compared to the first illustration.

In the drawing of a jet engine in figure 3-2, the applied force results from the chemical reaction between the air and the fuel. Observe that the resulting acceleration given to the weight of working fluid is the change in velocity of the working fluid from v_1 to v_2. The thrust that drives the engine forward is equal in magnitude but opposite in direction to the force that drives the working fluid rearward. The drawing of the jet engine in figure 3-2 illustrates a typical situation in which an engine develops thrust in accordance with Newton's second law of motion.

In the case of a rocket-propulsion engine where the working fluid, or fuel, is stored within the unit, the original velocity (v_1) is zero. The thrust formula for a rocket when the pressure of the working fluid (exhaust gas) at the nozzle exit is equal to the ambient pressure is

$$T = \frac{w}{g} v_2 \qquad (8)$$

In a rocket engine the exhaust gas leaving the nozzle is frequently at a pressure other than ambient. Under this condition, another term must

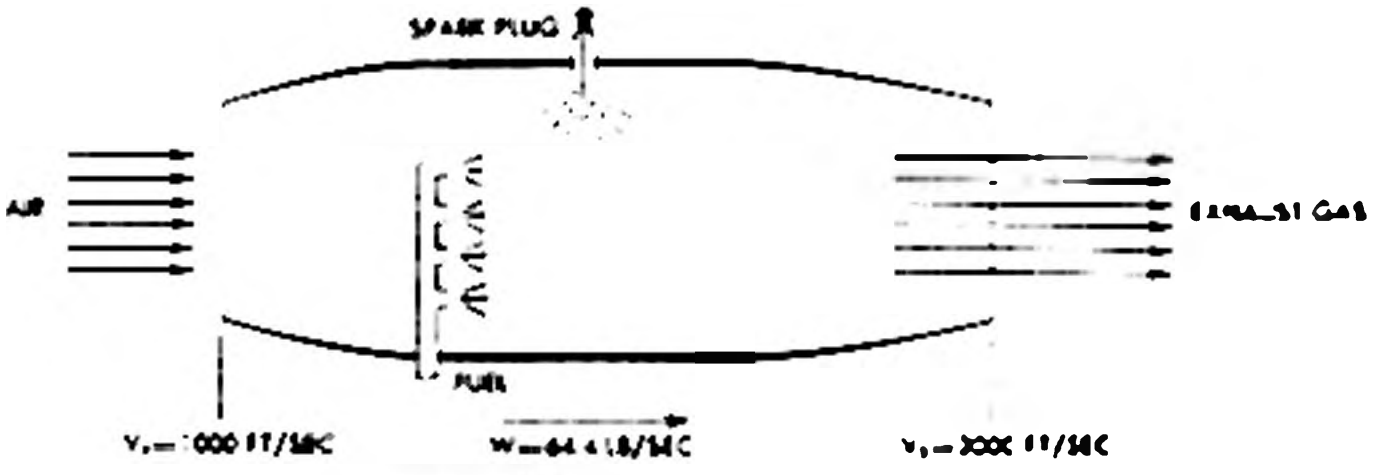


Figure 3-2. Practical Example of Newton's Second Law of Motion

be added to account for this additional thrust. Equation 8 then becomes

$$T = \frac{w}{g} v_2 + (P_e - P_a) A_e \qquad (9)$$

where P_e = pressure of the gas at the nozzle exit
P_a = atmospheric pressure
A_e = area of nozzle exit

A rocket motor which permits the expansion of the propellant products to the pressure of the surrounding atmosphere is referred to as a motor with optimum expansion ratio.

Since changes in the atmospheric pressure affect the pressure thrust, a variation of the rocket-motor thrust with altitude is to be expected. The change in pressure thrust due to altitude changes may amount to 10 to 30 percent of the overall thrust. The following conclusions may be stated:

- External pressure decreases thrust.
- Rocket would operate most efficiently in a vacuum, where atmospheric pressure (P_a) is zero.

Notice that thrust is always expressed as pounds of force and not in terms of work or horsepower. A jet engine that is being static-fired

is not moving through any distance and is doing no work, nor is it developing any horsepower. You can calculate the thrust horsepower for a moving missile powered by a jet engine by the following relationship:

$$\text{Thrust Horsepower (THP)} = \frac{\text{Speed of missile (mph)} \times \text{Thrust (pounds)}}{\text{One Horsepower (mile-pounds per hour)}} \quad (10)$$

The horsepower in the denominator is equivalent to 375 mile-pounds per hour and is derived from the more frequently used horsepower unit of 33,000 foot-pounds per minute. A missile of the V-2 type, if developing 56,000 pounds of thrust and traveling at 3750 miles per hour, would develop 560,000 horsepower. Using equation 10,

$$\frac{3750 \text{ mph} \times 56{,}000 \text{ lbs}}{375 \text{ mi-lb/hr}} = 560{,}000 \text{ THP} \quad (11)$$

Consider the case of a tactical missile equipped with an air-breathing jet engine and apply the same formula, because turbojet engines are also rated in thrust horsepower (THP) rather than in units of horsepower.

The missile, making a test flight, is flying at a velocity of 642 mph. Its turbojet engine is developing 6000 pounds of thrust. What is the THP?

$$\frac{642 \text{ mph} \times 6000 \text{ lbs}}{375 \text{ mi-lb/hr}} = \text{THP}$$

$$\frac{3{,}852{,}000}{375} = 10{,}272 \text{ THP}$$

Since the air density decreases with increasing temperature or increasing altitude, either or both of these variations will result in a decrease in the mass flow rate of air through a jet engine. If all other factors remain unchanged, this will result in decreased thrust.

Suppose a tactical missile is launched on a mission involving a distance of approximately 1000 miles. Wide variations in weather can take place over a 1000-mile range. As the missile flies along at a preset altitude of 25,000 feet, it is certain to fly through air masses at different temperatures. If it is flying in a cold air mass, the temperature decreases, and the density of that particular mass increases. Because of this increase in density, the THP increases. If, at a later period, the missile encounters a warm air mass with a resulting increase in temperature, the density of that mass will decrease. With the decrease in density the THP decreases.

A common misconception in the case of jet engines, especially the rocket, is that the exhaust gases impart thrust to the missile by pushing against the outside air. The air in no way helps drive the rocket. It acts only to impede the rocket's motion.

CLASSIFICATION OF JET PROPULSION SYSTEMS

This section discusses jet-propulsion systems of two general classifications: rocket and air-breathing propulsion systems.

Rocket Systems

Rockets are self-contained thrust-producing systems. A rocket-propelled vehicle derives its thrust from ejection of hot gases generated by the combustion or decomposition of propellants carried in the vehicle. Rockets may be either the liquid or solid propellant types. The necessary propellants are fuel and oxidizer as shown in figure 3-3. An oxidizer is a substance which supports combustion when combined with a fuel and thereby permits operation of the rocket independent of the atmosphere.

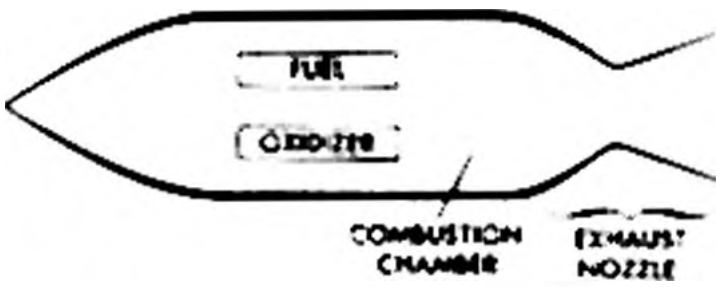


Figure 3-3. Rocket Containing Fuel and Oxidizer (Air Independent)

Air-Breathing Engine Systems

The air-breathing engine system provides a means of jet propulsion and thrust augmentation. The medium (air) through which the missile passes is forced into the duct (primarily by the ramming motion of the missile through the air), and the intake (ram) air, in the case of one such type system, is compressed by a mechanical compressor unit, heated, and allowed to expand through an exhaust nozzle. The compression is

usually achieved by some form of rotary compressor wheel driven by a turbine.

Study the sketches in figures 3-4 and 3-5 of the turbojet and turboprop engines as you read the next few lines. Hot gases that result from the combustion of fuels and air in the combustion chamber strike the turbine wheel causing it to rotate. The rotary motion is transferred by means of a shaft to the compressor wheel which compresses the air.

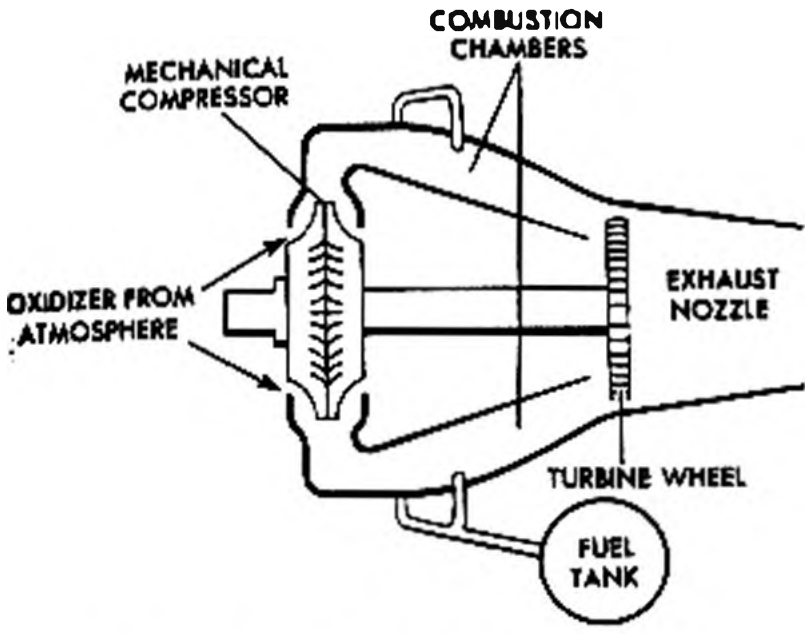


Figure 3-4. Turbojet (Air Dependent)

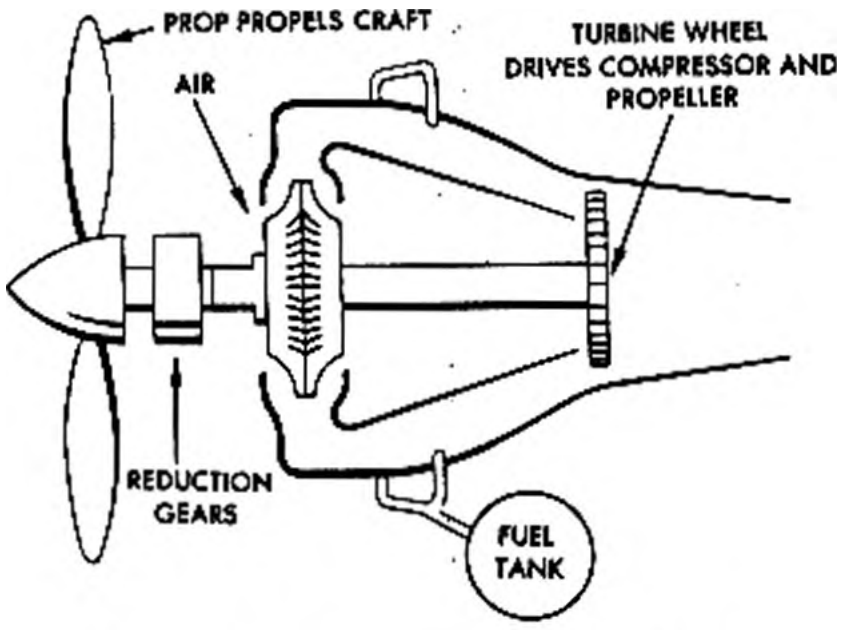


Figure 3-5. Turboprop (Air Dependent)

The ramjet engine is another example of an air-breathing engine system. In this type, compression of the working fluid (air) is obtained by proper design of the intake section of the duct and the high forward velocity of the vehicle. The intake air is compressed to create additional pressure for assisting combustion. The gases which result from the combustion process expand to a high velocity through the exhaust nozzle. The ramjet engine pictured in figure 3-6 is of this type.

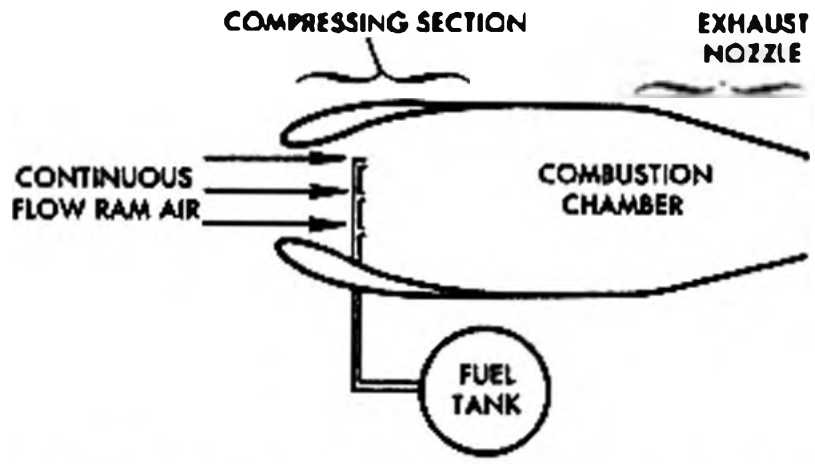


Figure 3-6. Ramjet Duct Unit (Air Dependent)

A third type of air-breathing engine system is the intermittent, or pulsejet, unit. This engine also obtains pressure and velocity changes in the fluid medium by proper design of the intake section of the duct. It differs from the two previous types in

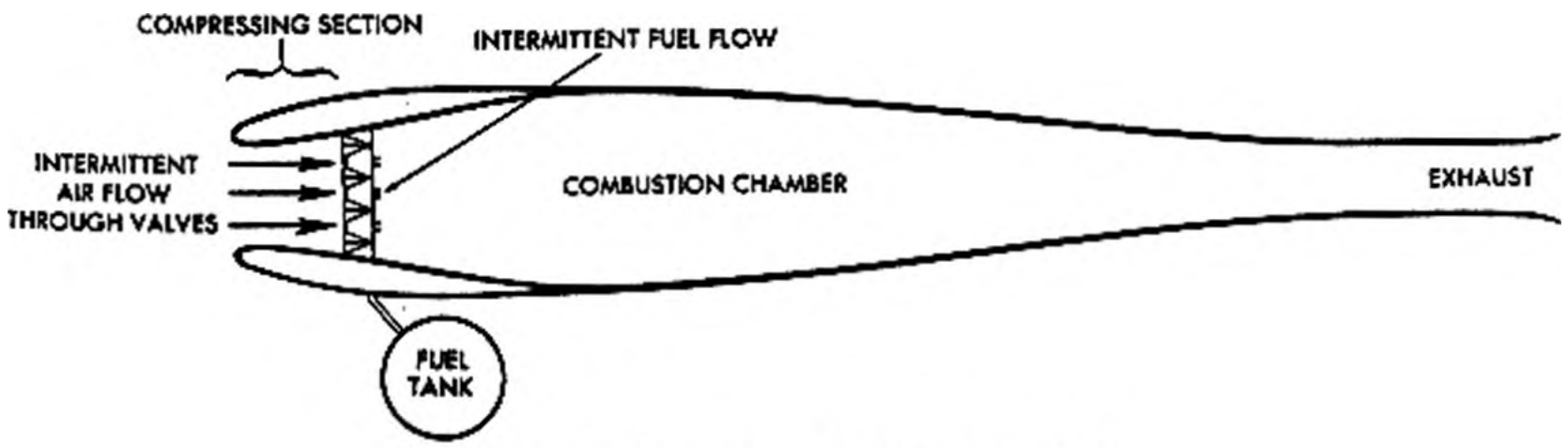


Figure 3-7. Intermittent Duct (Air Dependent)

that the flow of air and heat-giving fuel is intermittent, or in cycles. Notice the illustration in figure 3-7.

Another form of an air-breathing system is typified by a rocket engine which has an augmenting shroud encircling the exit of the thrust chamber nozzle. This feature increases the thrust of the shrouded rocket engine by the momentum of the additional mass of air accelerated into and ejected from the inside of the encircling duct by the exhaust stream.

A variety of engines have been proposed that combine a rocket and a ramjet in an attempt to obtain a jet-propulsion system that would have better low speed characteristics than a ramjet and a higher specific impulse than a rocket.

The various types of jet-propulsion systems are discussed in detail later in this chapter. Consider first the types of fuels used by these systems.

GUIDED MISSILE PROPELLANTS

Earlier you read that thrust is developed as a result of energy released in the jet-propulsion engine. A large quantity of readily available energy is found in certain materials. When these materials react in a jet engine, they impart thrust to propel the missile. These materials and the oxidizers with which they react are called propellants. Large quantities of high-pressure and high-temperature gases are produced by the chemical reaction of a fuel and an oxidizer at the proper time and rate in the combustion chamber. The heat energy thus made available is converted into kinetic energy in the exhaust nozzle, or tail pipe.

When you read of an engine that can travel faster than an artillery shell, operate in a vacuum, deliver more energy than a reciprocating engine, and do so with few or no moving parts, you may get the idea that some very complex chemical mixture must be used as the propellant. Such is not the case. Jet-propulsion systems can operate on such common, cheap fuels as kerosene, gasoline, alcohol, gunpowder, and coal dust. However, obtaining a desired result depends upon, among other things, the proper choice of a fuel-oxidizer combination on the basis of energy-releasing capabilities and convenience of use.

From the standpoint of physical state, propellants may be either solids, liquids, gases, or various combinations of these. Generally the propellants are either solid or liquid. Hybrid rockets use a solid fuel and a liquid oxidizer. Gases are rarely used for propellants unless they have first been liquefied since it is difficult to carry the required mass of gas in a tank of reasonable size and weight.

Propellant Performance Terms

There are various methods of rating the performance of different propellants. A liquid or solid propellant rocket system may be rated by what is referred to as *total impulse*. The total impulse is simply the product of the thrust in pounds times the firing duration in seconds, or stated in formula:

$$Th \times t = I_t$$

Where:

Th = thrust developed (lbs)
t = firing duration (sec)
I_t = total impulse (lb-sec)

(Note that total impulse is in lb-sec and not in lbs per sec.)

Rocket propulsion systems and propellants are rated on the basis of *specific impulse*. Specific impulse is defined as the amount of thrust produced per pound of propellant consumed per second or more simply the thrust produced divided by the propellant flow rate.

$$I_{sp} = \frac{Th}{w}$$

Where: I_{sp} = specific impulse (pound-seconds per pound)
Th = thrust (lb)
w = propellant flow rate (lb per sec)

When I_{sp} is used to rate a propellant, the thrust used in the equation is the thrust that would be theoretically produced by burning the propellant at 100 percent combustion efficiency under specified operating conditions, usually 1000 psi chamber pressure, expanded to sea level pressure.

(Note that the specific impulse is pound-seconds per pound or lb-sec/lb.)

For example, if the specific impulse of white fuming nitric acid (WFNA) and furfuryl alcohol is 214 pounds per pound per second (lbs/lb/sec), it simply means that 214 pounds of thrust are

obtained from each pound of propellant burned each second.

The term specific impulse is usually restricted to rocket propulsion systems and includes both the fuel and oxidizer. Pure air-breathing engines use the oxygen in the air to burn their fuel and it would be misleading to calculate I_{sp} for these systems in the same manner as for a rocket. Consequently, the term *specific fuel consumption* is used for air breathers. It is defined as the fuel flow rate in pounds per hour divided by the thrust.

$$\text{Specific fuel consumption} = \frac{w}{Th}$$

Where: w = fuel flow rate in lb per hr

Th = thrust in lb

Other terms are mixture ratio and exhaust velocity.

Mixture ratio designates the relative quantities of oxidizer and fuel used in the propellant combination. It is numerically equal to the weight of oxidizer flow divided by the weight of fuel flow or:

$$\frac{W_o}{W_f} = \text{mixture ratio}$$

Exhaust velocity is determined theoretically on the basis of the energy content of the propellant combination. The actual velocity of the exhaust gases is less than this theoretical value. The term "effective exhaust velocity" is sometimes used and is determined on the basis of thrust and propellant flow. The actual, or real, exhaust velocity is that obtained when the kinetic energy of the gas flow produces actual thrust. The actual exhaust velocity is determined by using real values for thrust and total weight flow rate of propellants, or by using calculated values that are modified by a "velocity correction factor," combining all necessary corrections into one value.

Solid Propellants

Solid propellants are chemicals, in a plastic or rubber-like form, which, when ignited, produce hot, high pressure gases which can be used to provide the reaction force for rocket propulsion. A solid propellant formulation usually includes two or more of the following constituents:

1. Oxidizer (nitrates or perchlorates)
2. Fuel/Binder (organic rubbers, resins, or plastics)
3. Fuel (aluminum, powder)
4. Chemical compound combining fuel and oxidizer qualities (nitrocellulose or nitroglycerin)
5. Additives (to control fabrication process, burn rate, etc.)

These substances are mixed to produce a solid propellant of desired chemical and physical characteristics. The finished product is called a solid propellant grain.

Ideally, a solid propellant would possess the following desirable characteristics:

1. High energy content
2. High density
3. Manufactured from easily obtained substances
4. Safe and easy to handle
5. Easily stored; stable to shock and temperature changes
6. High physical (mechanical) strength properties
7. Ignites and burns uniformly
8. Nonhygroscopic (will not absorb water vapor)
9. Smokeless
10. Flashless
11. Non-toxic

Note that none of the propellants used today has all these desirable properties; in fact, every one of them has several drawbacks. It is most often the case that a specific desirable property in a solid propellant is obtained at the expense of other properties.

SPECIFIC SOLID PROPELLANTS. Modern solid rocket propellants may be divided into two general classes, double-base propellants and composite propellants. A third important class, which combines features of both, is the composite-modified double-base (CMDB) propellants.

Double-base propellants have as their principal components nitrocellulose and an energetic plasticizer, such as nitroglycerin. The term "double-base" differentiates this type of propellant from many gunpowders which have been traditionally based on nitrocellulose. Other materials may be added in smaller proportions to serve as stabilizers, non-explosive plasticizers, coolants, opacifiers, and burning rate modifiers, or otherwise confer desirable properties on the end product.

The propellant is prepared in the form of a rigid plastic having a nearly homogenous structure.

A typical double-base formulation, that of ballistite, is as follows:

Chemical	*Percent*	*Purpose*
Nitrocellulose	51.38%	Propellant
Nitroglycerin	43.38%	Propellant
Diethylphthalate	3.09%	Plasticizer
Potassium Nitrate	1.45%	Flash Suppresser
Diphenylamine	.07%	Stabilizer
Nigrosine Dye	.10%	Opacifier

Ballistite has a specific impulse at sea level of approximately 210 lb-sec/lb and is characterized by a smokeless exhaust. A variety of processes are used to manufacture double-base propellants. In the extrusion process, the blending and mixing are accomplished by running the propellant through heated differential rollers or other mechanical shaping devices. The material is then extruded through a suitable die at very high pressures. The extruded grain is then heated to relieve internal stresses. Although the extrusion process requires costly equipment, it permits the exact control of grain size and shape. In the casting process of double-base propellants, nitroglycerin liquid is cast into an evacuated mold with small pellets of nitrocellulose. The mold is heated and the nitrocellulose forms a nearly homogeneous solid with the liquid mixture.

Composite propellants are made by embedding a finely-divided solid oxidizing agent in a plastic, resinous, or elastomeric matrix. The matrix material usually provides the fuel for the combustion reaction. Composite propellants have been made in a great variety of compositions. Oxidizing agents which have been used extensively include ammonium nitrate, sodium nitrate, potassium nitrate, ammonium perchlorate, and potassium perchlorate. Asphalt, natural and synthetic rubbers, aldehyde-urea and phenolic resins, vinyl polymers, polyesters, and nitrocellulose are among the matrix materials which have been employed as binders. Minor constituents may be added to modify the properties of the binder or to catalyze the burning process. In addition, metallic fuels, such as aluminum, have been used in composite propellants. The addition of metallic particles increases the combustion temperature, thus increasing the specific impulse. Metallic particles also increase the density and alleviate certain types of combustion instability.

The common characteristic of the varied composite propellants is a markedly heterogeneous structure with adjacent regions of oxygen-rich and oxygen-deficient materials. Most solid propellants being manufactured today for missiles are made up of a rubber binder with ammonium perchlorate oxidizer and aluminum powder fuel. Such propellants are used in all stages of the Minuteman III and Poseidon missiles.

The composite-modified double-base propellant essentially uses a double-base propellant as the binder and incorporates solid fuel and oxidizer particles. Such propellants have a higher specific impulse than a conventional double-base propellant and have been used in the Polaris and Minuteman missiles.

A comparison of the three classes of propellants is shown in table 3-1.

SOLID PROPELLANT BALLISTICS. The burning rate of a solid propellant is the velocity at which the grain is consumed. The rate is a measure of the linear distance burned, in inches per second, in a direction perpendicular to a burning surface.

The thrust of a rocket motor depends on the mass rate of flow and change in velocity of the working fluid. For high thrust levels, large mass rate of flow is necessary. This high mass rate of flow can be obtained by either a large burning rate or a large burning area. A smaller burning area produces less mass flow and less thrust. By varying the geometrical shape of the grain, the thrust developed by a given amount of propellant in a given combustion chamber can be greatly influenced.

Figure 3-8 shows the two major grain configuration types. The top drawing shows an end-burning grain. It is a solid cylinder which completely fills the combustion chamber and burns only on the end face. The thrust is directly proportional to the cross-sectional area and the burning duration is directly proportional to the length or web thickness of the grain. The web thickness is defined as the distance which the burning thickness must recede in order to consume all of the propellant. The second drawing is a center perforated grain and the web thickness is the distance between the inside and outside surfaces. Usually the outside surface is inhibited or bonded directly to the case and burning occurs only on the inside surface.

Table 3-1. General Comparison of Solid Propellants

	DOUBLE-BASE (Ballistite)	COMPOSITE (Representative Formulation)	COMPOSITE-MODIFIED DOUBLE-BASE
Specific Impulse Lbsec/lbs	210	265	270
Propellant Density lb/in³	.058	.065	.067
Safety	Explosive Hazard	Fire Hazard Only	Explosive Hazard
Mechanical Strength	Acceptable for small motors	Acceptable for all size motors	Acceptable — motors up to 65" diameter



END BURNER



CENTER PERFORATED

Figure 3-8. Solid Propellant Configurations

The ends can either be inhibited or allowed to become a burning surface.

Burning of solid propellants depends on the size and shape of the grain. One propellant grain may burn in such a manner that the burning area remains constant. This type of burning is known as neutral burning and an end-burning grain is an example of a neutral burning grain. Progressive burning is exhibited when the burning area of a grain increases as burning progresses. The third class of burning is regressive burning. This type of burning is exhibited by an ever-decreasing burning area as burning progresses. A star-shaped center perforated grain exhibits regressive burning.

SOLID PROPELLANT LIMITATIONS. There are two characteristics peculiar to solid propellant rockets which limit their use.

Temperature sensitivity. The initial temperature of a grain noticeably affects its performance. A given grain will produce a higher level of thrust on a hot day than on a cold day. Temperature changes affect the equilibrium pressure and the burning rate. The percentage change of thrust per degree fahrenheit temperature change is referred to as the temperature sensitivity for a particular propellant. This characteristic of a grain emphasizes the need for storage under certain temperature conditions until time of usage.

Temperature limits. Some solid propellants are limited to a fixed temperature range for reasons other than temperature sensitivity because temperature also affects the physical properties of

solid propellants. At extremely low temperatures, solid propellant grains become brittle and are subject to cracking. Cracks in a grain may have a catastrophic result because increases in burning area increase the combustion chamber pressure. If this pressure exceeds the pressure for which the chamber was designed, a fracture or explosion results. A propellant exposed to high temperatures before firing will cause the propellants to become weak and exhibit a plastic behavior. The propellant grain, therefore, may be unable to withstand the loads caused by a sudden rise in chamber pressure or accelerations which the vehicle might undergo. The subsequent failure again is due to an undesirable increase in burning area. The temperature under which propellants can function is, therefore, limited. Typical operating temperature limits for an asphalt base thermoplastic propellant is +25° F to 120° F; for a double-base propellant, −20° F to +140° F; and for a modern castable composite propellant, −60° F to +160° F.

Liquid Propellants

In contrast to solid propellants which are an integral part of the rocket motor, liquid propellants are injected into the combustion chamber from tanks. Liquid propellants permit longer firing duration and intermittent operation. Combustion can be stopped and started at desired intervals by controlling propellant flow. Throttle thrust control can also be obtained, if desired.

Many liquid-propellant combinations are in use and others have been experimentally investigated, but as in the case of solids, no combination has been found possessing all the desired characteristics for all applications. Desired characteristics are as follows:

1. Adequate availability of low cost raw materials and ease of manufacture.
2. High heat of combustion per unit of propellant mixture to give high performance.
3. Low molecular weight of the gaseous products of the reaction. Generally, propellants having a large weight percentage of hydrogen will best meet both "2" and "3" because of the much higher heat of combustion per pound of hydrogen than carbon. Also, the hydrogen-containing products of the reaction have lower molecular weights than the carbon-containing products.
4. A low freezing point and high boiling point (low vapor pressure), permitting a wide range of operation.
5. High specific gravity, allowing a large weight of propellant to occupy a given space and thus permitting smaller missile construction with less structural weight as a result of smaller tanks.
6. Low toxicity.
7. Non-corrosive and thermally stable, thus simplifying storage problems.

CLASSIFICATION OF LIQUID PROPELLANTS. Liquid propellants may be grouped under either monopropellants or bipropellants. Bipropellants can be further identified as cryogenic or storable. Heterogeneous propellants (containing a solid component in a liquid carrier) and tripropellant systems have also been investigated.

Monopropellants can consist of a single compound that will decompose, or of a fuel and oxidizer combined into a single fluid that is combusted. Hydrazine (N_2H_4) and hydrogen peroxide (H_2O_2) are examples of compounds that can be catalytically decomposed to produce hot gases. Iridium on alumina granules is the catalyst commonly used for hydrazine. Silver screens (solid or plated) are used for hydrogen peroxide monopropellant systems. Monopropellants consisting of fuels and oxidizers pre-mixed, such as nitric-acid/amine or hydrogen peroxide/glycol, have not gained acceptance because of their sensitivity to shock, or ability to propagate a detonation.

Cryogenic propellants must be kept refrigerated. Liquid hydrogen and liquid oxygen are allowed to boil-off, thus keeping the remainder of the liquid cold. Insulated tanks must be used. Propellant losses in a missile must be made up to avoid sacrifice of range. Such missiles must be serviced shortly before launch and can be kept in a flight-ready condition for only a few days. Liquid fluorine requires the use of a refrigerant, such as liquid nitrogen, since boil-off of toxic fluorine gas cannot be permitted. The entire system must be chilled prior to servicing, or provisions made to collect the vaporized fluorine.

For military applications, *storable propellants* offer the advantage of instant readiness. Nitrogen tetroxide, nitric acid, chlorine pentafluoride, hydrazine, etc., can be loaded into the missile and stored in a ready condition for several years. Small missiles can be filled at the manufacturing plant

(pre-packaged) and shipped and stored without any further propellant-servicing operation required.

Aluminum (Al) and beryllium (Be) are high performance fuels. To use their high energy, and yet retain the flexibility offered by liquid propellants, heterogeneous fuels have been evaluated. Fine particles of Al or Be (10-25 micron) are suspended in a liquid fuel such as hydrazine or kerosene. A gelling or thickening agent (about 2-5%) is added to prevent the metal from settling. Extreme care must be exercised in the formulation of such *gelled or slurry propellants* to assure that requirements of storage stability and flow characteristics are met.

The highest performing chemical propulsion system is the *tripropellant combinations* of hydrogen, beryllium, and oxygen, with an Isp of about 525 sec. Hydrogen, lithium, and fluorine combination is also a very high energy system. More development work is required before these concepts will be practical.

Rocket propellants, by their very nature, are reactive chemicals. They must be stored and transferred with equipment constructed of compatible materials. Handbooks listing suitable materials, components, design criteria, proper handling procedures, and safety guidelines must be consulted. The properties and characteristics of operational propellants are well-defined and must be understood by those working with them. Properly handled, liquid propellants present no undue hazard to personnel or facilities.

The commonly used liquid propellants, and some that are in an advanced development state, are discussed below.

LIQUID FUELS. *Ethyl alcohol* (C_2H_5OH) is a clear liquid, lighter than water, and has a boiling point of about 173° F and a freezing point of −178° F. It is stable to shock and temperature changes. This fuel is readily available because of its wide commercial market in the chemical and liquor industries. Alcohol is not ordinarily considered toxic but proves so if taken internally in excess. This fuel used along with liquid oxygen constituted the propellant of the German A-4 (V-2) rocket.

Unsymmetrical dimethylhydrazine [$(CH_3)_2N_2H_2$], commonly known as UDMH, is a colorless, watery liquid with a "decaying fish" odor. UDMH has a boiling point of approximately 145° F, a freezing point of −71.5° F, and a specific gravity of 0.792. UDMH is similar to hydrazine. It is a toxic, volatile, and corrosive liquid. It is also a highly flammable liquid which will flash at the extremely low temperature of −6° F. Because of its low flash point and high flammable properties, UDMH presents a fire hazard wherever it is handled or stored. Although liquid UDMH is not shock sensitive, its vapors in air can be exploded by an electric spark or flame.

UDMH has two main advantages over pure hydrazine; it is thermally stable and has a very low freezing point. It does not provide quite as high a specific impulse as hydrazine. Mixing of the two substances offers a suitable compromise, thus a 50-50 (wt%) mixture was selected for the Titan II and Titan III systems.

UDMH is stored in tanks or drums in areas remote from other activities. Oxidizers or acids should not be stored in the same storage area used for UDMH. Adequate water supplies must be available for firefighting purposes and for use in washing down spills and leaks.

Personnel working with or handling UDMH must wear approved fire-resistant protective suits, gloves, hats, boots, and face shields. Approved self-contained breathing apparatus must also be worn when working in high concentrations of UDMH vapors or when exposure will be for a long period of time. Emergency showers and eye baths are provided in areas where UDMH is handled or stored. These are required safety items for all areas where toxic propellants are handled or stored.

Hydrazine (N_2H_4) is a clear, colorless, oily liquid with an odor similar to that of ammonia. It is a strong reducing agent, weakly alkaline, and very hygroscopic. It will react with carbon dioxide and oxygen in air. Hydrazine is a stable liquid, boiling at 236° F, and freezing at 35° F. Since it contracts upon freezing, tanks and components will not be damaged. Thermal decomposition begins at about 320° F, but will occur at lower temperatures in contact with catalytic materials such as copper, cobalt, iron, and molybdenum, their alloys and oxides. Because of the susceptibility of hydrazine to decompose, it is used as a monopropellant in small engines, such as attitude control thrusters. Decomposition of hydrazine

with Shell 405 catalyst (iridium on a porous alumina substrate) is spontaneous as 40° F and above. Most storable oxidizers (nitric acid, hydrogen peroxide, nitrogen tetroxide, chlorine pentafluoride) are hypergolic with hydrazine.

Liquid hydrazine, if spilled onto the skin or into the eye, can cause local damage or burns. In addition, it can penetrate the skin to cause effects similar to those produced when the fuel is swallowed or inhaled. Repeated or gross exposures may cause toxic damage to the liver and kidneys and anemia. Protective clothing must be worn during any handling operation. Respiratory protection must be worn if spills or leaks are likely.

Hydrazine is stable to shock, but the vapor propagates detonation within the flammable limits, and at high temperatures. Water is the most effective fire-fighting agent.

High purity aluminum and 300-series stainless steels are generally acceptable for hydrazine service.

Liquid hydrogen (H_2) is non-toxic and noncorrosive. It forms combustible and explosive mixtures with oxidizers. If allowed to evaporate, it becomes highly combustible with air over a wide range of mixtures. It ignites very easily with oxygen and spontaneously with fluorine and chlorine trifluoride. Liquid hydrogen gives a specific impulse of over 400 lb-sec per lb when combined with oxygen or fluorine. It is also considered as the working fluid in nuclear rocket propulsion designs because of its low molecular weight. When liquefied a few degrees above absolute zero, the liquid has a density of only 4.8 lb/cu ft.

High purity liquid hydrogen is a transparent, colorless, odorless liquid. It has a boiling point of –423° F, and it freezes at –435°F. Its density at –423° F (as a liquid) is 0.59 lbs/gal. It is usually boiling vigorously when in an observable condition because of its low boiling point and creates a voluminous vapor cloud when exposed to the atmosphere. It is chemically stable but is physically stable only when stored under suitable conditions, again because of its low boiling point.

Liquid hydrogen presents many problems in its use, storage, and handling. As with most cryogenic fluids, the effects of low temperatures on material properties are extremely important. Materials such as mild steel and ferrous alloys lose ductility at liquid hydrogen temperatures and become brittle. The severe temperature changes to which equipment associated with liquid hydrogen use is subjected can cause stress concentrations that must be considered when designing the equipment.

When no impurities are present, hydrogen burns in air with an invisible flame. With unconfined hydrogen-air mixtures, rapid burning generally takes place when initiated by heat, spark, or flame. In confined areas or when ignition is accomplished by a shock source equivalent to a blasting cap or a small explosive charge, an explosion of the mixture can occur.

Liquid hydrogen may be stored in either fixed or mobile tanks of proven design and materials. The containers used for shipping, storage, and transfer are insulated or vacuum-jacketed similar to those used with liquid oxygen to prevent rapid vaporization of the liquid hydrogen.

Personnel working with liquid hydrogen are exposed to the possibility of serious "freezing" burns of the skin and eyes if they accidentally come in contact with the material through splash, spillage, or by contact with pipes and valves at liquid hydrogen temperatures. Personnel must wear prescribed protective clothing to prevent injury. Emergency showers and eye baths are provided in storage and work areas for immediate use by persons who have been splashed with this substance. In a confined area, gaseous hydrogen is an asphyxiant as a result of the exclusion of oxygen.

Kerosene is used in large quantities in certain ICBM missiles and space-launched vehicles. Kerosene is derived from crude petroleum by refinery processes and is substantially a 100 percent hydrocarbon mixture. The nonhydrocarbons present in kerosene rarely exceed 0.1 percent and are principally sulphur, oxygen, or nitrogen compounds. The sulphur content is limited by specifications.

Kerosene is a clear liquid, ranging in color from water white to a very pale yellow before a dye is added. Specific red dyes are added to the fuel to make leaks readily detectable.

Kerosene has a maximum freezing point of –40° F and a boiling range from 350° F to 525° F. The handling and storage practices for petroleum products are well established.

Kerosene, like all liquid hydrocarbon fuels, burns only in the form of a vapor or gas and then

only when mixed with oxygen or air within a definite range of proportions. Kerosene in a liquid state is not explosive.

Although no special clothing is required for handling any of the hydrocarbon fuels, personnel who are exposed to this material should wear protective clothing such as goggles, a rubber or plastic apron, and rubber or plastic gloves when splashing of fuel is likely. Fuel-splashed clothing should be changed and body contacts washed with soap and water.

The health hazards of hydrocarbon fuels at rocket launch sites are essentially the same as those in normal operations of aircraft and motor vehicle fueling.

LIQUID OXIDIZERS. Oxidizers provide the necessary oxygen to support the burning of fuel. In a narrow sense, an oxidation is a chemical addition of oxygen uniting with a fuel to provide or to speed up combustion. When fluorine (F_2) unites with ammonia (NH_3), no oxygen is involved, and yet chemically the process is so similar that it too is called oxidation.

Liquid oxygen (O_2) is made by liquefying air and separating nitrogen and other gases. This bluish-looking liquid has a boiling point of approximately –297° F and a freezing point of –363° F. It is heavier than water with a specific gravity of 1.14. Such a low boiling point causes an extremely high rate of evaporation. For this reason, storage and shipment to launching areas are problems, resulting in appreciable loss. When poured on metal at ordinary temperatures, liquid oxygen reacts like water dropped on a red-hot stove. Evaporation loss in the Atlas amounted to 7 pounds for every minute of time that elapsed between fueling and firing of the missile.

The extremely low freezing temperature of liquid oxygen causes water vapor from the surrounding atmosphere to collect and freeze on pipes and valves. Liquid oxygen is noncorrosive and nontoxic, but contact with the skin may produce severe burns. The term "burns" is used here in the same sense as you speak of burns resulting from dry ice.

Nitric acid (HNO_3) has been used for many years and is still used in the Agena, Lance, and other rocket systems.

Most frequently used type is called inhibited red fuming nitric acid (IRFNA), consisting of concentrated nitric acid in which 14 percent of nitrogen dioxide (NO_2) is dissolved. IRFNA varies in color from orange to brick red and got its name from the reddish color of the nitric oxide fumes. It is easily obtained because of the large quantities commercially produced for explosives and fertilizers. Nitric acid is highly corrosive, so stainless steel or high purity aluminum is used for storage tanks and delivery pipes. However, the corrosivity of IRFNA is reduced considerably by the use of an inhibitor (0.7% hydrofluoric acid). The fumes of IRFNA are toxic and severe burns result from bodily contact with the liquid.

Nitrogen tetroxide (N_2O_4) is an equilibrium mixture of nitrogen tetroxide and nitrogen dioxide. N_2O_4 is colorless in the solid state, and in the liquid state the color of the equilibrium mixture varies from light yellow to deep reddish brown. The color of N_2O_4 in the gaseous state is reddish brown. It has a boiling point of 70.01° F and freezes at 11.8° F.

This compound reacts with water to form nitric acid and nitrous acid. The nitrous acid decomposes immediately forming additional nitric acid and evolving nitric oxide. Nitrogen tetroxide is hypergolic with fuels such as hydrazine, aniline, UDMH, and furfuryl alcohol.

The fumes of N_2O_4 are highly toxic, and approved respiratory protection and clothing must be worn by personnel engaged in handling or using nitrogen tetroxide. The toxicity of N_2O_4 is essentially the same as that of nitric acid.

Because nitrogen tetroxide is noncorrosive when dry, mild steel containers are used for shipping and storage. Storage areas are well supplied with easily accessible deluge-type showers and eye wash fountains.

Nitrogen tetroxide is a stable, nonflammable, shock-insensitive compound. It will not burn alone, but its vapors will support combustion. If spilled, it may cause a fire by coming in contact with organic materials such as sawdust, excelsior, wood scraps, or cotton waste. Toxic vapors would also be generated from such spills.

Fluorine (F_2) is one of the most vigorous and powerful of the oxidizers. It has a boiling point of –306° F and freezes at approximately –363° F. Its specific impulse when combined with hydrogen is one of the highest of any of the liquid propellant combinations.

Fluorine is highly toxic and corrosive in liquid or vapor form. Contact of fluorine with the skin will cause burns and blisters. It requires all the care needed in handling and storing of a cryogenic liquid, as well as the precautions necessary for corrosive materials.

Chlorine pentafluoride (ClF_5) and *chlorine trifluoride* (ClF_3) are high-energy, highly reactive oxidizers. In the vapor state these compounds are nearly colorless. The liquid is clear and faintly greenish-yellow in color. Their odor has been described as both sweetish and pungent, similar to chlorine or mustard.

ClF_3 and ClF_5 are toxic and corrosive agents, similar to elemental fluorine in reactivity. They attack soft glass and asbestos, will occasionally ignite fluorocarbon polymers (i.e., Teflon), and react with most materials, including concrete. Both oxidizers are hypergolic with most fuels.

These compounds are low boiling liquids (ClF_3, +53° F; ClF_5, +18° F) so that at ordinary temperatures considerable hazard can be present if the liquid is inadequately enclosed. No refrigeration is required since ClF_3 and ClF_5 will remain liquid under their own vapor pressure.

Stainless steel and high nickel alloys are preferred metals for ClF_3 and ClF_5 systems; however, the dry product is shipped and stored in clean, passivated mild steel cylinders. In proper containers, these oxidizers are storable for many years and are ideal for pre-packaged (loaded at the manufacturing plant) propulsion systems.

No materials are known which will provide complete personal protection, thus remote operation is highly recommended. Some clothing items are available that will provide limited protection for short periods.

Although both ClF_3 and ClF_5 react with water, high pressure fog can be used to fight fires. Sodium bicarbonate base (dry powder) extinguishing agent is suitable.

Hydrogen peroxide (H_2O_2) is a colorless liquid with freezing point at 14° F and boiling point of 288° F at 87 percent concentration. As sold at the corner drugstore, it is a mild liquid of about 3 percent concentration. As a monopropellant or oxidizer for rocket units, the concentration ranges from 90 to 99 percent.

When in contact with some catalysts, such as silver or potassium permanganate, it decomposes, forming steam and gaseous oxygen. When 90 percent hydrogen peroxide decomposes, about 42 percent of the total weight of the decomposition products is gaseous oxygen. It is also used as an oxidizer with JP-4 in super performance rockets or fighter aircraft.

Another use of this substance is as a gas generator; in this case, the gaseous products of decomposition are used for driving auxiliary power units for generation of electrical or hydraulic power. The X-15 research aircraft used H_2O_2 to generate electrical power, to drive the main engine fuel and oxidizer pumps, and to provide thrust for attitude control.

Containers for hydrogen peroxide are made of aluminum alloys, stainless steel, and certain plastics. It deteriorates in strength by only 0.1 percent per year when stored at moderate temperatures.

Splashes of concentrated hydrogen peroxide on the skin may produce chemical burns, depending on exposure time.

COMPARISON OF LIQUID PROPELLANTS. Table 3-2 gives a general comparison of a few liquid propellants. The specific thrust and exhaust velocities are calculated values, assuming an exhaust gas expansion to one atmosphere of pressure.

Keep in mind that no one propellant can be considered the best for all missiles. The intended mission of a missile or vehicle determines the kind of propellant used.

High Pressure Gases

Although not propellants, helium and gaseous nitrogen are also used extensively in servicing missiles.

HELIUM (He). Helium is an odorless, colorless, and chemically inert gas, which is not corrosive, explosive, or flammable. Its principal use in missile operations is for pressurizing rocket propellant and hydraulic systems, inflating high altitude balloons, and for use in heliarc welding operations. Because of its inert characteristics, helium is also used to detect leaks in missile systems. Helium under high pressure presents a constant hazard, and suitable provisions for eye and ear protection are maintained for personnel working in areas where the gas under high pressures is being handled. No special protective clothing is required for personnel. High concentrations of

Table 3-2. General Comparison of Liquid Propellants

OXIDIZER/FUEL	MIXTURE RATIO	BULK DENSITY (g/CM³)	CHAMBER TEMPERATURE °F (Tc)	Isp††
	OPTIMUM EXPANSION, P_e = 1000 psia		P_e = 14.7 psia	
LO_2/Hydrogen	4.00	0.28	4910	391.2*
LO_2/Kerosene	2.60	1.02	6164	300.0*
N_2O_4/Hydrazine	1.30	1.22	5406	292.2*
N_2O_4/50% N_2H_4/50% UDMH	2.00	1.21	5610	289.2*
IRFNA/UDMH	3.2	1.27	5139	272.0***
H_2O_2 Monopropellant (100%)	.....	1.47	1823	144.01**
Hydrazine Monopropellant	.....	1.00	1141	191**
OF_2/Hydrazine	3.0	1.492	7079	307.4†
OF_2/Hydrogen	9.0	0.468***	6919	409**
OF_2(70%) LO_2(30%)/Kerosene	3.80	1.20	7820	342.5*

*Rocket Engine Liquid Propellants Rocketdyne (Pub 573-A-2, Rev 3-64)

**Liquid Propellants Handbook, Vol 3 Battelle Memorial Institute

***Propellant Performance Data, Gallery Chemical Co (1961)

†Isp Program (23 Apr 68)

††All Readings For Shifting Equilibrium

helium may create an oxygen deficiency in the immediate area, making the use of breathing apparatus necessary.

NITROGEN (N_2). Gaseous nitrogen is colorless and odorless, and is not corrosive, explosive, or flammable. It is used primarily in missile operations for purging liquid oxygen lines and fuel tanks, for pressurizing liquid rocket propellants, and for blanketing containers of highly active oxidizers and fuels.

LIQUID NITROGEN (LN_2) is used at missile sites primarily as a coolant, and for operational checking of liquid oxygen lines and valves without incurring danger of fire or explosion. It is also used to provide nitrogen gas.

High purity liquid nitrogen is a faint yellow, transparent liquid which is slightly lighter than water. It is noncorrosive, highly inert, and will undergo chemical reactions only at very high temperatures. Liquid nitrogen has a boiling point of –320° F and freezes at –346° F. It is stable to shock, heat, and electric spark.

The main hazards of liquid nitrogen to personnel arise from its extremely low temperature and its ability to cause asphyxiation. As with liquid oxygen, serious "burns" and frostbite can

occur when LN_2 comes in contact with exposed areas of the body. Nitrogen is not toxic (78% of the atmosphere is N_2), however, nitrogen can displace oxygen in the atmosphere. Numerous deaths have resulted from asphyxiation. Because nitrogen has no color, asphyxiation is a real hazard. Personnel must wear approved insulated loose-fitting mittens, protective suit, helmet, boots, and face shield when engaged in the handling, transfer, or use of liquid nitrogen. Respiratory equipment is required in confined areas.

BASIC COMPONENTS OF JET-PROPULSION SYSTEMS

In any conventional jet-propulsion system, a working fluid must be compressed, heated, and expanded, converting some of the thermal energy to kinetic energy and thereby increasing the momentum of the working fluid. In all types of jet engines, the heating takes place by a chemical reaction in the combustion chamber, and the transformation of thermal to kinetic energy takes place in the exhaust nozzle. The method of compressing the working fluid varies widely and is a primary factor in distinguishing one engine type from another. In a ramjet the compression occurs in the diffuser as a result of the deceleration of high velocity air. In the turbojet a rotating compressor accomplishes the compression. In a solid rocket, the compression results from the combustion of the propellant; while in a liquid rocket the propellant is pressurized by the feed system before entering the combustion chamber.

Combustion Chambers

A combustion chamber is the enclosure in which the transformation of energy from chemical to thermal form occurs. Geometric shapes such as the cylinder and sphere are most common in the design and manufacture of the chamber. Although the ideal combustion chamber is in the shape of a sphere, those used in liquid fuel motors are generally oval shaped or in the shape of a cylinder.

The design of a combustion chamber is affected by many factors. For example, the chamber must have a large enough volume to permit combustion to take place smoothly and continuously. Smooth burning is just as important here as it is in an automobile or aircraft engine. The length and diameter of the chamber must be such as to produce a chamber volume most suitable for complete and stable combustion. Depending upon the type of propellant used, the combustion chamber may also contain an injection system and an ignition system. Various types of these are discussed later.

One of the most important factors for good combustion is how well a fuel and oxidizer are atomized after entering the chamber. The manner in which the propellants enter the chamber is largely controlled by injector design and position. By keeping the exposed wall area of the chamber to a minimum, the load is reduced. A small combustion chamber diameter means higher combustion chamber gas velocities which results in higher heat transfer to the chamber wall.

When propellants enter the combustion chamber by various types of injectors, they do so at a high velocity and pressure. They enter the area of combustion in the form of liquid droplets which quickly vaporize because of the intense heat. The propellants, then in a gaseous form, thoroughly mix and what is known as the "gas phase reaction" takes place. The high exhaust velocities needed are produced by expanding these gases in the exhaust nozzle.

INJECTORS. The function of an injector is similar to that of a carburetor in an internal-combustion engine. It atomizes and mixes the propellants in such a manner that a correctly-proportioned fuel-oxidizer mixture results.

In the multiple-hole impingement injector illustrated in figure 3-9, oxidizer and fuel are injected through a varied arrangement of separate holes in such a way that the jet-like streams impinge (intersect each other at some predetermined point), breaking up into fine vapor-like droplets.

A spray injector, second illustration in figure 3-9, has oxidizer and fuel holes arranged in circles so as to produce conical or cylindrical spray patterns which intersect, thus becoming atomized and well mixed.

A non-impinging type injector, lower illustration in figure 3-9, is one in which the oxidizer and fuel do not impinge at any specific point but mix mainly by the turbulence and the formation of propellant vapors. The V-2 used such an injector head,

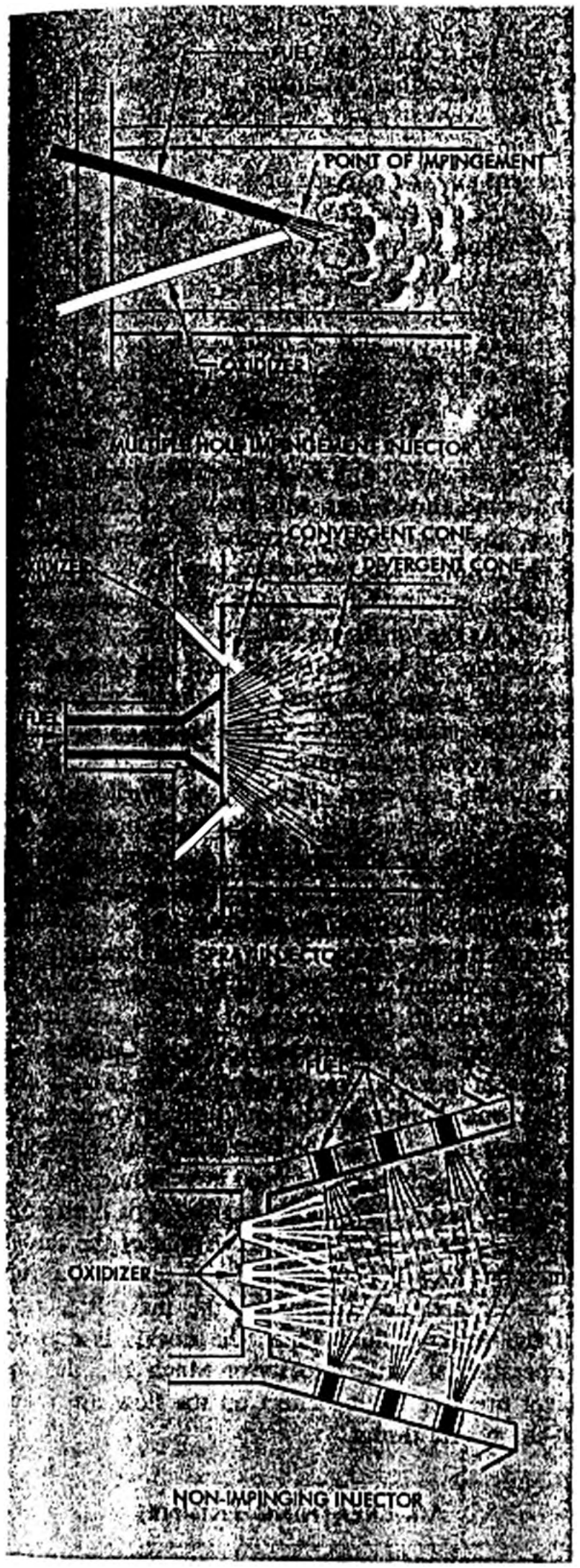


Figure 3-9. Types of Injectors

using fine droplets of alcohol mixed with gaseous oxygen.

IGNITION SYSTEMS. To initiate combustion, nonspontaneously ignitable propellants must be activated by absorbing energy beyond that which they already contain. This energy is supplied by an ignition system. An igniter must be located in such a manner that it receives a satisfactory starting mixture that readily ignites. Too great a collection of fuel and oxidizer in the combustion chamber before ignition may result in an uncontrolled explosion.

A spark-plug ignition system has been used successfully by locating the spark plug in a region where initial fuel and oxidizer vapors form an ignitable mixture. This type of ignition has been used on all types of jet engines. Sometimes a spark plug is used to ignite a smaller combustion chamber which in turn ignites the main chamber.

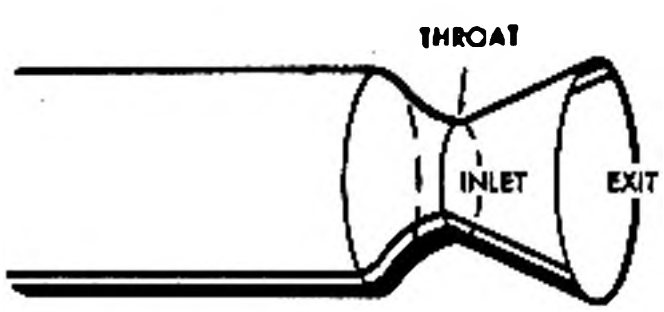
SPONTANEOUS IGNITION. Spontaneous ignition takes place when hypergolic propellants such as nitrogen tetroxide and hydrazine are used. In this case, merely bringing the two propellants into contact with each other is sufficient to initiate combustion. A slug start system is used to obtain spontaneous ignition when the main propellants are not hypergolic. In this case, a small quantity or slug of another propellant that is hypergolic with one of the main propellants is used to obtain ignition. Usually the starting slug is injected through the same injector passages as one of the main propellants—obviously not the one with which the starting slug is hypergolic.

A powder-charge ignition system is used mainly for solid propellant charges. It consists of a powder squib which can be ignited electrically and then burned for a short time with a flame hot enough to ignite the main propellant charge.

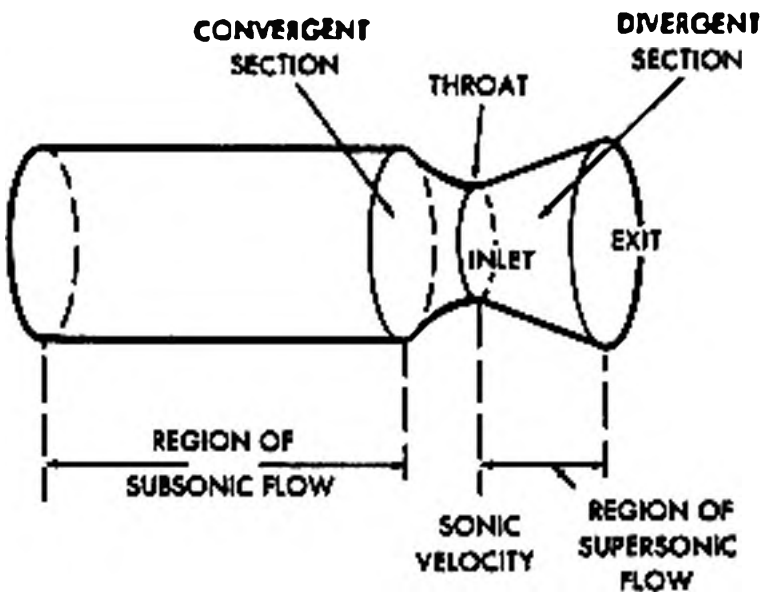
A catalytic ignition method employs a solid or liquid catalytic agent which activates a chemical decomposition of the propellant, producing high-pressure exhaust gases.

EXHAUST NOZZLES. The function of a nozzle is to increase the velocity of the exhaust gas. This is accomplished by converting pressure and thermal energy to kinetic energy or velocity. In a convergent nozzle where the smallest cross sectional area is at the exit, the maximum velocity is limited to sonic or Mach 1. As the pressure ratio (static pressure at the nozzle inlet divided by the

static pressure at the nozzle exit) is increased, the exit velocity will increase until it reaches Mach 1. The pressure ratio at which Mach 1 is achieved will vary somewhat with the gas composition but is in the vicinity of 1.5 to 2 (1.892 for air). At this point, further increase in the inlet pressure will not result in any increase in exit velocity, but will result in the pressure at the exit plane exceeding the ambient pressure. Under these conditions, the momentum thrust does not increase further but the excess pressure at the nozzle exit does increase the total thrust. Whenever the pressure ratio exceeds the value required to produce Mach 1 at the exit, a convergent-divergent (DeLaval) nozzle shown in figure 3-10 can be used to advantage. Under these conditions the exhaust gas will continue to accelerate above Mach 1 and its static pressure will continue to decrease. For optimum expansion (maximum thrust) the divergent section of the nozzle must terminate at the point where the exhaust gas has expanded to ambient pressure. Further expansion would increase the exhaust velocity and the momentum thrust, but the nozzle exit pressure would then be below ambient and the adverse pressure area term would cause the pressure area term in the thrust equation to become negative. This would result in a decrease in the actual thrust. For a given nozzle and nozzle pressure ratio, there is only one nozzle exit area that results in optimum expansion. The nozzle of a rocket on a long range ballistic missile is usually designed to be overexpanded at launch; that is the gas pressure at the exit of the nozzle will be expanded to a lower pressure than ambient and some thrust loss will result. As the missile rises, the atmospheric pressure decreases and at some altitude the nozzle expansion is optimum (nozzle exit pressure equals atmospheric pressure). As the missile altitude continues to increase, the nozzle will be under-expanded and the gas will leave the nozzle at a pressure higher than ambient with that altitude. Note that although the nozzle is not producing as much thrust as it would if the expansion ratio were optimum, the thrust will continue to increase as the atmospheric pressure decreases due to the pressure term of the thrust equation. If a convergent-divergent nozzle is used in a situation where the pressure ratio is insufficient to produce sonic velocity at the throat, it will act as a venturi. The gas will be accelerated as it passes through the convergent section and decelerated in the divergent section. Convergent-divergent nozzles are therefore used in rockets, supersonic ramjets, and some turbojets where the pressure upstream of the nozzle is relatively high. The convergent nozzle is used in subsonic ramjets, pulsejets and most turbojet engines. For both nozzle types the most critical area is that of the throat since this area in conjunction with the combustion chamber pressure determines the mass flow rate of the working fluid. Once the nozzle is choked (operating at sonic velocity or Mach 1), the ambient pressure has no effect on the flow upstream of the nozzle throat.



A LOCATION OF NOZZLE COMPONENTS



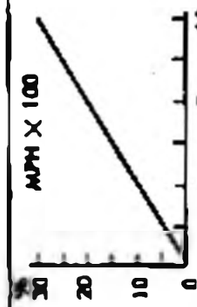
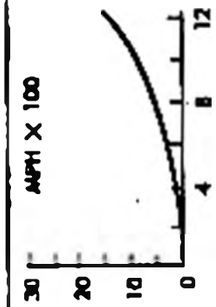
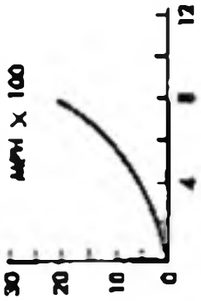
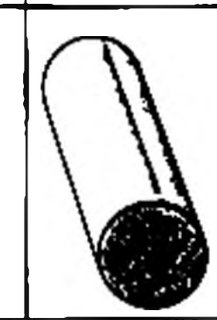
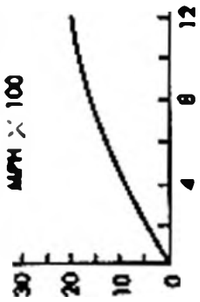
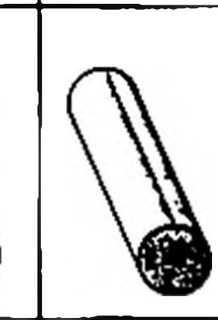
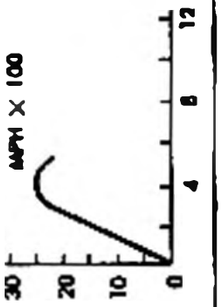
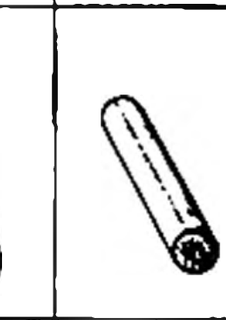
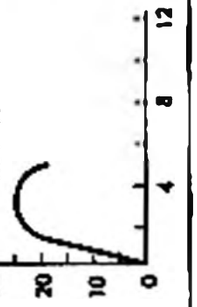
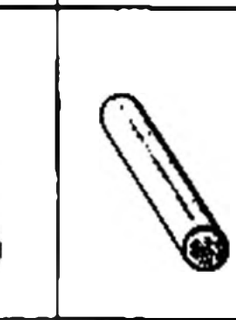
B CONVERGENT-DIVERGENT, OR "DeLaval," NOZZLE

Figure 3-10. Typical Exhaust Nozzles

AIR-BREATHING ENGINES

As mentioned in chapter 1, any jet-propulsion system that obtains oxygen from the surrounding

Table 3-1. Operational Characteristics of Various Jet-Propulsion Units

PROPULSION METHOD	DIAGRAM OF UNIT	OVERALL EFFICIENCY (%)	RELATIVE WEIGHT OF FUEL FOR GIVEN DURATION	RELATIVE FRONTAL AREA (DRAG)	PROBABLE RANGE OF FLIGHT SPEEDS
ROCKET					SUPERSONIC
RAMJET					SUPERSONIC
PULSEJET					SUBSONIC
TURBOJET					SUBSONIC AND SUPERSONIC
TURBOPROP					SUBSONIC
SUPERCHARGED RECIPROCATING ENGINE					SUBSONIC

atmosphere to support fuel combustion is known as an air-breathing engine. Pulsejets, ramjets, turbojets, turboprops, and turbofans all fall in this category. The effective operation of these engines is limited by the amount of oxygen available. They can operate only at altitudes where the oxygen content of the atmosphere is adequate for engine operation. The upper limit varies with different engines.

Before taking up the types of air-breathing engines separately, look at table 3-3. The table gives a general comparison of some of the operational characteristics of various jet-propulsion units.

Pulsejet Engines

Pulsejet engines are not currently in use by the Air Force, however, the engine is an interesting application of jet-propulsion principles and could conceivably be used again.

Pulsejet engines received their name from the pulsating manner in which the combustion process took place. This type of engine first drew international attention when used to propel the German V-1 missile, often referred to as the "Buzz Bomb." The American version of the V-1 was known as the JB-2.

COMPONENTS OF A JB-2 ENGINE. The four major parts of a pulsejet engine are the diffuser, grill assembly (air valves, air injectors, and fuel injectors), combustion chamber, and tailpipe as shown in figure 3-11.

A diffuser is defined as a duct of varying cross section, designed to convert a high-speed gas flow into low-speed flow at an increased pressure. The pulsejet air intake increases in cross section from the mouth to the grill assembly. As air flow through this section, it undergoes a decrease i velocity and an increase in pressure. This increase pressure at the inlet of the grill assembly resul in a higher combustion chamber pressure an finally in a higher thrust.

The *grill assembly unit* consists of a honeycom arrangement of air intake valves, three starting ai injectors, and nine fuel injectors.

The design of air intake valves, or "flapper" valves as they are frequently called, is the mos important feature of the grill assembly. Thes valves consist of V-section supporting member which are fitted with strips of spring steel. The open ends of the Vs are toward the diffuser section, and the closed ends point into the combustion chamber. The spring steel strips are spring-loaded in a normally closed position. They exert pressure against their respective supporting members. When the air pressure in the diffuser section is greater than combustion-chamber pressure, the flapper valves are forced open. When combustion-chamber pressure is equal to, or greater than, atmospheric ram pressure, the valves are closed. Figure 3-12 shows the construction and operation of one of the many air valves located in the grill assembly. The flapper-valve unit is so designed that any number of layers can be assembled to form a grill of desired dimensions.

The starting air injectors are connected to a pressurized external air supply. The air injector provides an atomized mixture of fuel and air for initial combustion. The three starting air injectors operate only during the starting phase—not during flight.

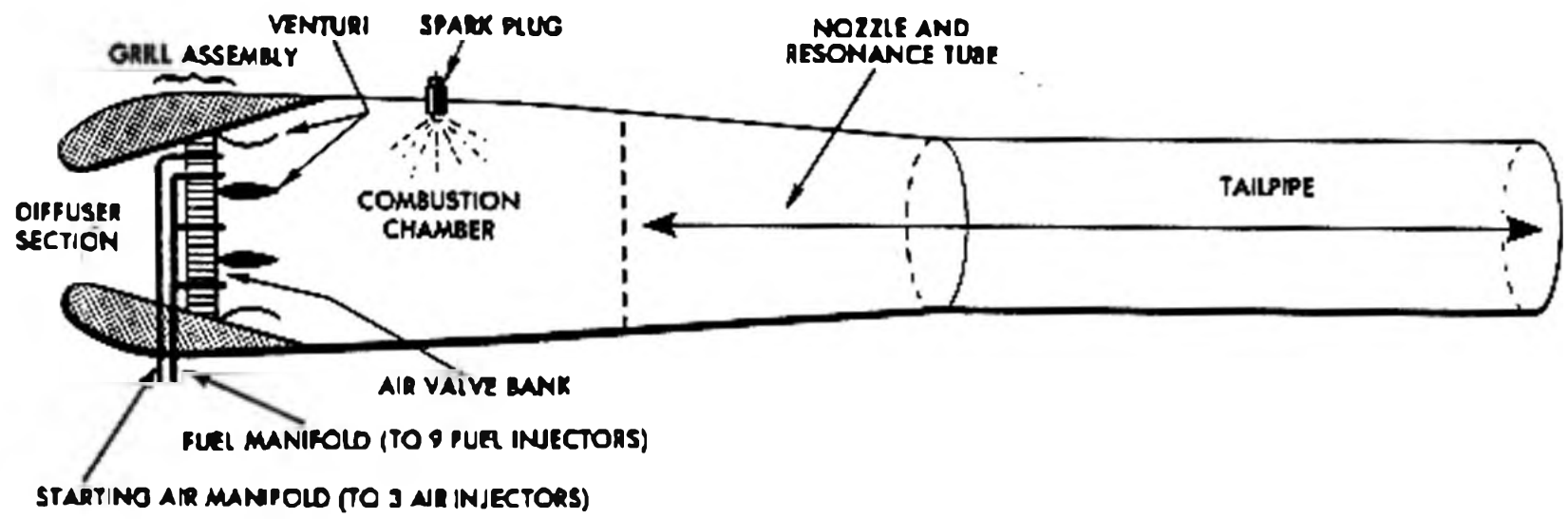


Figure 3-11. Structure of Pulsejet Engine

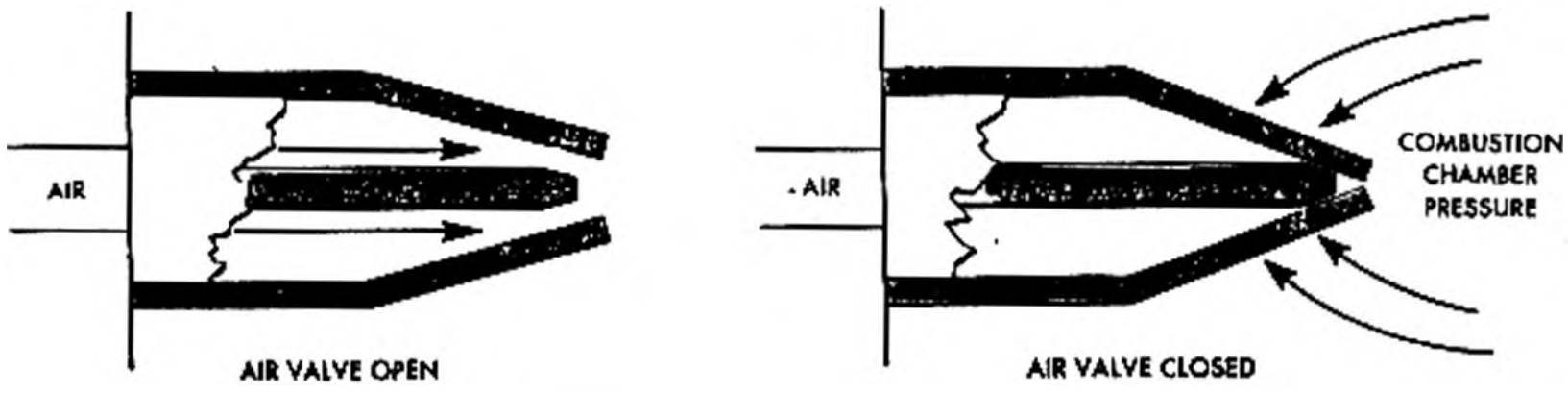


Figure 3-12. Air Valve Positions

The nine fuel injectors are fed through a delivery tube connected to the pressurized fuel supply stored within the missile.

Venturi sections, immediately aft of the grill assembly, insure proper fuel atomization and increased mixing of atomized fuel with air that enters through the flapper valves.

The *combustion chamber* contains the spark plug which provides initial ignition. Once started, combustion occurs periodically without spark-plug action.

The *tailpipe* of a pulsejet increases the velocity of the exhaust gases and determines the frequency of the combustion cycle. The frequency of operation is expressed by the following relationship:

$$\text{Frequency} = \frac{\text{Velocity of Sound}}{4 \times \text{Length of Tailpipe}}$$

The frequency of operation of the JB-2 pulsejet was approximately 50 cycles per second. In small engines with short tailpipes, the combustion cycle frequency could easily exceed 200 cycles per second.

ENGINE AND FUEL-SYSTEM OPERATION. In operation, a fuel air mixture in the combustion chamber is ignited and burns explosively causing an abrupt rise in the combustion-chamber pressure. The flapper valves close, forcing the combustion products to flow from the engines through the tailpipe. As the tailpipe is considerably smaller in cross sectional area than the combustion chamber, the gases are accelerated to a high velocity. The combustion-chamber pressure rapidly decays as the gas flows through the tailpipe. When the combustion-chamber pressure reaches ambient, the gas in the tailpipe still has a significant velocity. The momentum of this gas causes it to continue to flow for a short interval which, in turn, causes the combustion-chamber pressure to momentarily drop below ambient. Under these conditions, the flapper valves open admitting fresh air. Upon the addition of fuel to the fresh air, the mixture is ignited by the residual flame from the burning of the preceding charge. The fuel in the tank is forced through a filter, a shutoff valve, and metering unit to the fuel injectors.

Although the pulsejet is the simplest and cheapest jet engine that is capable of producing static thrust, its limitations have restricted its use. The durability of the flapper valves is limited to a few hours at most; the pulsejet is not suitable for the propulsion of supersonic vehicles and its specific fuel consumption is higher than most other jet engines.

Ramjet Engines

A ramjet engine derives its name from the ram action which makes possible its operation. This engine is sometimes referred to as the athodyd, meaning aerothermodynamic duct.

Theoretically, ramjet operation is limited to altitudes below 90,000 feet because atmospheric oxygen is necessary for combustion. The velocity that can be attained by a ramjet engine is unlimited. Theoretically, the faster it travels the better it operates and the more thrust it develops. Actually, its speed is limited at approximately Mach 5.0 because the skin temperature, as a result of frictional heating, has a harmful effect on the metals used in construction. The major disadvantage of a ramjet power plant is that the higher the operational speed for which it is designed, the higher the speed to which it must be boosted before initial starting of the engine. The speed range in which ramjets are designed to operate is the

basis for their classification. They are classified as subsonic or supersonic ramjets.

COMPONENTS OF A RAMJET ENGINE. Basically, a ramjet consists of a cylindrical shaped tube open at both ends, with a fuel-injection system inside. The engine is extremely simple in design and has no moving parts. Even though all ramjets contain the same basic components, the structure of these components must be modified to produce satisfactory operation in various speed ranges. The necessary modifications are explained later in the discussions of the subsonic, low-supersonic, and high-supersonic ramjets.

The component parts of a ramjet engine are

- Diffuser section.
- Combustion chamber with the following parts:
  - Fuel injectors.
  - Spark plugs.
  - Flame holder.
- Exhaust nozzle.

The *diffuser section* serves the same purpose in the ramjet as it does in the pulsejet. It decreases the velocity and increases the pressure of the incoming air. Since there is no wall or closed grill in the front section of this engine, the pressure increase of the ram air must be great enough to prevent the escape of the combustion gases out of the front of the engine. Diffusers must be especially designed for a given entrance velocity, or predetermined ramjet speed. The desired pressure barrier is developed only when air is entering the diffuser at the speed for which that particular diffuser was designed. The ideal situation would be a ramjet using a diffuser design which could be automatically altered in flight to conform with any ram air entrance velocity.

The *combustion chamber* is the area in which burning occurs and high-pressure gases are formed. In contrast to the pulsejet, the ramjet uses a continuous flow of fuel and air, and combustion is continuous. The fuel injectors are connected to a continuous-flow fuel-supply system, adequately pressurized to permit fuel flow in the midst of high-pressure gases which exist in the forward section of the combustion chamber. Combustion is initially started by spark-plug ignition, but once started it is continuous and self-supporting. The flame holder prevents the flame front from being swept too far toward the rear of the engine, thus stabilizing and restricting the actual burning to a limited area. The flame holder also insures that the combustion-chamber temperature will remain high enough to support combustion.

The *exhaust nozzle* performs the same function as in the case of the pulsejet or any other jet-propulsion engine.

SUBSONIC RAMJET STRUCTURE AND BASIC OPERATION. A subsonic ramjet engine cannot develop static thrust; therefore, it cannot take off under its own power. If fired at rest, high-pressure combustion gases would escape out the front as well as out the rear. For satisfactory operation, the engine must be boosted to a suitable subsonic speed so that the ram air entering the diffuser section develops a pressure barrier high enough to

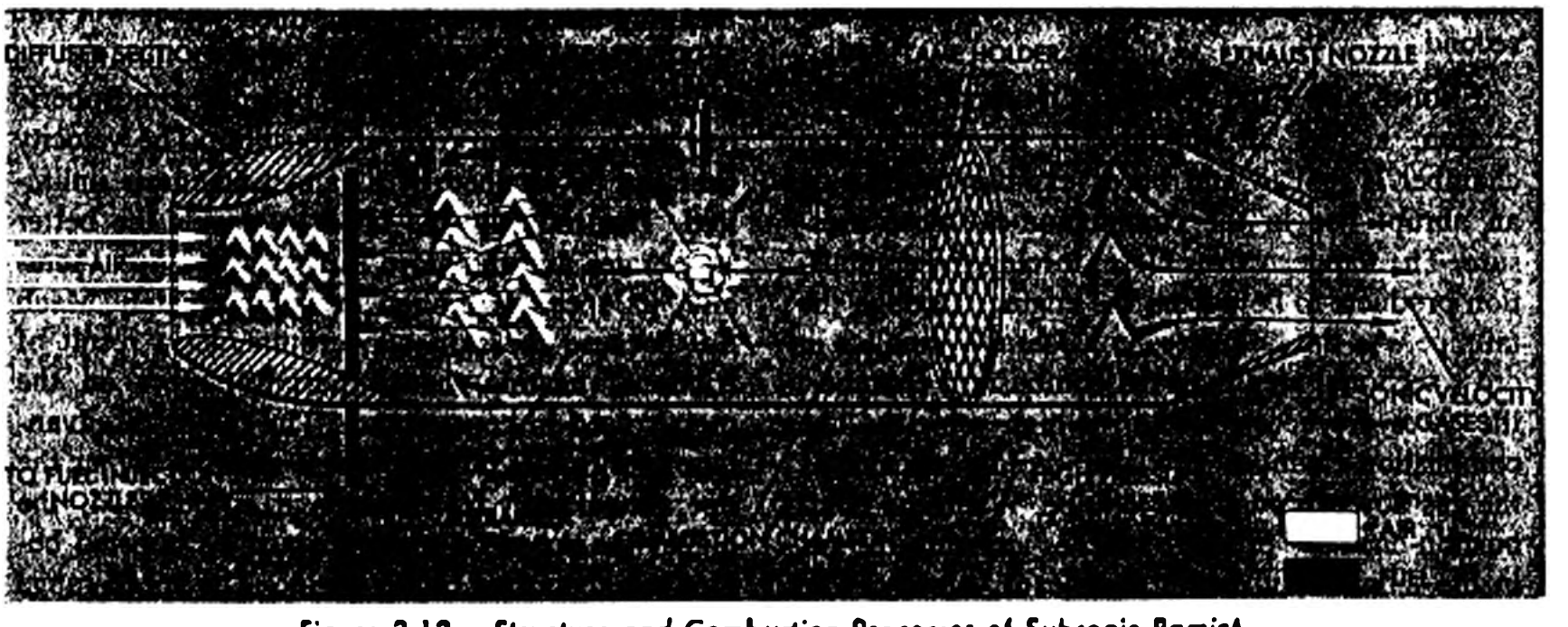


Figure 3-13. Structure and Combustion Processes of Subsonic Ramjet

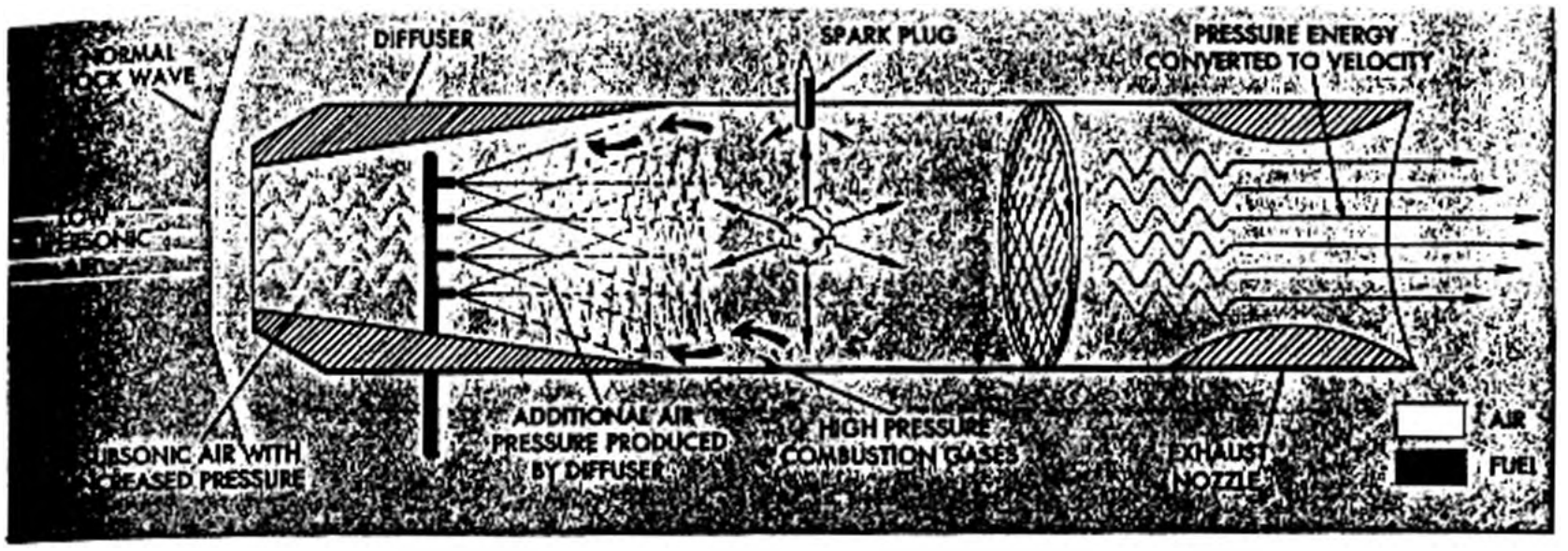


Figure 3-14. Structure and Combustion Processes of Supersonic Ramjet

restrict the escape of combustion gases to the rear only as shown in figure 3-13.

As ram air passes through the diffuser section, the velocity of the air decreases while the pressure increases. This is brought about by the increase in cross section of the diffuser (Bernouilli's theorem for incompressible flow). Fuel is sprayed into the combustion chamber through the fuel injectors. The atomized fuel mixes thoroughly with the incoming air, and the mixture is ignited by the spark plug. After initial ignition, burning is continuous and no additional spark-plug action is necessary.

The gases which result from the combustion process expand in all directions as shown by the arrows in the central portion of the combustion chamber. The gases, expanding in the forward direction, are stopped by the barrier of high-pressure air and the internal sloping sides of the diffuser section, as indicated in the diagram by the short, wide, black arrows. The only avenue of escape remaining for the combustion gases is through the exhaust nozzle, and here another important energy conversion occurs. The pressure energy of the combustion gases is converted to velocity. To be more specific, the high-pressure combustion gases enter the exhaust nozzle with a velocity which is below the local speed of sound. While passing through the convergent-type nozzle, the pressure energy of the gases decreases and the velocity increases up to the local speed of sound at the exhaust nozzle exit.

Thrust is developed in the ramjet as a result of the unbalance of forces acting in forward and rearward direction. The bombardment of combustion gases against the sloping sides of the diffuser and the ram air barrier exerts a force in the forward direction. This forward force is not balanced by the combustion gases which escape through the exhaust nozzle. The unbalanced force determines the thrust.

Low-Supersonic Ramjet Structure and Basic Operation. To operate, a low-supersonic ramjet must be boosted to a supersonic speed approximately equal to operating speed before ignition. When the forward speed of the ramjet becomes supersonic, a normal shock wave forms at the entrance to the diffuser section as indicated in figure 3-14. On the upstream side of this normal shock wave, the free-stream air is moving at a low-supersonic velocity. As the supersonic air passes through the shock wave, its velocity drops abruptly to a subsonic value with a corresponding increase in pressure. The shock-wave formation produces a valuable increase in air pressure at the diffuser entrance. As the compressed subsonic air flows through the diverging-type diffuser section, an additional increase in pressure and decrease in velocity is produced. The highly compressed air is now ready for combustion.

The combustion process is essentially the same as the combustion process in a subsonic ramjet. Fuel is mixed with the highly compressed air, the mixture is ignited initially by means of a spark plug, and burning is continuous thereafter. The potential energy possessed by the combustion gases is converted into kinetic energy by the exhaust nozzle.

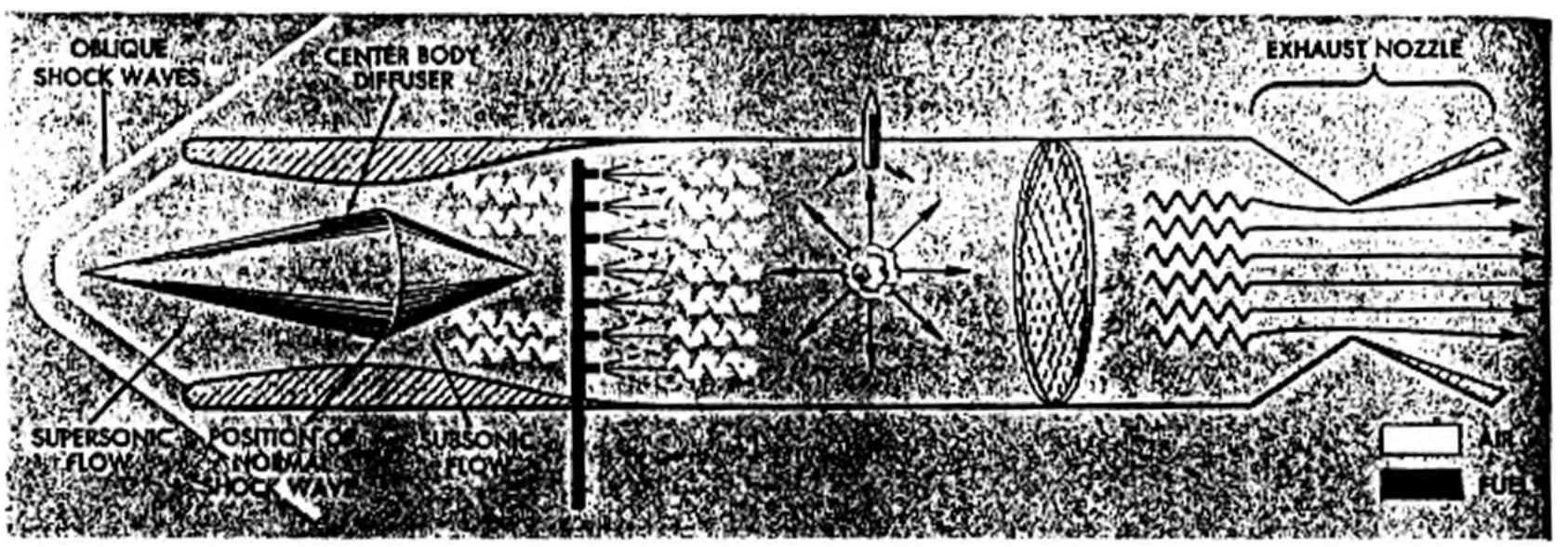


Figure 3-15. Structure and Combustion Processes of High Supersonic Ramjet

The convergent-divergent nozzle shown in the diagram allows the gases to exceed the local speed of sound. With proper design modifications, the ramjet engine can travel efficiently at supersonic speeds.

HIGH-SUPERSONIC RAMJET STRUCTURE AND BASIC OPERATION. Assume a ramjet is desired that will travel at much higher supersonic speeds, say Mach 2.0. At speeds in the neighborhood of Mach 2.0, shock waves formed at the diffuser inlet would be oblique and not normal. Air velocity in front of an oblique shock wave is high supersonic. When supersonic free-stream air passes through an oblique shock wave, an increase in pressure and a decrease in velocity occur, but the velocity is still supersonic. For example, air with a free-stream velocity of 1500 mph may pass through an oblique shock wave and still have a velocity of 900 mph. Also, when supersonic air flows through divergent-type diffuser sections, the velocity of that air increases and the pressure decreases. The diffuser design for high-supersonic ramjets must be modified so that in progressing from diffuser inlet to combustion-chamber entrance, the obliqueness of the shock waves successively decreases until a normal shock wave followed by subsonic flow is realized.

The energy transformation is achieved by using a diffuser design of the type shown in figure 3-15. The center body diffuser decreases the obliqueness of the shock waves, allowing supersonic air to flow inside the diffuser inlet.

As supersonic flow passes through the convergent section of the diffuser, the velocity is steadily decreased and the pressure correspondingly increased. At some predetermined point in the diffuser, air velocity approaches the sonic value and a normal shock wave forms. When low-supersonic air flows through a normal shock wave an abrupt decrease in velocity and increase in pressure results. Now, the subsonic air produced by the normal shock wave flows through the divergent section of the diffuser where it undergoes an additional velocity decrease and pressure increase. Here again the diffuser has achieved a pressure barrier at the entrance to the combustion chamber. The combustion process is the same as that described for the subsonic ramjet. The exhaust nozzle shown in the diagram is the convergent-divergent type designed to produce supersonic flow at the exhaust-nozzle exit.

A ramjet is designed to operate best at some given speed and altitude. The pressure recovery process in a diffuser designed for oblique shock waves is more efficient than in the case of diffusers

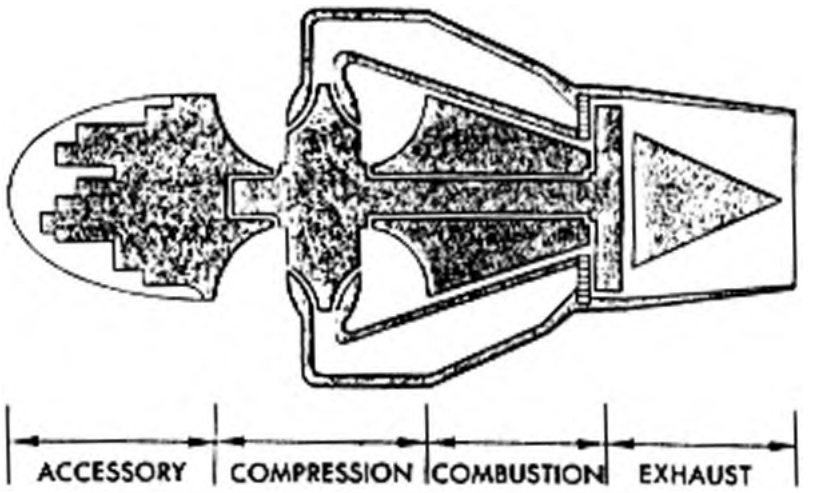


Figure 3-16. Centrifugal-Flow Turbojet

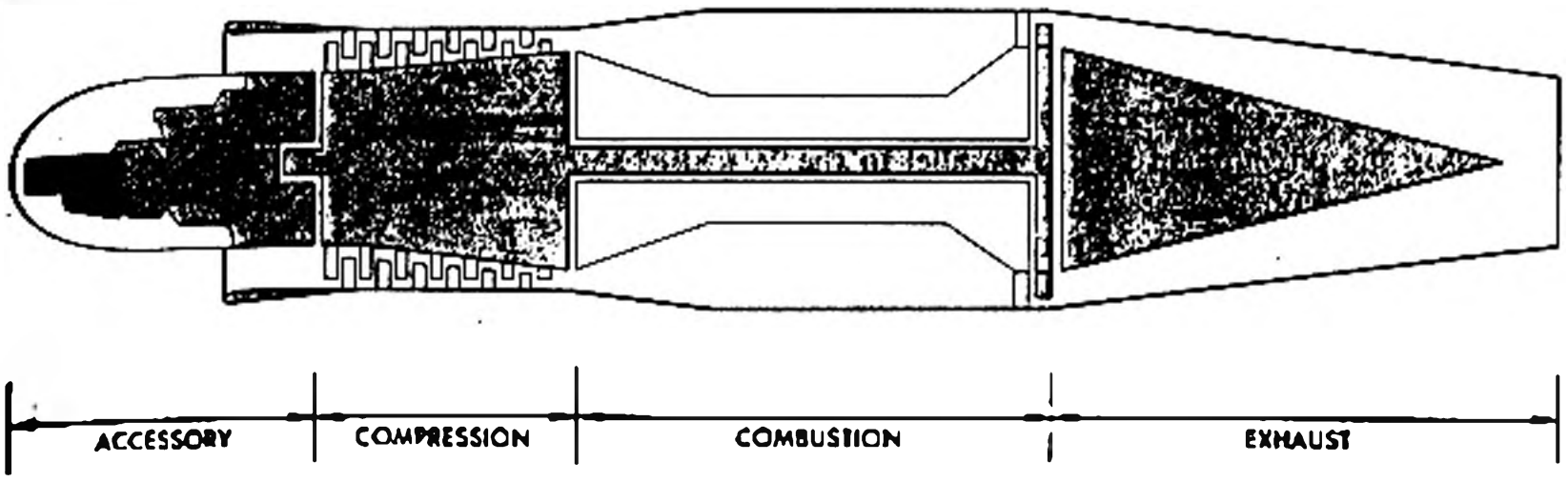


Figure 3-17. Axial-Flow Turbojet

designed for subsonic flow or single normal shock waves. The ramjet engine operates best at high supersonic speeds.

Turbojet Engines

A turbojet engine is an air-dependent thermal jet-propulsion device. The turbojet derives its name from its design, which uses an exhaust-gas-driven turbine wheel to drive the compressor.

CLASSIFICATION OF TURBOJETS. Turbojets are classified into two general groups, depending upon the type of compressor used. These types are the

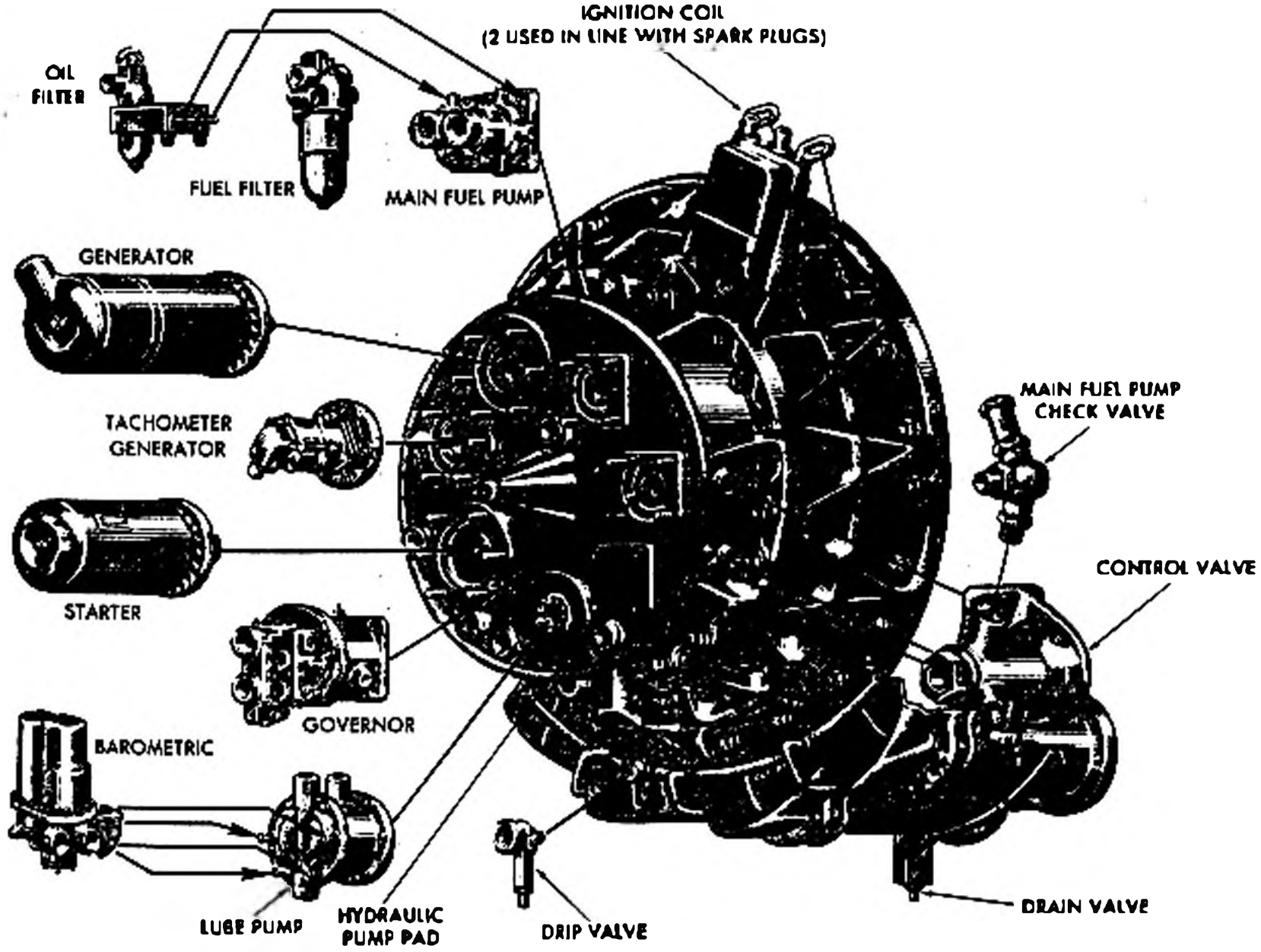


Figure 3-18. Accessories of Turbojet Engine

centrifugal-flow compressor, used when the direction of flow is perpendicular to the longitudinal axis of the engine, and the axial-flow compressor, used when the direction of flow is parallel to the longitudinal axis of the engine. The names, based upon these processes, are centrifugal-flow turbojets and axial-flow turbojets. Figures 3-16 and 3-17 show each type of turbojet engine.

COMPONENT PARTS OF TURBOJETS. The operation of both types of turbojet engines is basically the same. Both consist of the following major sections: accessory section, compressor section, combustion section, and exhaust section.

Accessory section. The accessory section serves as a mounting pad for accessories, various engine components, and the front engine balancing support. The accessories are those units not essential to engine operation, such as the generator, hydraulic pump, starter, and tachometer. The components are the units of the fuel system and oil system that directly affect engine operation. The accessory case serves as the engine oil reservoir and houses the accessory gear-train cage. The accessories comprising the accessory section of a typical turbojet engine are shown in figure 3-18.

Compressor section. The primary functions of the compressor section in turbojets are to receive, compress, and distribute the large masses of air to the combustion chambers.

The *centrifugal compressor* illustrated in figure 3-19 consists mainly of a stator, often referred to

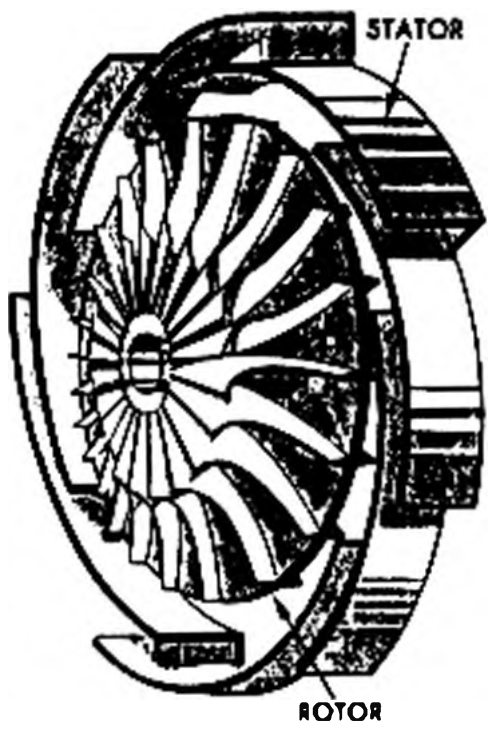


Figure 3-19. Centrifugal-Flow Compressor

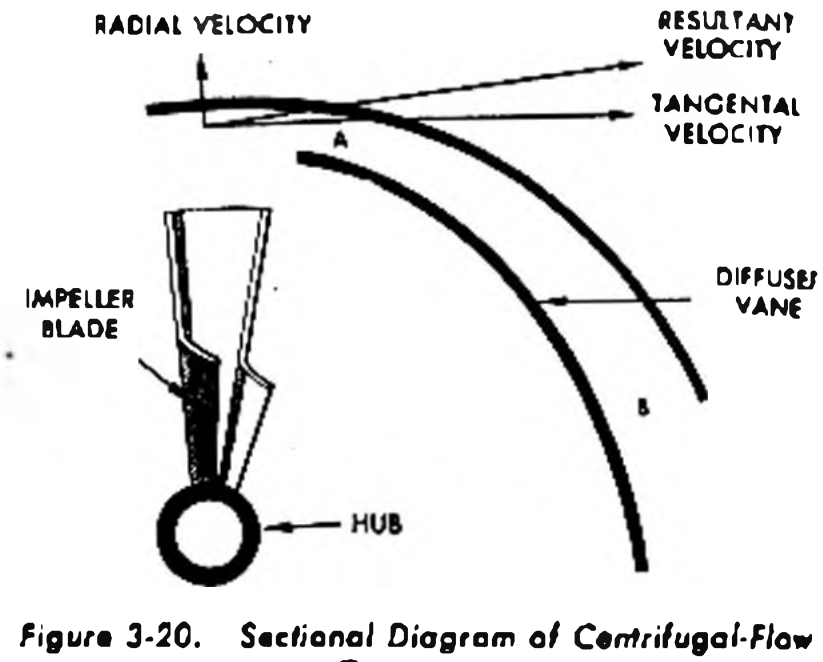


Figure 3-20. Sectional Diagram of Centrifugal-Flow Compressor

as a diffuser vane assembly, and a rotor or impeller. The rotor consists of a series of blades which extend radially from the axis of rotation. The double-faced type is commonly used, allowing air to enter on both sides. As the rotor revolves, air is drawn in, whirled around by the blades, and ejected at a high velocity created by centrifugal force. The stator consists of diffuser vanes which compress, as well as direct, the air into the various firing chambers. As the air leaves the impeller wheel, it has a large resultant velocity which directs the air into the diffuser vanes. The energy, which the masses of air acquire in the rotor as velocity, is converted into pressure, by the process of diffusion into a larger space. The air increases in pressure and decreases in velocity as it moves from A to B in figure 3-20.

The *axial compressor* is similar to a propeller. The rotor again consists of a series of blades which are set at an angle and extend radially from the central axis. As the rotor of the axial compressor turns, the blades impart energy of motion, in both a tangential and axial direction, to the ram air entering through the front of the engine. A typical stator for an axial-flow compressor consists of a series of blades arranged in a circle around the inside wall of the compressor casing immediately behind the rotor, and extending inward toward the central axis of the engine. The stator does not rotate. The blades are set in an angular fashion, so as to turn the air thrown off the trailing edge of the first-stage rotor blades and redirect it into

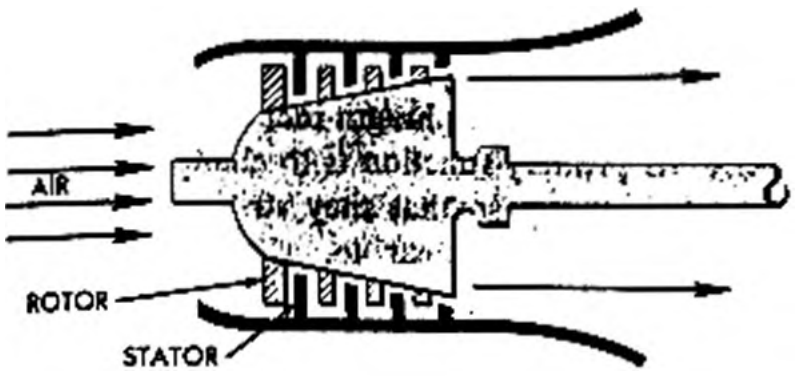


Figure 3-21. Four-Stage Axial Flow Compressor

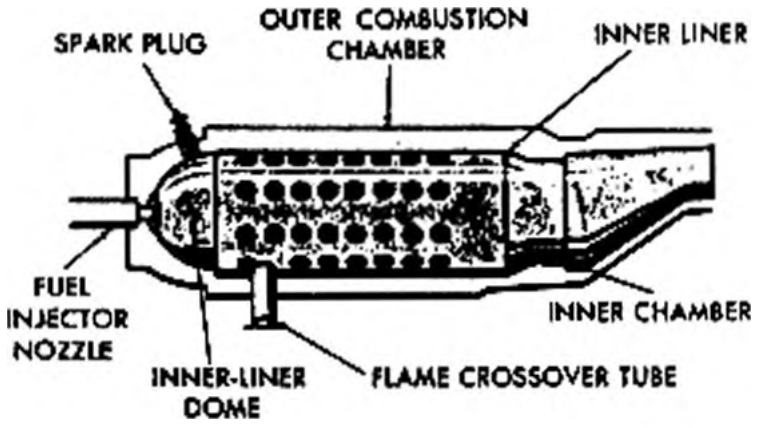


Figure 3-23. Basic Construction of Turbojet Combustion Chamber

the path of the second-stage rotor blades. One rotor and one stator constitute a single-stage compressor. A number of rotors and stators assembled alternately make up a multistage compressor. An example of a multistage compressor is illustrated in figure 3-21. Air from the first row of compressor blades is accelerated and forced into a smaller space. The added velocity gives the air greater impact force. This impact forces air into a smaller space causing the air to become denser. This results in greater static pressure. The above cycle of events is continued through each stage. By increasing the number of stages, the final pressure can be raised to any desired value. For visual purposes, the stator is shown in figure 3-22 with blades extending from a central drum outward. The most frequently used construction is with the blades extending inward as previously described.

In both centrifugal- and axial-flow compressors, the pressure ratio of the gases is stated as exhaust pressure to intake pressure. As the speed of the compressor rotor is increased, the volume of the air passed by the compressor unit increases. At high-pressure ratios, the volume of air, passed for a given rotor speed, decreases. Above a given pressure ratio, the efficiency of operation for a given speed drops sharply. This drop is due to the development of pressure pulsations set up in the air as it passes through the compressor.

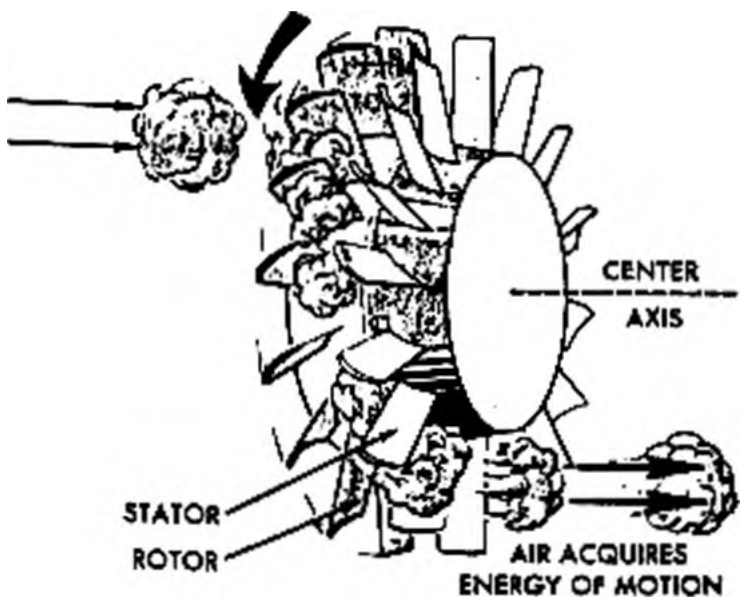


Figure 3-22. Single-Stage Axial-Flow Compressor

Combustion section. The combustion section includes combustion chambers, spark plugs, a nozzle, diaphragm, and a turbine wheel and shaft. The combustion chambers, or burners, in both types of turbojet engines have the same function and produce the same results. They do differ in size and number, depending upon the type of engine. One particular centrifugal-flow turbojet has 14 chambers with a spark plug located in chamber No. 7 and another in chamber No. 14. A widely used axial-flow engine has eight chambers with two spark plugs located in chambers directly opposite each other. In either case, each combustion chamber has the following parts: outer combustion chamber, inner liner, inner-liner dome, flame crossover tube, and fuel-injector nozzle. Figure 3-23 points out these components.

The outer combustion chamber serves to retain the air so that a high-pressure supply is available to the inner liner at all times. This air also serves as a cooler jacket. The inner liner houses the area in which fuel and air are mixed and burned. Many round holes in the inner liner allow the air to enter and mix with the fuel and high-temperature combustion gases. The forward end of the inner liner is allowed to slide over the dome to accom-

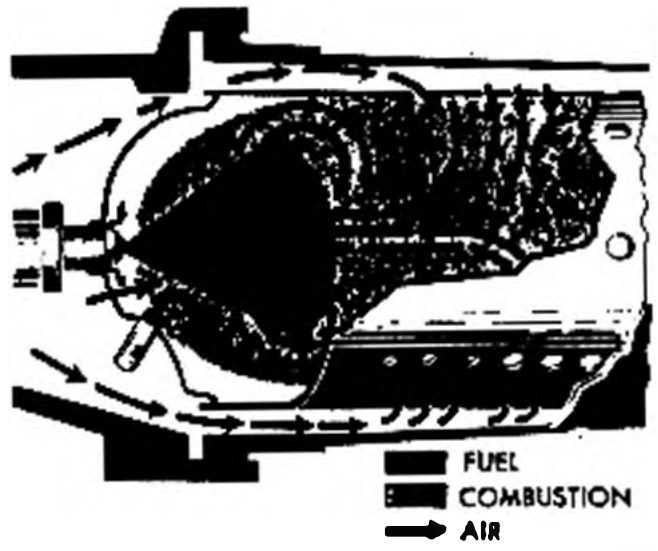


Figure 3-24. Propellant Flow and Combustion Process in Combustion Chamber of Turbojet

modate expansion and contraction. The aft ends of the burners are convergent to increase the velocity of the gases just before they pass through the nozzle diaphragm. The flame crossover tube connects one chamber to the next, allowing ignition to occur in all chambers after the two chambers containing spark plugs have been fired. Figure 3-24 illustrates the injection paths of fuel and air into the combustion chamber as well as the burning area.

Exhaust section. The exhaust section, shown in figure 3-25, consists primarily of a nozzle and an inner cone. This assembly straightens out the turbulent flow of the exhaust gases caused by rotation of the turbine wheel and conveys these gases to the nozzle outlet in a more perfect and concentrated gas-flow pattern. The exhaust-nozzle diaphragm is composed of a large number of curved blades standing perpendicular to the flow of combustion gases and arranged in a circle in front of the turbine wheel. By acting both as a restrictor and director, this diaphragm increases the gas velocity. Its primary function is to change the direction of the gases so that they strike the turbine-wheel vanes at, or near to, a 90° angle. This diaphragm does not rotate. The impact of the high-velocity gases against the buckets of the turbine wheel causes the wheel to rotate. The turbine-wheel shaft is coupled to the compressor-rotator-assembly shaft. Part of the energy of the exhaust gases is transformed and transmitted through the shaft to operate the compressor and engine-driven accessories. Note figure 3-26.

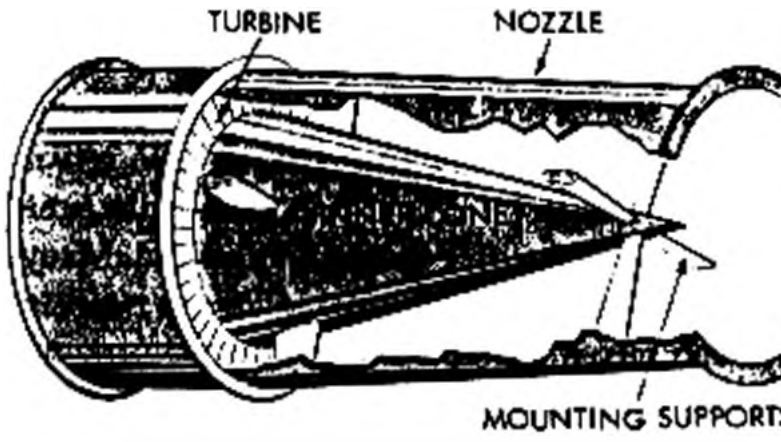


Figure 3-25. Exhaust Assembly of Turbojet

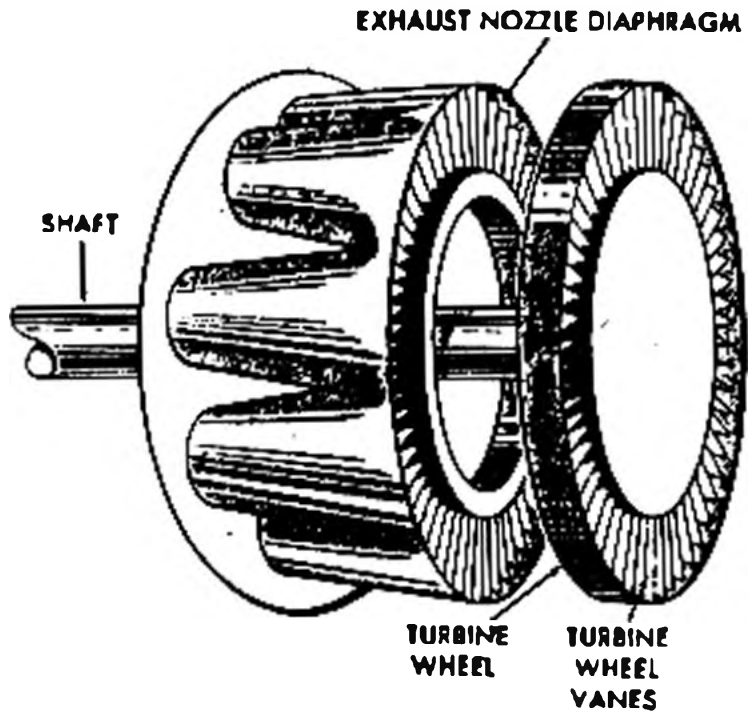


Figure 3-26. Exhaust-Nozzle Diaphragm and Turbine Assembly

TURBOJET OPERATION. In general, turbojet operation may be summarized as follows: The rotor unit of the compressor is brought up to maximum allowable speed by the starter unit, which is geared to the compressor shaft for the starting phase. Air is drawn in from the outside, compressed, and directed to the combustion chambers. Fuel is injected through the fuel manifold under pressure and mixes with the air in the combustion chambers. Ignition occurs first in the chambers containing the spark plugs and then in the other chambers an instant later by way of the flame crossover tubes. High-pressure combustion gases, plus the unburned coolant air, pass through the exhaust-nozzle diaphragm and strike the turbine blades at

the most effective angle. The greater portion of the energy of the exhaust stream is absorbed by the turbine, resulting in a high rotational speed. The remainder is thrust. The turbine wheel transmits energy through the coupled turbine and compressor-rotor shafts to operate the compressor. Once started, combustion is continuous.

AFTERBURNERS. Afterburners were developed to satisfy special conditions in which large bursts of extra thrust are needed for a short period of time. For instance, an afterburner is needed for acceleration during takeoff runs and in climbs. This additional thrust may be obtained by burning fuel in the tailpipe section. That portion of air which served only as a coolant for the main combustion chambers is sufficient to support combustion of the additional fuel. The added thrust is large, but the overall efficiency of the turbojet decreases because the specific fuel consumption is greatly increased. The basic construction of a typical afterburner is shown in figure 3-27. Note the use of the clam-type exhaust nozzle, the area of which is adjustable to compensate for the changing flow characteristics that occur when the afterburner is cut in. An afterburner boosts thrust nearly 30 percent for takeoff and from 70 percent to 120 percent when the missile is traveling approximately 600 mph.

TURBOPROP ENGINE. The turboprop engine is a jet engine geared to a propellor. In operation, air enters through the air inlets or scoops and is ducted directly to the compressors where it is compressed, heated, and expanded to drive the turbine. The hot gases then pass out through the exhaust stack. The propeller and compressors are driven by the turbine. The turbine drives the propeller through a gear train. The compressors are keyed to the turbine drive shaft.

TURBOFAN ENGINE. The turbofan engine has the same basic design as the axial flow turbojet. The difference is in the addition of a fan and the provision for a secondary flow of air. This addition provides increase in thrust and decrease in fuel consumption for greater operating efficiency. The fan is located ahead of the first compressor stage. Air drawn through it is compressed slightly and is thereby heated. A portion of this air is ducted directly to the outside, bypassing the main compressor intake. The heated air expands as it passes through the duct and imparts thrust as it is exhausted into the airstream. The major portion of air entering the inlet and passing through the fan is conducted directly to the main compressor intake where it is further compressed and heated and passed on in the normal turbojet operating cycle.

While air-jet engines are limited in area of operation because of their dependency upon the surrounding atmosphere for oxygen, rockets are not restricted by such conditions.

ROCKETS

Any missile for which the power for operation depends only on the mass and energy stored within the vehicle itself is called a rocket. In rocket operation, chemical reaction occurs at a very rapid rate. This results in higher temperatures, higher operating pressures, and higher thrust development than in the case of the jet engines. Control over this action is desirable and is sometimes necessary to distribute the available energy over a longer period of time. Because of the high pressures which are

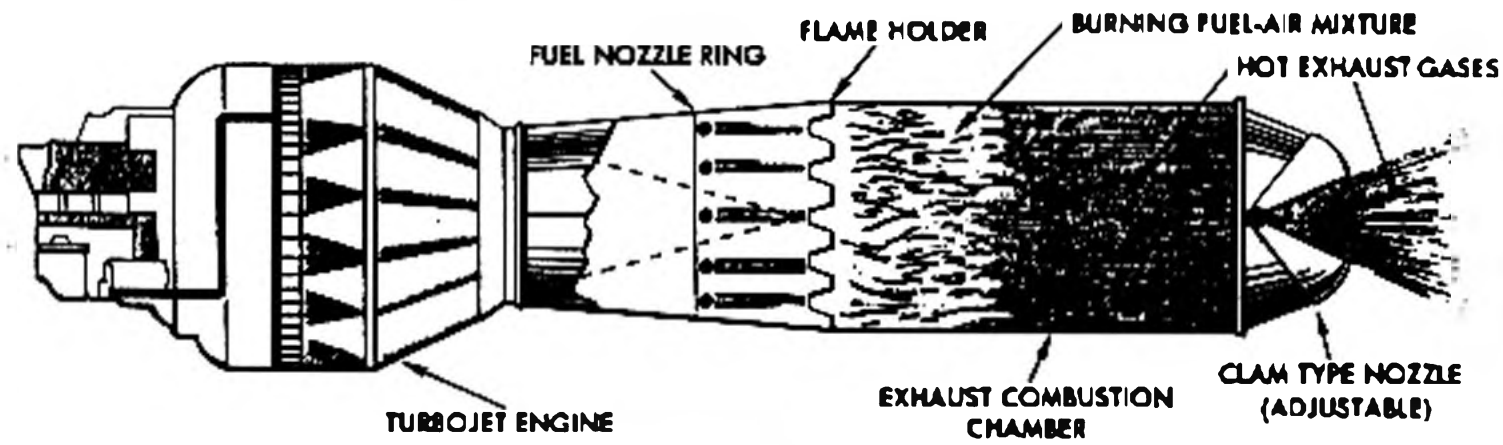


Figure 3-27. Turbojet Afterburner

developed in rockets, the convergent-divergent (DeLaval) type nozzle is used so that more of the energy can be extracted from the gases after they have passed the throat section. The basic principles involved in the action of other jet-propulsion units also apply to rockets.

Classification of Rockets

The physical state of the propellant used determines whether a rocket is of the solid or liquid type. By definition, a rocket is a thrust-producing system which derives its thrust from the ejection of hot gases generated from material carried in the system. When the thrust-producing system is self-contained, enclosed within a housing of some type, and is designed to be propelled through aerospace to a target, it is considered to be a missile.

The term rocket engine is applied to a rocket propulsive device that is relatively complicated in its workings, as distinguished from a simple rocket motor of the solid fuel type. The large liquid fuel rocket has elaborate pumping equipment, pressure chambers, and fuel lines, with electrical and hydraulic equipment to govern its operation. The term rocket engine is often applied to the large solid fuel rocket also, although the term "motor" is associated with the solid rocket that is characterized by a short burning time, simple design, heavy construction, and non-intermittent operation. Such a system is primarily used as a booster unit (JATO) or as a powerplant for short-duration, high-speed missiles. In this manual, the liquid propellant rocket propulsion system is refered to as an engine while a solid propellant rocket propulsion system is referred to as a motor without regard to the size or complexity of the system involved.

SOLID ROCKETS. A solid rocket motor consists of four major components; propellant grain, case, igniter, and exhaust nozzle. A typical solid rocket motor illustrating these components is shown in figure 3-28.

The solid propellant rocket motor is a comparatively simple powerplant. The case serves two purposes. First, it acts as a storage container for the propellant, and secondly, it serves as a chamber in which burning takes place. The products of the combustion process or reaction gases are exhausted through one or more exhaust nozzles. Cases are usually cylindrical in shape with elliptical or spherical ends. The case has to withstand starting pressure, surges, and often severe heating.

The temperature of the wall, which carries the pressure load, determines the physical property requirements of the stress-carrying members and, therefore, largely determines the design and weight of the case. As the wall becomes hotter its ability to carry loads diminishes. It is good design practice to limit the heating of the walls because this will reduce the required wall thickness. Therefore, the internal case wall is normally thermally protected by an insulating material.

The case design must also provide for the attachment of a nozzle, one or more support points for mounting handling hooks, and provisions for loading the grain. The primary criterion for case

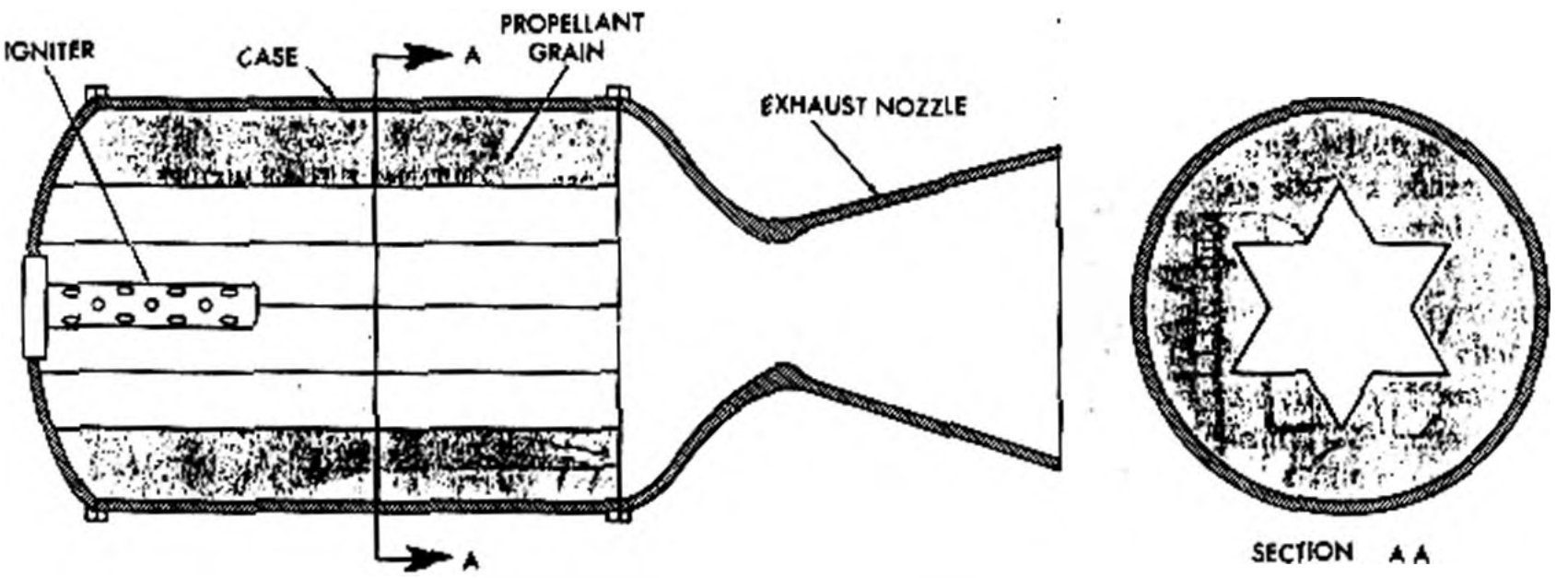


Figure 3-28. Construction of Solid Rocket Motor

material selection is normally the strength to weight ratio of the material. Currently, three of the more popular case materials are titanium, AISI 4130 steel, and Ladish D6aC steel. In addition, fiberglass filament-wound cases have been manufactured for rocket combustion chambers.

The igniters used for solid propellant rockets are almost exclusively of the pyrotechnic type. Three main components comprise the igniter, namely, some form of electrically fired squib, main igniter charge, and the case. The squib consists of two lead wires which are insulated from each other and which are connected by a fine, high-resistance wire. The latter is surrounded by a primer propellant which is sensitive to temperature and will ignite readily and burn when heated. The main igniter charge is immediately adjacent to the primer. It produces a hot flame which ignites the main rocket motor propellant grain. Various materials have been used as the main igniter charge, from black powder to a mixture of metals with oxidizers such as potassium perchlorate.

The exhaust nozzles of solid propellant rockets, like the cases, are uncooled. For low-temperature propellants and short-burn durations, simple steel nozzles are adequate. For longer durations and high-temperature metallized propellants, a composite construction is normally used. Reinforced plastics and phenolics, which carbonize slowly or char, are commonly used in the inlet section as well as the diverging section of the nozzle. Inserts of special materials are put into the throat section. Common throat materials are tungsten, graphite, and various carbides.

The structural support shell is normally constructed from steel. To prevent the entry of moisture and to aid ignition, a blow-out closure across the exit section of the nozzle is usually provided.

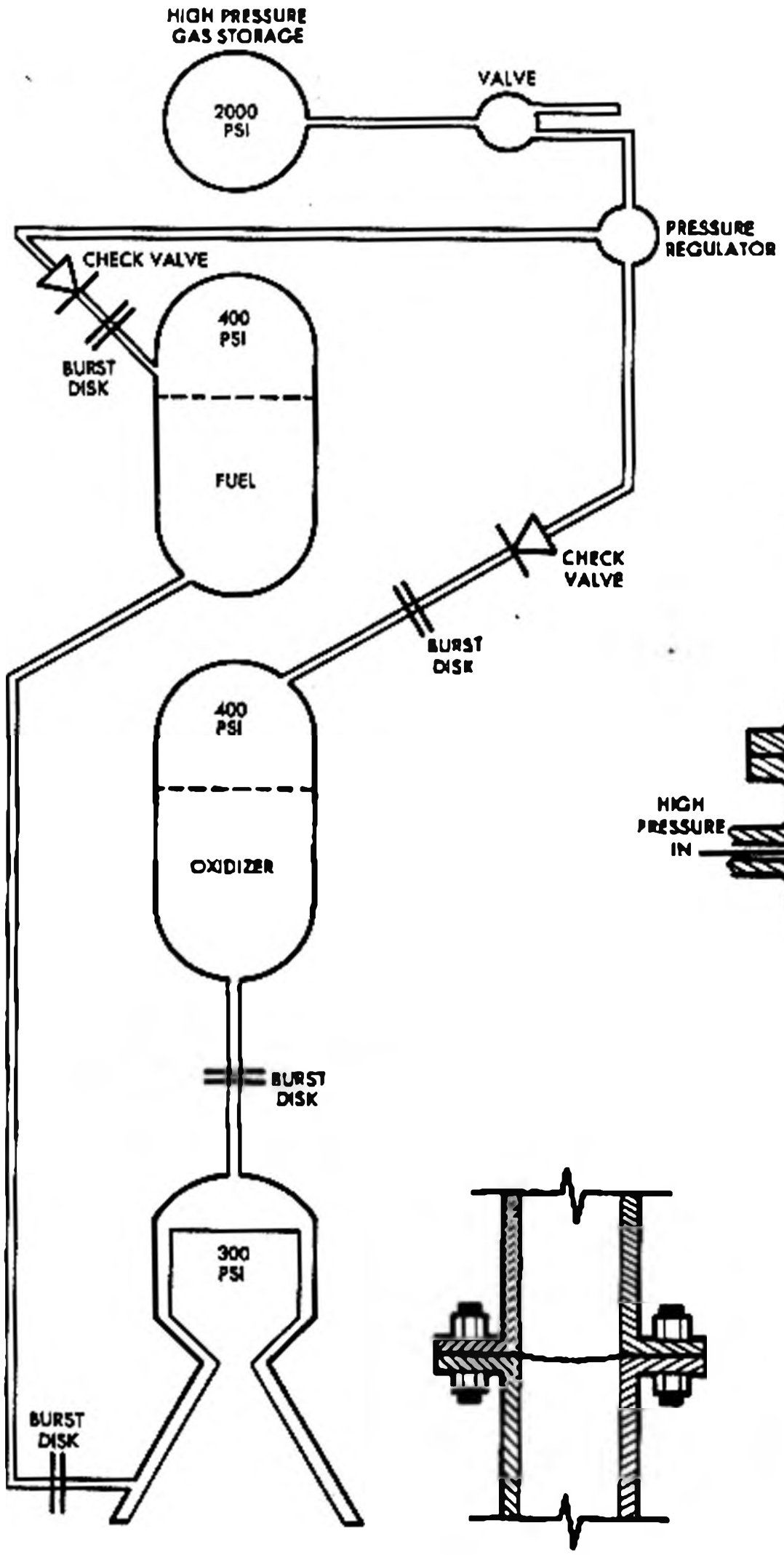
In addition to the four major components of a solid rocket motor, there are numerous secondary components which may or may not be used. Many applications of solid propellant rockets, such as upper stages of ballistic missiles, require that the thrust be precisely terminated at a specified flight velocity. Several thrust termination schemes exist, the most common using the opening of holes called blowout ports in the forward closure wall of the chamber.

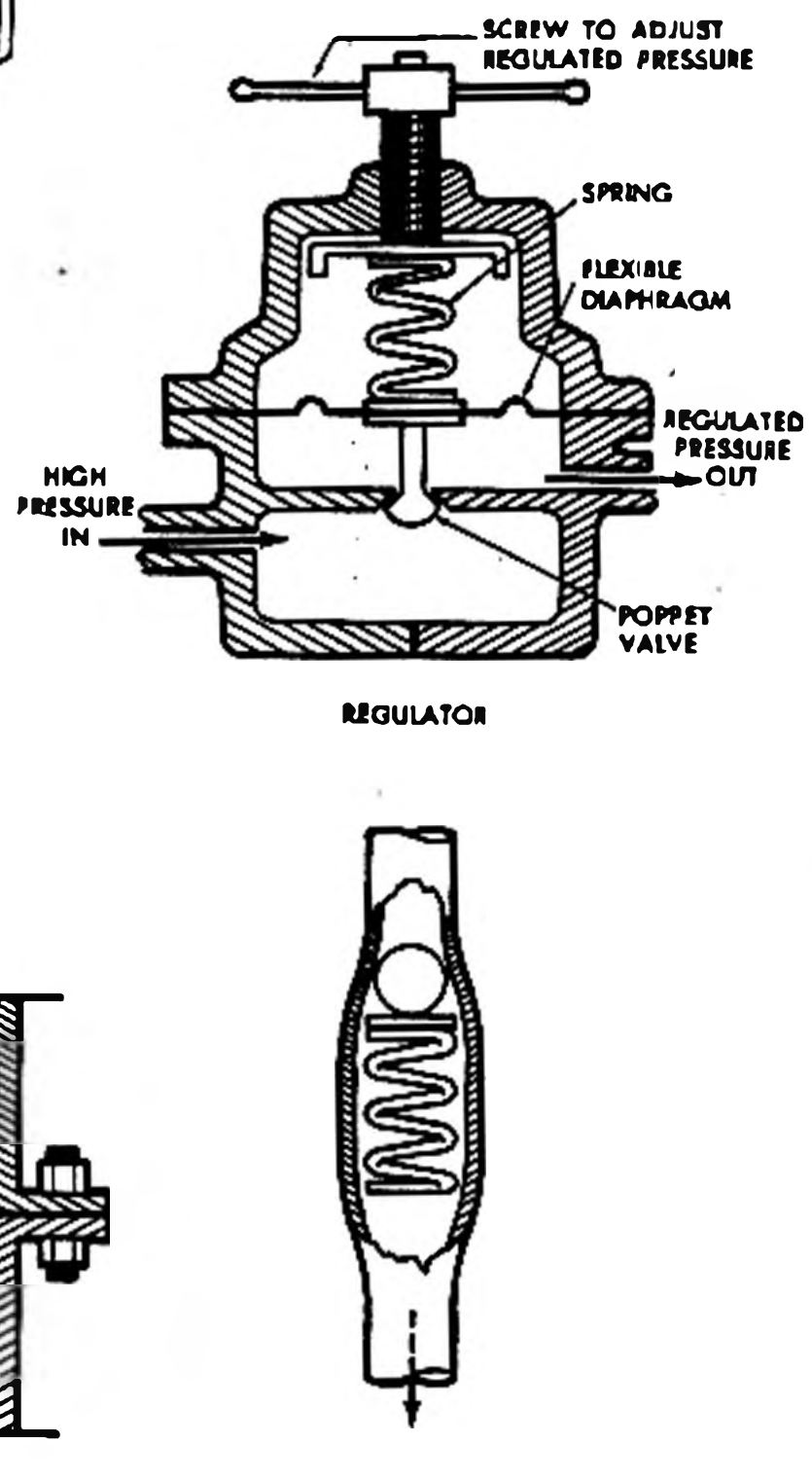
Several different methods of controlling the thrust vector direction of solid propellant rockets have been used in the past. The most common methods include gimbaling or hinging of the nozzle, the use of mechanical flow spoilers, and secondary fluid injection into the exhaust gas stream.

It is also possible to control the thrust magnitude of a solid propellant motor. The most common technique of controlling the thrust magnitude is through the use of a pintle valve located in the throat of the exhaust nozzle. Modulation of the valve changes the throat area and, therefore, affects the thrust level through chamber pressure variations.

LIQUID ROCKETS. The major components of a liquid rocket system are: the propellants, propellant tanks, feed system, injector, thrust chamber, and exhaust nozzle. If the propellants are non-hypergolic, an ignition system is also required. Liquid propellant rocket systems are generally categorized by the type of feed system, as pressure-fed or pump-fed.

Pressure-fed systems. In a pressure-fed system, illustrated in figure 3-29, the propellant tanks themselves are pressurized to a pressure which is a few hundred pounds per square inch higher than the desired thrust chamber operating pressure. This pressure may be provided by a non-reactive gas such as helium or nitrogen which is stored at a pressure of several thousand psi. The high-pressure gas passes through appropriate pressure regulators and check valves before being admitted to the propellant tank. Frequently the propellant is isolated from the pressurization system by burst disks, thin metal diaphragms in the pressurization lines which rupture under a high differential pressure when the system is activated. The burst disks prevent corrosion or contamination of the pressurization system by the propellant vapors which in many cases are highly corrosive. Burst disks or diaphragms, as they are sometimes called, are frequently employed on the propellant discharge lines as well. In this case they are sometimes mechanically ruptured by the actuation of the propellant valves. Burst disks are usually scored with grooves radiating from the center in such a manner as to cause them to fold against the pressurization or propellant line without any pieces coming loose. Since the pressurizing gas tanks are quite heavy, larger pressurized systems generate high pressure gas as needed rather than





REGULATOR

BURST DISK

CHECK VALVE

Figure 3-29. Stored Pressure-Feed System

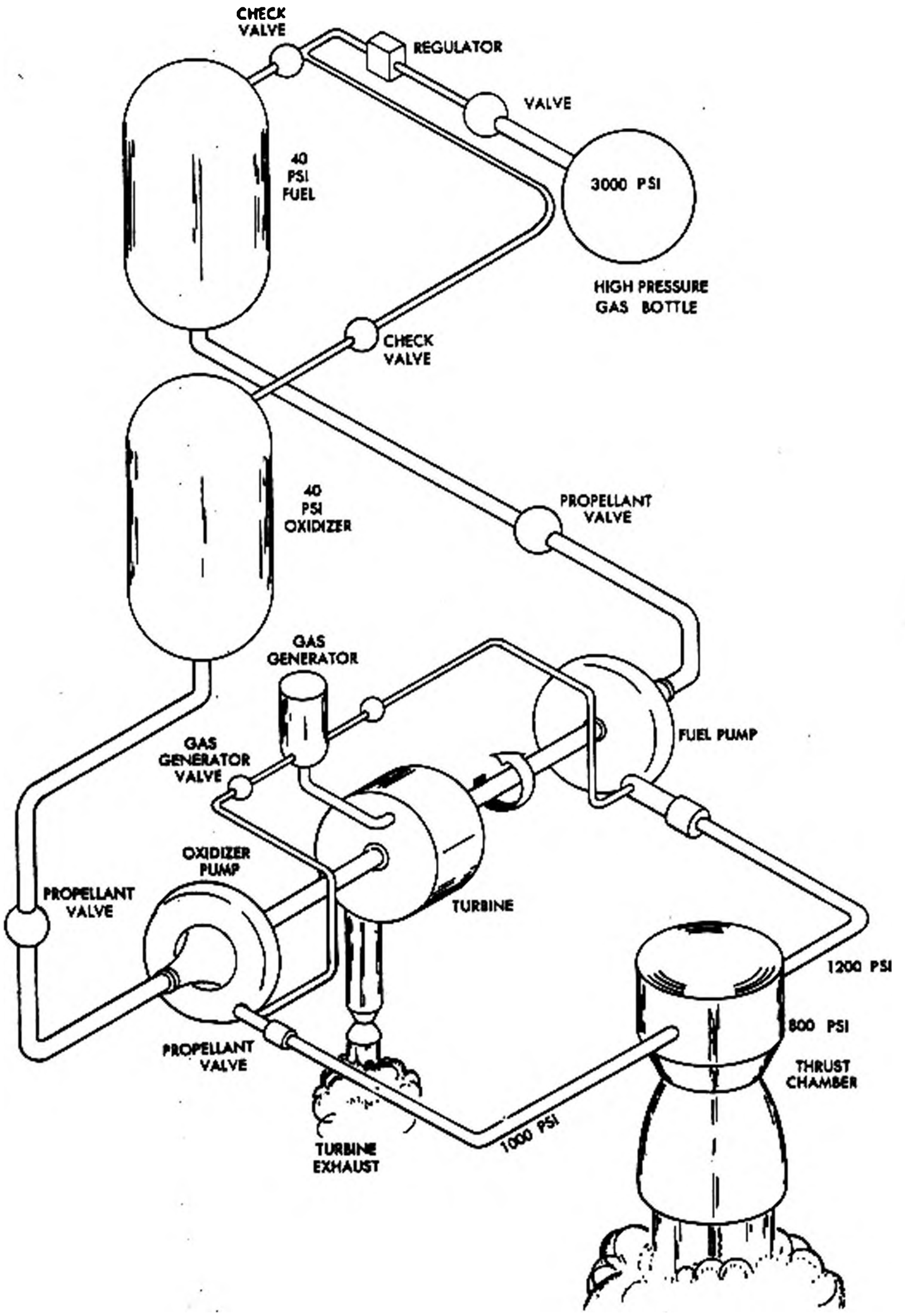


Figure 3-30. Typical Pump-Feed System

storing it. Either solid or liquid propellant may be used to produce the gas. Since most solid propellants produce a fuel-rich gas; extreme care must be exercised when they are used directly to pressurize an oxidizer tank. If the fuel-rich gas reacts immediately with the oxidizer vapor without causing excessive temperature rise, the system can be used. The liquid propellant Bullpup and liquid Sparrow missiles use this system. A more conservative approach employs a thin flexible metal diaphragm to separate the propellant and pressurizing gas. Liquid propellant gas generators are usually adjusted to operate fuel-rich for pressurizing the fuel tank and oxidizer-rich for the oxidizer tank. One of the more interesting methods of pressurizing hypergolic propellant combinations is Main Tank Injection or MTI as it is called. In this case, a carefully controlled quantity of fuel is injected into the oxidizer tank and likewise oxidizer injected into the fuel tank. The pressurizing gas is thereby generated within the propellant tank itself. Since the propellant tanks of a pressure-fed system must withstand pressures significantly higher than the thrust chamber, those tanks tend to become quite heavy, particularly in the larger missiles such as ICBMs. To permit the use of light-weight propellant tanks, pump-fed engines are used.

Pump-fed systems. Figure 3-30 is an illustration of a pump-fed system. Since tank weight no longer limits the chamber pressure, pump-fed rocket engines tend to operate at higher chamber pressures than pressure-fed systems. Pressures of 1000 psi and above are normal. Experimental systems exceeding 5000 psi have been successfully operated. Pumps are usually of the centrifugal variety and, except for the very high pressures, are normally single stage. The high pressure is obtained by operating them at very high speeds. A speed of 9000 RPM is representative of an ICBM propellant pump. The power required to drive rocket engine propellant pumps is impressive. The pump drive turbine of the Titan II missile develops over 5000 horsepower. The turbine of the F-1 engine (used in the Saturn V first stage) develops over 63700 HP. Liquid hydrogen presents a special problem and multistage pumps, both centrifugal and axial types, are in use. In all cases the propellant pumps are driven by a gas turbine. The gas to drive the turbine is usually provided by a gas

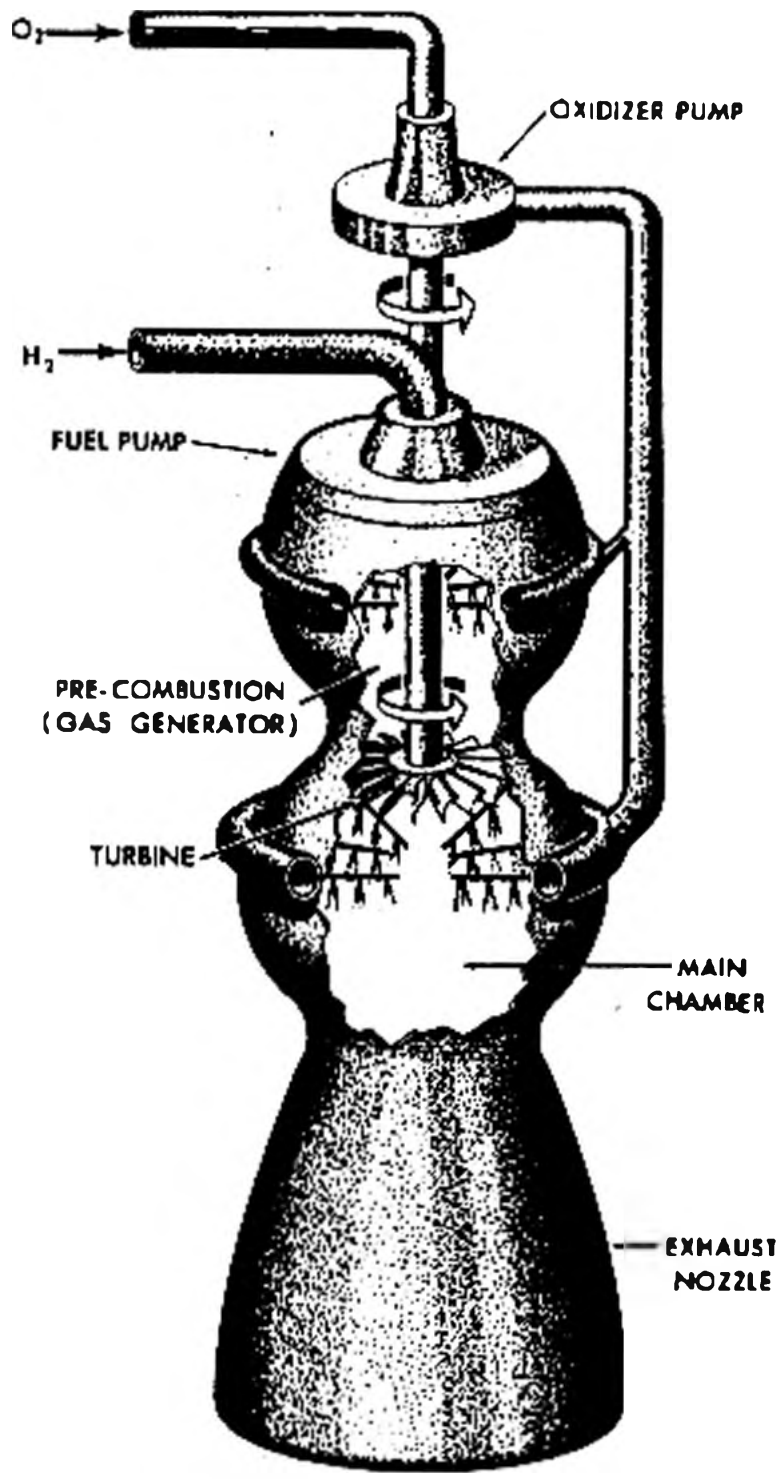


Figure 3-31. Staged Combustion Cycle

generator which operates with the same propellants as the thrust chamber, such as in the Titan and Saturn engines, or from a monopropellant gas generator, such as in the German V-2, the Redstone, and X-15 rocket engines. Once started, these gas generators operate from propellant furnished by the engine's own propellant pumps. This method of operation is known as "boot strapping." It is possible to tap off gas from the thrust chamber to drive the pump turbine. This is known as a bleed turbine cycle. In the past, some small pump-fed engines used what is termed a blast turbine. In this case, the turbine was so positioned that part of the turbine wheel extended into the

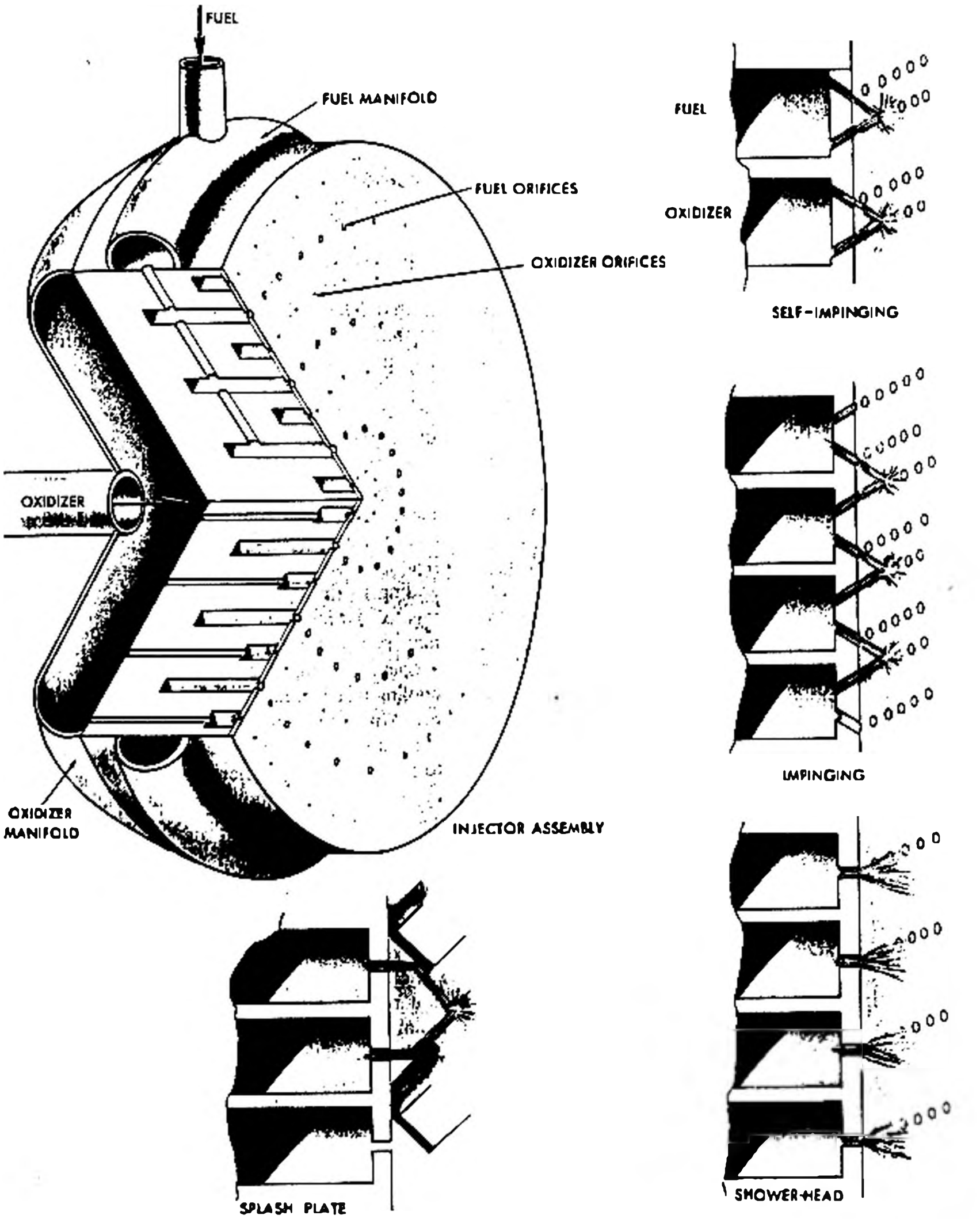


Figure 3-32. Fuel Injection System

exhaust nozzle exit and was driven directly from the exhaust jet. For a system using the main engine propellants to drive the pump turbine, the percentage of the over-all propellant flow becomes excessively large as the chamber pressure exceeds 3000 psi. This is particularly true in the case of an engine using liquid hydrogen since the pump power required is a function of the volume of propellant flowing rather than weight flow. The density of liquid hydrogen is approximately 1/10 that of most other fuels; therefore, the pump power required for the same flow and pressure is approximately 10 times that required for most propellants. In this case, more propellant is required for the gas generator to drive the pump than is furnished to the thrust chamber. Under these conditions, a staged combustion or topping cycle is used. All of the hydrogen from the pump flows through the gas generator where it is partially burned by adding a small amount of oxygen. The exhaust product from the gas generator passes through the pump turbine, then to the main thrust chamber where it is burned with additional oxygen. This system, shown in figure 3-31, is very efficient since all of the propellant is discharged through the main exhaust nozzle at a high Isp, whereas other pump-fed systems dump the turbine drive gas overboard where it delivers thrust less efficiently, i.e., lower Isp.

Injectors. Of all the components of a liquid propellant rocket engine, the injector has by far the greatest effect on the engine performance and durability. The injector serves as the forward end of the thrust chamber and may range in shape from flat to nearly hemispherical. The function of the injector is to admit the propellants to the thrust chamber in such a manner that they will burn stably and efficiently. The combustion process is accompanied by considerable turbulence and recirculation. The injector design, shown in figure 3-32, must control the atomization and mixing of the propellants in such a way that they do not overheat any part of the chamber or injector. The number of injector orifices range from two or three in small attitude control chambers to thousands in the larger thrust chambers. Injector orifices may be impinging, where the fuel and oxidizer streams intersect; self-impinging, where two or more streams of the same propellant intersect; or the non-impinging or showerhead type. Occasionally swirl cups are used to produce a hollow conical spray combined with slots to produce thin sheets of spray. The propellant is sometimes impinged on solid surfaces, called splash plates, to enhance droplet breakup.

Cooling. Because of the severe heating of the thrust chamber wall in the vicinity of the injector, the outer row of injector orifices frequently spray fuel directly on the chamber wall to help cool it by a process known as film cooling. Film cooling can be accomplished anywhere along the length of the thrust chamber and nozzle as may be required. The V-2, for example, made extensive use of film cooling. Film cooling is not desirable from the performance standpoint since much of the fuel used for film cooling does not enter into the combustion process and therefore decreases the engine Isp. The most efficient cooling method from the performance standpoint is regenerative cooling. In this method, one of the propellants is passed through a cooling jacket around the nozzle and chamber before entering the injector. All of the heat absorbed by the propellant enters the chamber adding to the heat of combustion. One disadvantage of the regenerative cooling is that the propellant pump must provide additional pressure to account for the increased pressure drop in the cooling jacket. This can amount to several hundred psi. An advanced form of film cooling that is sometimes employed in critical areas of high performance engines where other cooling concepts are inadequate is known as transpiration or sweat cooling. In transpiration cooling, the coolant, normally one of the propellants, is injected through a porous wall. The coolant may be either a liquid or a gas. Unlike film cooling, which cools only the surface, transpiration cooling takes place within the wall itself after which the coolant acts in the same manner as normal film cooling. Short duration rockets have been built with uncooled steel chambers and nozzles. Refractory metal, ceramic, or graphite linings have been used, as well as ablative materials for longer duration systems; however, all of the uncooled concepts have a definite firing duration limit. Radiation cooling has been employed in applications where the heating rate is not severe.

Ignition. The ignition of a rocket engine is extremely critical. Because of the large flow rates, any delay in the ignition sequence would permit

the accumulation of a mixture of unburned oxidizer and fuel, the subsequent ignition of which could rupture the thrust chamber. Hypergolic propellant combinations such as UDMH and nitrogen tetroxide ignite spontaneously on contact, making the ignition sequence relatively simple. Non-hypergolic combinations such as LOX-RP are frequently ignited by a pyrotechnic device which is ejected from the chamber upon ignition of the rocket propellants. Such devices were used in the Atlas, Titan I, and F-1 engines. When you want to reignite an engine with non-hypergolic propellants several times during flight, a pyrotechnic device is not practical and an electric spark plug is usually employed. With engines employing liquid hydrogen as the fuel, the hydrogen vaporizes readily and is sufficiently easy to ignite so that the spark plug alone is adequate. For other non-hypergolic propellant combinations, a separate igniter chamber is used to provide adequate ignition energy. The LR-99 engine in the X-15 aircraft used three spark plugs (for redundancy) to ignite gaseous oxygen and gaseous ammonia in the first stage igniter, which in turn ignited liquid propellants in the second stage igniter. The second stage igniter, which was a small rocket engine in its own right, ignited the main chamber. Another ignition system known as a slug start has been used with non-hypergolic propellants. A small quantity (slug) of a third propellant that is hypergolic with one of the main propellants is placed in the propellant line of the other propellant. When the propellant valves are opened, the two hypergolic propellants mix in the thrust chamber and ignite. By the time the slug is exhausted, normal combustion has been established.

Monopropellant systems. Monopropellant rockets are generally restricted to the smaller thrust sizes due to their poorer performance in terms of Isp. Their primary advantage is their simplicity. Most satellite attitude control systems employ a number of attitude control thrusters operating from a single propellant tank and feed system. The most common monopropellants are hydrogen peroxide and hydrazine. Hydrogen peroxide is usually decomposed by passing it over a catalyst pack composed of a stack of silver-plated screen wires. In the case of hydrazine, the catalyst consists of alumina particles with a thin coating of iridium. In either case, the catalyst pack is constructed integral with the thrust chamber. Thrust levels for these systems range from ounces to several hundred pounds.

AIR-AUGMENTED ROCKETS AND ROCKET RAMJETS. The term Air-Augmented Rocket has been applied to a variety of powerplants in which the working fluid consists of a mixture of rocket exhaust products and air. The simplest form is the shrouded rocket in which a circular duct or shroud is installed around the exit of the rocket nozzle so that air is entrained and accelerated by the rocket exhaust jet. The final exhaust velocity is lower than that of the rocket alone, but the mass flow rate is substantially higher. The net result is an increase in thrust over most of operational speed range of the system. Although there may be some afterburning or combustion of the rocket exhaust products and air in the shroud, this system depends primarily on the momentum exchanged between the rocket exhaust and the air.

The other main type of air-augmented rocket is the ducted rocket in which the rocket exhaust is introduced between the diffuser and combustion chamber of a ramjet engine. The rocket may be operated fuel-rich or additional ramjet fuel may be added. The velocity of the rocket exhaust entrains the air and with a properly designed duct results in a significantly higher combustion chamber pressure than would be possible with the ramjet diffuser alone. Some ducted rocket configurations have been designed so that the rocket provides thrust for the missile to accelerate to the ramjet operating speed, after which the rocket may be shut off and the vehicle operated as a ramjet. This has also been referred to as a rocket-ramjet. In some designs, a solid rocket burning a very fuel-rich propellant merely furnishes vaporized fuel for the ramjet. This is also frequently referred to as a ducted rocket but is more correctly termed a solid fuel ramjet. Although experimental models of these combined cycle engines have been designed and tested over the last twenty-five years, none are in current use for missile propulsion.

ADVANCED PROPULSION SYSTEMS

Although at the present time missile propulsion by other than systems such as those previously described is considered unlikely, it may be of

interest to examine some advanced non-chemical propulsion schemes.

Nuclear Propulsion

A nuclear reactor can be used as a heat source for both air-breathing and rocket engines. In the case of a nuclear turbojet or ramjet where the working fluid is air, the range attainable is for all practical purposes unlimited. In the case of a nuclear rocket, the working fluid must be carried on board and, therefore, limits the total impulse available. The working fluid selected for a nuclear rocket is usually hydrogen; because of its low molecular weight, it yields a very high Isp. Reactors suitable for both air-breathing and rocket powerplants have been built and tested on the ground; however, none have been flight-tested. Several problems have prevented the general acceptance of nuclear propulsion for aircraft or missiles. The primary problem is the intense gamma radiation from the reactor. Extremely heavy shielding is required to protect the surrounding structure and crew members. Even though the attainable specific impulse of a nuclear rocket is roughly twice that of the best chemical propellants, the weight of the reactor and shielding is such that the vehicle performance for an ICBM mission or low earth orbits is inferior to a good chemical system. Advanced nuclear rocket reactors have been proposed in which the reactor temperature is so high that the nuclear fuel is liquified or in some cases vaporized. Theoretically, a reactor can be constructed that will produce temperatures higher than that of the surface of the sun. None of these systems have yet been demonstrated as they involve some formidable engineering problems.

Electrical Propulsion

Electrical propulsion has been a dream for some time, and various approaches to the problems have been made. Three concepts for electrical propulsion systems have evolved. These propulsion systems have been publicized as the electrostatic or ion engine, electrothermal or arc-jet engine, and the electromagnet or plasma engine.

The acceleration of charged particles is the thrust-producing media used by the electrostatic or ion engine. The propellant, an element such as cesium, is vaporized and impinged on a heated tungsten emitter. Positive ions are generated as a result and are then accelerated through electrostatic fields and ejected to provide thrust.

The electrothermal or arc-jet engine provides thrust by heating a gaseous propellant with an electric arc. The superheated propellant is then expanded through a conventional nozzle to produce thrust.

The operating principle of the electromagnetic or plasma engine is similar to that of a synchronous motor with its rotating magnetic field. In this engine, a gaseous mixture—the plasma, which has been ionized into an approximately equal division of electrons and positively charged ions—responds to magnetic and electrical fields. The gas is accelerated by the effect of the energized electrical and magnetic fields which are mutually perpendicular.

The sun is also regarded as a source of power. Research is being conducted for a system that will use solar heating of a working fluid as a means of propulsion.

MISSILE LAUNCHING METHODS

Missiles could not be used effectively without proper launching equipment and facilities. Each type of missile has its own launching system. Missiles are launched from the air, ground, underground, and underwater. Most missiles also require a booster assembly for takeoff.

Booster Assemblies

The powerplants of certain interceptor and tactical missiles do not develop enough static thrust for independent takeoff. Missiles in this category must be given an initial acceleration to insure proper engine operation, development of lift, and aerodynamic stability at launch. To satisfy these requirements, a booster unit is used to supplement the missile propulsion system.

A booster assembly is an auxiliary propulsion system which imparts thrust to a missile during the initial phase of flight. It generally consists of a solid rocket motor and an attaching device. A solid rocket is best suited for this purpose because of its simple construction and operation and its ability to develop a high thrust in a short time. In some instances, a booster assembly includes an

aerodynamic stabilizing surface. Usually a booster is attached directly to the missile.

A booster of this type is used with missiles such as the CGM-13B Mace. When a missile employing such a booster has been made ready for launch, the engine of the missile is started and accelerated to full throttle and the booster rocket fired. The combined thrust of the booster rocket and engine propels the missile to the airspeed and altitude required to sustain flight. The booster rocket is ejected from the missile after burnout.

There are cases in which the auxiliary thrust-producing system is incorporated in the launching structure. Such a system was once referred to as a catapult, but the use of this term has been discontinued by the Air Force.

Launcher Types

Some means must be employed to support a missile in the desired attitude prior to launch. Such a device is known as a launcher. Basically, it is a mechanical structure which supports and "constrains" the missile so that it moves in the desired direction during the initial phase of flight. Mobility, trainability, and high rate of fire are desirable characteristics of some missile launchers. The degree of importance of these and other characteristics depends upon the type of missile and the tactical use for which the launcher was designed.

ZERO-LENGTH LAUNCHER. A zero-length launcher exhibits no appreciable constraining action on a missile. A fraction of a second after initial motion begins, physical contact between the missile and launcher is broken. In order that the missile may attain flying speed as soon as possible a powerful booster assembly must be attached to the missile.

Zero-length launchers designed for missiles such as the Mace are constructed in the form of a parallelogram having support members that collapse when the missile moves forward. A zero-length launcher by definition orients the missile initially but has no significant effect once the missile has started to move. The length of such an airborne or mobile missile launcher is too short to impart guidance but long enough to hold the missile in a launch position. Figure 3-33 illustrates the zero-length launcher used by the MGM-29A Sergeant missile.

SILO LAUNCHING. A ballistic missile installation

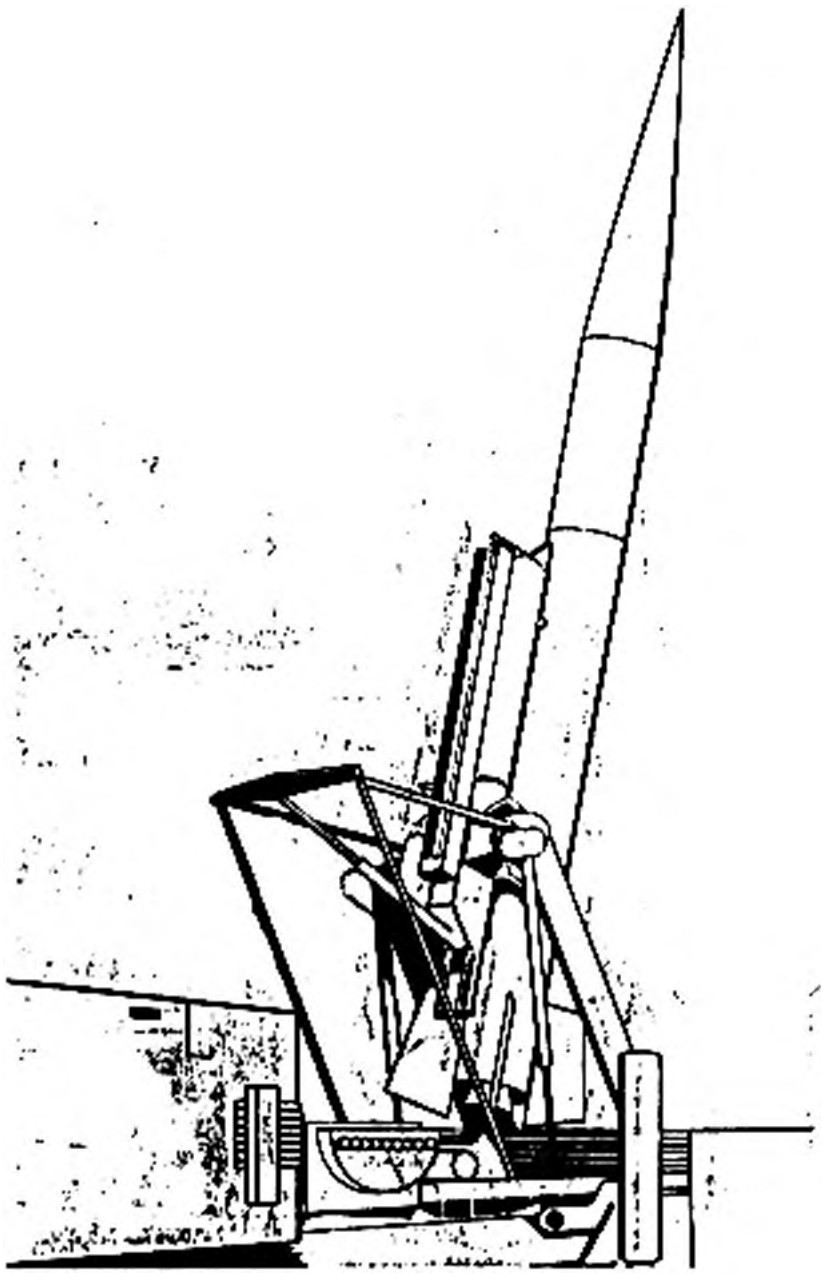


Figure 3-33. Zero-Length Launcher

becomes a hard site when an inclosure such as a silo is constructed beneath the surface of the earth to contain the missile and related aerospace ground equipment (AGE). Hard site is the term used to describe an operational missile site so constructed that it is capable of withstanding the tremendous overpressure resulting from a near-miss nuclear detonation, thereby protecting the ready missiles from damage or destruction.

The silo itself derives its name from a configuration which resembles a typical farm silo. A crib, built and suspended within the silo, provides mounting and securing facilities for the launcher platform and accessory equipment. The manner of mounting the launcher platform and accessory equipment (AGE) depends upon the type of missile to be handled. The Titan II (LGM-25C) is launched directly from within the silo.

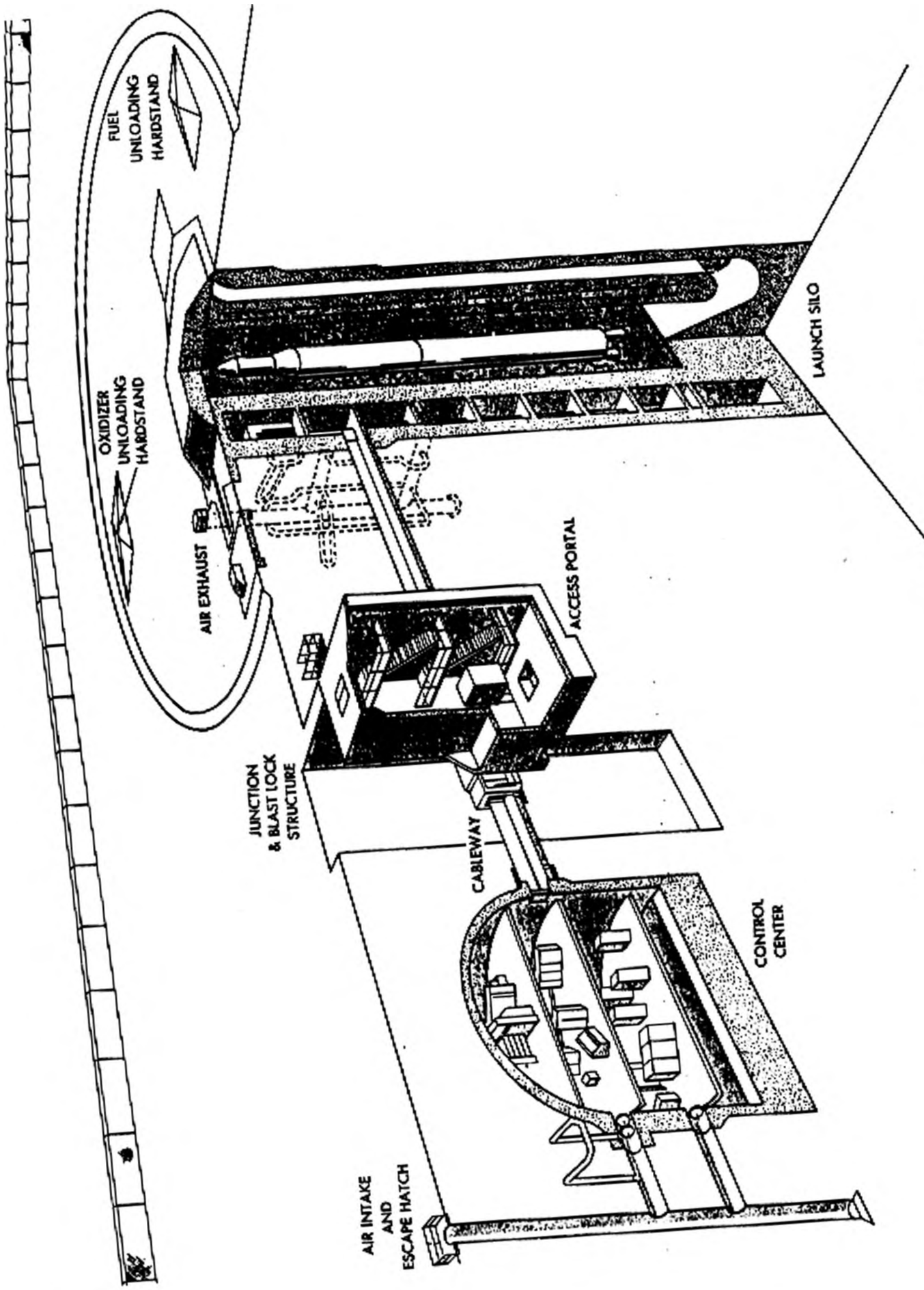


Figure 3-34. Titan II (LGM-25C) Launch Site Structures

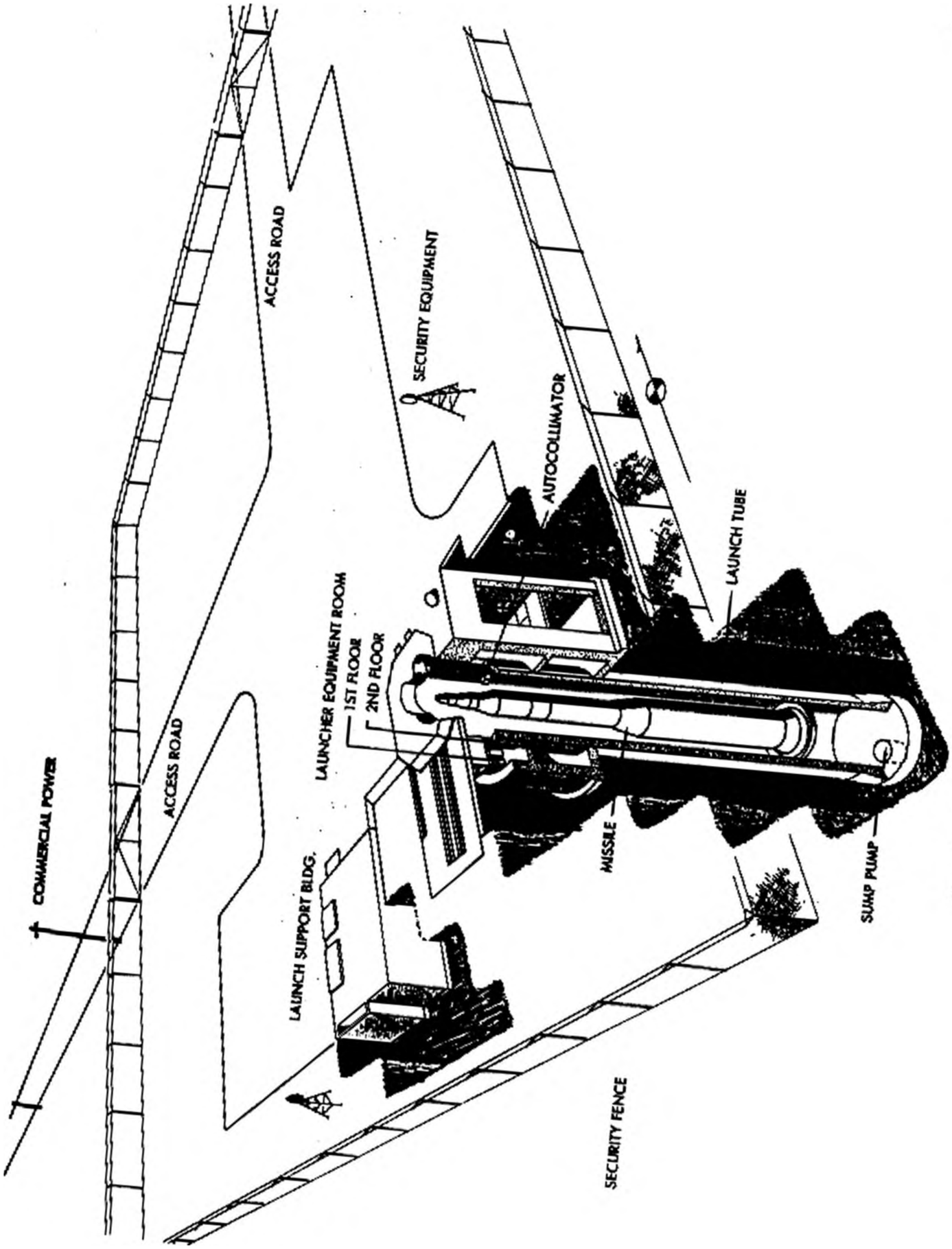


Figure 3-35. Minuteman (LGM-30) Launch Facility

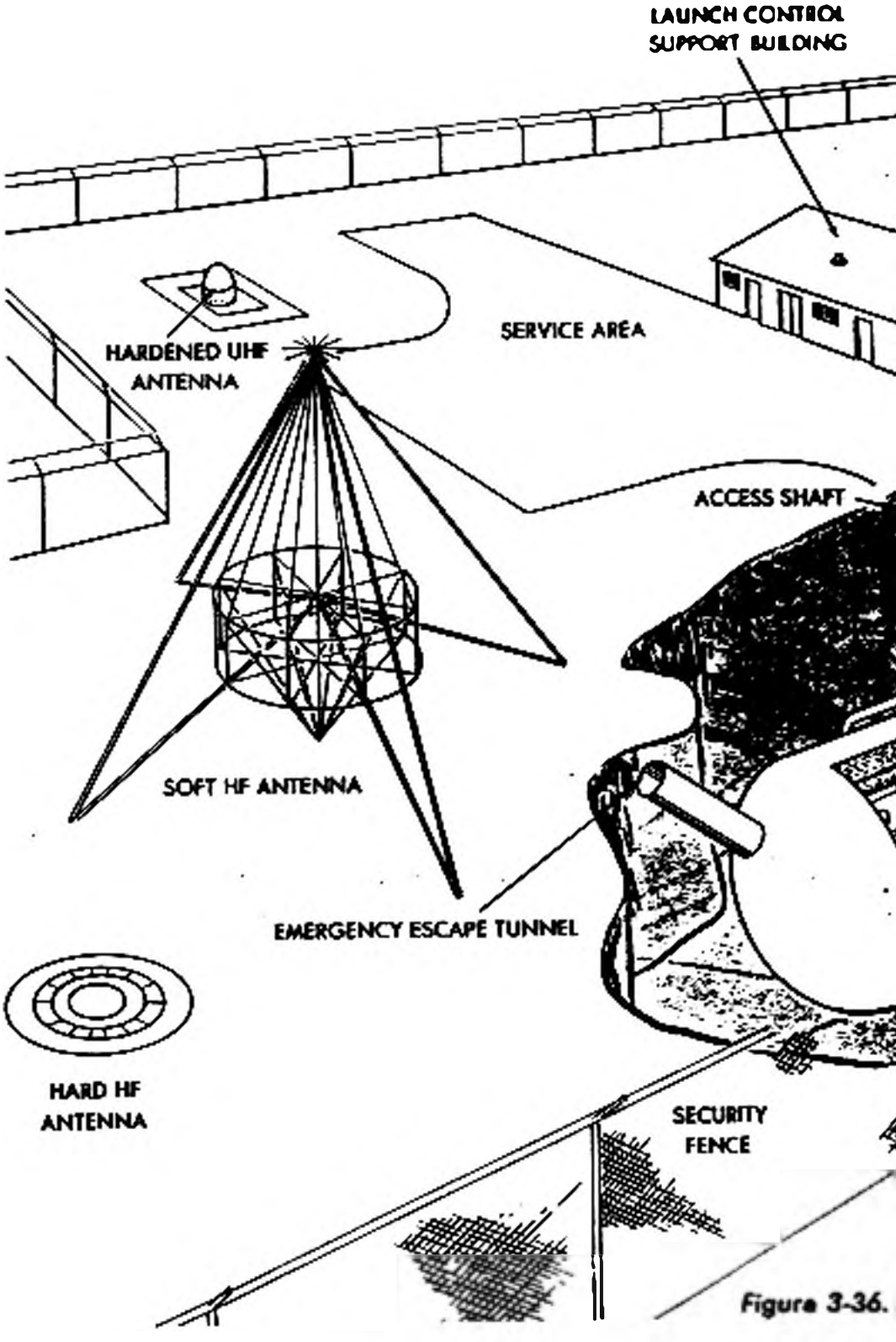
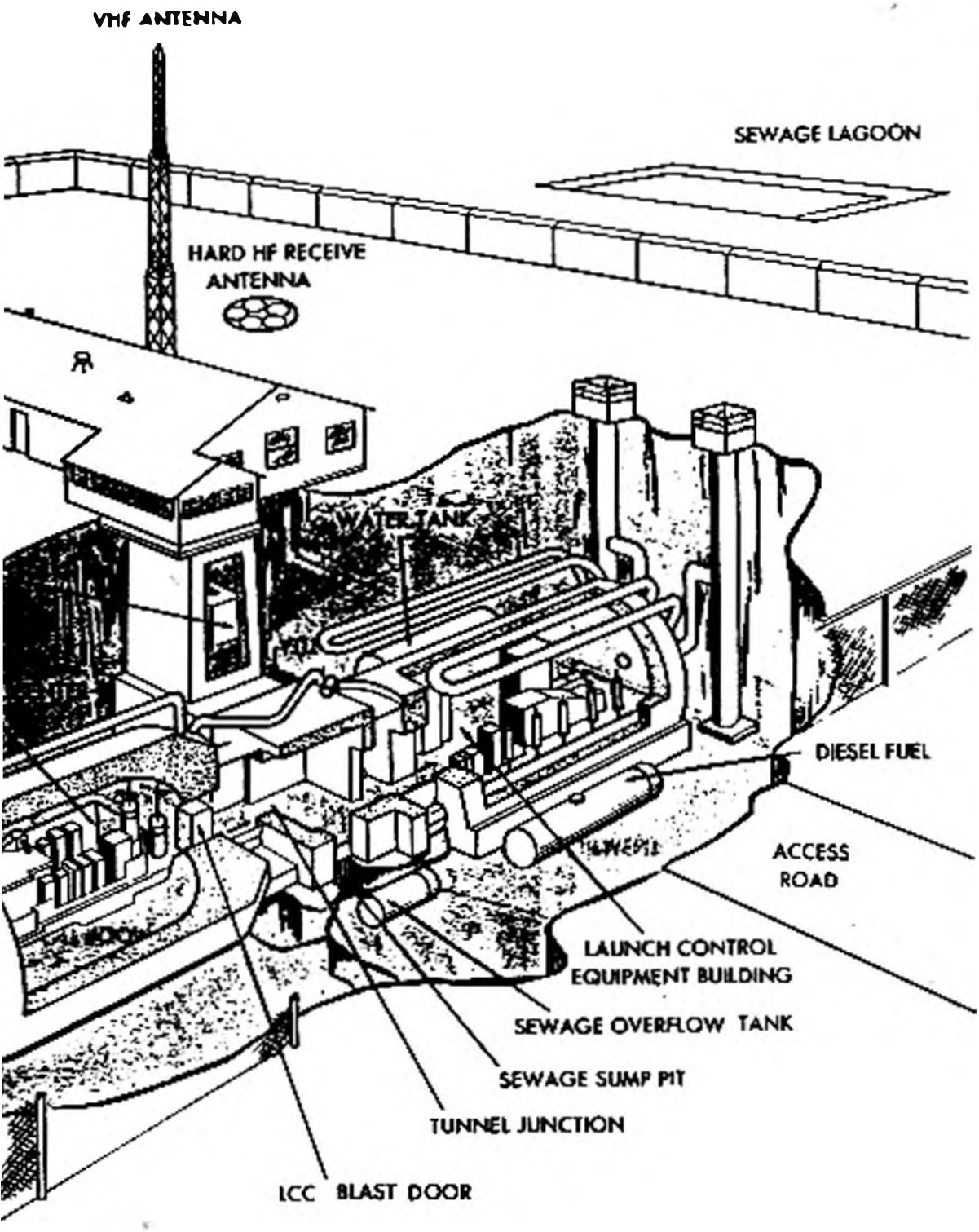

LAUNCH CONTROL
SUPPORT BUILDING
SERVICE AREA
HARDENED UHF
ANTENNA
ACCESS SHAFT
SOFT HF ANTENNA
EMERGENCY ESCAPE TUNNEL
HARD HF
ANTENNA
SECURITY
FENCE

Figure 3-36.



LGM-30 Launch Control Facility

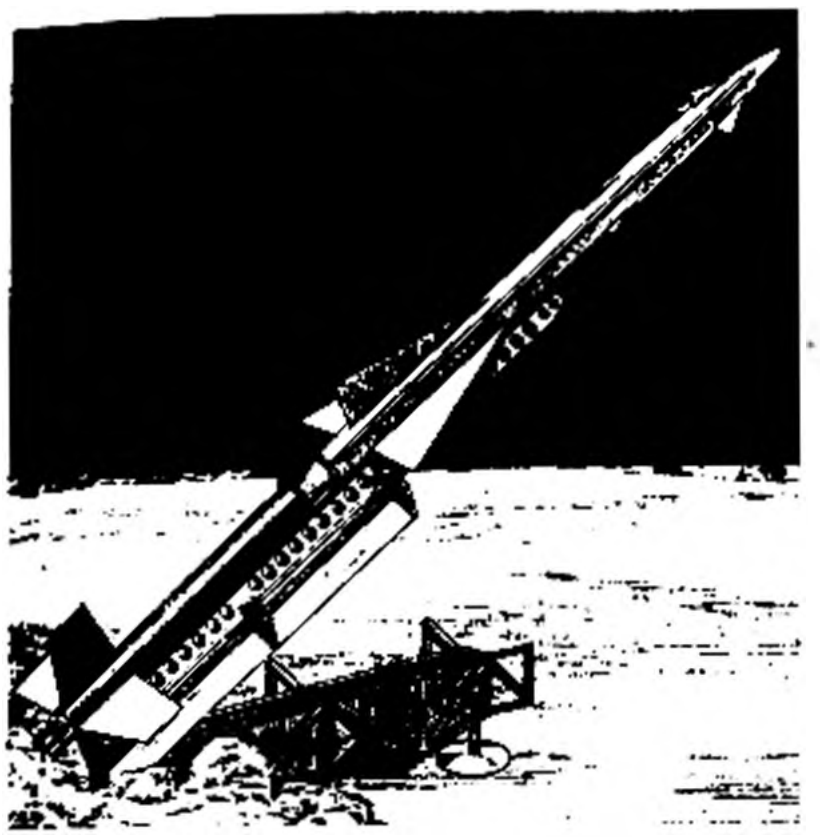


Figure 3-37. Monorail Launcher

angle of approximately 90° and rotated through 360°. It is an ideal surface-to-air launcher because of its trainability and rapid rate-of-fire capabilities. Its short rail length limits use to missiles with powerful rocket-propulsion systems. Contact between missile and rail is maintained by use of a slotted attaching device.

The launch site structure arrangement for the Titan II is shown in figure 3-34. This silo is 155 feet deep. The missile rests on a heavy steel ring supported by heavy steel springs.

The launch facility for the Minuteman (LGM-30) is shown in figure 3-35. The arrangement of structures in figure 3-36 of the LGM-30 complex gives complete on-site control of this complete system.

MONORAIL LAUNCHER. The monorail launcher illustrated in figure 3-37 can be elevated to an

A variation of this launcher consists of replacing the single rail with a slotted trough. The bottom surface of the missile is then equipped with runners to decrease friction and to constrain the missile.

TOWER LAUNCHER. This type of launcher has a derrick-type construction which holds the missile in a vertical, or near vertical, position during the launching phase. Some tower launchers are trainable through a few degrees of tilt to compensate for wind or other factors which may affect the desired trajectory. They must have enough height to allow the missile to attain adequate speed for aerodynamic stability before breaking contact with the tower. The major use of tower launchers is for high-altitude research rockets which fly a ballistic-type trajectory.

MULTISTAGE LAUNCHING. Multistage launching, as shown in figure 3-38, consists of two or more missiles, or missile sections, each containing a separate propulsion system. These sections are called "stages." As an example, in a 3-stage system, the first stage develops takeoff thrust and propels the entire missile system until its propellant is exhausted. At this point the first stage falls away and the second-stage powerplant ignites automatically, increasing the velocity already attained to a higher value. The second stage is then disconnected, and the third stage further increases the velocity imparted to it by the two previous stages. In this way, much higher velocities and greater altitudes are reached than would be possible with a single-stage missile using one propulsion system. One of the first examples of this system was the Wac Corporal which was successfully launched from the nose of a V-2 (2-stage "Project Bumper"), enabling the Corporal to reach an altitude of 250 miles.

Multistage launching is very efficient in the launching of satellites and placing manned vehicles in outer space.

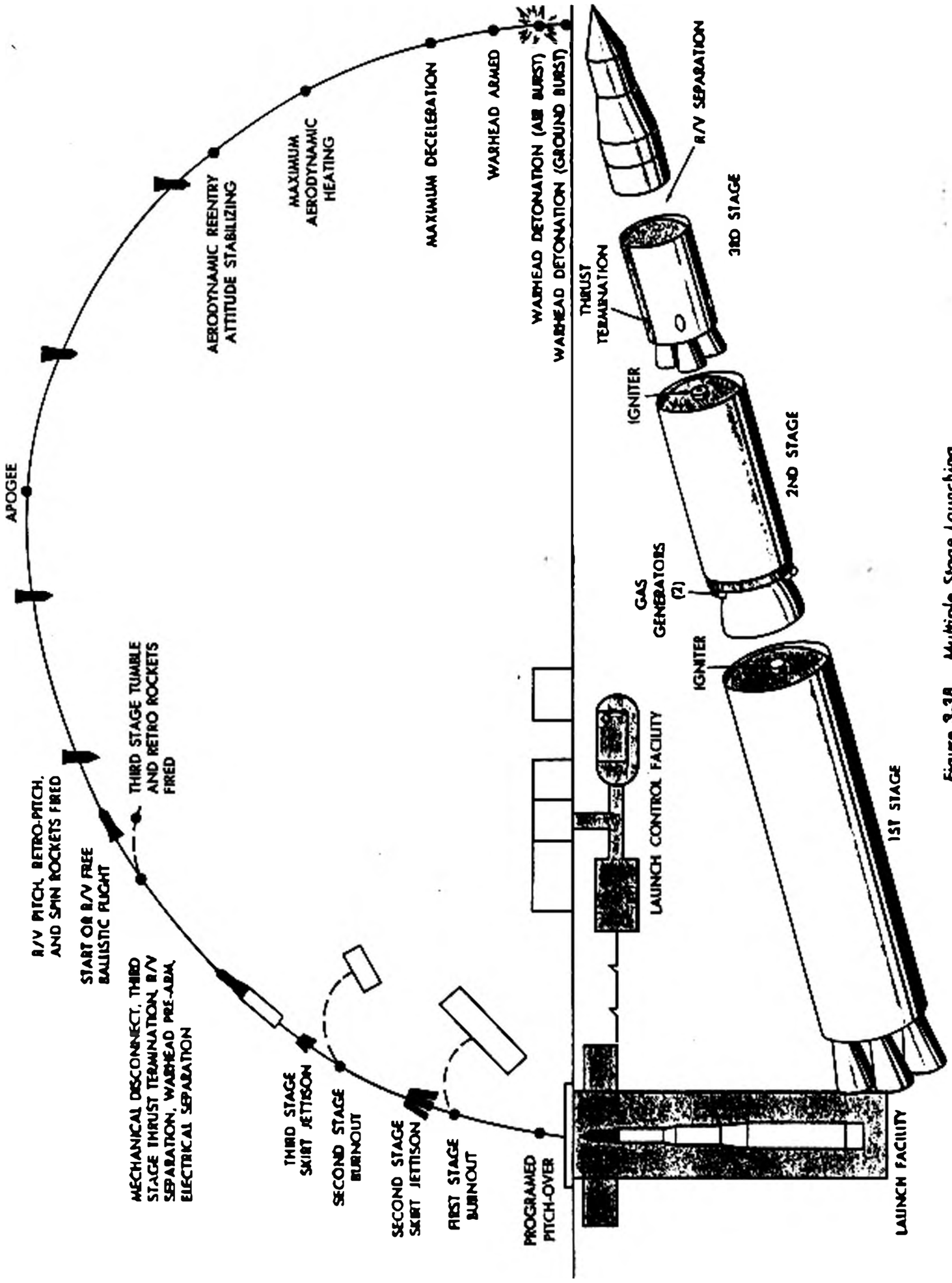


Figure 3-38. Multiple Stage Launching

CHAPTER 4

Physics Involved In Guided Missile Design

You will find the configuration of missiles and the design of all components used in their fabrication, as well as the systems used for their guidance and control, are all governed by basic laws and principles of physics.

The function of each electronic or mechanical component of a missile involves some physical principle and/or property which has been found applicable to the specific requirement. Many of the components are simple in form and perform a single fundamental operation; other components may operate simultaneously or in conjunction with other simple components to perform highly complex operations, as in the case of the components and systems used for missile guidance, control, and instrumentation.

In this chapter you will become familiar with those areas of physics as they apply to missile design. You will learn how the function of electronic or mechanical components involves physical principles and/or properties. The chapter also explains how optical and electronic principles work together. It presents the physics of transistors, and the final section of the chapter covers in detail the modulation of carrier waves. A knowledge of the material covered in this chapter will help you to understand the functions of components and systems employed in missile operation. The next chapter begins the explanation of such components and systems.

A REVIEW OF NEWTON'S LAWS OF MOTION

Newton's laws of motion are specific examples of basic laws of physics which are involved in guided missile operations. Even though you are already familiar with these laws, they are important enough to the study of missiles to be repeated here.

Newton's first law of motion states, "A body at rest remains at rest and a body in motion continues to move at constant speed along a straight line, unless the body is acted upon in either case by an external force."

Newton's second law of motion states, "An external force acting on a body causes the body to accelerate in the direction of the force, and the acceleration is directly proportional to the external force and inversely proportional to the mass of the body." The formula for this law reads:

$$a = \frac{F}{m}$$

where a = acceleration of the body
F = the external force acting upon the body, and
m = the mass of the body

Newton's third law of motion states, "For every action there is an equal and opposite reaction, and the two are along the same straight line." In stating this law, the term "action" refers to the force which one body exerts on a second body, and "reaction" refers to the force which the second body exerts on the first.

The principles stated in the foregoing paragraphs are involved in missile design and operation. To control the flight of the missile, devices must be employed to act with or to react against the natural forces present along the flight path of the missile.

In some cases, it is possible to cause one undesired force to react against another to cancel or nullify the effects of the undesired forces. Such

operations may become complex, depending upon the number of interacting forces involved. In cases involving a large number of interacting forces, complex systems, such as those employed in missile navigation, are needed. Some forces that missile navigation systems must deal with are weight, inertia, aerodynamic forces, engine thurst, Coriolis forces, and magnetism.

PHYSICAL PRINCIPLES AND PROPERTIES INVOLVED IN MISSILE OPERATION

To understand fully the functions of the basic components and systems employed in missile operation, you should have further information pertaining to the physical principles and properties involved. You need to be familiar with such terms as inertia, gravity, circular motion, centripetal and centrifugal forces, rotational motion, radius of gyration, and motion of precession.

These terms and others are explained in the following paragraphs. It is neither practical nor desirable to present complete analyses of the meaning of the terms here. Instead, the relationships among the various physical principles and properties to which the terms refer are explained. You need to know these relationships so you may more readily recognize those which may be involved in the basic units and operational systems of missiles.

Inertia

Regardless of the specific purpose or design of a missile component and regardless of whether it be electronic or purely mechanical, the operation of the component depends upon one or more physical properties. One physical property inherently common to matter is inertia.

Inertia is defined as a property of matter by which it will tend to remain at rest or if in motion will tend to follow the same straight line or direction unless acted upon by some external force, as stated in Newton's first law of motion. You may also think of inertia as the opposition offered by a body to any change of motion; that is, an external force is required to overcome inertia and to give it linear acceleration.

The property of inertia implies a broader concept of the definition of mass than mere quantity of matter. You may think of *mass* as the property of an object by virtue of which it possesses inertia.

A tendency, or measure of a tendency, to produce motion about an axis is referred to as a *moment*. A *moment of inertia* of a body is equal to the mass of one particle of the body multiplied by the square of its distance from the axis of rotation, plus the mass of another particle of the body multiplied by the square of its distance from the axis of rotation, plus—and so on, until all the particles of the body have been included. In other words, the masses of the particles of which a body is composed and the squares of the individual radii that extend from these particles to the axis determine the moment of inertia of the body.

Moment of inertia is often expressed in terms of the aggregate mass of the body and a single fictitious radius called the radius of gyration. The *radius of gyration* of a body is defined as the distance from the body's axis of rotation to a point at which the entire mass of the body may be considered concentrated without altering the moment of inertia.

Velocity

You already know that velocity is a representation of speed and a specific direction of motion. The velocity of an object may be constant or varying. If velocity is varying, the variation may be in magnitude, in direction, or both. When variation takes place in direction, the term angular velocity is applied to describe the motion. *Angular velocity* is the time rate of change of angular displacement (i.e., the rate of change of angular position).

The angular velocity of a rotating body may be represented by an axial line having a length which indicates the numerical value of the velocity.

The direction of the vector, as shown in figure 4-1, represents the direction in which an ordinary

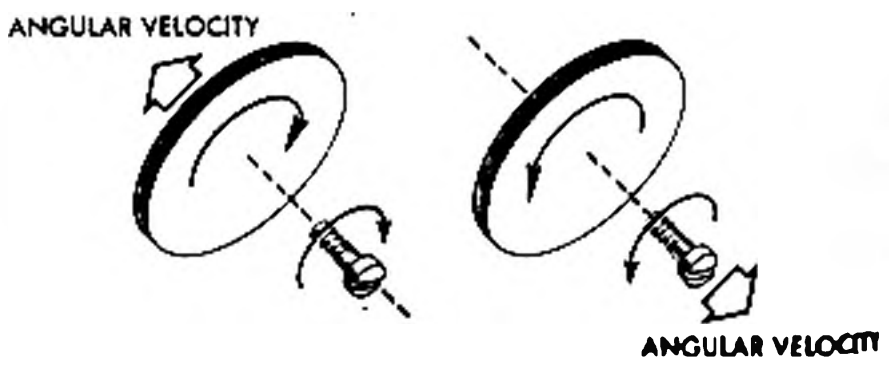


Figure 4-1. Angular Velocity as Vector

right-hand threaded screw would travel if turned in the direction in which the body (wheel) rotates.

When a rotating body is given an additional angular velocity about the same axis, the resulting velocity will be their algebraic sum. For example, assume that a mass is rotating at 10 radians per second in a clockwise direction. A reverse torque is applied to it which, if the rotating mass was at rest, would impart to it an angular velocity of 2 radians per second in a counterclockwise direction. The resultant velocity in this instance is 8 radians per second in the clockwise direction. A *resultant velocity* is obtained by vector addition if a velocity is added about a different axis than that of original rotation.

A radian is a unit of measurement of angles. In any circle, the size of the central angle in radians is the ratio of the arc subtended by this angle to the radius of the circle. When the length of arc is equal to the radius, the angle subtended by the arc is one *radian*. A radian is equal to 57 degrees 17 minutes 44.8 seconds. There are 2 π radians in a circle.

Acceleration

Acceleration is the time rate of change of velocity. It represents motion in which the velocity changes from point to point.

When the velocity of a mass moving in a straight line changes by equal amounts in equal intervals of time, the acceleration is said to be constant and the motion *accelerated uniformly*.

Moving objects rarely exhibit constant acceleration; however, a free-falling body in a vacuum near the surface of the earth is one example of uniformly accelerated motion. Experiment has shown that this constant acceleration resulting from gravity is 32 ft/sec^2. It is represented in physics formulas by the letter, g.

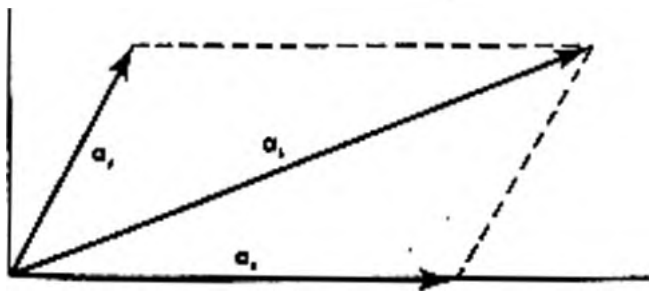


Figure 4-2. Vector Resultant of Two Accelerations

When acceleration is not known to be constant, the average acceleration may be expressed by the equation:

$$a = \frac{V_f - V_o}{t - t_o}, \text{ where}$$

V_f = final velocity

V_o = initial velocity

$t - t_o$ = time interval during which velocity changed from V_o to V_f

The units of acceleration depend upon the unit of velocity as well as upon the unit of time during which the velocity changes.

Based on the above equation, acceleration is equal to the difference of two vectors (velocities) divided by the scalar quantity (time) and is a vector quantity which may be determined as the resultant by the common parallelogram method of vector addition. For example, if a mass is subjected to accelerations, a_x and a_y, the resultant acceleration, a_r, may be obtained as shown in figure 4-2. The dotted lines are extended along a path opposite and parallel to a_x and a_y. The point at which the dotted lines intersect determines the resulting acceleration, or resultant.

Note that acceleration is not the time rate of change of speed. *Acceleration* is the time rate of change of velocity treated as a vector quantity. A mass moving with constant speed in a circle is accelerated despite the fact that the time rate of change of speed is zero. The velocity of the mass is changing continuously since the direction of motion is changing. The time rate of change of velocity is not zero. Such acceleration is termed normal acceleration and is explained later in this chapter with reference to circular motion.

For any kind of motion, the distance (s) of an object from its starting point is obtained as the product of average velocity (V_{av}) and time (t).

$$S = V_{av}t$$

When the motion is uniformly accelerated (a is constant), the relationship between the average velocity (V_{av}) and the initial velocity (V_o) and final velocity (V_f) may be expressed by the equation

$$V_{av} = \frac{V_o + V_f}{2}$$

because in terms of initial and final values the average of any two quantities is equal to one-half their sum. By substituting and transposing from the equation

$$a = \frac{V_f - V_o}{t - t_o}$$

and the equation

$$V_{av} = \frac{V_o + V_f}{2}$$

to the equation

$$S = V_{av}t \quad \text{or} \quad \left(V_{av} = \frac{S}{t}\right)$$

you can derive the equation

$$S = V_o t + \frac{at^2}{2}$$

Here's how the latter equation is derived:
from

$$a = \frac{V_f - V_o}{t - t_o}$$

in which $t_o = 0$.
you get $V_f = V_o + at$.
If in equation

$$S = V_{av}\, t$$

you substitute for V_{av} its value

$$\frac{V_o + V_f}{2}$$

you obtain

$$S = \frac{(V_o + V_f)\, t}{2}$$

Combining this equation with equation

$$V_f = V_o + at$$

will result in equation

$$S = \frac{(V_o + V_o + at)\, t}{2}$$

or

$$S = \frac{(2V_o + at)\, t}{2}$$

By further reduction you obtain

$$S = V_o t + \frac{at^2}{2}$$

which is the original formula.

From this equation the average acceleration (a) required to displace an object a distance s in t seconds, assuming zero initial velocity, is found to be

$$S = \frac{1}{2}\, at^2$$

where

S = distance in feet
a = acceleration in ft/sec/sec
t = time in seconds

If initial velocity (V_o) is zero, the term

$$V_o\, t = o$$

and equation

$$S = V_o t + \frac{at^2}{2}$$

reduces to

$$S = \frac{at^2}{2} \text{ or } S = \frac{1}{2}\, at^2$$

A practical example of the application of these formulas pertaining to accelerated motion may be illustrated by assuming that a rocket is launched from a position of rest, and it attains a velocity of 200 feet per second during the next second, and so on. It can be said that the rocket has an acceleration of 200 feet per second in each second. By the end of 3 seconds, the rocket will have a velocity of 600 feet per second, and so on, as long as it can maintain the same acceleration.

During this period of increasing velocity, every value of velocity from zero up to the maximum velocity attained by the rocket is passed through at some instant of time.

If you plotted a graph of the instantaneous values of velocity, you would find that they would describe a curve as shown in figure 4-3.

The line is straight from "O" to "Y," showing that during the first 3 seconds the rocket gains velocity at a uniform rate. Acceleration is constant.

Beyond "Y," the line slopes off indicating a reduction in acceleration until it becomes horizontal at "X," indicating zero acceleration or constant velocity.

If the acceleration of a body is considered as the change of its velocity during any interval of time divided by the duration of that interval, the definition is based upon change of velocity and not upon distance traveled. The definition may be stated in the form of an equation by supposing the

motion of the body to be observed for some specific interval of time.

Taking V_o as the initial velocity of the body at the beginning of the period of observation and V_f as its final velocity at the end of the period, the change in velocity is

$$V_f - V_o$$

And if this change occurs in an interval of time, t, the average acceleration of the body throughout that interval is

$$av = \frac{V_f - V_o}{t}$$

Referring to figure 4-3 showing the acceleration curve of a rocket, the velocity of the rocket under observation increases from 600 to 900 feet per second during the time interval from 3 seconds to 6 seconds after launching. During this period the velocity does not increase uniformly as it does in the period from the instant of launching to 3 seconds after launching; therefore, the acceleration varies from instant to instant beyond point "Y." By using the formula

$$a_{av} = \frac{V_f - V_o}{t}$$

the average acceleration of the rocket may be computed over the interval of time between the third to the sixth second as follows:

$$a_{av} = \frac{900 \text{ ft/sec} - 600 \text{ ft/sec}}{6 \text{ sec} - 3 \text{ sec}} = \frac{300 \text{ ft/sec}}{3 \text{ sec}}$$

$$= 300 \text{ ft/sec} \times \frac{1}{3} \text{ sec}$$

$$= 100 \text{ ft/sec}^2$$

Similarly, by choosing any time interval during the first 3-second period after launch (period of uniform acceleration), a result of 200 ft/sec² for the acceleration is obtained.

Referring again to the curve, if you solve for the average acceleration throughout the time interval between ½ second after launching (V_o) and 2 seconds after launching (V_f), you will find

$V_o = 100$ ft/sec, $V_f = 400$ ft/sec, $t = 1\frac{1}{2}$ sec $(2 - \frac{1}{2})$.

Using these values in the formula, the equation becomes

$(400 \text{ ft/sec} - 100 \text{ ft/sec}) \div (1\frac{1}{2} \text{ sec})$
or $(300 \text{ ft/sec}) \div (1\frac{1}{2} \text{ sec}) = 200 \text{ ft/sec}^2$

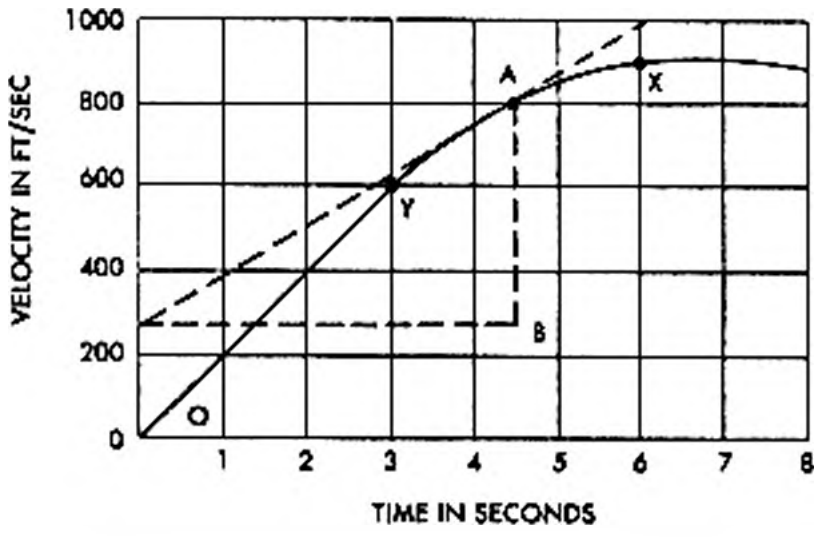


Figure 4-3. Acceleration Curve of Rocket

As the time interval is made shorter and shorter, the average acceleration approaches nearer and nearer to the instantaneous acceleration.

In the limit, for an infinitesimal change of velocity, dv (dee v), occurring in an infinitesimal time interval, dt (dee t), the instantaneous acceleration, a, may be expressed as:

$$a = \frac{dv}{dt}.$$

Concisely defined, "*Acceleration* is the time rate of change of velocity," and its value at any specific instant is represented graphically by the slope of the velocity-time curve at the corresponding point. From figure 4-3 (the acceleration curve of a rocket) you can find the instantaneous acceleration at any time after launching by drawing a tangent to the curve at the point where it crosses the specific time ordinate and determining the slope of the tangent. For example, if you wish to find the instantaneous velocity of the rocket computed on the acceleration curve at 4½ seconds after launching, you can draw a tangent to the curve at point A, where the curve crosses the 4½-second time ordinate.

Since AB = 800 ft/sec – 300 ft/sec or 500 ft/sec, the slope is 500 ft/sec ÷ 4½ sec = 111 1/9 ft/sec².

This value is the instantaneous acceleration of the rocket at the specific time of 4½ seconds after launching.

CENTRIFUGAL AND CENTRIPETAL FORCES. In

accordance with Newton's second law, a body in motion will not deviate from a straight line unless an external force is exerted upon it. When a locomotive arrives at a curve in the track, its forward motion causes the flanges on its wheels to press outwardly against the edge of the outer rail; and conversely, the rail presses inwardly against the flanges. The locomotive changes direction as a result of this inwardly directed force and follows the curved track.

The lateral force required for motion along a curve also may be illustrated by a ball attached to a cord, as shown in figure 4-4. When the ball



Figure 4-4. Forces Acting on Ball Whirled Around at End of Cord

is whirled around at the end of the cord, the cord becomes taut and exerts an inward pull, *centripetal force*, on the ball; at the same time the ball pulls outwardly, exerting a *centrifugal force* on the cord.

The ball, traveling along a circular path at constant speed, is pulling against the cord with a constant force directed outwardly along a radial line from the center of its path of rotation. Since for every action there is always an equal and opposite reaction, the cord exerts an equal inward force upon the moving body.

Any motion along a curve involves *centrifugal* and *centripetal* forces which are equal and opposite, and both forces are exerted in the plane in which the curve lies. Although equal, these forces cannot balance each other because they are not exerted upon the same object.

An unbalanced force always produces acceleration. The centripetal force acting upon a body in circular motion continually accelerates it toward the center of the circle. The body moves in a circle as a result of this inward motion combined with its forward motion.

Note again the fact that when a body follows a circular path at constant speed, its velocity is changing continually, not in magnitude but in direction.

Assume a body to be moving at a constant speed, v, around a circle of radius, r, which is centered at O, as shown in diagram A of figure 4-5. Assume also that in an interval of time, t, it moves a distance PQ = vt, as shown in diagram B of figure 4-5. Its velocities at P and Q may be shown by vectors V_X and V_Y tangent to the circle at points P and Q, respectively. The velocities in magnitude are equal to v.

These velocities differ in direction, indicating that some additional velocity must have been given to the body in moving from P to Q to change its velocity from V_X to V_Y. To determine this additional velocity, V_X and V_Y can be drawn from a common point, K, as in diagram C of figure 4-5. Then V_X is subtracted from V_Y. This subtraction is done by reversing the direction of V_X and adding in accordance with the parallelogram method of vector addition as explained previously. As indicated, the change in velocity is V_Z.

As the velocity change from V_X to V_Y occurs during time, t, the average acceleration through-

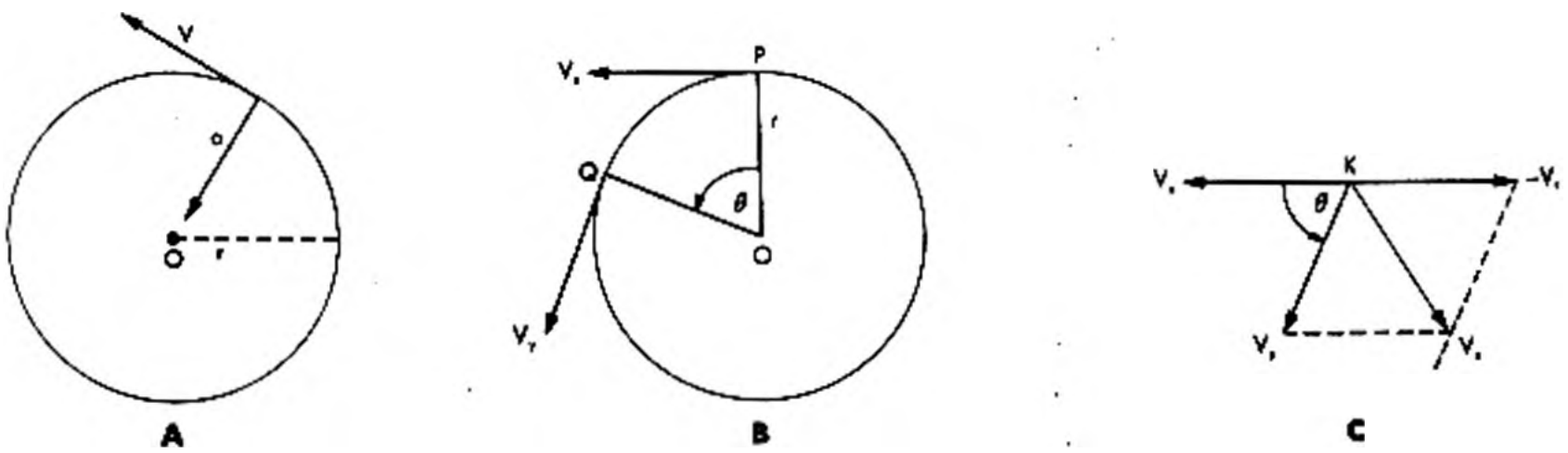


Figure 4-5. Determining Acceleration of Body Moving in a Circle

out time, t, is

$$a = \frac{V_x}{t}$$

To determine the instantaneous acceleration, first notice that the angles designated θ (theta) are equal. Now, by assuming progressively shorter time intervals, the angles indicated by θ would simultaneously become progressively smaller. When zero time as a limit is approached, the sector POQ in figure 4-5B more nearly resembles the isosceles triangle, the equal sides of which form the angle θ of figure 4-5C. This resemblance increases until finally

$$V_x/V_y = PQ/OP$$
or
$$at/V_y = vt/r$$

Since $V_y = v$ numerically, you can derive the magnitude of the acceleration a by substituting v in the second term of the equation $at/V_y = vt/r$. You get

$$at/v = vt/r$$

By cross-multiplying, you get

$$atr = v^2t$$

After canceling the t's, you have

$$ar = v^2$$

Now by transposing r, you get the simplified equation

$$a = v^2/r$$

The above process shows that acceleration is a vector quantity. The direction of this vector quantity can be found by noting that the shorter the time interval taken, the more nearly V_y becomes perpendicular to V_x and V_y, until at the time limit ($t = \theta$), it is perpendicular to both vectors V_x and V_y. In this manner you show that the *centripetal acceleration is directed toward the center of the circle.*

The equation

$$F = \frac{W\,a}{g}$$

in which the acceleration, a, is caused by an unbalanced force, F, acting on a body of weight, W, and the acceleration, g, is due to gravity, may be used to show the force that must be exerted upon the body to produce the acceleration.

The centripetal force acting upon a body of W pounds when moving at speed, v, feet per second around a curve of radius, r, feet may be expressed

$$F = \frac{W\,v^2}{g\,r}$$

Note from this expression that the centripetal force acting on a body which is moving in a circular path varies directly as the square of the speed and inversely as the radius of the circle.

The force with which a body is attracted toward the earth is diminished as a result of the rotation of the earth. The effect of this rotation is most noticeable at the equator where velocity resulting from the earth's rotation is greatest. The influence of gravity is about 0.0035 or 7 pounds per ton at the equator. Velocity is zero at the poles. An object which weighs 2000 lbs at the equator weighs about 2007 lbs at the poles.

ANGULAR ACCELERATION. For purposes of illustrating the combined action of forces, consider how angular acceleration is produced by torque.

Torque (a moment of force) is a rotational effect on a body and is measured by the product of the force present and the perpendicular distance from the axis of rotation to the line of action of the force. The perpendicular distance is referred to as the *lever arm.* Torque may be expressed by the equation T = FL, where L stands for lever arm.

Referring to figure 4-6, suppose that mass, m,

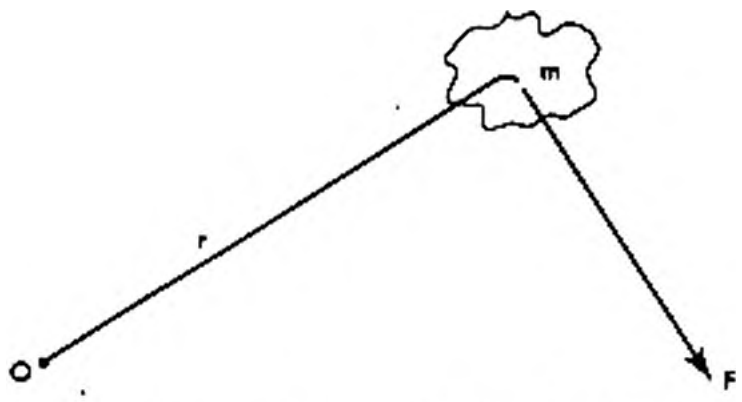


Figure 4-6. Angular Acceleration Produced by Torque

is attached to the end of a crank which in itself has no weight and is pivoted at O. Suppose also that force, F, which acts upon the mass is a steady force so directed that it is always at right angles (tangential) to the crank. Under these conditions,

as the mass moves, the crank pulls it toward the center and causes it to travel in a circular path. At the same time, the tangential force, F, causes it to move along the circle with increasing speed, giving it an acceleration along this path equal to F/m. Thus, a = F/M.

The crank will have an increasing angular velocity, and if the crank is of length, r, its angular acceleration can be expressed as α (alpha) = a/r, where α is the instantaneous angular acceleration.

The angular acceleration of a body is defined as its time rate of change of angular velocity, or in formula: $\alpha = dw/dt$, where dw is the infinitesimal change of angular velocity occurring in the infinitesimal time interval, dt.

The force, F, acting on the mass, m, produces a torque, T, about the axis, 0, which may be expressed by the equation, T = Fr.

Precession

When two angular velocities about separate axes are added to a body, an angular motion about a third axis is produced. This new motion is called *precession*. This type of motion is illustrated in figure 4-7 in the form of a wheel mounted loosely between collars on an axle, A. Assume that the wheel is rotating. The shaft does not rotate with the wheel, and one end of it is set horizontally upon a pivot on the vertical standard, C. The other end of the shaft is unsupported. If the wheel were not revolving, the free end of the shaft would drop, and the entire assembly would fall off the vertical support.

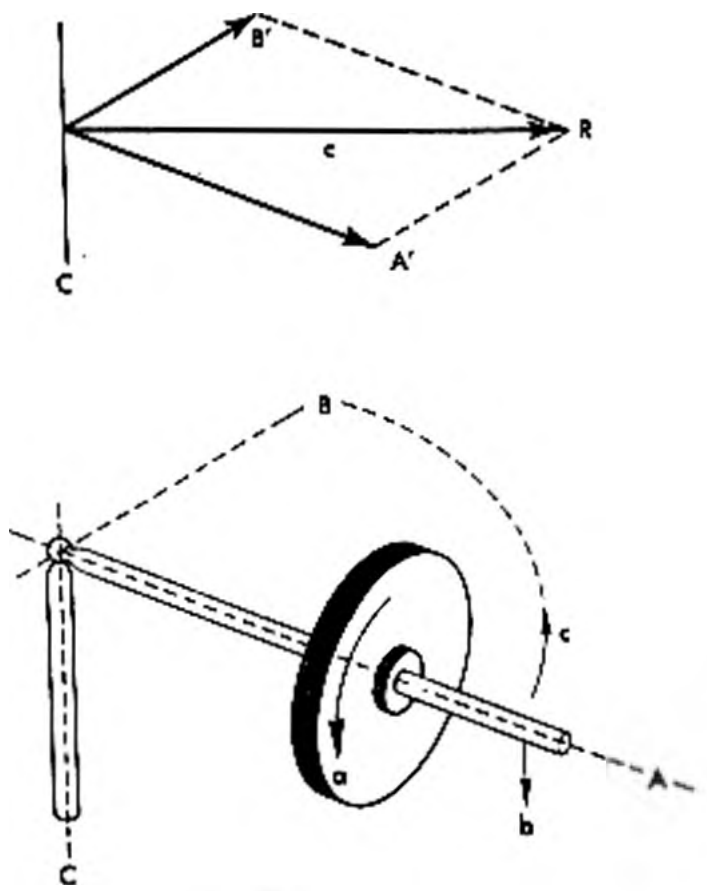


Figure 4-7. Precession of Revolving Wheel About One End of its Axle

But with the wheel revolving, the tendency for the free end of the shaft to drop causes the wheel and the shaft to describe horizontal circles about the pivot. In other words, the wheel precesses in the horizontal plane.

You may now determine how precession develops. The angular velocity of the wheel rotating in the direction, a, about axis A is represented by vector A parallel to shaft A. The gravitational pull of the earth upon the wheel produces a torque which tends to pull the entire assembly about axis B in the direction indicated by arrow b.

This torque produces an angular acceleration and gives the body an additional angular velocity about axis B which is represented by the vector B. If this velocity B is added to the rotational velocity A, both being in the same horizontal plane, the resultant will be R. Shaft A will shift its position to point in direction R, turning about the vertical axis C as indicated by arrow c. This motion of precession takes place about axis C which is perpendicular to both axis A and axis B.

The precession continues, for as soon as the shaft reaches position R, the assembly is subject to another torque due to the tendency of the free end to drop. The corresponding change in angular velocity is at right angles to R and a new resultant is formed to which the shaft next progresses, and so on. This progression of the shaft is not in discrete steps, but it is actually a continuous procession of infinitesimal angular shifts, producing a uniform velocity of precession.

It can be shown that the angular velocity of precession, Ω (omega), of a wheel is equal to the torque, T, which tends to change the direction of the axis, divided by the product of the angular velocity, ω, of the wheel and its moment of inertia, I:

$$\Omega = \frac{T}{\omega I}$$

The above physical principles are embodied in the theory and operation of a gyroscope which

finds a multitude of applications in missile guidance and control systems.

A gyroscope exemplifies the physical principles and forces to a greater degree than any other single device. One form of gyroscope that you are familiar with is the earth. The earth resembles a huge gyroscope rotor spinning freely in space about an imaginary axis through its poles. Because of the rapidity of its rotation, the earth maintains its position in the universe and remains constantly in the same plane of rotation with its poles pointing approximately in a constant direction in space.

Aside from variations in physical form and certain variations resulting from bearing-friction and structural unbalance, the gyroscope represents the same physical phenomena in action as the earth, and the same forces are involved.

The gyroscope basically consists of a wheel or disc so mounted as to be able to spin rapidly about an axis, which can be called the axis of symmetry. The wheel also is mounted so as to be free to rotate about one or both of two axes which are perpendicular to each other and to the axis of spin.

If, when the wheel is spinning, a torque or twisting moment is applied about one of these two axes (axis of torque), the moment will produce precession about the other axis which then becomes the axis of precession. Under certain conditions, depending upon their orientation, these two axes may be referred to as the horizontal and vertical axes.

Another force which acts on a body to a small degree is Coriolis force, named for the French engineer G. G. Coriolis, who first brought it to attention.

Coriolis is a deflecting effect acting on a body in motion and occurring as a result of the different surface speeds, due to the earth's rotation. The effect diverts north or south heading horizontal motions over the earth's surface to the right in the Northern Hemisphere, and north or south headings to the left in the Southern Hemisphere. This effect is not present in stationary bodies.

Gyroscopes are classified into three basic types: balanced, unbalanced, and Foucault. These types are shown in figure 4-8.

A balanced gyroscope is one placed at one end of a horizontal rod and balanced on a pivot by means of a counterweight near the other end.

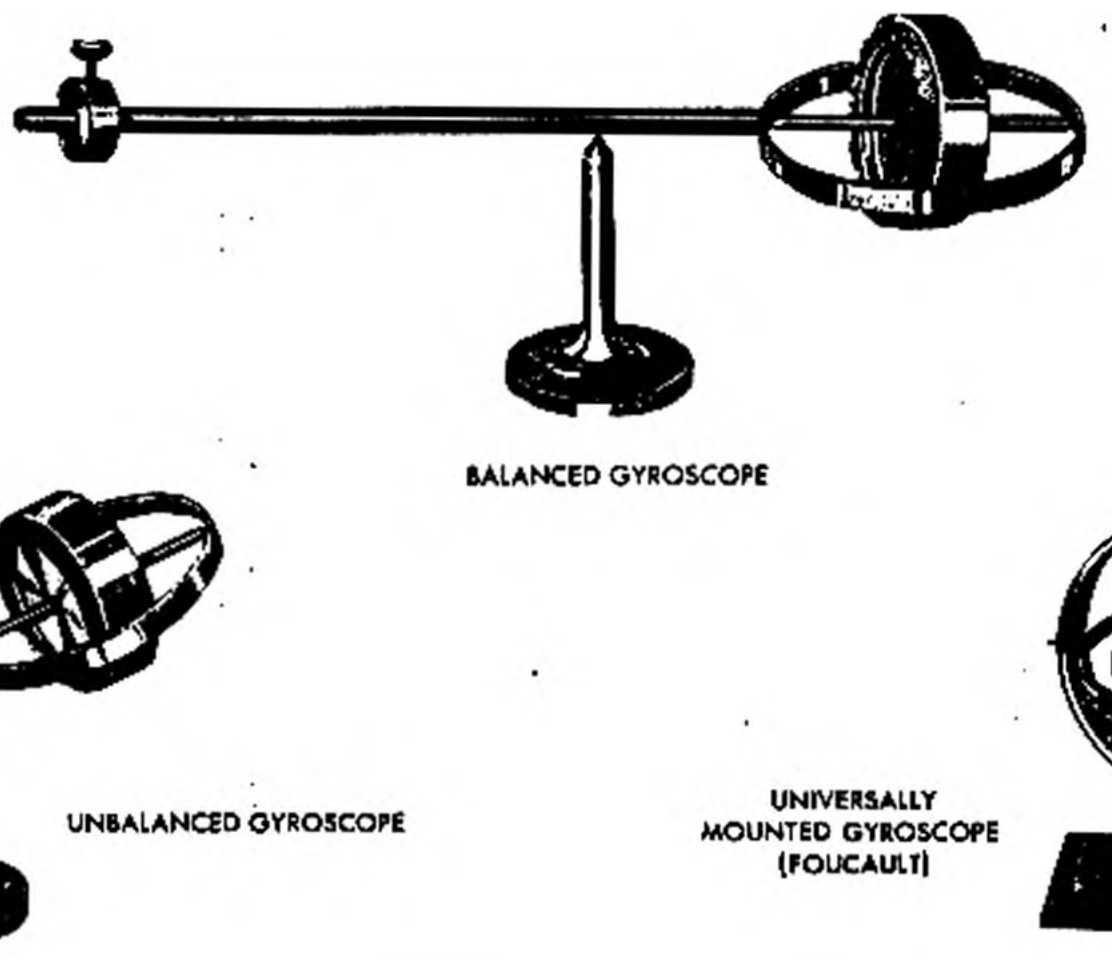


Figure 4-8. Basic Types of Gyroscopes

An unbalanced gyroscope is one in which the rod and counterweight beyond the pivot are omitted.

The Foucault gyroscope is one in which the spinning wheel is supported by gimbals so that it can turn freely about any axis, and the parts are so balanced that gravity does not exert a torque upon the wheel.

The spinning gyroscope offers considerable resistance, depending on the angular momentum, to any torque which would change the direction of the axis of spin. This property, known as *rigidity*, makes the gyroscope suitable for use as a stabilizer in such cases as resisting the roll of a ship or airplane, maintaining equilibrium of a moving body, and functioning as a steering device.

If the motion of a Foucault gyroscope about one of its axes is restrained and it is given a rotational motion about another axis, the axis of spin will place itself parallel to the latter axis. This principle is exemplified in the gyro-compass.

Gyroscopes in many forms and modifications are used in the control and guidance systems of guided missiles. These are treated at greater length and detail in connection with chapters on those subjects.

This section can be summarized by considering a gyroscopic element as a system with the following properties:

1. Is nonpendulous, and has freedom to rotate about a point.
2. Includes a symmetrical body spinning at a constant high rate about its axis of symmetry (axis of spin).
3. Has all angular momentum of the system effectively concentrated about the spin axis.
4. Is effectively a viscous damper with one end fixed in inertial space.

Inertial space may be conceived as a region in which all accelerations are assumed to be zero and all gravitational forces (mass attractions) are in a state of equilibrium.

With acceleration assumed to be zero, the unaccelerated space must also be nonrotating since rotation always involves acceleration.

Inertial space conforms to Newton's laws of motion, as exemplified in applied force being equal to time rate of change of linear momentum. Linear momentum is a vector quantity representing momentum of translation, and it is equal to the product of the mass and the velocity of the center of the mass.

In our concept of inertial space, which is generally associated with the "fixed" stars, the vector sum of all forces acting throughout the region is equal to zero.

Energy

A body possesses *energy* if it is capable of doing work. When work is performed by a body, its energy is reduced by an amount exactly equal to the work done. If a body is set in motion, it will exert a force and consume energy in coming to rest. Thus it has done work.

A moving body always possesses energy by virtue of its motion: such energy is called *kinetic energy*. A body at rest may possess *potential energy* by virtue of its position. A body possesses such energy when it is at a position from which it is capable of doing more work.

The law of *conservation of matter* states that matter can neither be created nor destroyed. Accordingly, the total amount of matter in the universe remains constant.

From the two laws of physics presented above, we can deduce that when work is performed, energy is released or set in motion. Energy is not created. In the same process, matter may be changed in form; it is not destroyed.

Another form of mechanical work occurs when one body moves over another. Work is done against friction which exists between the bodies. When mechanical work is done on a body, it can be entirely accounted for by one or more of the following effects:

1. Increase in the kinetic energy of the body.
2. Increase in its potential energy.
3. Production of heat due to friction.

Equilibrium

A body which continues in a state of rest is said to be in *static equilibrium*. The simplest condition of equilibrium is that in which a body is acted upon by only two forces. In such a case, one of these forces must be equal and opposite to the other, as indicated in figure 4-9.

The first condition of equilibrium requires that the vector sum of all the forces acting upon a body along any direction must equal zero.

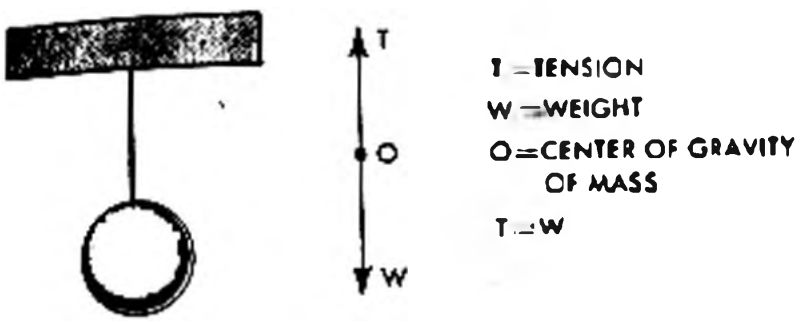


Figure 4-9. Simple Equilibrium

The second condition of equilibrium requires that the torques acting upon a body shall be balanced. The clockwise torques must be equal to the counter-clockwise torques. Notice in figure 4-10 that the algebraic sum of torques equals 0.

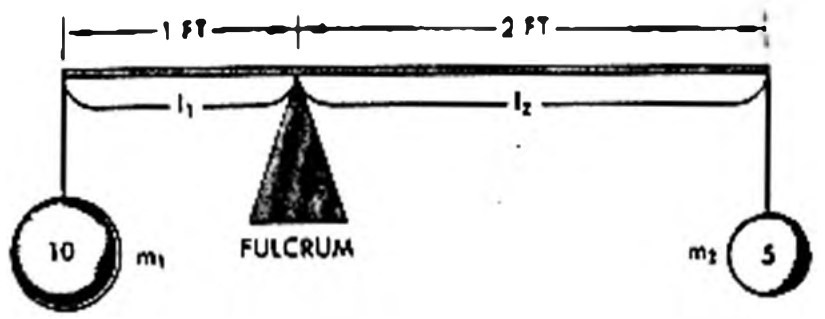


Figure 4-10. Fulcrum-and-Lever Example of Balanced Torque

In the diagram $m_1l_1 = m_2l_2$. Substituting, you have 10 X 1 = 5 X 2.

An important theorem, which follows, is based upon the second condition of equilibrium. "If three nonparallel forces acting upon a body produce equilibrium, their lines of action must pass through a common point."

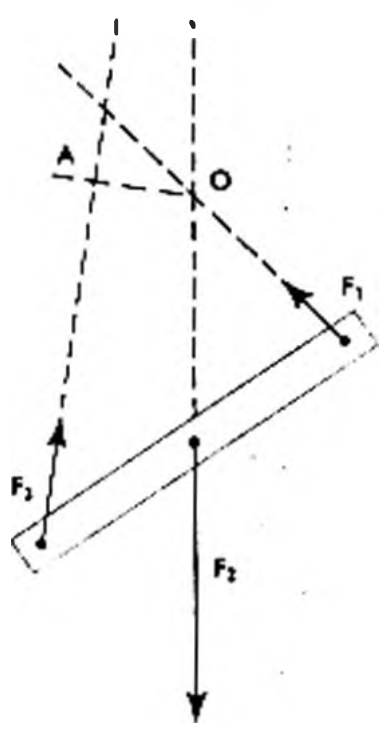


Figure 4-11. Body Under Action of Three Forces

This theorem can be demonstrated by taking moments of forces about the intersection of two of the lines of action and showing that the lever arm of the remaining force must be zero. Assume that a body is in a state of equilibrium under the action of three forces, as shown in figure 4-11.

Since their lines of action are not parallel, any two, if extended, will meet at some point. The lines of action of F_1 and F_2 meet at 0. From this point draw a line, 0A, perpendicular to the third force, F_3, meeting its line of action at A.

Apply the second condition of equilibrium (algebraic sum of torques = 0) by taking the moments of all forces about 0, and setting their sum equal to zero. The moments of F_1 and F_2 are both zero, since the lines of action of these forces pass through 0, and thus their lever arms are zero.

Since the body is assumed to be in equilibrium, you would normally expect the moment of the remaining force F_3 about 0 to be zero. Since the force F_3 is not zero, its lever arm 0A must be zero instead of as shown. The line of action of F_3 must pass through 0. The lines of action of three forces in equilibrium pass through a common point. In this example, the assumption is incorrect because the body cannot be at rest under the given conditions.

You may deduce from the foregoing example that when a body is in a state of equilibrium under the action of three forces, the resultant of any two of these forces is equal and opposite to the third force.

The same principle employed in the preceding

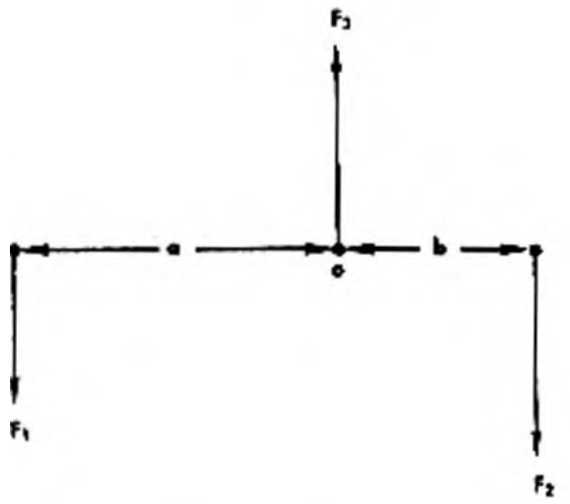


Figure 4-12. Body in Equilibrium under Action of Three Parallel Forces

example may be used in determining the resultant of two parallel forces.

Suppose that the body shown in figure 4-12 is in equilibrium under the action of three parallel forces, F_1, F_2, and F_3. From the first condition of equilibrium, $F_3 = F_1 - F_2$. The resultant of the forces F_1 and F_2 is equal and opposite to F_3. From the second condition of equilibrium $F_2b - F_1a = 0$, when the moments of force are taken about point 0. This equation shows an important relation between the forces and their lever arms: $F_2/F_1 = a/b$.

The resultant of two parallel forces has the same direction as the forces and is equal to their sum, and the resultant's line of action divides the distance between them into two parts which are inversely proportional to the respective forces.

The resultant of any number of parallel forces may be found by first determining the force which, included with the forces given, will produce equilibrium. This force is called the *equilibrant* of the given forces, and *the resultant is equal and opposite to the equilibrant and acts along the same line.*

The attraction which the earth exerts upon a body extends to every particle of matter of which the body is composed. The weight of a body may be considered as an aggregation of parallel forces acting upon the individual particles in the body. If these parallel forces are replaced with their resultant, this single force is exactly equivalent to them.

For any given object there is a particular point through which the resultant of the weight forces will always pass regardless of the orientation of the object. This point is called the *center of gravity* (C.G.) of the body. The weight of an object, although actually a system of parallel forces acting upon all its component parts, can be correctly represented by a single force acting downward at the center of gravity.

There are instances when we want to find the center of gravity of a missile. For example, the center of gravity of a missile has shifted due to adding weight to the craft.

In figure 4-13, if the weight of the missile, less the warhead, is 12,000 pounds and its center of gravity is at point A, the addition of the 2,000-pound warhead with a center of gravity at point B will produce a resultant downward force of 14,000 pounds at a new center of gravity as indicated at point C. If the distance between points A and B is 7 feet, the new center of gravity will be 1 foot, a, forward of the original C.G. as shown at point A. The 1-foot change of position of the C.G. is found in the following manner:

$F_2/F_1 = a/b$, $F_1 \times a = F_2 \times b$

$12{,}000 \times a = 2{,}000\ (7 - a)$ $a + b = 7$ feet

$12{,}000\ a = 14{,}000 - 2{,}000\ a$

$14{,}000\ a = 14{,}000$ or $a = 1$

Harmonic Motion

The principles of harmonic motion are closely associated with those of circular motion. These principles pertain to both mechanical and electrical systems.

Simple harmonic motion is defined as a vibratory motion in which the acceleration of the vibrating body and the restoring force acting upon it are proportional to its displacement from

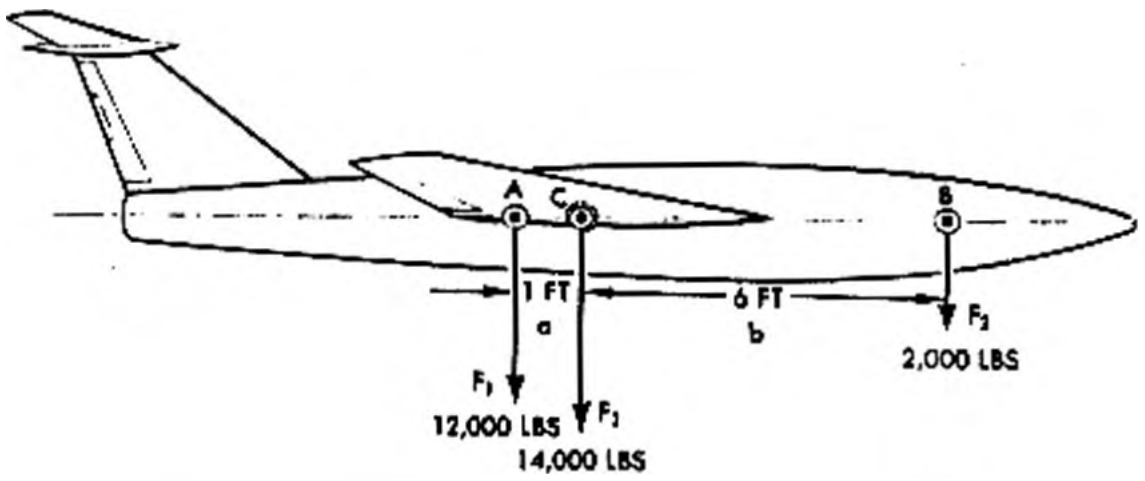


Figure 4-13. Change of Center of Gravity Resulting from Added Weight to a Missile

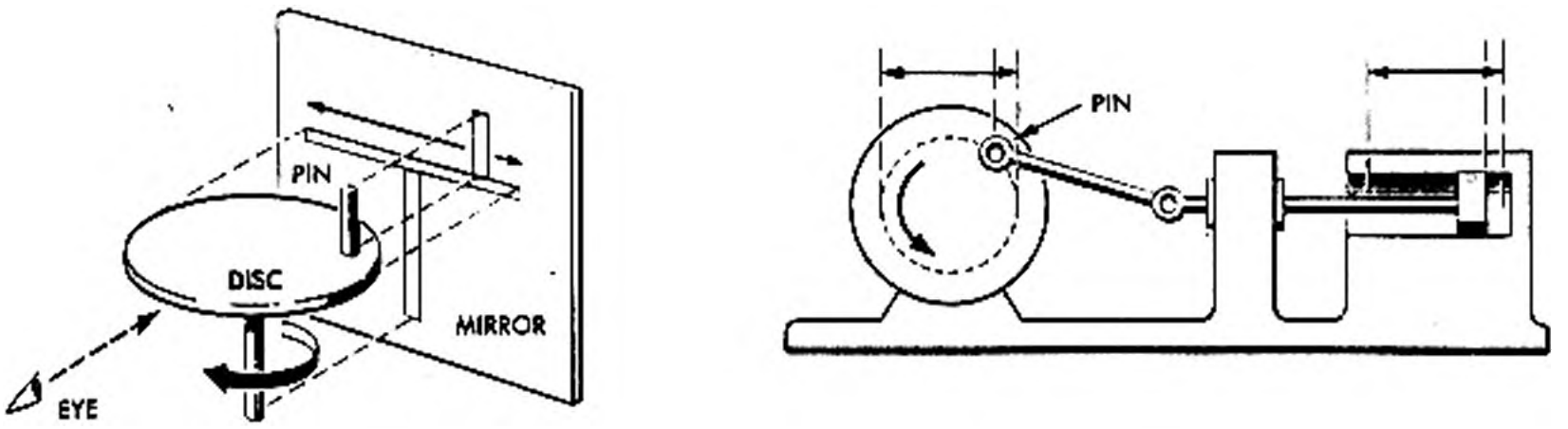


Figure 4-14. Harmonic Motion Compared to Circular Motion

the midpoint of its path and are directed toward that point.

The relationship between simple harmonic motion and circular motion is exemplified in figure 4-14. If the pin on the rim of the rotating disc is viewed along the plane of the disc it appears in the mirror to be vibrating to and fro along a path equal in length to the diameter of the disc. If the disc rotates at constant speed, the projection cast by the uniform circular motion in the plane of the disc is simple harmonic motion.

Similarly in the example of the engine, note that the circular motion of the wheel is applied to a piston which is so mounted that its motion is restricted to a single plane. In this case, when the wheel is rotated at constant speed, the drive-pin travels in a circle at a uniform rate, but the piston receives only horizontal projection of this motion. It slides to and fro in the cylinder through a distance equal to the diameter of the circle described by the pin. The to-and-fro movement of the piston is simple harmonic motion.

You must bear in mind that not all to-and-fro motion is harmonic motion. Many machines employ parts which move back and forth repeatedly along the same path in equal time intervals but do not conform to the definition of harmonic motion. Such motion may be classed as *periodic motion* since it recurs in equal periods of time. But such motion is not harmonic since the rate of change of the motion is not linear. It does not necessarily follow that a body which travels in a circular motion describes harmonic motion.

Pendulums

As you study further the makeup of a missile, you will come to realize that many basic units used in missile control and guidance systems use the principle of the pendulum. Examples of such units are inertial switches, stabilized platforms, timing standards, and compensating devices.

A *pendulum* is defined as a body so suspended from a fixed point as to swing freely to and fro under the combined action of gravity and momentum.

Usually, the theory of the pendulum is explained in terms of a simple or mathematical pendulum, which is shown in figure 4-15. The *simple pendulum* is a purely theoretical device

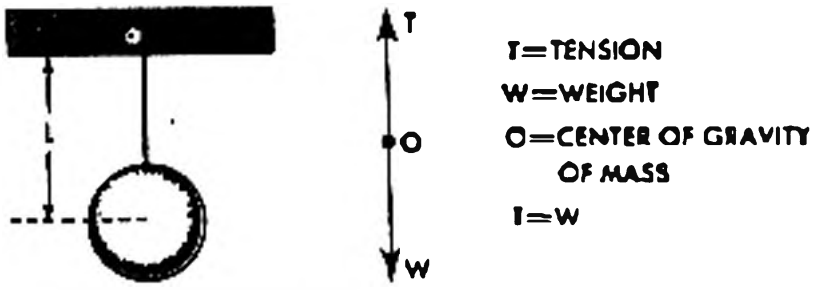


Figure 4-15. Simple or Mathematical Pendulum

consisting of a particle or material point suspended by a thread without weight and oscillating without friction. By visualizing the pendulum as existing under these conditions, the time of vibration can be determined by considering the pendulum to be displaced slightly and then by studying the forces acting upon it. An analysis will show that the unbalanced force is proportional to the displacement and opposite in direction. The vibration (oscillation) of the pendulum is harmonic, and the equations of harmonic motion can be used to determine the period of vibration, T.

$$T = 2\pi \sqrt{\frac{L}{g}} \qquad T = \text{Period}$$

The time of vibration of a pendulum is determined from the amplitude of vibration (or the acceleration due to gravity, g, at the point where the pendulum is located) and its length, L (or the distance between the axis of suspension and the axis of oscillation).

By using a pendulum of known length and measuring its period of vibration, the value of gravity can be determined. The equation

$$T = 2\pi \sqrt{\frac{L}{g}}$$

applies only to small values of displacement angle; the values need not be constant. As the vibration of the pendulum continues, the amplitude of oscillation continually decreases, but the period remains practically unchanged. This property of the pendulum makes it especially suitable for controlling the escapement mechanisms of time-keeping devices.

Another type of pendulum, the *physical or compound pendulum*, is a body which vibrates in the manner of a pendulum but in which the mass is distributed and not concentrated as in the simple pendulum.

The *center of oscillation* of a physical pendulum is that point at which the concentration of the whole mass of the pendulum would produce no change in its period of vibration. If the mass were so concentrated, the physical pendulum would be identical to a simple pendulum having the same period of vibration.

$2\pi \sqrt{\frac{L}{g}}$ represents the period of the simple pendulum.

$2\pi \sqrt{\frac{L}{Mgh}}$ represents the period of the physical pendulum.

Equating the two preceding expressions, the length of the equivalent simple pendulum is

$L = \frac{I}{Mh}$, where I = moment of inertia;

h = distance between axis of suspension and the center of gravity.

The center of oscillation of a physical pendulum is separated from the axis of suspension by the distance, I/Mh. This expression also shows the length of the simple pendulum which has the same period of vibration as the physical pendulum. The center of oscillation may be interchanged with the center of suspension without affecting the period. These characteristics of the pendulum lend themselves to the measurement of vibration periods, centers of gravity, and other factors of missile configuration and their probable effects upon the flight characteristics of the missile.

The center of oscillation is also the *center of percussion* because at this point the pendulum can be struck with no resulting jar on the axis of suspension. This property is of value in the determination of correct suspension of components for absorbing shock. When struck at any point other than the center of percussion, the pendulum not only tends to rotate but the axis of suspension force receives a motion of translation.

Still another type pendulum, a *torsion pendulum*, is illustrated by figure 4-16. It is a suspended body that vibrates, not by swinging but by rotating

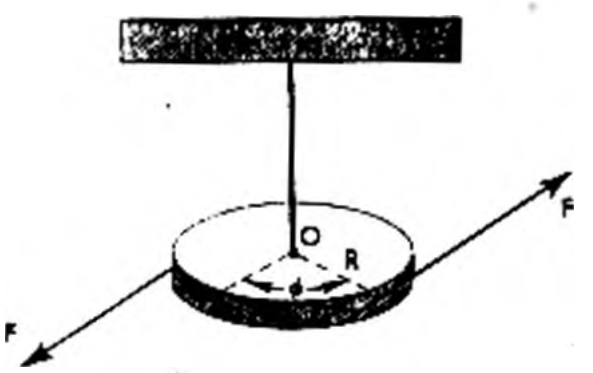


Figure 4-16. Torsion Pendulum

with an alternate twisting and untwisting movement.

The characteristics of this type pendulum make it especially suitable for use in timing devices, balance wheels of watches, torque-measuring instruments, torque-compensation devices, magnetometers, strain-gages, and others.

You can consider a torsion pendulum as a weight attached to a vertically suspended wire. When the wire is twisted and released, it describes a series of angular vibrations similar to the linear vibrations of harmonic motion.

When the suspension wire is twisted, an oppositely directed restoring torque which is proportional to the angular displacement is set up within the wire. Just as in linear harmonic motion, a restoring force is set up which is proportional

to the linear displacement. The two types of motion are closely analogous, and angular harmonic motion can be represented most simply by comparing it directly to linear harmonic motion.

You can take the equation

$$\frac{F}{x} = 4\pi^2 n^2 m$$

which expresses linear harmonic motion, and replace each linear quantity with its corresponding angular quantity and derive a corresponding expression for angular harmonic motion:

$$\frac{T}{\phi} = 4\pi^2 n^2 I$$

where T represents the restoring torque, ϕ is the corresponding angular displacement, n is the frequency, and I is the moment of inertia of the vibrating body about the axis of rotation.

The ratio of the torque to the corresponding twist is a constant determined by the stiffness of the suspension wire and is essentially negative (opposite to the applied force). You can replace the ratio T/θ with τ (tau) and rearrange the expression. The new arrangement shows the frequency of angular vibration to be:

$$n = \frac{1}{2\pi}\sqrt{\frac{\tau}{I}}$$

Consequently, the period is expressed by

$$T = 2\pi\sqrt{\frac{I}{\tau}}$$

which is analogous to

$$T = 2\pi\sqrt{\frac{L}{g}}$$

This latter equation expresses the period of linear vibration.

The torsion principle can be used to determine the moment of inertia of a body by supporting the body as a torsion pendulum and measuring its period and the angle of twist produced by a measured torque.

By transposing equation

$$T = 2\pi\sqrt{\frac{I}{\tau}}$$

the moment of inertia I of the suspended body can be determined:

$$I = \frac{\tau T^2}{4\pi^2}$$

In figure 4-16, the forces F acting at the distance R from the center O produce a torque 2RF.

MECHANICS OF GASES

Gas is one of the three basic forms in which matter exists, the other two forms being liquid and solid.

Gas can be defined as an aeriform fluid which has neither independent shape nor volume but tends to expand in an unlimited manner. This tendency to expand results from the fact that the molecules of a gas are relatively far apart and are continuously in motion throughout the entire space occupied by the gas.

In many respects, gases resemble liquids. Since both are capable of flowing, they are commonly referred to as fluids.

The two major differences between gases and liquids are in respect to compressibility and expansion. Gases are highly compressible, while most liquids are but slightly so. Gases tend to completely fill any closed vessel in which they may be contained, while most liquids, like solids, fill a container only to the extent of their normal volume.

Both gases and liquids exert pressure upon the surfaces with which they are in contact, and each exerts an upward buoyant force that is in accordance with the principle of Archimedes, which states that "a body submerged wholly or partially in a fluid is buoyed up by a force equal to the weight of the fluid which it displaces."

Flowing gases tend to conform to Bernoulli's theorem on liquids when their compressibility is considered as a factor. The velocity of their effusion through an orifice may be calculated in the same manner as for liquids. Bernoulli's theorem states that if no work is done on or by an incompressible liquid as it flows, the total head remains unchanged. "Total head" refers to the total energy of a moving liquid at a given point. From this theorem, we can deduce that when a compressible fluid flows, the total head varies in proportion to the degree of compressibility.

The mathematics of Bernoulli's theorem are simple when applied to hypothetical fluids considered as being incompressible or to water which

is but very slightly compressible. But the mathematics become complex when applied to gases which are highly compressible in varying degrees and whose densities differ with differences in pressure.

This general effect applies to all fluids. In respect to gases, it is enough to keep in mind that as a flowing stream of gas increases in velocity, its pressure decreases, and vice versa. This effect is noticeable in the case of lift exerted upon a moving airfoil and governs the design of the airfoil for specific applications. Lift is discussed in chapter 2.

Gases, like liquids, adapt themselves to the shape of the vessel in which they may be contained and, having no elasticity of shape, are unable to exert shearing stresses other than those due to their viscosity.

A gas which can be liquefied by pressure alone is termed a vapor. At room temperatures, steam and carbon dioxide are called vapors. Air, hydrogen, and nitrogen are called gases.

Kinetic Theory of Gases

The simple structure of gases makes them readily adaptable to mathematical analysis from which has evolved a detailed theory of the behavior of gases. This is called the *kinetic theory of gases*. The theory assumes that a body of gas is composed of identical molecules which behave like minute elastic spheres, spaced relatively far apart and continuously in motion.

The degree of molecular motion depends upon the temperature of the gas. Since the molecules are continuously striking against each other and against the walls of the container, an increase in temperature with the resulting increase in molecular motion causes a corresponding increase in the number of collisions between the molecules. The increased number of collisions results in an increase in pressure because a greater number of molecules strike against the walls of the container in a given unit of time.

If the container were an open vessel, the gas would tend to expand and overflow from the container. If the container is sealed and possesses elasticity (such as a rubber balloon), the increased pressure will cause the container to expand.

You may have noticed when making a long drive on a hot day, the pressure in the tires of your automobile increases and that a tire which appeared to be somewhat "soft" in the cool morning temperature may appear normal at a higher midday temperature.

Such phenomena as these have been explained and set forth in the form of laws pertaining to gases, and they tend to support the kinetic theory.

At any given instant, some molecules of a gas are moving in one direction, some in another direction. Some are traveling fast, while some are traveling slowly. Some may even be in a state of rest. The combined effect of these varying velocities corresponds to the temperature of the gas. In any large amount of gas, there are so many molecules present that in accordance with the "laws of probability," some average velocity can be found which, if it were possessed by every molecule in the gas, would produce the same effect at a given temperature as the total of the many varying velocities.

To raise the temperature of a substance requires energy, so you can assume that the temperature of a gas is directly proportional to the mean kinetic energy of the gas molecules. You must realize that this mean velocity must be such that it imparts the same kinetic energy to a given number of molecules of specific mass as would be imparted by their various individual velocities.

Atmospheric Pressure

A gas that is vital to missile flight is air. The mass of air surrounding the earth and held to it by gravitational attraction exerts a pressure, p, upon the earth's surface:

$$p = hdg$$

where: h = the height of the air layer
d = the density of the air layer
g = the gravitational attraction of the earth

This atmospheric pressure can be measured by any of several methods. The common laboratory method employs the *mercury column (barometer)*. A mercury column consists of a glass tube approximately 34 inches in length, sealed at one end, then completely filled with mercury, and

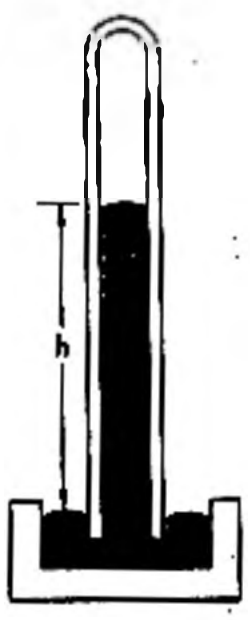


Figure 4-17. Mercury Column for Measuring Atmospheric Pressure

inverted in an open cup partially filled with mercury, as shown in figure 4-17.

The mercury in the tube settles down, leaving an evacuated space in the upper end of the tube. The height of the mercury column serves as an indicator of atmospheric pressure.

At sea level and at a temperature of 0° C, the height of the mercury column is 76 cm or approximately 30 inches, which represents a pressure of 14.7 pounds per square inch. The 30-inch column is used as a reference standard.

At higher levels, the atmospheric pressure on the surface of the mercury in the open cup is less than at sea level; hence the column of mercury in the tube settles lower. These variations in the height of the mercury column represent changes in atmospheric pressure which may be calibrated in terms of altitude with reference to sea level.

Another device used to measure atmospheric pressure is the *aneroid barometer*, pictured in figure 4-18. This barometer consists of a small sealed metal drum that has been partially evacuated and which has one side or end made of thin corrugated metal. This thin metal moves in or out with the variation of pressure on its external surface. This movement is transmitted through a system of levers to a pointer. The pointer is pivoted so that it sweeps across a graduated scale calibrated either in pounds per square inch or in feet to designate an altitude corresponding to the indicated pressure. This device is used in aircraft altimeters and as an end instrument or transducer in telemetering circuits to obtain information pertaining to pressures and altitude.

The atmospheric pressure does not vary uniformly with altitude. It changes more rapidly at lower altitudes because of the compressibility of the air which causes the air strata close to the earth's surface to be compressed by the air masses above them. This effect is partially counteracted by contraction of the upper strata due to cooling. The cooling tends to increase the density.

Altitude can be approximately estimated from a knowledge of the corresponding atmospheric pressure. When proper compensation is made for humidity, temperature, and any other variables, fairly accurate results can be obtained from the barometric altimeter.

Atmospheric pressures are quite large, but in most instances practically the same pressure is present on all sides of objects so that no single surface is subjected to a great load. To exemplify the tremendous atmospheric pressure which may exist on a surface, consider the pressure present on the face of an ordinary 5-inch cathode-ray tube as used in oscillographs. At sea level the pressure on the face of this tube is approximately 300 pounds. Consider also an average 17-inch rectangular television tube. It has a pressure of almost 2400 pounds on its screen surface at sea level.

The force produced by atmospheric pressure is calculated by the use of the equation

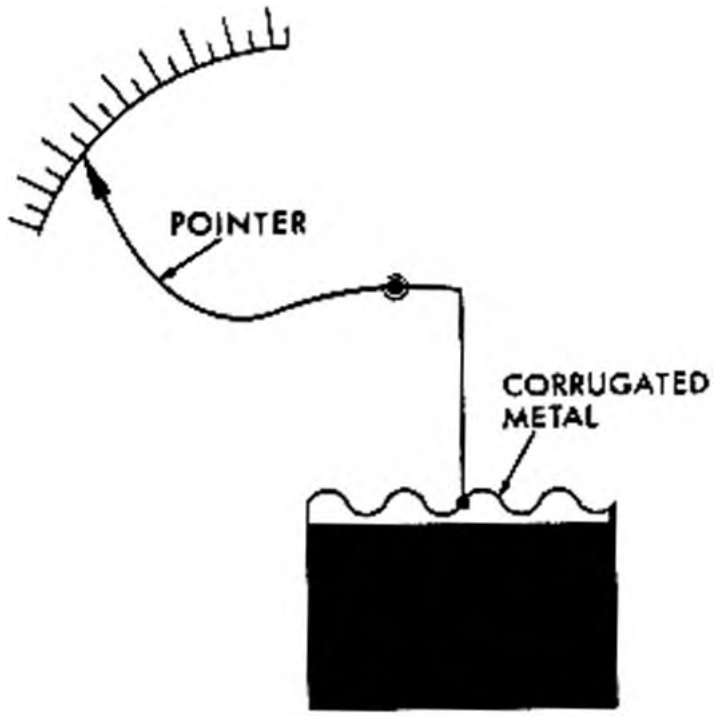


Figure 4-18. Aneroid Barometer

$$F = pA,$$

where: F = total pressure on a surface
p = atmospheric pressure per square inch under specific conditions of altitude and density
A = the area of the surface in square inches

Compressibility of Gases

Compressibility is an outstanding characteristic of gases. The simple relationship between the pressure of a gas and its volume is stated in *Boyle's law*, which states that "the volume of a confined body of gas varies inversely as the absolute pressure, provided that the temperature remains constant."

This law can be demonstrated by confining a quantity of gas in a cylinder which has a tightly fitted piston. A force is then applied to the piston so as to compress the gas in the cylinder to some specific volume. When the force applied to the piston is doubled, the gas is compressed to one-half its original volume, as indicated in figure 4-19.

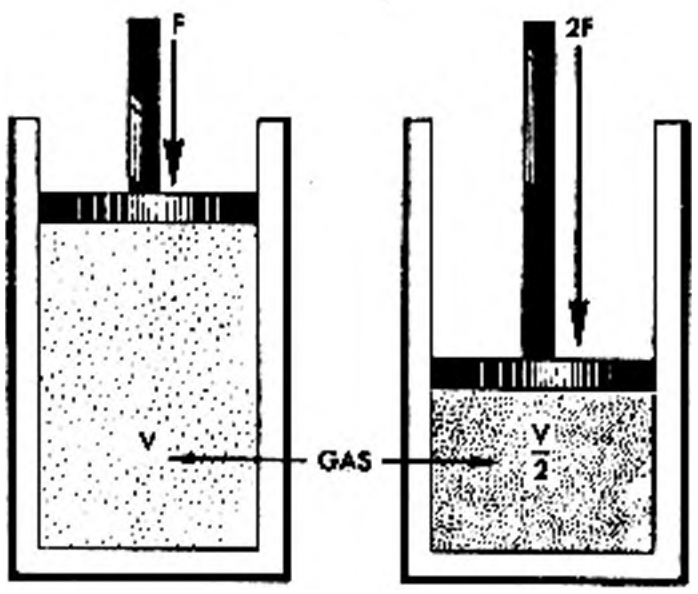


Figure 4-19. Gas Compressed to Half its Original Volume by Double Force

If p_A and V_A represent the pressure and volume under one set of conditions, and p_B and V_B represent the pressure and volume under another set of conditions, then

$$V_A : V_B = p_B : p_A$$
$$p_A V_A = p_B V_B$$

A gas which conforms to Boyle's law is termed an ideal gas. When pressure is increased upon such a gas, its volume decreases proportionally and its density is increased. The density of a gas varies directly as the pressure, if temperature remains constant as in the case of an ideal gas. Density also varies with temperature, since gases expand when heated and contract when cooled.

The specific gravity of a gas is computed with reference to air as a standard. The density of air has been determined accurately. One liter (1000 cubic centimeters) of air at 0° centigrade and 76 cm of mercury weighs 1.293 grams. The density of air is 0.081 pounds per cubic foot under standard conditions.

By dividing the density of a gas by the density of air, the relative specific gravity of the gas is found. Table 4-1 shows the density and specific gravity of some common gases.

Table 4-1. Comparison of Density and Specific Gravity

GAS	DENSITY		SPECIFIC GRAVITY
	gm/liter	lb/cu ft	
Air 0° C 76 cm Hg	1.293	0.081	1.000
Air 20° C	1.205	0.0755	0.932
Steam 100° C	0.598	0.037	0.462
Hydrogen	0.090	0.0056	0.069
Helium	0.179	0.011	0.138
Nitrogen	1.251	0.078	0.967
Oxygen	1.429	0.089	1.105
Carbon dioxide	1.977	0.123	1.529

Dalton's Law

If a mixture of two or more gases which do not combine chemically is placed in a container, each gas expands throughout the total space and the absolute pressure of each gas is reduced to a lower value called its *partial pressure*. This reduction is in accordance with Boyle's law. The pressure of the mixed gases is equal to the sum of the partial pressures. This fact was discovered by Dalton, an English physicist, and is set forth in Dalton's law which states that "a mixture of several gases which

do not react chemically exerts a pressure equal to the sum of the pressures which the several gases would exert separately if each were allowed to occupy the entire space alone at the given temperature and in the same volume as each is present in the mixture."

Avogadro's law

An Italian physicist, Avogadro, conceived the theory that "at the same temperature and pressure, equal volumes of different gases contain equal numbers of molecules." This theory was proven by experiment and found to agree with the kinetic theory. It has come to be known as Avogadro's law.

Viscosity

Another property common to both gases and liquids is *viscosity*, which is defined as *resistance to flow*. Viscosity is due to molecular friction and is less pronounced in gases than in liquids because of the greater space between molecules in the gases.

The viscosity of gases noticeably retards their flow through pipes or tubes and also retards the passage of other bodies through a stationary body of gas. Increasing the temperature of a gas increases its viscosity as a result of the increase in the movement of its molecules.

Viscosity can be considered as the resultant of cohesion and adhesion. *Cohesion* is the molecular attraction which tends to unite all particles of a body throughout its mass. *Adhesion* is the molecular attraction exerted between the surfaces of bodies which are in contact. Airfoils are streamlined to reduce the effect of this air friction, thereby increasing the velocity of the aircraft of which they may be components.

All of the phenomena of gases mentioned in this section and the laws to which they conform must be considered in the design of aircraft from the viewpoints of both airframes and propellants. Specific applications of the mechanics of gases are treated in detail in other chapters of this manual in connection with subjects to which they are pertinent.

Physics Thus Far

Quickly review what has been covered in this chapter so far. Keep in mind that *inertia* is a property of matter by which matter tends to remain at rest or, if in motion, tends to remain in motion in the same direction unless acted upon in either case by some external force.

Velocity is a representation of speed and a specific direction of motion. *Angular velocity* refers to a variation in direction of a moving object. *Acceleration* is the time rate of change of velocity. It represents motion in which the velocity changes from point to point.

Centrifugal and centripetal forces can be illustrated by a ball attached to a cord. When the ball is whirled at the end of the cord, the cord becomes taut and exerts an inward pull or centripetal force on the ball. The ball pulls outwardly, exerting a centrifugal force on the cord.

Angular acceleration is produced by torque. *Torque* is a rotational effect exerted on a body and is measured by the product of the force acting on the body and the perpendicular distance from the axis of rotation to the line of action of the force.

When two angular velocities about separate axes are added to a body, an angular motion about a third axis is produced. This new motion is called *precession*. The operation of a gyroscope exemplifies the phenomena of precession.

Moving bodies possess *kinetic energy* by virtue of their motion. A body may possess *potential energy* by virtue of its position. A body possesses potential energy when it is at a position from which it can do work.

A body which continues in a state of rest is said to be in *static equilibrium*. Equilibrium of a body depends upon two conditions. The first condition requires that the vector sum of all the forces acting upon a body along any direction must equal zero. The second condition of equilibrium requires that the torques acting upon a body shall be balanced.

Simple harmonic motion is defined as a vibratory motion in which the acceleration of the vibrating body and the restoring force acting upon it are proportional to its displacement from the midpoint of its path and are directed toward that point. The to-and-fro motion of a pendulum describes simple harmonic motion.

Remember that gases are highly compressible and that they tend to completely fill any closed

vessel in which they may be contained. Also remember that as a flowing stream of gas increases in velocity, its pressure decreases, and vice versa. The temperature affects the pressure that a gas exerts against the walls of a confining container. An increase in temperature results in an increase of molecular motion which, in turn, causes an increase in pressure. This behavior is referred to as the *kinetic theory of gases.*

A phase of physics which is important in the missile field is optics or, in other words, the science of light. Optics are discussed in the following section.

SECTION A
Optics Involved in Guided Missile Operation

Some of the systems of navigation and homing used in modern missiles are based upon the behavior and phenomena of light or some form of radiant energy possessing characteristics similar to those of light. Examples of such systems are the heat-seeking and light-seeking homing and automatic celestial navigation systems.

To use the properties of light and radiant energy in missile systems, the radiant energies must be transformed or adapted to a form of energy suitable for actuating or controlling the mechanical or electrical components employed in guiding and controlling missiles. This adaptation generally involves the use of an optical system in conjunction with the electrical and/or mechanical units used to control the flight of a missile or to determine its position. Your primary concern is the method by which the sources of light or radiant energy are employed as reference points in homing and navigation systems.

For you to understand the theory and operation of homing and navigation systems, some knowledge of the properties of light and of radiant energy is essential.

Modern theory considers light, both visible and invisible, as consisting of *quanta* (bundles) of energy which move as if guided by waves. The statistical behavior of the quanta depends upon the assumption that the energy of the quanta at any point is on the average equal to the intensity of the wave system at that point.

The *wave theory of light* assumes that light is transmitted from luminous (light-emitting) bodies to the eye and other objects by an undulatory or vibrational movement. The velocity of this transmission is approximately 186,300 miles per second, and the vibrations of ether (conducting medium of light in space) are transverse to the direction of propagation of the wave motion. These waves vary in length from approximately 3.85 to 7.60 ten-thousandths of a millimeter.

The impression of *color* produced when the light energy impinges on the retina of the eye varies in a complex way with the wavelength, the amplitude of vibration, and various other factors and conditions, some of which are beyond the scope of this manual.

Waves of similar character, but whose lengths are above or below the limits mentioned in the preceding paragraph, are not perceptible to the average eye under normal conditions. The very short waves between 1.0 and 3.85 ten-thousands of a millimeter in length constitute *ultraviolet* light and are detectable by photographic or chemical action. Those waves which are longer than 7.60 ten-thousandths of a millimeter are the *infrared* waves and are detected by their thermal (heat) effects.

The *electromagnetic theory of light* as set forth by Maxwell, English physicist, holds that these waves, including those of light proper, are the same kind as those by which electromagnetic oscillations are propagated through ether and that light is an electromagnetic phenomenon.

The most important phenomena of light are reflection, refraction, dispersion, interference, and polarization. One or more of these phenomena of light act through a suitable optical system as a medium for missile homing and navigation.

Optics is the science pertaining to light, light's origin and propagation, the effects to which it is subject and which it produces, and other phenomena closely associated with it. *Geometrical optics* is concerned with the optical phenomena associated with reflection and ordinary refraction as they can be deduced mathematically from the simple laws of reflection which have been derived from observation and experimentation. *Physical optics* is concerned with the description and explanation of all optical phenomena in terms of physical theories, as wave theory in general, electro-

magnetic phenomena, quantum mechanics, and other light properties.

One of the mysteries pertaining to radiant energy such as light, heat, and electromagnetic waves is the medium through which it is conducted. Only because the medium absorbs the energy and changes it to some other form are we able to recognize its existence and determine its characteristics. As indicated earlier, the word *ether* is used to name the medium through which radiant energy is conducted. But it is not known what the medium actually is.

The most obvious fundamental property of light is that it travels in straight lines when passing through a homogeneous (uniform in density and composition) medium. The rectilinear (straight-line) propagation of light supports the idea that a ray of light is the rectilinear path in a homogeneous medium along which light is propagated or transmitted.

By choosing one point on a luminous body and from that point drawing a straight line in the direction of the propagation of light, you can represent a *ray of light*. From this point-source of light, you can draw an infinite number of rays, as shown in figure 4-20. This collection of rays, or cone of light, is referred to as a *pencil of rays*. Two rays from such a group are enough to locate the point-source of the light by simple geometric means. The point-source is the point of intersection of lines extended along the paths of the rays.

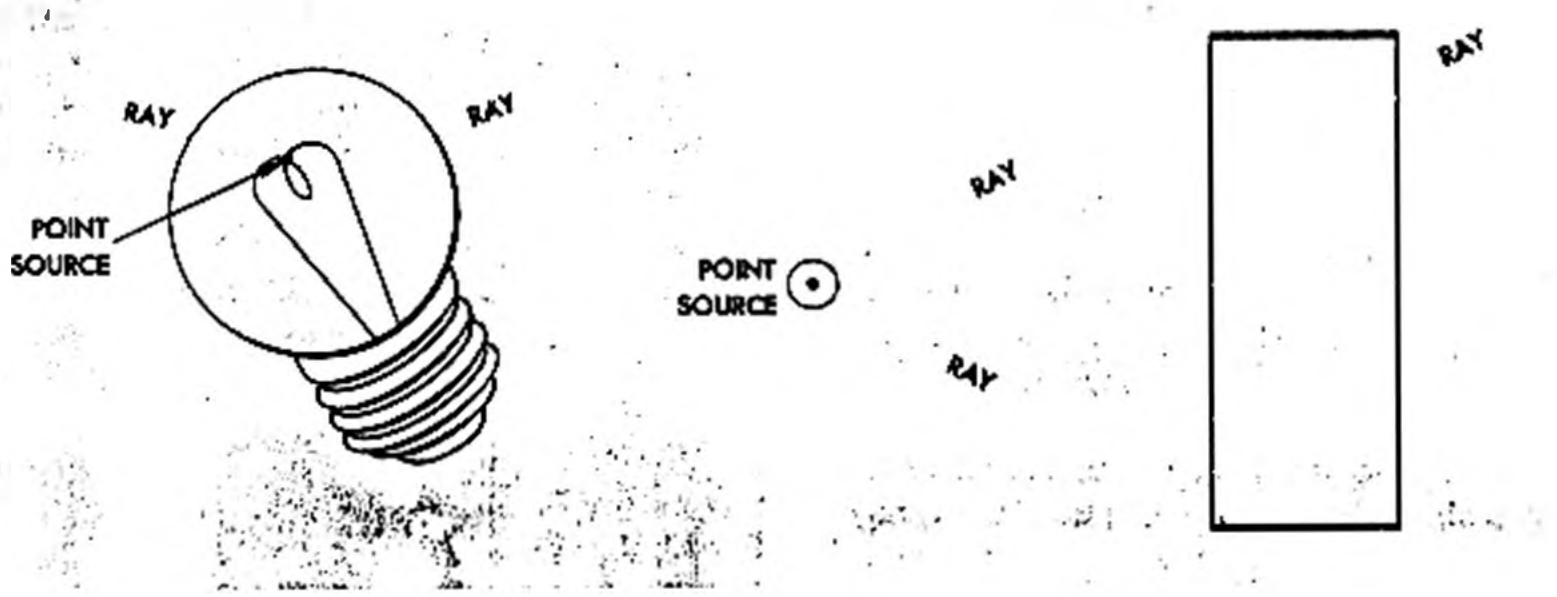


Figure 4-20. Light Rays from Point-Source

UNITS OF LIGHT INTENSITY

Luminous intensity or brightness of light represents the degree to which visible light is present in the radiant energy emitted by the source.

The retina in the human eye is sensitive to a relatively small portion of the radiant energy emitted by an incandescent body. To measure the relative intensities of visible light, special standards and techniques must be used. Such standards and techniques constitute the science of *photometry*.

Light can be considered as a flow of radiant energy or *luminous flux* (expressed in ergs per second). Because of the variations of sensitivity of the human eye to different colors (different wavelengths of light), luminous flux cannot be measured visually in ergs per second. In place of the erg, a unit called the lumen must be employed for this purpose.

A *lumen* is the amount of light flux radiating from a uniform 1-candlepower (1-cp) source throughout a solid angle of such size as to surround a unit area at a unit distance from the source.

Light flux refers to the rate at which a source emits light energy, evaluated in terms of its visual effect.

If you imagine the 1-candlepower source to be located at the geometric center of a hollow sphere having a radius of 1 foot, then each square foot of the interior surface of the sphere receives 1 lumen of light. Since the total area of the sphere

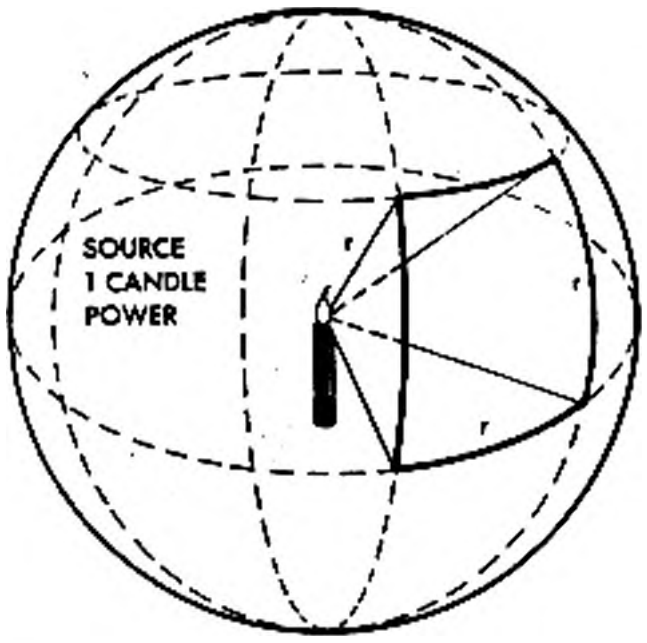


Figure 4-21. Light Falling on Area R² at Distance R from 1-cp Source Equals 1 Lumen

is 4 pi (4 π) square feet, the total light emitted by the 1-candlepower source is 4 π lumen. Figure 4-21 of a sphere and solid angle illustrates this example. Light from 1-candlepower source falling upon a unit area at unit distance represents 1 lumen.

Most sources of light have different luminous intensities along different directions. The average of the candlepowers measured in all directions about a source of light is called the *mean spherical candlepower*. If a source having a mean spherical candlepower of 1 cp emits 4 π lumens of light flux, the total flux, F, in lumens emitted by a source of mean spherical candlepower, I_o, may be expressed by the equation $F = 4\pi I_o$.

The primary standard of luminous intensity, developed by the National Bureau of Standards, consists of a glowing enclosure operated at the temperature of solidifying platinum (2046° kelvin or 2187° centigrade) and arranged as shown in figure 4-22. The platinum is contained in a crucible of fused thorium oxide or quartz, surrounded by a heat-insulating material. The unit is placed in an alternating magnetic field so that the platinum in the crucible is melted by the currents induced within it. A viewing tube of fused quartz or thorium oxide, containing some finely powdered thorium oxide, is enclosed by the molten platinum and serves as a "black body" radiator. The brightness within this tube is considered to be 60 candlepower per square centimeter when the metal, in cooling slowly, attains its solidifying temperature. The new *standard candle* is one-sixtieth of the luminous intensity of one square centimeter of a hollow enclosure at the temperature of solidifying platinum.

The amount of light flux which falls upon a surface and the area of the surface together determine the degree of illumination. The illumination is uniform only when a single source of light is employed and when all points on the illuminated surface are equidistant from the source.

A light intensity of 1 lumen per square foot is a foot-candle and is the amount of illumination received on a surface 1-foot distant from a standard candle. The illumination of a surface is measured by the number of lumens incident upon a unit of area. $E = I/d^2$, where E represents the total illumination, I equals the intensity of the source in candlepower, and d equals the distance from source to surface.

If you consider a surface of area, A, as receiv-

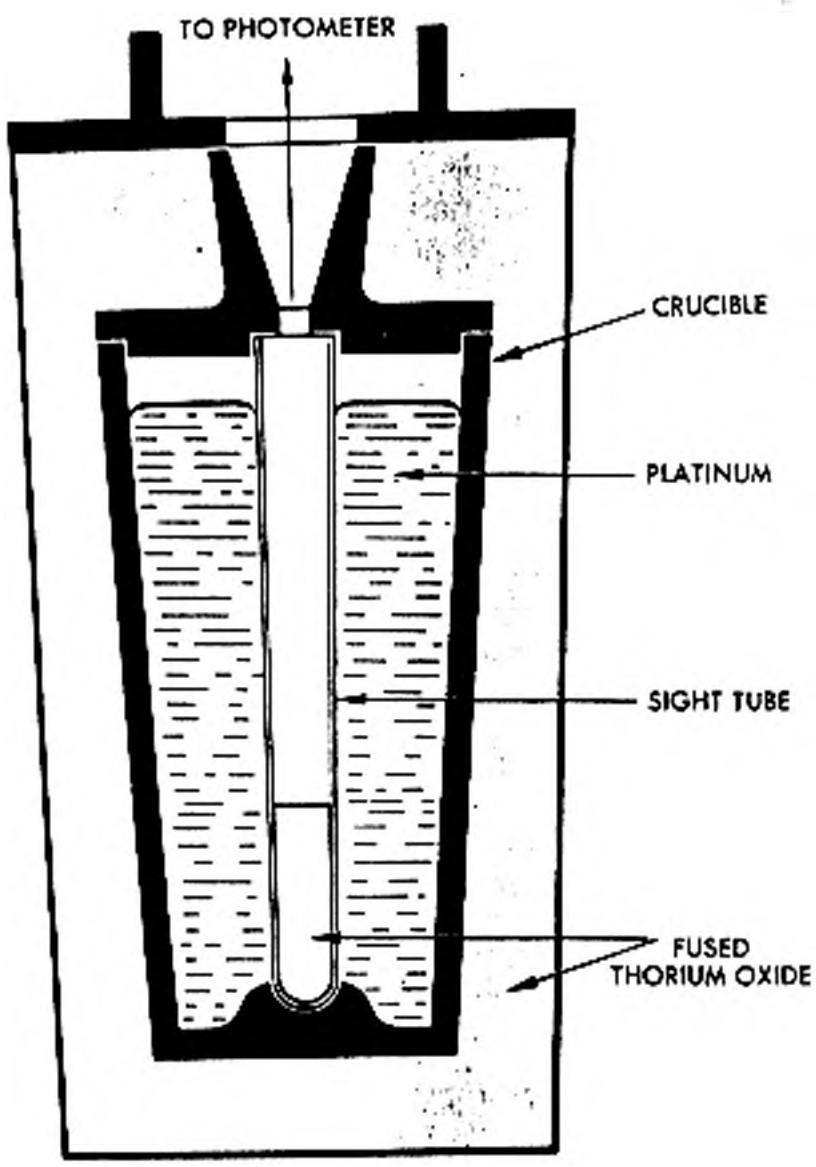


Figure 4-22. National Bureau of Standards Primary Standard of Luminous Intensity

ing a total light flux, F, you can express the illumination of the surface in terms of lumens per unit of area, such as lumens per square foot. Expressed as an equation, $E = F/A$.

The degree of illumination which a light source produces upon a given surface depends upon the intensity of the source and its distance from the surface, provided that the rays of light pass through a uniform medium and strike the surface normally.

Increasing the intensity of the source produces a proportional increase in the light flux falling upon the surface. Increasing the distance of the surface from the source decreases the illumination of the surface by an amount proportional to the square of the distance. Doubling the distance between the source of light and the illuminated surface will reduce the illumination of the surface to one-fourth its original value.

This effect is common to all forms of radiant energy and is expressed in the *inverse square law* which states that the "radiant flux density at any surface varies inversely as the square of the distance of that surface from the source of radiation."

The numerical value of illumination is identical whether expressed in foot-candles or in lumens per square foot.

Remember the equation $E = I/d^2$ can be used to calculate illumination only for a spherical surface with the source of radiation at its center. This condition is known as *normal illumination*.

For small surfaces where the distance from source to surface is large in comparison to the dimensions of the surface, the formula can be used with little error. Over large surfaces where all the flux paths are not perpendicular to the surface, the diffusion of light is not uniform. The flux intensity is reduced at points farther removed from the source by an amount proportional to the cosine of the angle of incidence. This factor must be included in calculations. The formula

$$E = \frac{I \cos i}{d^2}$$

is much more accurate under these conditions. Even with this formula the surface dimensions must still be small as compared to the distance from the source.

Up to this point, only the luminous intensity of a source in terms of point-sources have been considered. When referring to luminous intensities of larger surfaces which may or may not be self-luminous, the quantity, or term brightness, must be used to specify luminous intensities of unit area. *Brightness* is defined as the luminous intensity of a unit area of a surface in a given direction. Brightness is expressed in terms of candlepower per square unit of area. It is generally expressed in terms of square centimeters.

The difference between brightness and illumination can be illustrated by considering this page which you are reading. The page is uniformly illuminated (or nearly so), but the printed letters reflect less of the incident light and therefore are less bright than the white paper upon which they are printed.

Brightness of a surface and the illumination on the surface would be numerically equal only if the surface reflected all of the light that fell upon it. Table 4-2 shows the approximate values of brightness of some familiar self-luminous and non-luminous objects.

Table 4-2. Brightness of Common Objects

Sun's disk	153,000,000 cp./sq. ft.
Crater of a carbon arc	14,000,000 cp./sq. ft.
Tungsten lamp filament	465,000 cp./sq. ft.
Moon's disk	465 cp./sq. ft.
Clear blue sky	370 cp./sq. ft.
Newspaper	1.8 cp./sq. ft.

Generally the brightness of a surface depends upon the direction from which it is viewed, but there are some materials which scatter light in such a manner that their brightness is the same from all angles of view. Examples of such light-diffusing substances are magnesium-oxide and new-fallen snow. For surfaces of this type, a unit of brightness called the lambert is used.

A *lambert* represents the brightness of a perfectly diffusing surface which is emitting or scattering light in the amount of 1 lumen per square centimeter. When the reflected light is less than the incident light, the brightness in lamberts is equal to the product of the illumination and the reflection coefficient of the surface material.

MEASUREMENT OF LIGHT INTENSITY

The relative intensities of two or more sources are not discernible to the human eye by direct viewing, but whether or not two surfaces side by side are equally illuminated can be determined accurately. The matching of illumination on two adjacent surfaces is the basic principle of the *photometer*, a device employing two lamps located at some suitable distance apart with a screen located between them. Each side of the screen is illuminated normally by one of the sources. The flux paths from each source are perpendicular (or nearly so) to all points on the surface of the screen.

The screen is moved along the flux path between the two sources until the same degree of illumination is observed on both sides. The distances from the lamps to their respective sides of the screen are then measured. From the equation

$$E = \frac{I}{d^2}$$

you can establish the ratio

$$\frac{I_1}{d_1^2} = \frac{I_2}{d_2^2}$$

where I_1 and I_2 are the luminous intensities of the lamps in candlepower and d_1 and d_2 are their respective distances from the screen. If the value of either I_1 or I_2 is known, the value of the other may be readily computed. Figure 4-23 illustrates the application of the inverse square law as applied in the foregoing calculation.

If distance d_1 is found to be twice the value of d_2 and I_2 represents a source of 16-candlepower intensity, the intensity of I_1 must be four times that of I_2 to produce the same illumination of the translucent screen.

From the second equation above,

$$\frac{I_1}{d_1^2} = \frac{I_2}{d_2^2}$$

Therefore, if d_1 is equal to twice d_2 and if I_2 is 16 candlepower,

$$\frac{I_1}{2^2} = \frac{16}{1^2}$$

Thus, I_1 equals 4 times 16 cp; I_1 equals 64 candlepower.

The foregoing method of computing the intensity of one light source in relation to another is satisfactory when both sources are of the same color. A different procedure is necessary if the lamps are not the same color. Lamps of different colors can be matched with a standard lamp by varying the potential on the calibrated lamp. Lowering the potential produces a noticeably yellower color. Increasing the potential increases the whiteness of the color.

Another method of comparing lamps of different colors is to use a so-called *flicker photometer*, which employs a rotating prism. The rotating prism enables the viewer to see one side of the screen and then the other alternately in rapid succession. Upon increasing the frequency of alternation, a value will be found for which the flicker resulting from color difference disappears. The colors appear to blend into a single hue. If the

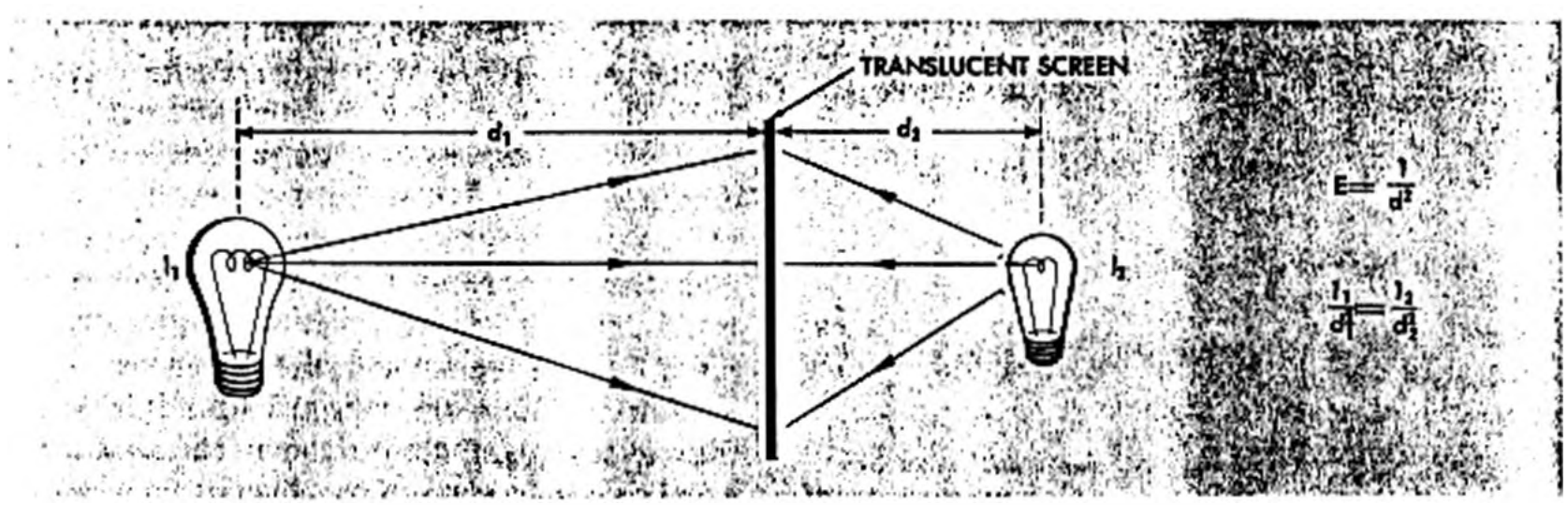


Figure 4-23. Practical Application of Inverse Square Law

frequency is not too high, the flicker resulting from illumination difference remains. The photometer screen is moved until this flicker also disappears. The lamps can then be compared in the usual way.

Photoelectric cells with suitable light filters can also be used to compare the candlepower of lamps or other sources of light of different colors.

Photoelectric current is proportional to the

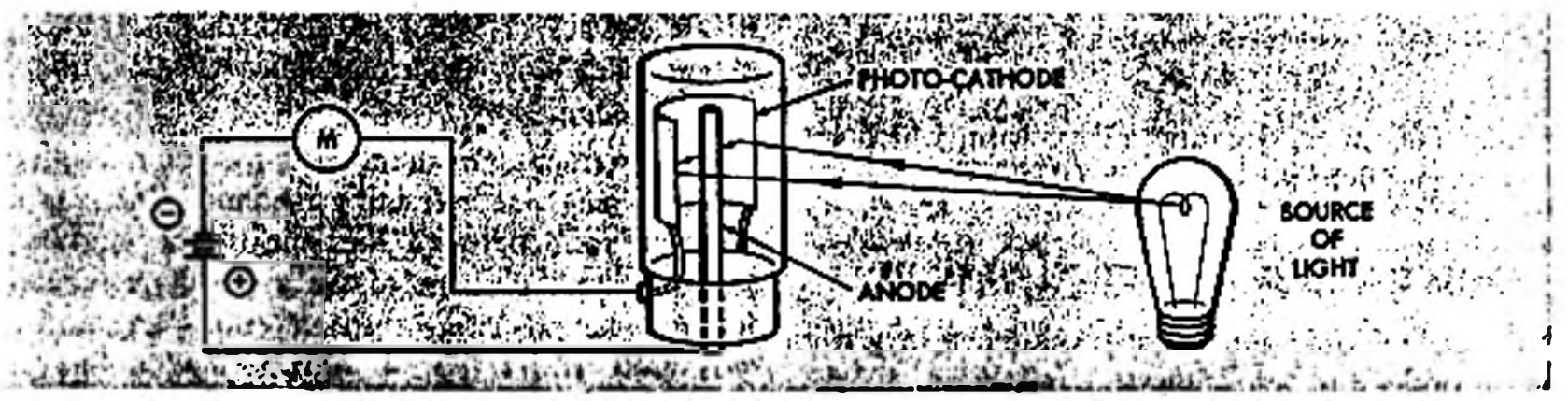


Figure 4-24. Photoemissive Cell used to Measure Light Intensity

illumination on the photo-cathode and the candlepower of the lamp under test can be expressed in terms of the current it produces as compared with that produced by a calibrated standard lamp. Figure 4-24 illustrates how a photoemissive cell may be used in this application.

Light from the source (lamp, etc.) falls upon the cathode of the cells, which is coated with some light-sensitive material. This material releases or emits electrons when struck by the incident light rays. These electrons are attracted to an anode which has been placed at a positive potential by means of a suitable battery. A meter connected in series with the cathode, anode, and battery registers the increase in current produced by the incident light. Filters may be placed between the source and the cathode so as to permit only light of a desired wavelength (color) to reach the cathode.

In selecting reference stars for celestial-navigation systems, the primary concern is the evaluation of intensity in one direction only. Photometric devices described in the preceding paragraphs can be adapted to this application.

VELOCITY OF LIGHT

Early experiments to determine the velocity of light gave inaccurate results because of time losses in the operation of the equipment used. But recent observations using improved optical devices and electronic timing controls have greatly reduced the error. The velocity of light is now considered to be approximately 186,300 miles per second. In general calculations, either 300,000,000 meters or 186,000 miles per second is used as the velocity of light and radio waves.

Modern methods for calculating the velocity of light usually employ some modified version of the rotating-mirror method used by the French physicist Foucault in 1850. Foucault's method involved the directing of a narrow beam of light upon a plane mirror rotating at high speed. A fixed mirror located at a considerable distance from the flashing (rotating) mirror received the momentary flash of light and reflected it back to the rotating mirror from which it was again reflected. During the interval of time required for the light beam to travel through the measurable distance between the rotating mirror and the stationary mirror and back, the flashing mirror rotated through a definite angle.

From the angular velocity of the rotating mirror and the distance between the mirrors, the time required for the light to travel from the flashing mirror to the fixed reflector and return was computed, and from these factors the velocity of the light was determined.

Foucault's method was improved by an American physicist, Michelson, who used a rotating octagonal mirror. Light from a source of high intensity was reflected from one face of the rotating mirror to a distant plane mirror. The returning light ray from the fixed mirror would strike

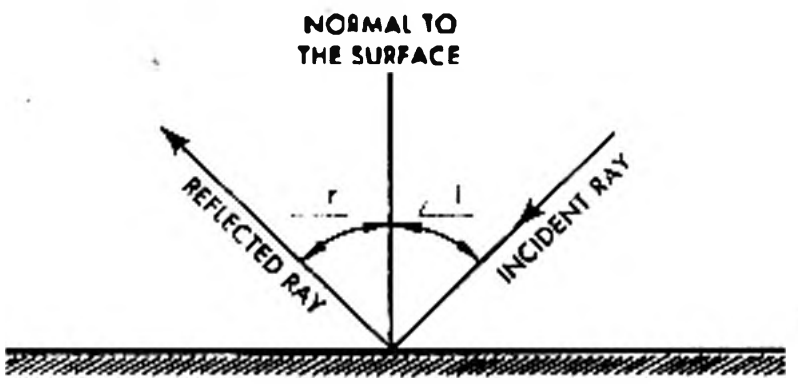


Figure 4-25. Angle of Reflection Equals Angle of Incidence

another face of the rotating mirror from which it was reflected into the observer's telescope.

REFLECTION AND REFRACTION OF LIGHT

Two of the most important phenomena of geometric optics are reflection and refraction. Each of these phenomena is characterized by a basic relationship or natural law and is present in all optical systems such as the human eye, lenses, prisms, telescopes, microscopes, etc.

Law of Regular Reflection

The fundamental law of regular reflection states that when a ray of light is reflected from a surface, the angle of reflection is equal to the angle of incidence. The reflected ray, the incident ray, and the normal (a line perpendicular to the reflecting surface at the point of reflection) all lie in one plane. Figure 4-25 illustrates this law.

The law of reflection applies when light is reflected at the interface (common boundary) between two unlike media such as air and a solid surface. The one medium (air) in this case is transparent and the other (solid) is opaque.

Refraction of Rays

If the second medium is not opaque, some of the light passes on through it and is refracted or bent in a direction which generally is different from the direction of the incident rays. The refracted rays still lie in the same plane as the normal to the surface. The degree to which this bending action takes place determines the *refractive index* of the medium.

Each refracting medium has a specific refractive index in any one refracting medium. There is a constant ratio between the sine of the angle of incidence and the sine of the angle of refraction in any one refracting medium. Both angles are measured with respect to the normal, a line in this case, perpendicular to the reflecting surface at the point of incidence and refraction.

According to Snell's law, a ray is bent toward the normal if the second medium has a greater refractive index than the first and is bent away from the normal if the second medium has a smaller refractive index.

The angle of refraction is smaller for some mediums than others. For example, the angle of refraction is smaller for glass than for water because the refractive index of glass is greater than that of water and thus tends to bend the refracted ray nearer to the normal.

Figure 4-26 illustrates examples of refraction. They show how a light ray is bent in passing from

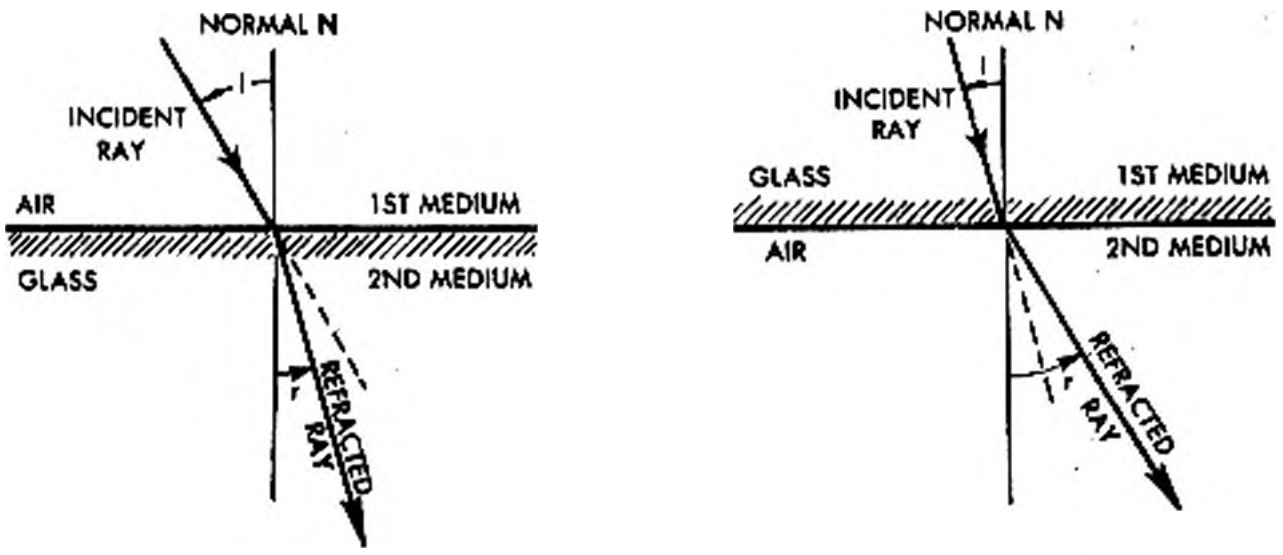


Figure 4-26. Refraction of Light Rays

one medium to another. The figure on the left illustrates the refraction of a light ray in passing from air to glass, and the right-hand figure illustrates the refraction of the ray in passing from glass to air.

You can conclude from this explanation that a ray will deviate toward the normal when its velocity is decreased, and it will deviate away from the normal when its velocity is increased. When a ray passes from one medium into another which tends to reduce its velocity to a greater degree than it was reduced in the first medium, it deviates toward the normal. The ray deviates away from the normal if it passes from a medium in which it met more opposition (velocity slowed) into another medium in which it encounters less opposition.

As noted above, the ratio of light velocities in two mediums which are in contact is a constant for those two mediums. As noted above, this ratio is referred to as the *refractive index* of the second medium with respect to the first. The refractive index for two mediums is represented by the symbol $\mu_{1,2}$ (mu sub 1, sub 2), with the order of the subscripts indicating the direction of the ray of light.

The *law of refraction* states that "when a wave travels obliquely from one medium into another, the ratio of the sine of the angle of incidence to the sine of the angle of refraction is the same as the ratio of the respective wave velocities in these mediums and is a constant for two specific mediums." Thus

$$\frac{\sin i}{\sin r} = \frac{V_1}{V_2} = \mu_{1,2}$$

The *absolute refractive index* of a medium is its index compared to the refraction of light in a vacuum, which is considered as unity (1.000). The refractive index of air is so small that for practical purposes it is used as a standard.

Another important phenomenon of refraction which you will encounter in the study of celestial-navigation systems is that when a ray of light passes through one or more parallel-sided media and emerges into the original medium, it is displaced laterally but its direction is unchanged.

In figure 4-27, i represents the angle of incidence, r the angle of refraction, and e the angle of emergence of the ray of light passing through the parallel-sided medium, glass.

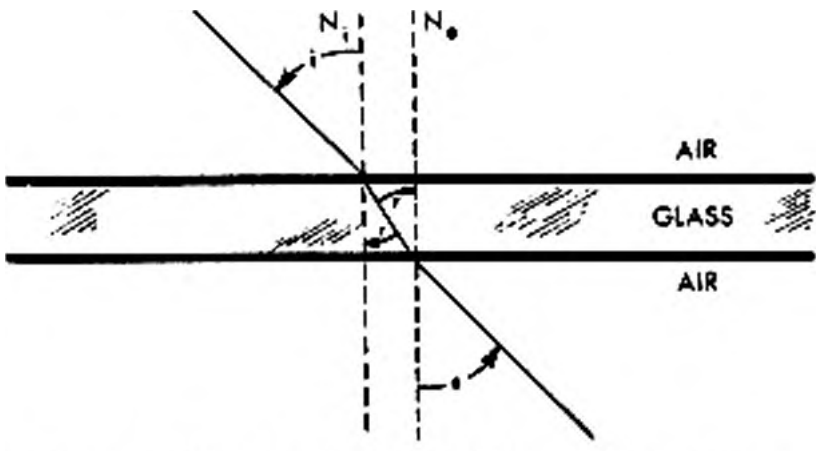


Figure 4-27. Passage of Light from Air to Glass to Air

A ray of light passing from air through a parallel-sided pane of glass and emerging into air follows the law of refraction for each surface of the glass. Thus, angles i and e are equal, and the incident ray is parallel to the emergent ray. Also,

$$M_{ag} = \frac{1}{\mu_{ga}}$$

which shows that the refractive index of glass, g, with respect to air, a, is the reciprocal of the refractive index of air with respect to glass.

These same principles hold true for a ray of light passing through more than one parallel-sided medium, as illustrated in figure 4-28. The figure shows how light is refracted in passing through several parallel-sided media. The refractive index of each medium is noted in the figure. Notice in

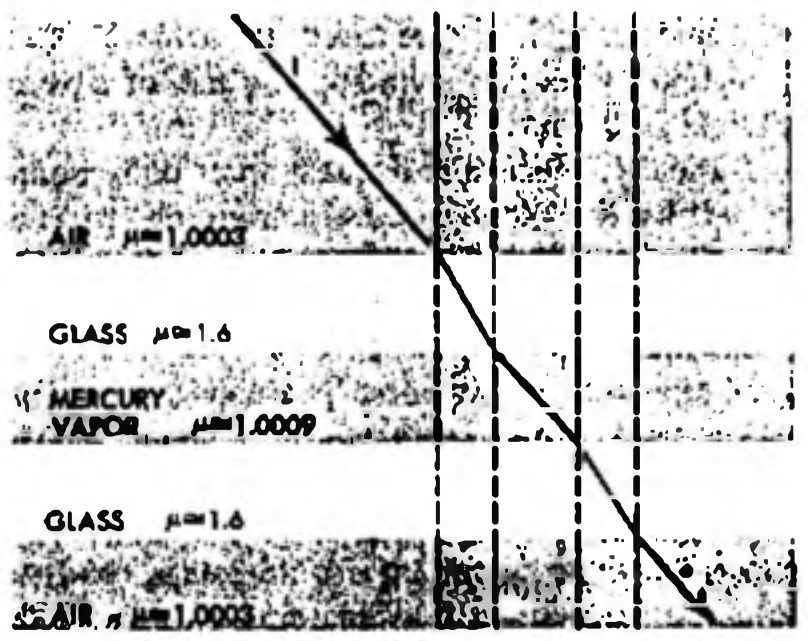


Figure 4-28. Refraction of Light in Passing Through Several Parallel-Sided Media

the illustration that the angle of refraction in the glass is less than in the air or gas.

A ray of light, passing through a medium with a high refractive index toward one with a lower refractive index, is refracted if the angle of incidence is not too large when it passes into the second medium. If the ray is inclined to an ever-increasing angle of incidence, it arrives at some position at which it no longer passes into the second medium. Instead, it is totally reflected at the common surface of the two mediums. Such a condition is known as the *critical angle of incidence*. The critical angle of incidence is the maximum angle at which the ray of light striking the surface of a medium passes through it.

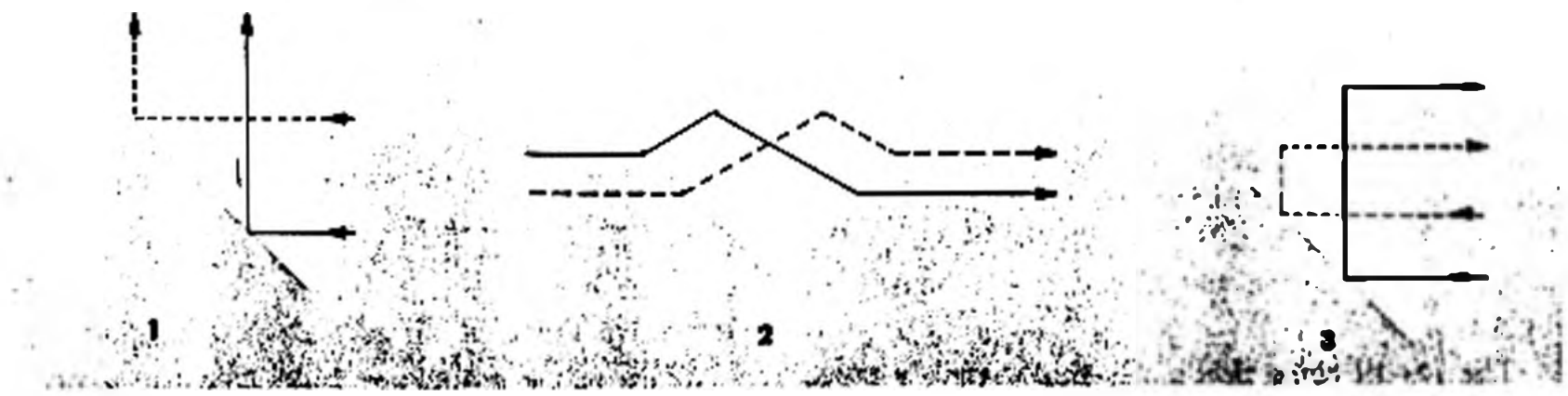


Figure 4-29. Total Reflecting Prism in Various Positions

When this critical angle of incidence is exceeded, the refracted ray grazes the surface with an angle of refraction of 90° in the medium, and the ray is totally reflected from the surface of the medium. This principle is often applied in optical instruments. Prisms are used to achieve "total reflection" whenever avoiding the use of silvered mirrors is desired.

Basically, a prism is a transparent body bound in part by two plane faces which are not parallel. The line in which these faces meet is called the edge of the prism. Prisms are taken up in more detail later.

Figure 4-29 shows three positions of a prism, which has 45° angles, in the path of light rays. The first position shows the prism functioning like a plane mirror tilted downward, the second position shows how the prism inverts an image, and the third position shows how the prism displaces the rays and reverses their direction.

It is possible to compensate for the displacement of a ray of light passing through one substance. The displacement is compensated for by causing the ray to emerge from the displacing substance into another material which has a refractive index that is the reciprocal of that of the first substance. This procedure tends to cause the final angle of emergence to equal the original angle of incidence. Such compensation may be necessary in cases where the desired light must pass through some secondary medium, such as a pressurized chamber, before it reaches the receiving device. To present the light to the receiving device in its original angle, the displacement produced by the secondary medium must be nullified. Problems similar to this example exist in the case of optical instruments used for celestial observations. The instruments must be housed within some protective container, such as the hull of a missile. The original light ray passes through a heavy glass pane and then perhaps into a zone of a different refractive index than the external atmosphere.

Variations in atmospheric conditions produce variations in the refraction of light from celestial bodies; therefore, in selecting a star as a fix for a celestial navigation system, choose one whose light has an angle of incidence as nearly equal as possible to the normal for the earth's atmospheric layer.

Light rays passing through successive layers of air of varying densities or through translucent media with varying indexes of refraction are bent or directed along a new path of radiation by each layer. When this process occurs at a visible rate, it produces an effect called *scintillation*, more commonly called glitter or twinkle. The twinkling

of starlight is produced in this manner. A similar effect can be detected with respect to electromagnetic waves passing through zones of conductance which vary in refractive characteristics for each specific wavelength of the radiations.

Light reflected from surfaces with many facets, such as cut diamonds, tends to produce the effect of flashing or scintillating. Electromagnetic waves, reflected from various objects and arriving at the receiver at intervals slightly later than the direct beam, tend to produce phase differences and frequency in varying degrees. This results in phase cancellation and interference, referred to as "fading," "swinging signal," or "ghosts."

In television, a "ghost" image on the picture screen occurs when the transmitted signal arrives at the receiver along more than one path, the paths being unequal in distance. This condition produces two or more images slightly displaced in phase and time.

Radar signals are subject to the same effect. The effect is present when the signals are reflected from moving surfaces or from more than one object, or when they are refracted by inequalities in terrain or atmosphere. To minimize this effect of fluctuation or scintillation, the radar beam is narrowed so that both the incident and reflected beams travel through a narrow zone and are less subjected to reflection from objects other than those directly along the principal axis of the beam. When scintillation is produced by a fixed or constant succession of media, the frequency of the scintillation may be used to identify the path or locate the source of the radiation.

Functions of Prisms

A simple optical device which may be used to compensate for light displacement or to produce some specific deviation in a beam of light is the prism, which was described briefly in the preceding section.

The amount of deviation produced by a triangular prism depends upon the angle of the prism, the refractive index of the material of the prism, and the angle of incidence of the light passing through the prism.

By controlling the position of a common triangular prism with respect to the source of light and by choosing a prism of proper angle, the desired deviation of the light beam can be produced even to the extent of producing total reflection.

Some light is absorbed or reflected by the transparent medium; consequently, the intensity of the light which emerges is less than that of the incident light even though the angle of incidence is 90°. When a light ray passes through air and strikes the surface of a glass window at a 90° angle (perpendicularly), approximately 4 percent of the incident light is reflected leaving 96 percent to pass into the glass. The degree to which light is absorbed in passing through a transparent medium depends upon the nature of the substance, its index of refraction, and the angle of incidence.

Spectra

The "rainbow" of color produced by sunlight passing through a crack and diffracting medium into a darkened room is a visible color spectrum. This phenomenon is the result of refraction and dispersion of light rays, and it exemplifies the basic principle of spectrography.

White light is composed of light rays of many hues blended together. When such light is passed through a narrow opening or slit into some diffracting medium, rays of each different color are diffracted at different angles and spread out into merging bands of six principal colors. These colors are red, orange, yellow, green, blue, and violet. The diffracting of radiations of light at different angles produces a "rainbow" or spectrum of light. The prism shown in figure 4-30 is acting as a diffracting medium, thus producing a spectrum of light.

A spectrum is composed of hundreds of hues which are grouped broadly into the six principal colors mentioned above. These colors appear in the order of increasing deviation from red to violet.

The color of light is determined by its frequency of vibration, the frequency being lowest for red and greatest for violet. The wavelength, λ (lambda), is longest for infrared light and shortest for ultra-violet light, and each color possesses certain characteristics and properties which may be used in light-sensitive or optical systems for missile guidance.

The deviation produced by a prism is greater for the higher-frequency components of white light than for the lower-frequency hues, but no

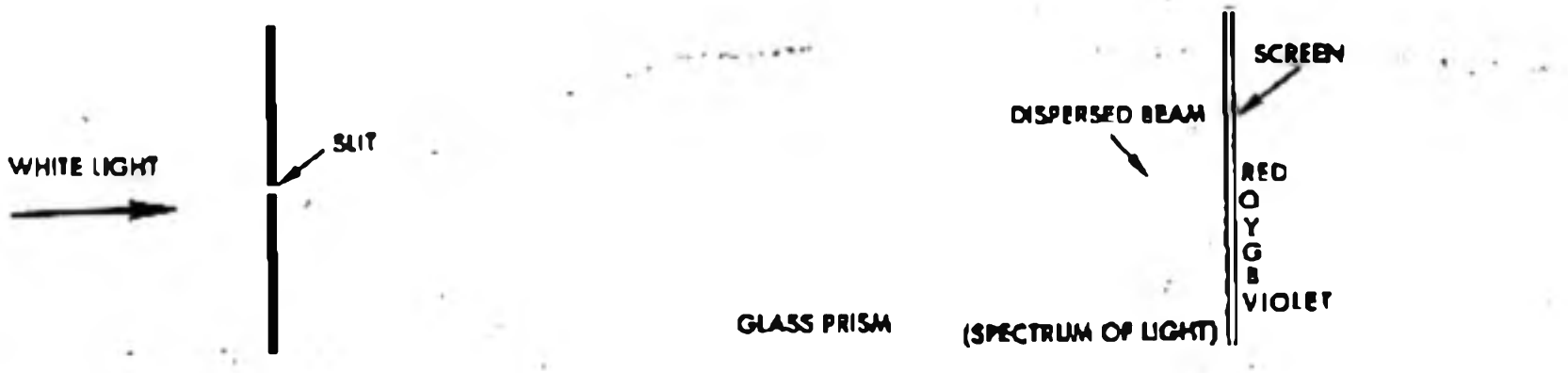


Figure 4-30. Dispersion of Light Through a Prism

definite relationship exists between frequency and deviation. Prisms of different materials spread the component colors of the spectrum in varying degrees.

Some materials produce anomalous dispersion; that is, some prisms are composed of materials that do not disperse white light in the regular sequence of colors. Such prisms deviate certain colors to a much greater degree than others and also absorb certain portions of the spectrum.

In a quantitative study of the spectrum, it becomes necessary to refer to each particular part of it with definiteness. This is done by specifying any hue by the vibration rate of the light source and its corresponding wavelength. As in all other forms of wave motion, the velocity of light is equal to the wavelength times the frequency,

$$V = f\lambda,$$

where f equals the frequency and λ the wavelength. When light is retarded by a medium such as glass, the frequency is unaltered. Since velocity decreases, λ must also decrease in direct proportion. Therefore, λ is not a constant quantity for a given vibration, but it depends upon the medium.

In general, values of λ for different colors of a spectrum are given as the wavelength in air or in a vacuum. Wavelengths are so short for visible light that a special unit of length shorter than the centimeter is commonly employed to measure them. This unit is known as the *Angstrom unit*, named in honor of a Swedish physicist. The Angstrom unit, A, is equal to one one-hundred millionth of a centimeter (10^{-8}cm).

In some cases of lower-frequency radiations, the micron may be used as a measuring unit. One micron is equivalent to 10,000 A.

Spectra are often classified into three general types: emission, absorption, and solar spectra.

EMISSION SPECTRUM. A spectrum produced by a glowing object is termed an emission spectrum. Its appearance depends primarily upon the composition and state of the luminous object.

Incandescent solids and liquids produce continuous spectra, extending from color to color without interruption. Luminous gases and vapors yield spectra consisting of definitely placed bright lines. Each bright line is an image of the slit through which the radiation is received.

Every gas emits radiation of particular wavelengths, and each spectrum is characteristic of the radiating substance. For example, sodium vapor yields two bright lines in the yellow part of the spectrum, while mercury vapor yields several bright lines, the most conspicuous being in the green and blue regions. Figure 4-31 shows the bright-line spectra of several elements over the range of the visible spectrum. The continuous spectrum at the top is an uninterrupted series of images of the illuminated slit.

The number of lines in a bright-line spectrum depends upon the amount of energy with which the atoms of the source are excited to produce glowing, as well as depending upon the nature of the source. The greater the excitation of the atoms of a substance, the greater the number of lines that appear in the spectrum of the substance.

ABSORPTION SPECTRUM. Absorption spectra occur when white light has passed through an absorbing medium before the light is dispersed. A glowing solid, or other source from which the light radiates, yields a continuous spectrum, but when the light passes through the absorbing medium, radiations of particular wavelengths are absorbed. The resulting spectrum is usually crossed

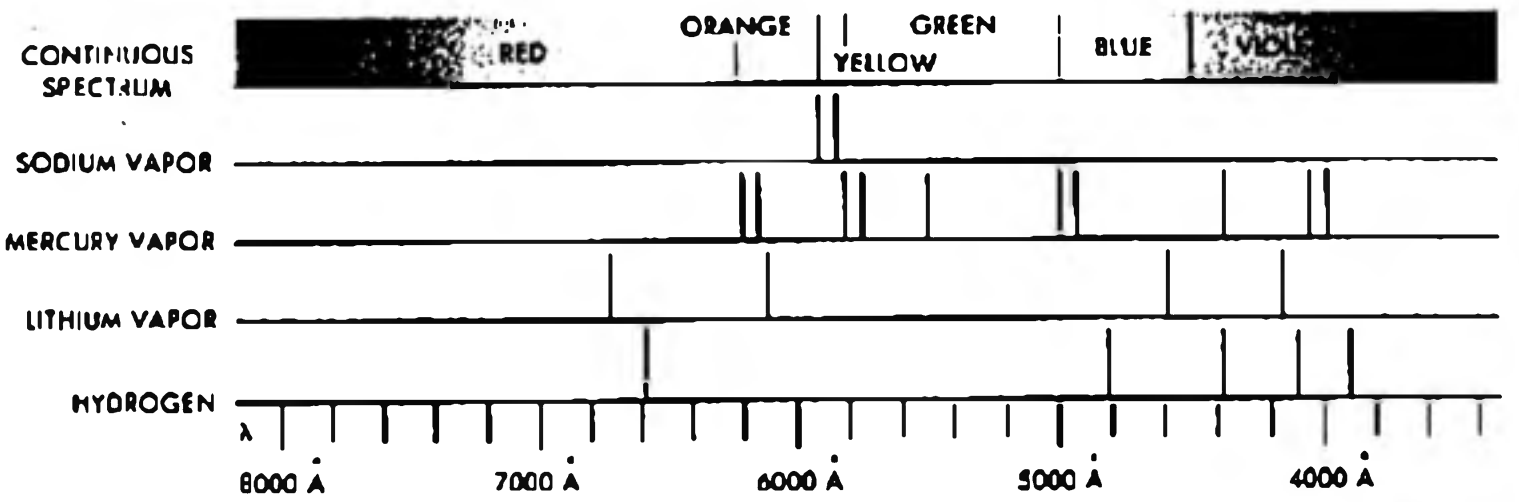


Figure 4-31. Continuous and Bright Line Spectra

by dark spaces because of the absence of the absorbed radiations. If the absorbing material is solid or liquid, these dark spaces appear as broad, structureless bands. If the material is gaseous, the dark spaces consist of dark lines which occupy the same positions as the bright lines in the corresponding bright-line spectrum.

SOLAR SPECTRUM. A spectrum formed by radiations from the sun is called the solar spectrum. This spectrum appears continuous from a casual inspection. A more critical examination shows that it is crossed by numerous dark lines. No doubt the sun's radiation comprises all wavelengths in the visible range, but in passing through the sun's "atmosphere," certain wavelengths are absorbed. The spectrum observed is in reality an absorption spectrum of the sun's atmosphere.

Continuous spectra are produced by light from incandescent solids and liquids. Bright-line spectra come from incandescent vapors or gases. Absorption spectra are produced by light passing from an incandescent solid or liquid through an incandescent vapor or gas.

POLARIZATION OF LIGHT

When the question arises as to whether light waves are longitudinal like sound waves or transverse like elastic waves, it is advisable to consider the phenomenon of polarization of light.

If a beam of light is passed through a substance composed of two colors, such as a crystal of tourmaline or a sheet of "Polaroid," the beam's passage is restricted to a particular plane of vibration. The beam is said to be *plane-polarized* and will pass through a second crystal of tourmaline or sheet of "Polaroid" only if the tourmaline or sheet is oriented exactly the same as the first. If the second crystal or sheet is rotated 90° with respect to the first, no light passes through.

The first crystal or sheet is called the *polarizer*, and the second the *analyzer*. The analyzer is so named because the angle of polarization may be determined from the angle through which the analyzer must be rotated between the points of light passage and cutoff. Figure 4-32 illustrates the principle of polarization of a transverse wave.

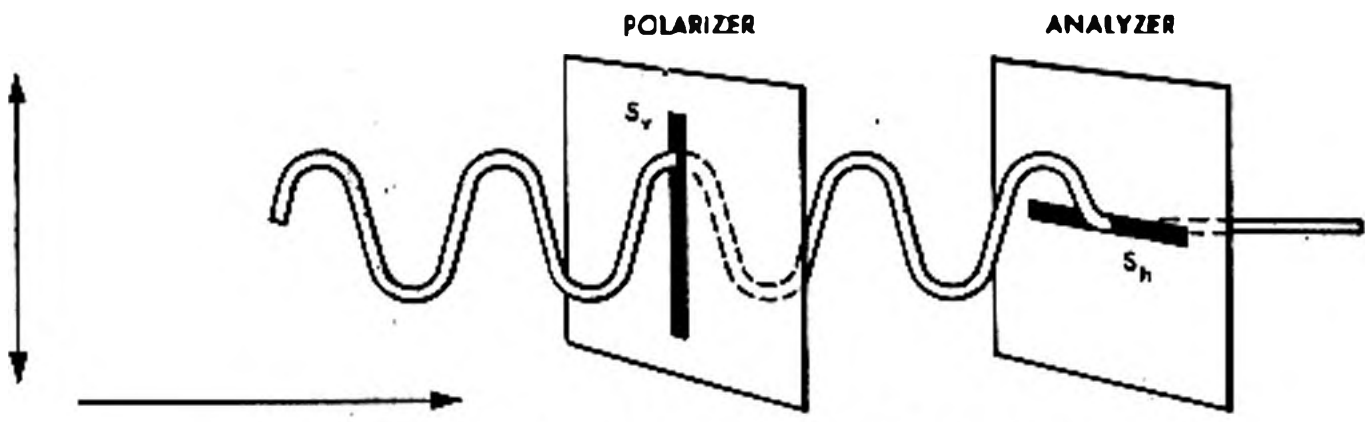


Figure 4-32. Analogy of Polarization of Transverse Wave

The vibrating rope corresponds to a beam of light in this analogy.

Vertical vibrations of the rope pass through a vertical slot (S_v), but they are stopped by the horizontal slot (S_h). The wave represented by the rope is polarized. It is vibrating in one plane only (in this example the vibrations are in the vertical plane). Light can be similarly polarized by certain optical substances such as tourmaline crystals and "Polaroid," which is the trade-name for a commercial product possessing this property.

If slot S_h were rotated 90° in the foregoing analogy, the waves would pass through it. S_v limits the wave motion to the vertical plane and is the polarizer. S_h is the analyzer. Since S_h must be rotated 90° to permit the waves to pass, it is evident that the degree of polarization produced by S_v is 90°. The degree could be determined by the required distance of rotation of S_h between points of passage and cutoff, even though S_v were not visible.

Both theory and experiment have shown that longitudinal waves cannot be polarized. We may conclude that light must have a transverse wave motion if it has the form of wave motion at all.

Light can be polarized to a considerable extent by reflection alone. All reflected light is polarized to a certain degree because of scatterings caused by dust and vapor particles. Scattering is more apparent in light of short wavelengths. This statement is supported by the color changes of the sky. The midday sky appears blue, but a sunset appears reddish. The sunset appears reddish because the light travels through a longer path of earth's atmosphere, resulting in the short wavelength blue light being dissipated by scattering. This scattering leaves a predominance of the longer wavelength reddish hues.

FLUORESCENCE AND PHOSPHORESCENCE

Some natural substances possess the property of emitting light when excited by an external force, such as "bombardment" by electrons of certain forms of radiant energy. In some cases, light is emitted by the substance only while the bombardment is taking place. In other cases the emission may persist for some interval of time after the external excitation has ceased. These properties are termed *fluorescence and phosphorescence*, respectively. The duration of light emission is referred to as the *persistence* of the material. Phosphorescence is commonly referred to as *afterglow*.

One of the fluorescent substances used in electronic application is willemite (zinc orthosilicate Zn_2SiO_4), a crystalline zinc compound varying in color from white, greenish-yellow, and green to shades of red and brown. The white and green varieties are frequently used in the coatings on the screens of cathode-ray tubes used in oscilloscopes and radar indicators. Willemite is used in television kinescopes in combination with other substances which impart the desired degree of persistence.

Generally when the emission is induced by some form of radiation, the luminescent substance emits light which is of a longer wavelength than the incident radiation. This phenomenon is encountered in the flourescent lamp in which certain phosphorus compounds are excited by ultraviolet radiation which is invisible. The compounds then emit visible light of various colors. The fluorescent compounds used in the "soft white" and "daylight" lamps are combinations of zinc beryllium silicate and magnesium tungstate.

Infrared radiations reflected from objects onto a photo-sensitive surface which is sensitive to infrared light may be used for seeing in the dark. A device called the snooperscope was developed during World War II for this purpose.

Essentially, the snooperscope consists of a source of infrared light and an image tube having a light-sensitive cathode. An infrared lamp projects a beam from which all visible light is filtered out. The reflected rays from the beam are caught upon a cesium cell, which is highly sensitive to infrared light and which forms the cathode of the image tube. The reflected radiations incident upon the cell cause it to emit photoelectrons which are then focused on a fluorescent screen to form a visible image, as in the cathode-ray oscilloscope.

You will find in later chapters that a knowledge of the emission spectrum of any radiating surface is of value in understanding a homing system to guide a missile toward a target presenting such characteristics of emission. This knowledge also is of value in understanding navigation systems employing light from certain fixed stars as a reference.

Such systems must be highly sensitive and selective and thus require the use of optical com-

ponents of the highest precision. Marked progress has been made in improving basic optical instruments and in adapting them to new applications and in developing electronic systems which employ the basic principles of optics. This progress has created systems which are extremely sensitive and selective with respect to heat, light, and other forms of radiant energy.

Many of the present-day missile guidance and navigation systems are of these improved types, and further research and development undoubtedly will produce revolutionary improvements in their accuracy and reliability as well as in their adaptation to industrial and commercial usage.

SECTION B
How Optical and Electronic Principles Work Together

A brief explanation of some of the basic optical and electronic components is given here to help you understand the guidance and navigation systems of missiles. In considering these basic components, you will gain a general idea of how light waves are used along with electromagnetic waves in electronic equipment. You will learn about some of the properties common to both types of waves. Devices using radiant energy (light) are taken up first.

SIMPLE OPTICAL DEVICES

To use the various phenomena of radiant energy, simple optical devices such as mirrors, prisms, and lenses are employed either singly or in combinations to suit the complexity of the requirements.

Functions of Mirrors

A mirror is a polished or smooth surface that forms images by means of reflected light. When an object is placed before a plane mirror, a right-side-up image is formed. The image appears to be just as far behind the surface of the mirror as the object is in front of it. You encounter this simple phenomenon whenever you approach a mirror.

Your image appears to move toward you from a point to the rear of and equally distant from the mirror surface. This phenomenon conforms to the laws of reflection and is an important one in regard to optics.

Figure 4-33 shows a plane mirror with an image that is direct and vertical and appears to be as far behind the mirror as the object is in front of it.

Curved reflecting surfaces are designed to give variations in the apparent position and dimensions of the image by deviating a beam of light, causing it to be more or less converging when reflected than it was when incident upon the mirror.

Spherical mirrors are classified as concave or convex. A concave mirror has its reflecting surface on the inside of the spherical shell while the convex mirror has its reflecting surface on the outer side of the shell.

A concave mirror is a converging reflector because it actually converges the light rays. A convex mirror is a diverging reflector because it makes the light rays appear to diverge.

The center of the spherical surface is called the center of curvature of the mirror. A line from the middle point of the mirror surface to the center of curvature is called the principal axis of the mirror. Figure 4-34 illustrates the two types

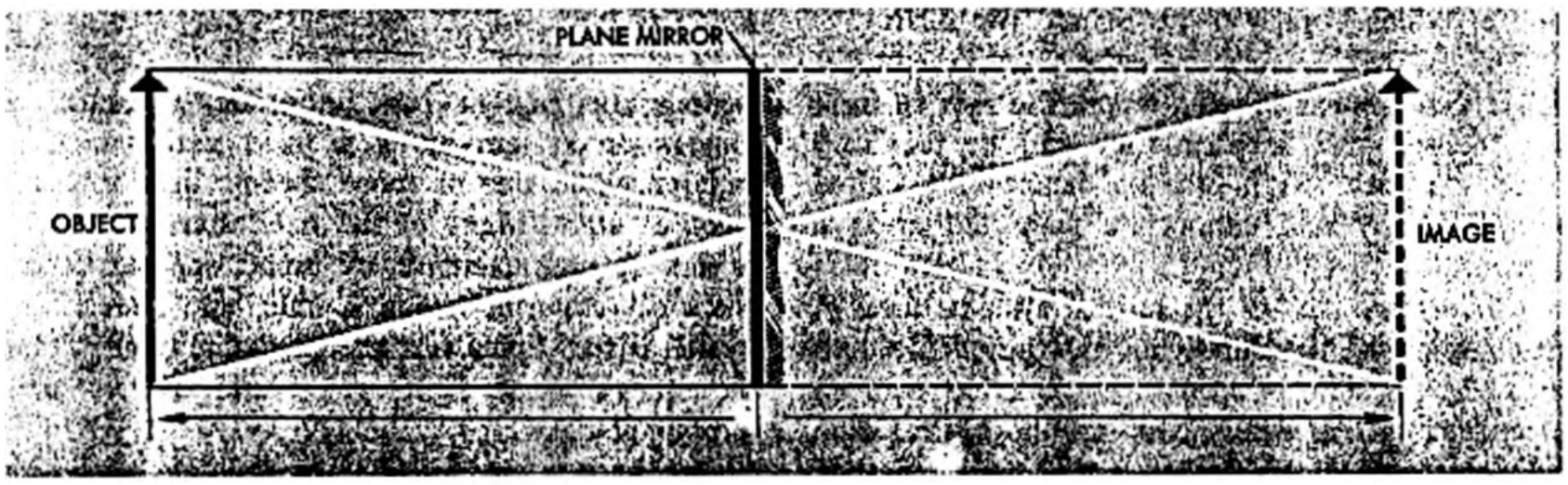


Figure 4-33. Object Distance Equals Image Distance

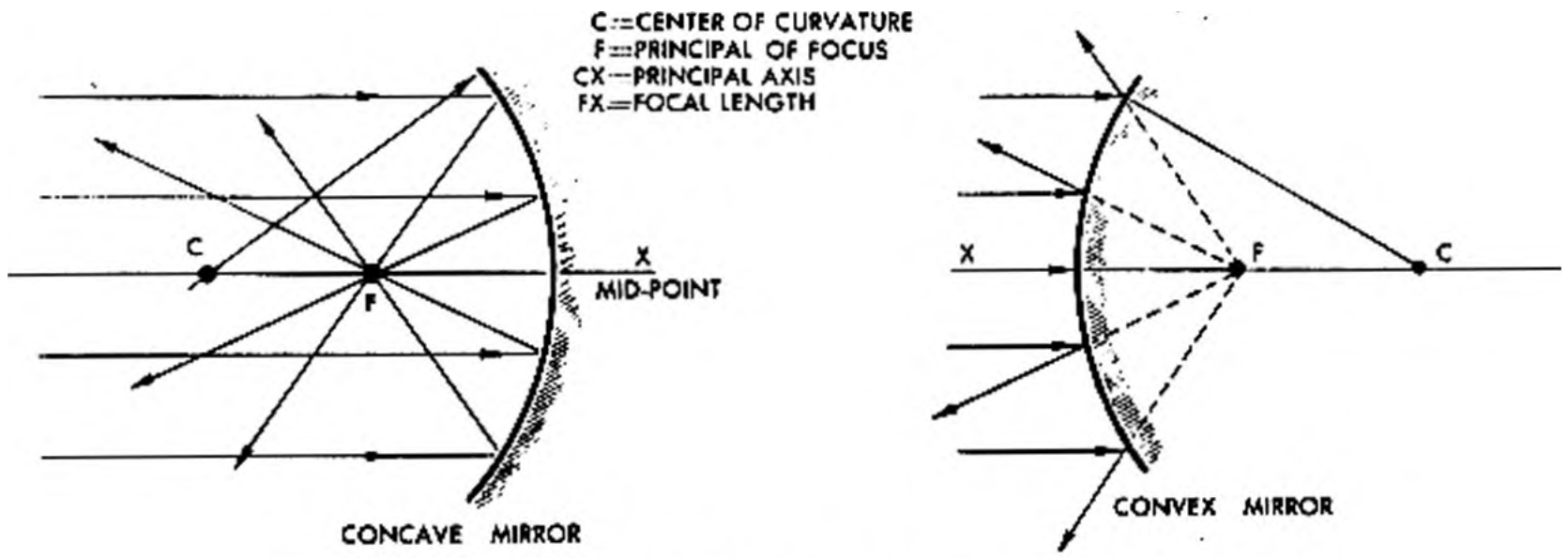


Figure 4-34. Reflection of Incident Light by Spherical Mirror

of spherical mirrors and how they reflect incident light.

Point "F" is the principal focus of the mirror. The distance of the principal focus from the mirror is the focal length. The principal focus of a spherical mirror is located on the principal axis half-way between the center of curvature and the mirror surface.

The ratio of the dimensions of a mirror with respect to the radius of curvature is referred to as the aperture of the mirror. Most optical mirrors are of small aperture and produce only a slight inclination of the incident rays with respect to the principal axis.

The ratio of the size of the image to the size of the object is referred to as the magnification of a mirror. A spherical convex mirror always causes the image to appear reduced in size but right side up. A spherical concave mirror can be made to produce inverted images which appear to stand out in space or to produce right-side-up images which appear to be behind the mirror.

The inverted images are called *real images* because they appear to exist where the rays of light are focused and because they can be localized upon a screen. Right-side-up images are called *virtual images* and cannot be projected upon a screen.

Mirrors are part of the optical systems used in navigation and in surveying instruments. Often an optical system is used in conjunction with an electronic system which converts the optical data into electrical data. The electrical data can then be transmitted over great distances. Television is an example of such a combined optical and electronic system.

Uses of Prisms

Prisms—covered previously in the section on light dispersion, reflection, and refraction—are used in binoculars, spectroscopes, refractometers, and many other optical devices which use the phenomena of light dispersion and refraction.

The characteristics of light and the chemical composition of light sources may be determined by methods which involve the use of prisms.

In celestial navigation systems, the light from the reference star is passed through a prism in such a manner that either the predominant color or any desired color present in the dispersed beam can be used to activate a suitable photocathode or light-sensitive cell. The photocathode or cell produces and maintains a voltage or current-output level which is proportional to the frequency (color) and/or intensity of the light which falls upon it.

Now you can visualize how such a constant voltage or current can be used to control a navigation system along a fixed path which is referenced to one or more specific light sources, such as stars. Many of the fixed stars emit light which is characterized by some specific color, such as Arcturus, which is blue, and Aldebaran, which is red. Spica is a spectroscopic binary; that is, Spica's white light is a blend of two predominant colors.

A star can be identified by its light spectrum once the spectrum has been tabulated. Spectro-

scopic equipment involves the use of prisms for the dispersion of the reflected light from the planet or the emitted light from the star.

Composition and Uses of Lenses

Usually, in addition to mirrors and/or prisms, lenses in some form are found in both optical and electronic-optical systems.

An optical lens is basically a piece of glass or other transparent material which has two opposite regular surfaces. Both surfaces may be curved, or one surface may be curved and the other plane.

Lenses are used singly or in combination with other lenses, prisms, or mirrors to perform specific functions. The primary function of a lens is to form an image by changing the direction of the rays of light. Such optical instruments as eyeglasses, cameras, microscopes, and telescopes are basically systems of lenses.

You will find that the curved surfaces of lenses generally are spherical, although in some rare instances cylindrical lenses may be encountered.

Spherical lenses may be classified broadly into six types as shown in figure 4-35.

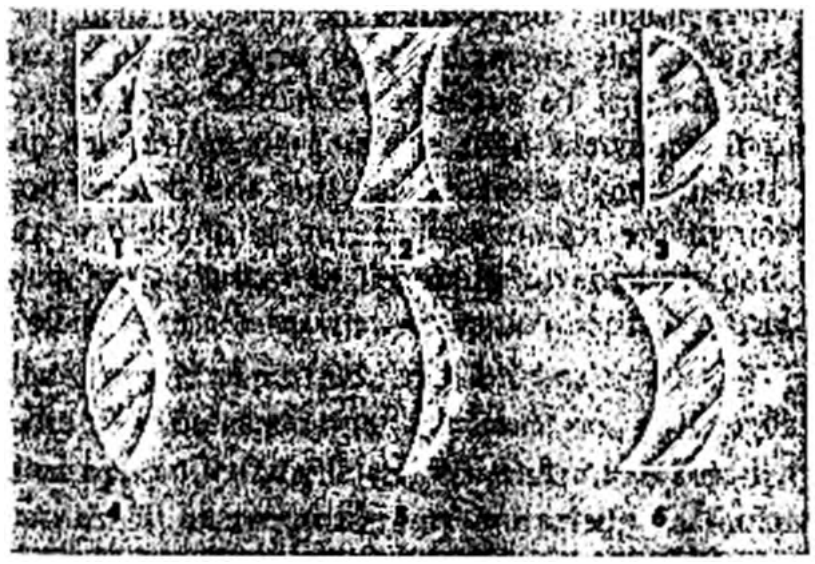


Figure 4-35. Six Types of Spherical Lenses

1. Plano-concave: one plane surface and one opposite concavely curved surface.
2. Double-concave (biconcave): two opposite concavely curved surfaces.
3. Plano-convex: one plane surface and one opposite convexly curved surface.
4. Double-convex (biconvex): two opposite convexly curved surfaces.
5. Converging concave-convex (converging meniscus).
6. Diverging concave-convex (diverging meniscus).

In all of these spherical lenses, the line joining the centers of curvature of the two surfaces is a line of symmetry of the lens and is called the *axis of the lens*.

A lens whose focus for parallel rays is real is called a converging lens, as in the case of mirrors. And a lens which has a virtual focus for such rays is called a diverging lens.

Many of the principles of lenses which apply to light also hold true for other forms of radiation such as electromagnetic waves, cathode rays, etc.

The first outstanding scientific application of the principles of optics was in the field of astronomy. A telescope developed by the Italian physicist, Galileo Galilei, at the beginning of the 17th century is generally regarded as the forerunner of optical instruments; however, it is probable that eyeglasses were used prior to Galileo's time and most certainly some of the properties of simple lenses and prisms had been known for centuries.

The accuracy of optical/electronic systems depends on the quality of the lenses and other optical components and the precision of the mechanical construction and calibration.

ABERRATION OF RAYS PASSING THROUGH LENSES. The glass from which lenses are made must be free from impurities and must be homogeneous in chemical structure. The radii of the spherical surfaces must be properly selected. These factors are important in minimizing spherical aberration (deviation of rays from a focal point) to the greatest possible extent. Spherical aberration occurs when rays of light parallel to the principal axis of a lens do not all converge at a common focal point but intersect, instead, along the principal axis at various points. The intersecting of rays produces a blurred image.

A plano-convex lens, used so that the incident light falls upon its curved surface, produces little spherical aberration, as illustrated in figure 4-36.

Chromatic aberration is another problem encountered in the use of lenses. Chromatic aberration occurs when the various colors present in the incident light converge as individual points along the principal axis of the lens. This results

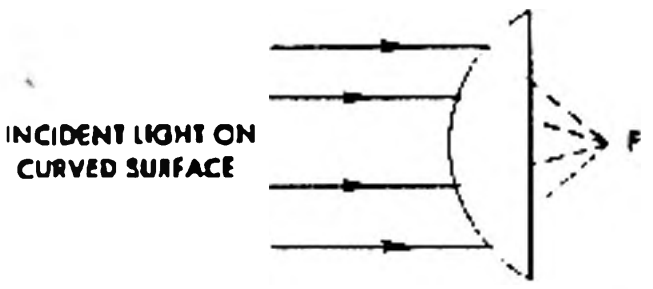


Figure 4-36. Plano-Convex Lens with Little Spherical Aberration

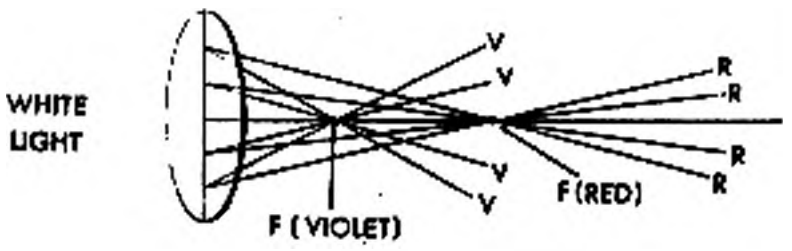


Figure 4-37. Bi-Convex Lens Showing Chromatic Aberration

from the variations in the refractive indexes of the colors. As shown previously in reference to the light spectrum, violet color has a greater refractive index than red; therefore, the focal length of the lens is less for violet light than for red. The violet rays converge at a point closer to the surface of the lens than the point at which the red rays converge. This phenomenon is illustrated in figure 4-37.

In passing through a lens, white light tends to disperse into its component colors which, in turn, tend to focus at different points along the principal axis of the lens. This phenomenon produces chromatic (color) aberration.

To overcome this effect, two or more lenses may be combined so that the divergence produced by one lens is nullified by the convergence of the other. Such a compounded lens is called an *achromatic lens* or doublet and generally consists of two lenses with opposite dispersion characteristics placed in contact with each other. Figure 4-38 illustrates a bi-convex lens (positive), a bi-concave divergent lens (negative), and two lenses combined to form an achromatic lens which is relatively free from chromatic aberration.

Indistinctness of an image produced by a spherical lens may result from *astigmatism* which becomes evident when rays of light pass through the lens obliquely nnd do not converge upon a common image point. This defect can be overcome by using two lenses suitably separated.

The high degree of accuracy required in the navigation of long-range guided missiles makes it obvious that any optical devices employed in missile navigation must be of the highest precision and free from the effects of undesired external influences.

Celestial navigation systems using fixed stars for references must be designed so that they are "sensitive" to only the selected stars and will not react to light rays of different intensity or color which might come within the fields of their telescopes. This means that all prisms used in the system must be accurately oriented with respect to the incident light reaching them through the "star-finding" telescope, so that the desired portion of the refracted spectrum is directed to the photo-cathode or photocell to maintain the standard reference voltage or current for the system.

All lenses so used must be free from chromatic aberration or carefully compensated by other lenses or color filters. They also must be free from spherical aberration and astigmatism since any

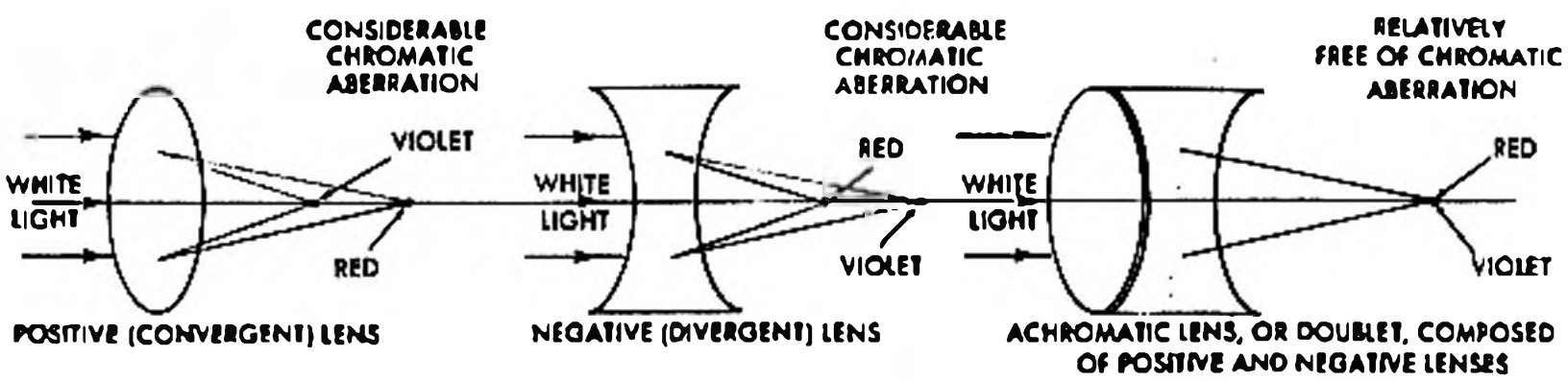


Figure 4-38. Chromatic Aberration Caused by Different Types of Lenses

minute deviation of the incident light beam within the optical system may produce a large "position error" in the navigation system.

RECTILINEAR PROPERTY OF LIGHT. Most astronomical and celestial navigation systems use the rectilinear property of light. Distances and angles are computed on the basis of straight-line measurements to the reference stars; therefore, any factor in the system which would introduce deviation or bending of light rays would cause an error in the straight-line computations. "Star-finding" telescopes are usually mounted so that the reference star, when seen along the principal axis of the lenses in the telescope, is also in a plane normal (perpendicular) to the earth's atmospheric envelope. This positioning is maintained by means of a stabilized-platform mounting for the celestial navigation optical components and by accurate parallel-sided windows in the hull of the missile. The windows serve as a passageway through which the light from the celestial body must pass before reaching the tracking telescopes.

Telescopes

For our purposes, a telescope can be defined as an optical instrument which employs a combination of lenses or a combination of lenses and mirrors (or prisms) to magnify the apparent size of a distant object.

An object at a great distance from the eye will subtend a small visual angle, while the same object viewed at closer range will subtend a larger visual angle. This phenomenon is illustrated in figure 4-39.

Magnifying power of an optical instrument may be defined as the ratio of the size of the image formed on the retina of the eye when the object is viewed with the aid of the instrument to the size of the image formed on the retina without the aid of the instrument.

The relative size of the images can be expressed in terms of the angles subtended by the images on the retina at a point along the optical axis of the eye and at the crystalline lens in the pupil of the eye.

The size of the image on the retina depends upon the distance between the object and the eye. There is a limit to the size of the image that can be formed on the retina. By using a converging lens or magnifier, the image size on the retina is increased because the visual angle is increased as a result of a real image being formed closer to the eye than the object.

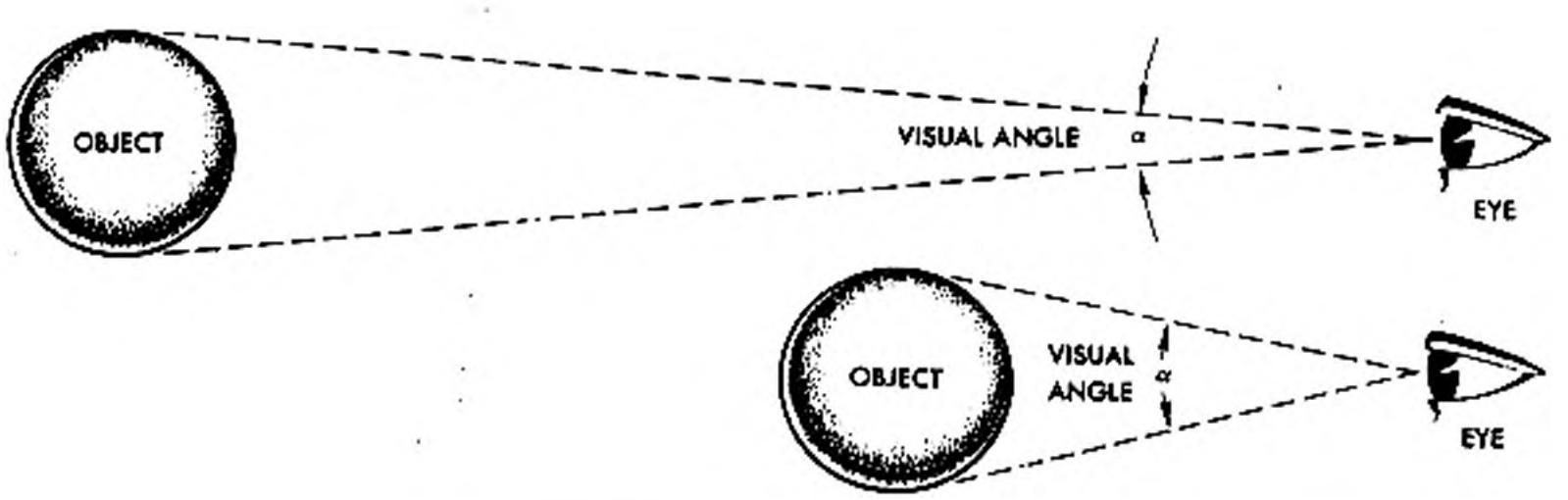


Figure 4-39. Relationship of Distance and Visual Angle

For high magnification, short focal-length eyepieces must be used when objective forward lenses of long focal length are used. The short focal-length eyepieces widen the angle of vision, thus rendering magnification.

Astronomical telescopes aid in viewing and photographing distant objects such as stars and planets by collecting and concentrating a larger beam of light, making the object visible or more distinct. There are many varieties of astronomical telescopes, but generally they are classified in two types: refracting telescopes and reflecting telescopes.

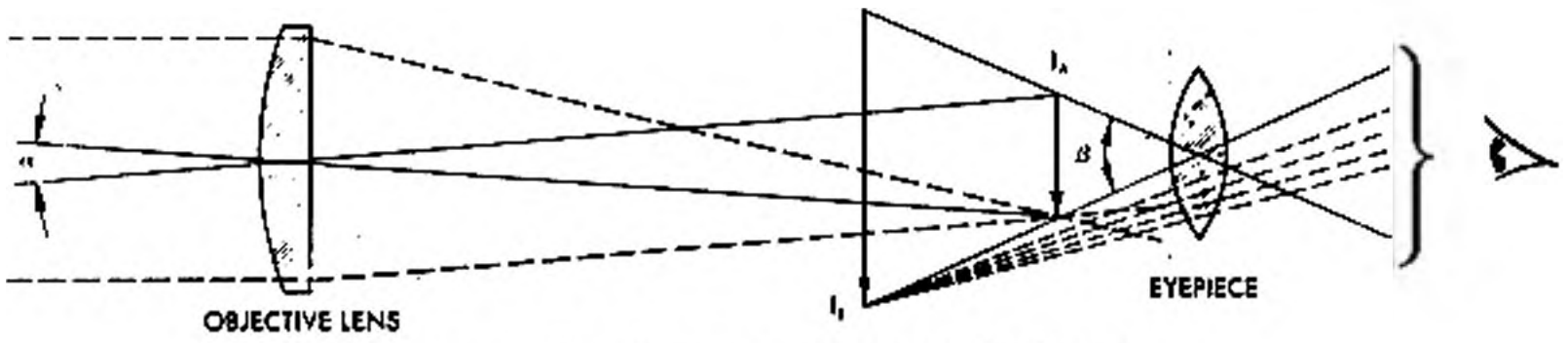


Figure 4-40. Refracting Astronomical Telescope

REFRACTING TELESCOPES. A refracting telescope is the type which employs two lenses or systems. One system includes the objective lens which serves to collect the beam of light and form the image. The image is then magnified by the other lens known as the eyepiece.

In this telescope, the objective lens has a focal length as long as possible (limited by the length of the telescope barrel) and produces a real image of the object. The eyepiece has a short focal length (10 inches or less) and produces a virtual image of the image produced by the objective lens. As mentioned before, the eyepiece magnifies the image by widening the angle of vision.

Optical arrangement of a refracting telescope is illustrated in figure 4-40.

The dotted lines represent rays from the upper end of a distant object. The rays converge at the lower end (point of arrow) of image I_A by the action of the objective lens.

These rays continue on and are deviated by the eyepiece so that they appear to come from the upper end of the inverted image I_B.

The solid lines represent the rays from the extremities of the object and image through the centers of the lenses.

Without the telescope, the distant object would produce an image on the retina subtending the angle of vision, α, but with the telescope, the larger image subtending the angle, β, is produced on the retina.

The objective lens of the refracting telescope is a carefully corrected achromatic lens, and the eyepiece is generally a compound lens.

REFLECTING TELESCOPES. The second type of astronomical telescope is called a reflecting telescope. This type employs a concave mirror for collecting the beam of light and for forming the image. It also has an eyepiece for magnifying the image.

Reflector-type telescopes may be designed with shorter overall length and wider fields of vision than the refractor type of equal magnifying power. Most astronomical telescopes of high power are reflector-type instruments.

Figure 4-41 illustrates a reflecting telescope

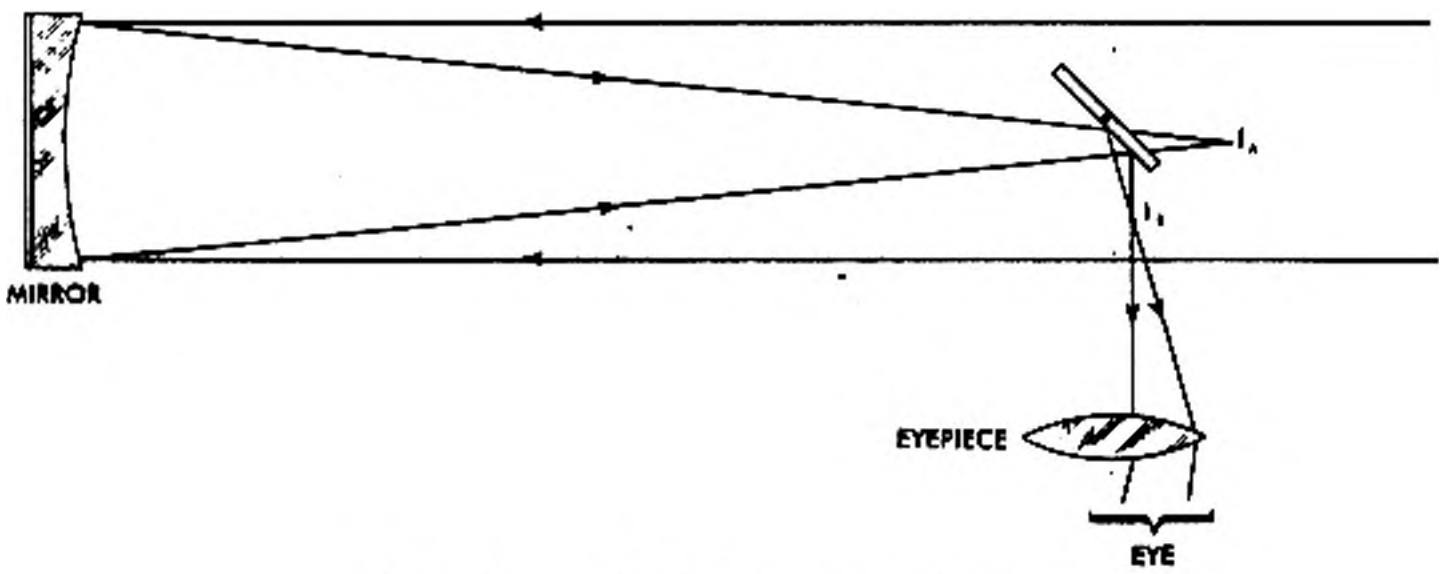


Figure 4-41. Reflecting Astronomical Telescope

designed by Newton. In this telescope, a small plane mirror (or total-reflecting prism) on the axis of the telescope shifts the real image formed by the concave mirror from position I_A to position I_B. From I_B, the image is enlarged by the eyepiece which is set at right angles to the plane of the concave collecting mirror.

The same basic optical principles are involved in both the refracting and reflecting telescopes.

If a telescope is used to make photographic records, the eyepiece is removed and a sensitized plate or film is placed at the point where the image is formed by the objective lens. The optical arrangement, in this case, is the same as that of a camera with a great focal length.

In missile navigation systems employing telescopes as star-trackers, refracting-type telescopes (with extremely accurate lenses of high magnification and relatively narrow fields of view) are used.

So far you have been given information on those optical devices in which only optical principles are involved. In the next few pages, the similarity of electromagnetic waves to light waves is pointed out. Television camera tubes, which employ electronic principles as well as optical principles, also are presented.

COMPARISON OF LIGHT AND ELECTROMAGNETIC WAVES

To illustrate the similarity between light and electromagnetic waves, your attention is directed to a comparison of optical systems and directional and beam-forming antenna systems, such as are employed with radio direction finders and radar equipment shown in figure 4-42.

Radiation of light from a point source in space is omnidirectional; that is, the light rays emanate in all directions along straight-line paths. By means of reflectors, lenses, and filters, light rays can be directed, diffused, concentrated, selected as to

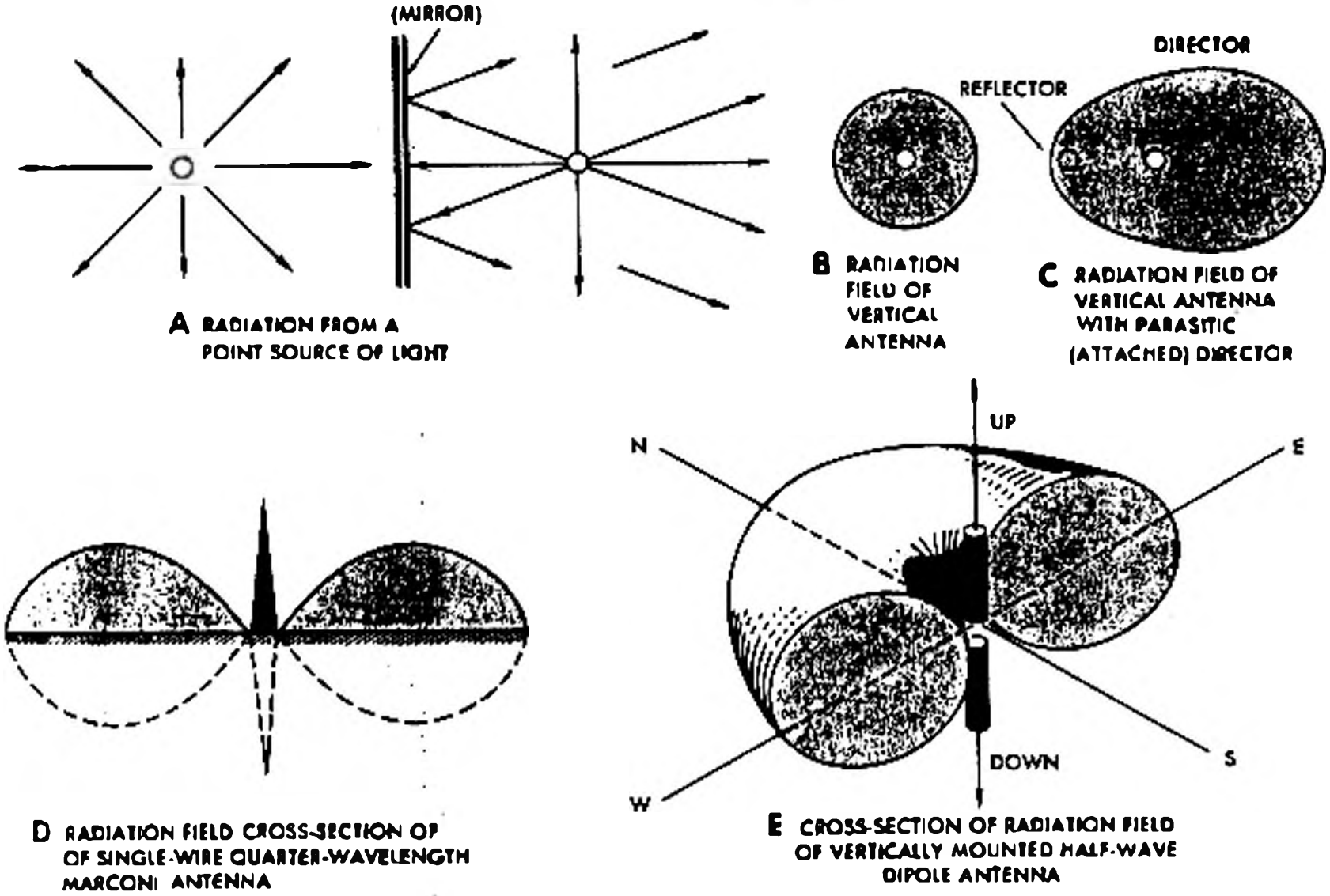


Figure 4-42. Radiation of Light Waves and Electromagnetic Waves

wavelength (color), and focused upon specific points as desired. This is illustrated in figure 4-42A.

Radiation of electromagnetic waves from a single-wire vertical antenna (Marconi type) or from a vertically mounted dipole is in the form of concentric fields along the horizontal plane as illustrated in figures 4-42B and 4-42C.

By means of suitable reflector and director elements, the electromagnetic radiations are directed and concentrated (beamed) to produce a radiation pattern of the desired form. Also by *phasing*, or feeding two or more antenna elements in varying degrees of phase with respect to each other, the fields produced around each element can be made to reinforce or to cancel each other. The reinforcing or canceling is done in such a manner that the individual fields combine vectorially to produce a radiation field of maximum intensity along the desired paths in either or both horizontal and vertical planes.

Where a narrow and intense beam of energy is required, as in tracking radar applications, parabolic reflectors or combinations of reflector and director elements may be used to form the beam. The beam is formed in much the same manner that the reflector and lens in a "bulls-eye" lamp or focusing flashlight shape the beam of light.

Basically, any *antenna* is a conductor or system of conductors used to radiate or receive energy in the form of electromagnetic waves. The height of the radiating element above ground, the conductivity of the earth below it, the wavelength of the radio propagation, and the shape and dimensions of the antenna all affect the pattern of the field radiated into space. By controlling these factors, it is possible to produce the desired radiation pattern.

If directing the radiation from a source of light in one general direction is desired, a reflecting surface behind the source directs all the rays toward the object to be illuminated. The rays which normally would follow paths in directions away from the object to be illuminated are reflected back toward the object and reinforce the rays normally traveling toward it.

If the object is small, or a maximum intensity of illumination on a specific point is desired, a greater number of omnidirectional rays can be concentrated on the desired point by using a curved reflector which would cause the light rays to converge at the desired point. A reflector of specific curvature located at a fixed distance from a source causes the light rays to converge at a given point along its principal axis.

If it should be undesirable or impractical to vary either the source or the reflector, a movable lens may be employed between the object and the source to focus the reflected beams, as illustrated in figure 4-43.

The foregoing effects are obtained with either light or high-frequency electromagnetic waves by

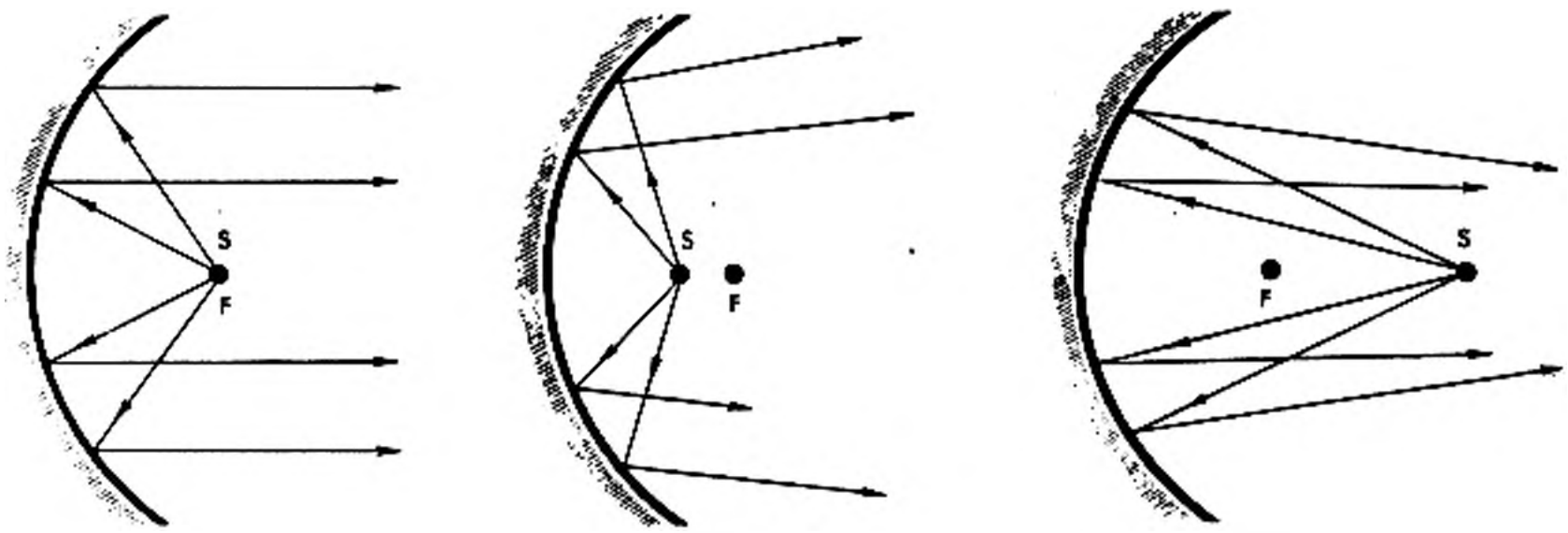


Figure 4-43. Reflecting Beam Focused by Employing Movable Lens

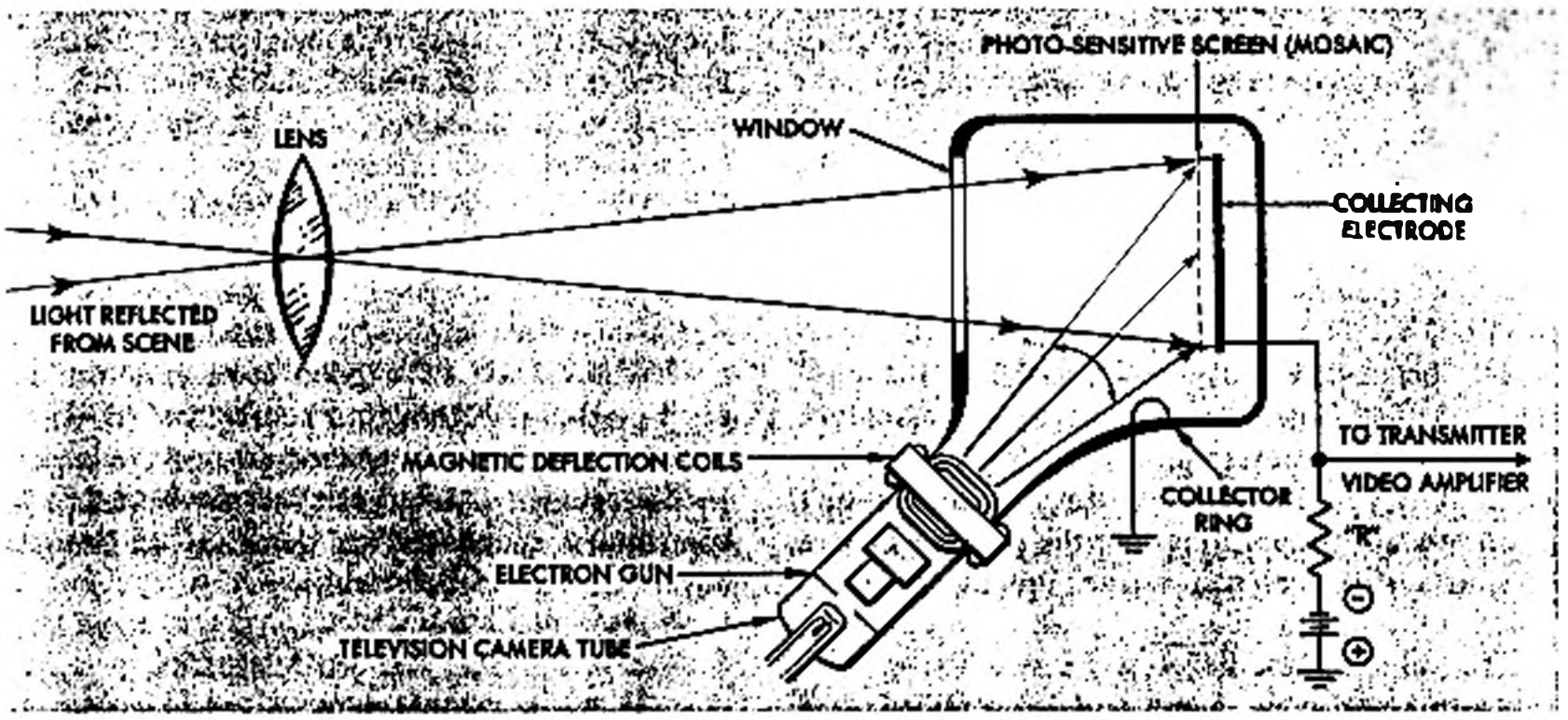


Figure 4-44. Iconoscope Picks up Image of Scene to be Televised

the use of parabolic reflectors. Parasitic (attached) reflectors and directors may be used to cause divergence or convergence of electromagnetic waves, serving as forms of electron lenses.

ELECTRONIC AND OPTICAL PRINCIPLES OF CAMERA TUBES

A good example of a device that employs both optical and electronic principles is a television-camera tube, used to pick up the scene to be transmitted. Several types of camera tubes are in use, and while they vary in mechanical details, their principles of operation are basically the same.

First consider the "Iconoscope," a camera tube widely used today. The "Iconoscope" consists of a glass envelope shaped somewhat like a saucepan with an enlarged handle. The side of the envelope corresponding to the top of the pan is a transparent window through which the image of the object is permitted to fall upon a screen or mosaic, composed of thousands of photosensitive globules. An electron gun similar to that of the common cathode-ray tube is used to project a beam of electrons over the mosaic.

This beam, known as a scanning beam, removes the charge from each section or globule of the mosaic in a fixed sequence. And in so doing, it produces discharge currents in proportion to the quantity of light received upon each section of the mosaic. These currents serve to amplitude modulate the transmitter and develop the picture intelligence in the television system.

The illustration, in figure 4-44, of the "Iconoscope" shows some of the associated components used to convert variations in intensity of the light reflected from the object being televised into variable currents.

The object or scene being televised is focused on the photosensitive mosaic which contains the thousands of minute globules. The globules are electrically insulated from each other.

The photosensitive screen consists of a thin sheet of mica. On one side of the mica sheet are the globules made up of a light-sensitive compound of silver and cesium. On the reverse side of the mica sheet is a thin metal coating that serves as a collecting electrode. Each light-sensitive globule is effectively one electrode of a tiny capacitor, while the metal coating on the back of the mica sheet serves as a common electrode for all the globules. The mosaic comprises many tiny capacitors all having one common electrode, and when the optical image is focused upon it, the reflected rays strike the tiny photocells, causing them to lose electrons through secondary emission. The number of electrons lost by each tiny element is proportional to the intensity of the light ray which strikes it. These released electrons are drawn to a collector ring (graphite coating) on the inner surface of the tube. The globules

from which they were released now have a positive charge with respect to the backplate or collecting electrode of the mosaic.

The portions of the mosaic which are most brilliantly illuminated emit the greatest number of electrons and have the highest positive charge.

After exposure, the mosaic is scanned sectionally in a regular sequence by a magnetically deflected beam of electrons from the electron gun. This scanning beam neutralizes each element of the mosaic and causes current to flow through the collector load resistor ("R" in the illustration) in an amount proportionate to the positive charge existing on each element struck by the scanning beam.

The voltage pulse developed across the resistor represents the video signal and is amplified to a suitable level for modulation of the transmitter's carrier wave. The picture intelligence is transmitted to the receiver in the form of an amplitude-modulated carrier.

The receiver employs a conventional amplifying and demodulation system to separate the intelligence from the carrier. This separation is in the form of a varying voltage which is applied to the signal grid (control grid) of the cathode-ray-type picture tube or kinescope, on the screen of which the image is reproduced.

A scene is transformed from variations in light to variations in current and back to variations in light again by means of combined optical and electronic processes. In this manner, a scene may be transmitted a long distance.

An *optical system* controls radiant energy by means of mirrors, reflectors, prisms, and lenses so that the radiant energy may be converted into a form suitable to the requirement of the application; an *antenna system* also controls electromagnetic radiations in part by means of reflectors and lenses (directors). When you study the control and navigation systems of missiles, you will find this information helpful toward a clear understanding of the systems.

SECTION C

Solid State Devices

Since the invention of the transistor in 1948, semiconductor devices have been developed and improved at an unbelievable rate. Solid state devices made from semiconductor materials offer compactness, efficiency, and versatility, heretofore unattainable. These devices have invaded virtually every field of science and industry and are being used in increasing numbers for communications and aerospace applications.

As its name implies, this solid state material is classed neither as a good conductor nor as a good insulator. Although it bears physical likeness to both, its electrical characteristics are unique. In the span of a few years, its development has contributed much to the demands of the ever advancing electronic world.

PHYSICS OF TRANSISTORS

Before the operation of a transistor can be understood, the semiconducting material must first be examined.

A semiconductor, as stated before, is neither a good conductor nor a good insulator. Germanium and silicon, which are used in most semiconductors, are substances that fall into this category.

To understand the theory of the transistor it will be helpful to consider briefly the structures of atoms and crystals. All matter is composed of one or more elements, and each element is composed of atoms. The atom is composed of the smallest units of matter called *electrons*, *protons*, and *neutrons*. Electrons possess a *negative charge*. Protons possess an equivalent *positive charge* and are more than 1800 times heavier than the electron. The neutron possesses *no charge* and has the same mass as the proton.

The *atom*, which is the smallest subdivision of an element possessing all the properties of the element, is composed of a nucleus or core of protons and neutrons around which the lighter electrons revolve. In the normal atom there are as many electrons (negative charges) outside the nucleus as there are protons (positive charges) within the nucleus. The atom is balanced, or electrically neutral.

Atoms of one chemical element differ from those of another element only in the number of electrons, protons, and neutrons in their structure. With reference to transistors, we will consider only a few elements. These elements are germanium, silicon, antimony, arsenic, aluminum, gallium, indium, and thallium.

The study of atomic structure has shown that a large portion of the electrons around the nucleus are tightly bound to it and do not enter into chemical reactions or transistor physics. The nucleus and the tightly bound electrons comprise an inert core with a net positive charge around which the less tightly bound electrons revolve.

In transistor physics, you are concerned with the net charge on the core and the electrons surrounding the nucleus. For example, each atom of germanium has 32 protons in its nucleus and 28 tightly bound electrons around it; thus the atoms of germanium can be represented by a core with a net charge of +4 surrounded by four electrons. Each atom of silicon, which has 14 protons in its nucleus and 10 tightly bound electrons around it, is represented in the same manner as an atom of germanium, as shown in figure 4-45.

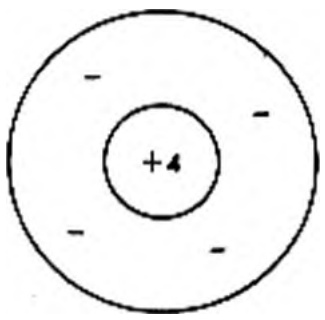


Figure 4-45. Atomic Structure of Germanium or Silicon

Pure germanium can not be used in producing transistors because germanium acquires the properties of rectification and amplification of current through the presence of impurities in the crystals. One type of impurity is known as *donor* and another type is called *acceptor*. They have these names because of the manner in which they affect the electron movement within the transistor.

Antimony and arsenic become donors when they join the crystal structure of germanium. The net charge on the cores of antimony and arsenic atoms is +5. Each atom has five electrons surrounding its core, as shown in figure 4-46.

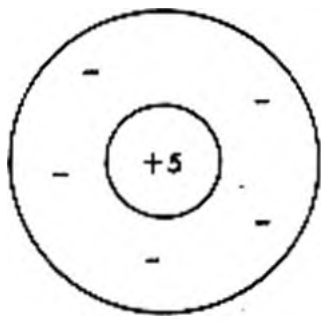


Figure 4-46. Atomic Structure of Antimony or Arsenic (Donors)

Aluminum, gallium, indium, and thallium become acceptors when they join the crystal structure of germanium. The net charge on the cores of their atoms is +3, and each of their atoms has three electrons surrounding its core. The atomic structure of these acceptors is pictured in figure 4-47.

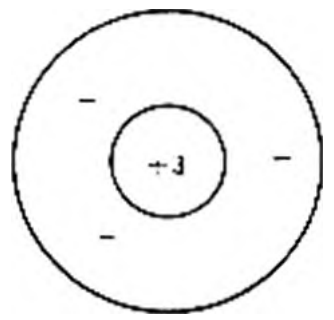


Figure 4-47. Atomic Structure of Aluminum, Gallium, Indium, or Thallium (Acceptors)

Figure 4-48 compares the atomic conditions existing in a pure germanium crystal with those existing in crystals containing donor and acceptor impurities.

In the pure crystal each atom has four neighbors which are equidistant from each other. Between the cores of the atoms and each of their neighbors are two electrons. These paired electrons form *electron-pair bonds* which come into existence when two or more atoms approach each other. Since the electrons are in constant motion around the core, these electron-pair bonds are formed when the movement of an electron from one atom becomes coordinated with the movement of an electron from another atom.

This coordination tends to attract the cores toward each other, but the cores' positive charges repel each other until they attain a perfect balance of attraction and repulsion. The atoms are then said to be in a condition of equilibrium.

As previously mentioned, pure germanium crystals can not be used as transistors. This condition exists because the atoms of the crystals would be in a state of equilibrium, in which form they constitute good insulators with a dielectric constant of approximately 16.

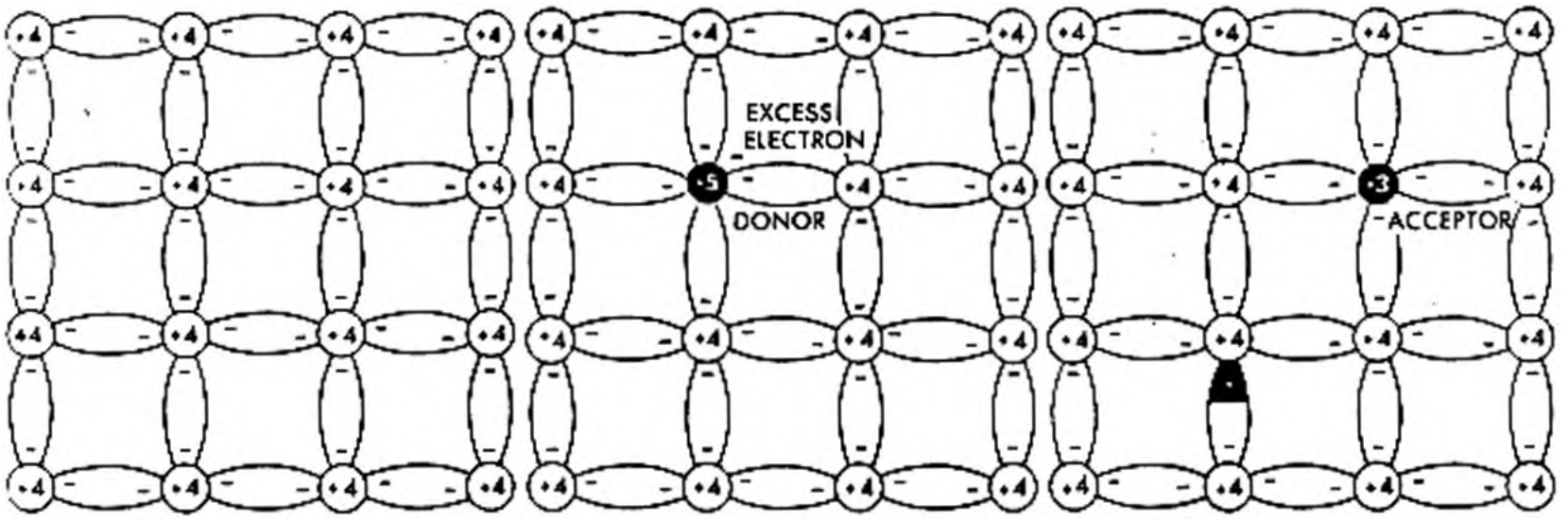


Figure 4-48. Atomic Conditions in Germanium Crystals

Effects of Donor Atoms on Germanium Crystals

When a donor atom (antimony or arsenic) joins the crystal structure of germanium, the atom must lose one of the five electrons surrounding its core because only four of its electrons can form electron-pair bonds with the electrons of its neighboring germanium atoms. The excess electron is free to move through the relatively wide spaces between the cores. It moves through the crystal as though it were in a vacuum.

If a battery were placed across the crystal, the electron would move toward its positive terminal and enter the battery at that point. Simultaneously, an electron would leave the negative terminal of the battery and enter the crystal. A continuous electron flow through the crystal would take place, but the cores of the germanium and donor atoms would remain undisturbed.

Germanium crystals containing donor impurities are known as N-type germanium.

The designation "N" results from the fact that conduction through the crystal is mainly a conduction of *negative charges* in the form of excess electrons from the donor atoms. It is this action which led to the term *donor* being applied to the element which when joined to the crystal structure of germanium gives off an excess electron.

Effects of Acceptor Atoms of Germanium Crystals

When an acceptor atom (aluminum, gallium, indium, or thallium) joins the crystal structure of germanium, it does so by accepting an electron from one of its neighboring germanium atoms. Atoms of acceptors have three electrons surrounding each of their cores, while the cores of germanium atoms have four electrons surrounding them. When this new electron-pair bond is formed, a *hole* is created in another electron-pair bond. This hole possesses the equivalent charge of an electron, but it is a *positive charge*.

Experiments have shown that the hole is free to move throughout the crystal structure. Conduction of current can take place through a germanium crystal containing an acceptor as readily as through one containing a donor. However, the conduction process is different.

If a battery were placed across this crystal, the hole would be attracted toward the negative terminal of the battery and an electron from this terminal would enter the crystal and fill the hole.

Simultaneously, an electron from one of the electron-pair bonds in the crystal and near the positive terminal of the battery would separate from its bond and enter the battery, creating another hole in the crystal. This action would repeat and maintain a continuous flow of current through the crystal.

Germanium crystals containing acceptor impurities are know as P-type germanium, the term "P" being derived from the fact that in such crystals the conduction through the crystal is mainly a conduction of *positive charges* (holes).

A *hole* can be defined as an incomplete group of electrons whose general properties are similar

to those of an electron except that it carries a positive charge instead of a negative one.

An *acceptor* is defined as an element which when it joins the crystal structure of germanium, produces a hole or excess positive charge within the crystal.

PN-TYPE JUNCTION TRANSISTORS

Germanium crystals are good conductors and can conduct current equally well in either direction. Rectification with a germanium crystal occurs if P-type germanium and N-type germanium are placed side by side. The plane at which the two types of germanium meet is called a PN junction, and the action which occurs at the junctions constitutes the basic action of transistor operation.

Figure 4-49 illustrates a PN-type junction in a

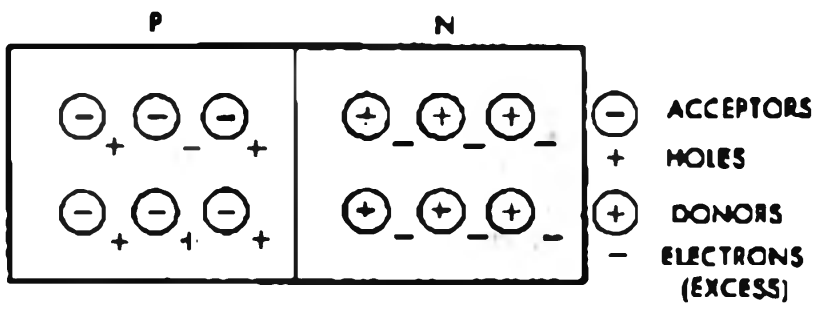


Figure 4-49. PN Junction in Condition of Equilibrium

state of equilibrium. Notice that the holes concentrate to the right of the P-type germanium atoms and the excess electrons concentrate to the right of the N-type germanium atoms. This phenomenon results from the distribution of electrostatic potential produced by the acceptor and donor atoms as illustrated in figure 4-50.

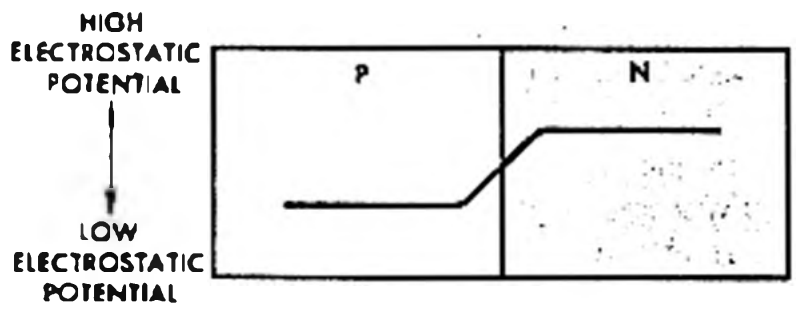


Figure 4-50. Electrostatic Potential

The electrons remain in the region of highest electrostatic potential, and the holes remain in the region of lowest electrostatic potential. When an electron is in the region of highest electrostatic potential, its potential energy is at a minimum as shown in figure 4-51. Since *potential energy* means

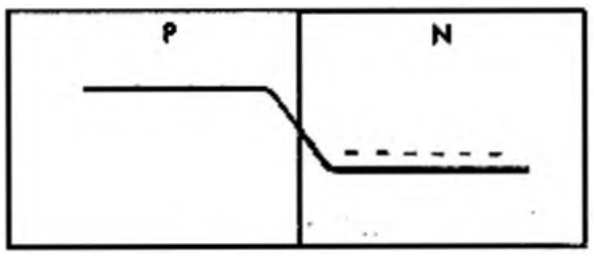


Figure 4-51. Potential Energy of Electrons

ability to do work and since the electron cannot move to do work after it reaches the point of highest electrostatic potential, it becomes apparent that the electron is in a region of low-potential energy.

Remember the preceding statement because potential energy diagrams are used often in illustrating transistor behavior.

The same statement can be applied with reference to holes. When the hole is in a region of lowest electrostatic potential (that is, low negative potential), its potential energy is at a minimum as illustrated in figure 4-52.

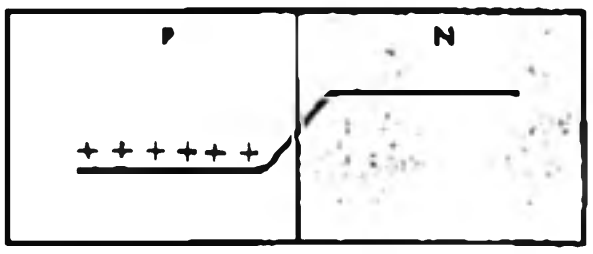


Figure 4-52. Potential Energy of Holes

Holes and electrons flow into regions of low-potential energy only. A low-potential-energy region for an electron is a high-potential-energy region for a hole, and vice versa.

PN Junction with Reverse Bias

If you connected a battery across a PN germanium crystal as shown in figure 4-53, there would be no conduction of current through the crystal. Such a connection is called *reverse bias*.

This condition occurs when the positive terminal of the battery is connected to the N-type germanium and the negative terminal is connected to the P-type germanium.

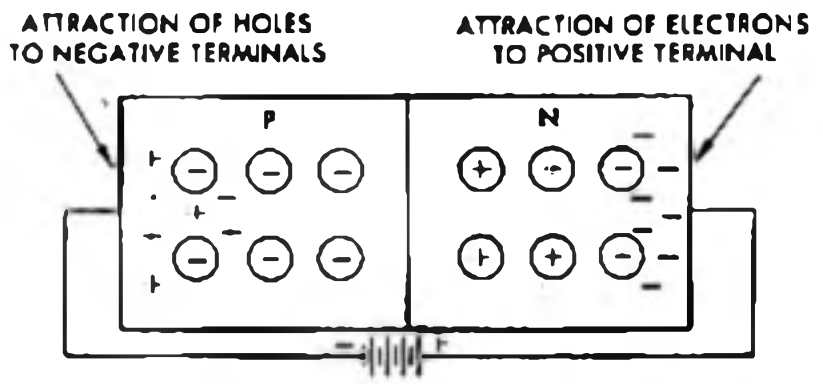


Figure 4-53. PN Junction with Reverse Bias

The positive terminal of the battery attracts the electrons and causes them to concentrate farther to the right than when the junction is in a condition of equilibrium. The negative terminal of the battery attracts the holes and causes them to concentrate farther to the left than when the junction is in a condition of equilibrium. As a result of these attractions, there is no flow of electrons to the left and no flow of holes to the right.

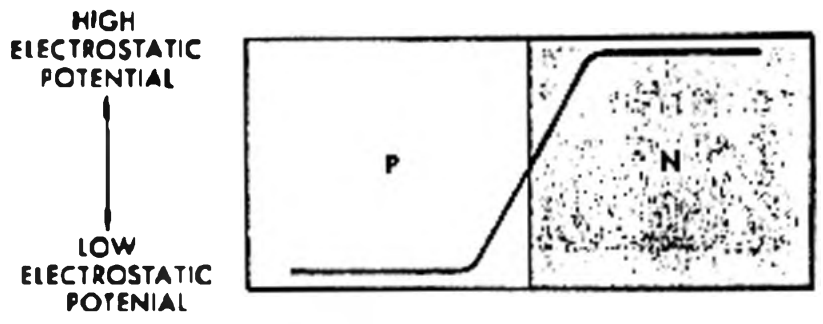


Figure 4-54. Electrostatic Potential Increased in PN Junction with Reverse Bias

The difference in electrostatic potential between the two types of germanium has been increased as shown in figure 4-54.



Figure 4-55. Potential Energy of Holes. (Hill has been increased due to reverse bias)

The *potential-energy hill* for the holes has been increased. They will not flow up the steep hill. This condition is shown in figure 4-55. The same condition applies to the *potential-energy hill* for the electrons and prevents their flow as shown in figure 4-56.

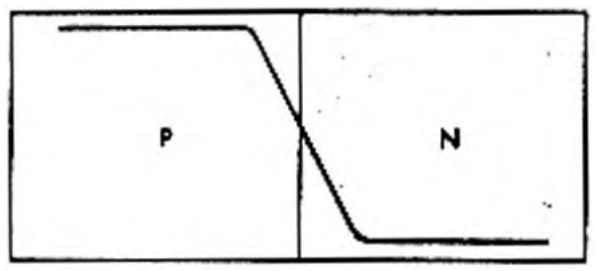


Figure 4-56. Potential Energy of Electrons. (Hill has been increased due to reverse bias)

PN Junction with Forward Bias

If you connected a battery across a PN germanium crystal with the positive terminal connected to the P-type germanium and the negative terminal to the N-type germanium, current would flow in proportion to the applied voltage. This type of connection is known as forward bias.

As shown in figure 4-57, the positive terminal

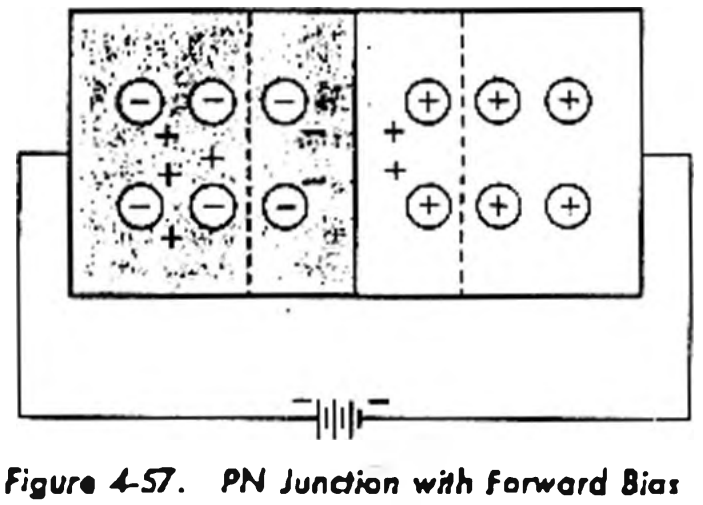


Figure 4-57. PN Junction with Forward Bias

of the battery repels the holes and causes them to move toward N-type germanium. Some of the holes enter the N area.

The negative terminal of the battery repels the electrons and causes them to move toward the P-type germanium. And, in this case, some electrons enter the P area.

Electrons and holes combine in a small area of diffusion on either side of the PN junction (between the dotted lines in the diagram). For each hole in the P region that combines with an electron from the N region, an electron from an

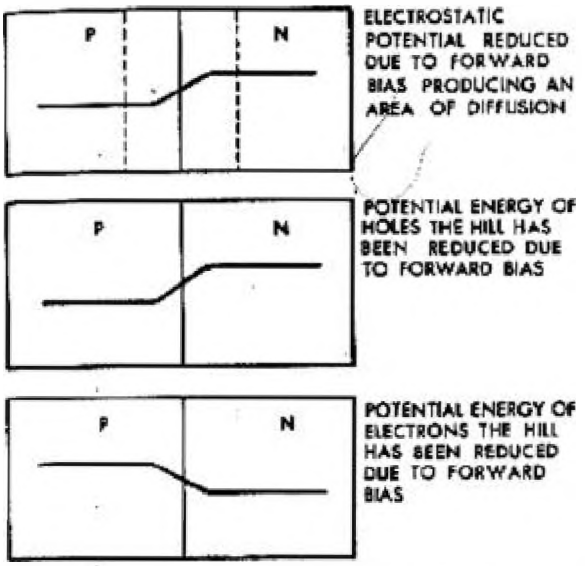


Figure 4-58. Potential Energy Conditions Existing under Application of Forward Bias

electron-pair bond in the crystal near the positive terminal of the battery enters the battery at the positive terminal. This action creates a new hole which moves toward the N-type germanium. For each electron that combines with a hole in the N-type germanium, an electron enters the crystal from the negative terminal of the battery. The current flow in the P region is mainly a flow of holes, while that in the N region is mainly a flow of electrons. The potential-energy conditions existing under the application of the forward bias are shown in figure 4-58.

The currents that flow through the crystal are shown in figure 4-59. The total current, I_T, is constant: current by holes, I_P, is shown by the solid line; and current by electrons, I_N, is shown by the dashed line.

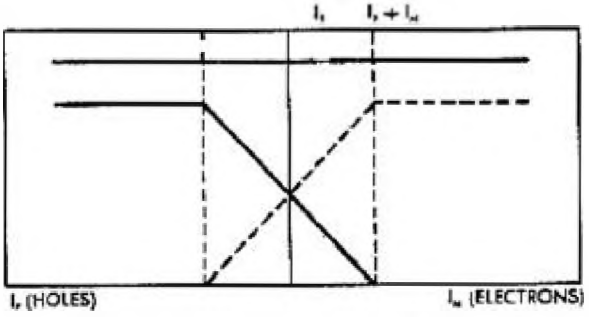


Figure 4-59. Current Flow through Crystal under Application of Forward Bias

From the above theory, you can understand more fully the process of rectification in germanium crystals; you can better understand how the crystal acts as a low resistance when forward bias (+ to P, – to N) is applied, and how it acts as a high resistance when reverse bias (– to P, + to N) is applied. These phenomena constitute the basic principles of transistor operation.

Transistors, as presently developed, fall into two general classifications: junction transistors and point-contact transistors. Junction transistors are presented first.

NPN AND PNP JUNCTION TRANSISTORS

The NPN transistor is constructed by placing a narrow strip of P-type germanium between two relatively long strips of N-type germanium. As the letters indicate, the PNP transistor consists of a narrow strip of N-type germanium between two relatively long strips of P-type germanium.

NPN Junction Transistors

An NPN junction transistor is pictured in figure 4-60. In this type, large surface contacts (low-resistance) are made to each strip. The N-type germanium, as shown at the left, is called the emitter, e. The N-type strip, shown at the right, is called the collector, c. The P-type section in the center is called the base, b.

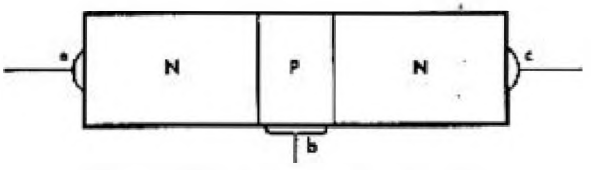


Figure 4-60. NPN Junction Transistor

Figure 4-61 shows the distribution of the do-

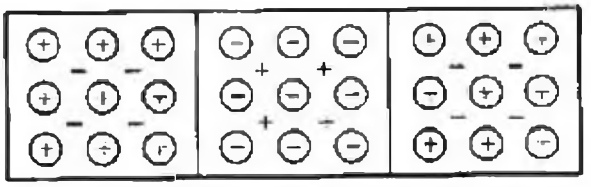


Figure 4-61. NPN Junction Transistor in State of Equilibrium

nors, acceptors, holes, and electrons in an NPN junction transistor under equilibrium conditions (no external voltages applied). The potential energy in this type of transistor is represented in figure 4-62. Figure 4-63 shows the potential energy region of the electrons.

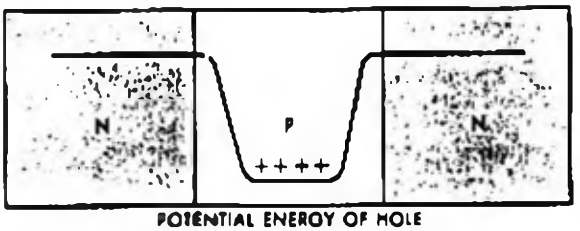


Figure 4-62. Potential Energy of Holes in NPN Junction Transistor

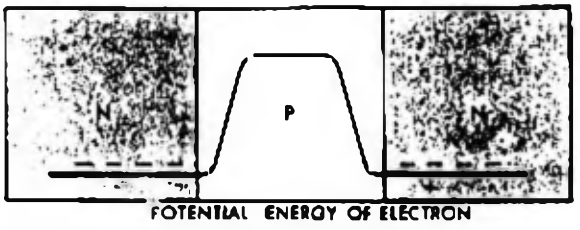


Figure 4-63. Potential Energy of Electrons in NPN Junction Transistor

The holes are concentrated in the region of lowest potential energy for them. They cannot climb the potential-energy hills to the right or left. The electrons are concentrated in the region of lowest potential energy for them and cannot climb the potential-energy hills to enter the P-type germanium. No current flows.

In practical applications, the NPN junction transistor normally is biased in the manner shown in figure 4-64. The PN junction between the emitter, e, and the base, b, is biased in the forward direction. The PN junction between the collector, c, and the base, b, is biased in the reverse direction.

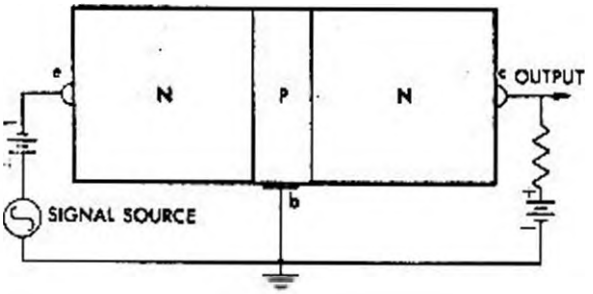


Figure 4-64. NPN Junction Transistor Operating Circuit

Figure 4-65 shows the potential energy condition for electrons, with no signal applied. In NPN transistors, the electrons are the major current components. Potential-energy diagrams for holes are not shown.

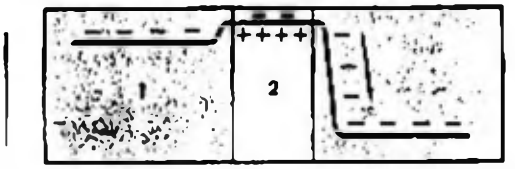


Figure 4-65. Potential Energy of an Electron (no signal)

The forward bias applied between the emitter and the base reduces the potential-energy hill at the left-hand PN junction. Some electrons climb this hill and enter the P-type germanium. Since the base strip is relatively thin, most of the electrons which enter it will not combine with holes. Instead they pass through the strip and readily go down the potential-energy slope at the right PN junction, as shown by the arrow in figure 4-65.

The steep potential-energy slope which permits easy entrance of the electrons from the base strip to the N-type germanium of the collector is produced by the application of reverse bias between the collector and the base.

When a signal is applied which opposes the forward bias on the emitter (makes base more negative with respect to emitter), the potential-energy hill between the emitter and the base is increased, and fewer electrons climb the hill to enter the P-type germanium. This condition is illustrated in figure 4-66. The electrons which do

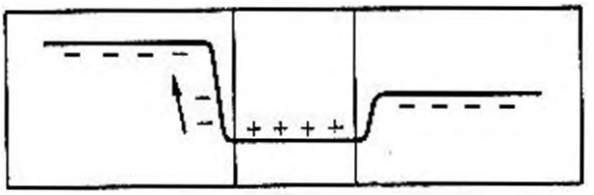


Figure 4-66. Signal Applied to NPN Transistor Making Base More Negative with Respect to Emitter

enter the P-type germanium do not recombine with holes but fall into the collector region of low potential energy.

When the polarity of the applied signal aids the forward bias (makes base more positive with respect to emitter), the potential-energy hill between the emitter and the base is decreased, and more electrons flow into the P-type region. Most of these electrons do not combine with holes but flow readily to the low potential-energy level of the N-type germanium of the collector. This condition is demonstrated in figure 4-67.

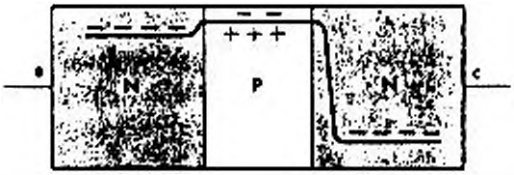


Figure 4-67. Signal Applied to NPN Transistor Making Base More Positive with Respect to Emitter

You can compare the operation of the NP junction transistor to the operation of the triode vacuum tube. The emitter is equivalent to the cathode, the base to the grid, and the collector to the plate. Practically all of the electrons which emerge from the emitter (cathode) go to the collector (plate). The base (grid) current is minute and consists only of a small number of electrons which represent the recombination number of holes and electrons in the P-type region.

Two types of current are present in the vacuum tube, and two types of current are involved in the transistor. The thermionic electrons (those emitted from the cathode) of the tube are equivalent to the excess electrons from the emitter.

The minute conduction current of electrons (those which flow in and out of the grid to control plate current) is equivalent to the small current of holes in the base region (which varies to control collector current).

PNP Junction Transistors

As stated before, the PNP junction transistor consists of a narrow strip of N-type germanium between two relatively long strips of P-type germanium. Large surface contacts (low-resistance) are made to each strip. The contacts are designated b, c, and e, representing the base, collector, and emitter, respectively, as in the case of the NPN transistor. This is illustrated in figure 4-68.

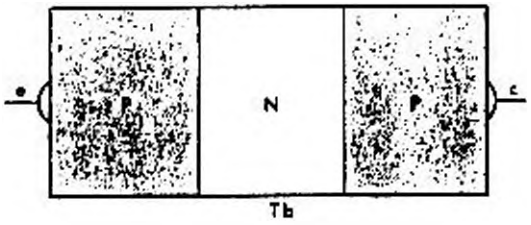


Figure 4-68. PNP Junction Transistor

The operation of the PNP junction transistor is analoguous to that of the NPN except that the hole constitutes the main current component instead of the electron. Pictured in figure 4-69 is the PNP transistor in a state of equilibrium.

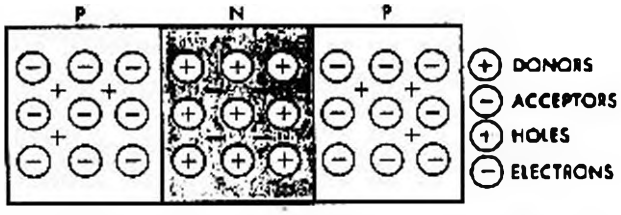


Figure 4-69. PNP Junction Transistor in State of Equilibrium

The excess electron is the key factor in the operation of the PNP junction transistor, but the hole is the main current component. Basic conditions determining the potential energy of the holes are shown in figures 4-70 and 4-71.

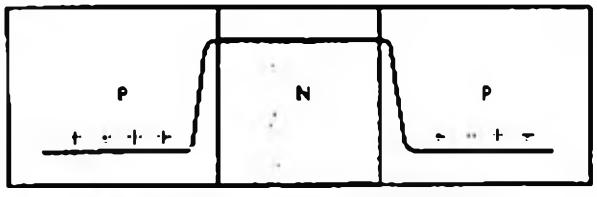


Figure 4-70. Potential Energy of Holes

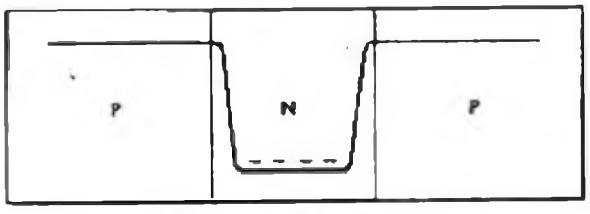


Figure 4-71. Potential Energy of Electrons

Figures 4-72 through 4-75 illustrate the conditions of operation for the PNP junction transistor in the same manner that the earlier figures illustrated the conditions of the NPN transistor.

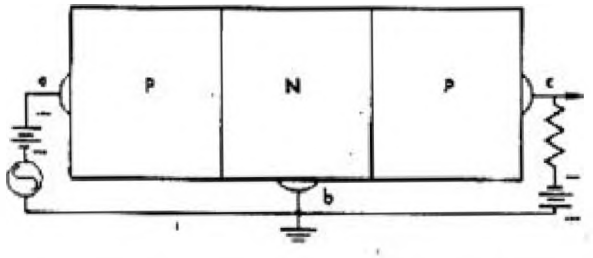


Figure 4-72. PNP Junction Transistor Operating Circuit

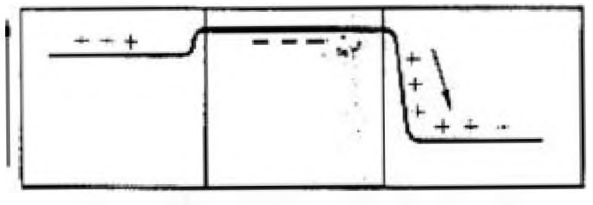


Figure 4-73. PNP Transistor (no signal)

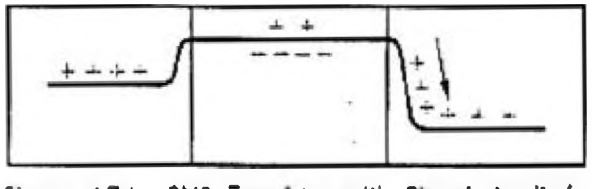


Figure 4-74. PNP Transistor with Signal Applied, making Base more Positive with Respect to Emitter

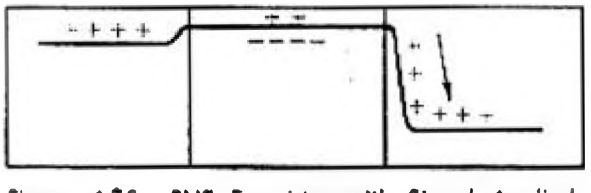


Figure 4-75. PNP Transistor with Signal Applied, making Base More Negative with Respect to Emitter

To bias the PN junction between the emitter and the base of the PNP transistor in the forward direction, the emitter is made positive with respect to the base. To bias the collector in the reverse direction, the collector is made negative with respect to the base.

The emitter, base, and collector connections to the crystal are low-resistance (large-area) connections in a junction transistor. In this respect, junction transistors differ from point-contact transistors, except with regard to the base connection.

POINT-CONTACT TRANSISTORS

In a point-contact transistor, the emitter and collector connections to the crystal are point-contact (relatively high-resistance, small-area) connections. Figure 4-76 illustrates a cutaway

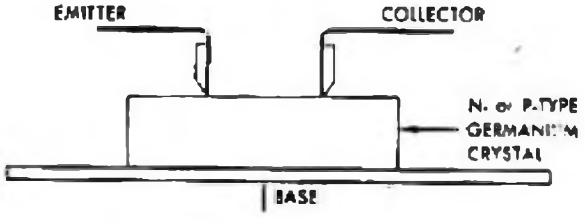


Figure 4-76. Point-Contact Transistor

view of this type of transistor. You can see that the base connection is a large-area, low-resistance connection.

As already explained, the emitter current is slightly larger than the collector current in a junction transistor, and the base current is extremely small. In a point-contact transistor, the collector current (un-like that of the junction transistor) is substantially larger than the emitter current and the base current is relatively large.

These differences between the junction transistor and the point-contact transistor are explained in the following paragraphs.

Although the point-contact transistor employs what is considered to be either an N-type germanium crystal or a P-type germanium only, actually in the N-type point-contact transistor, P-type layers occur and in the P-type point-contact transistors, N-type layers occur.

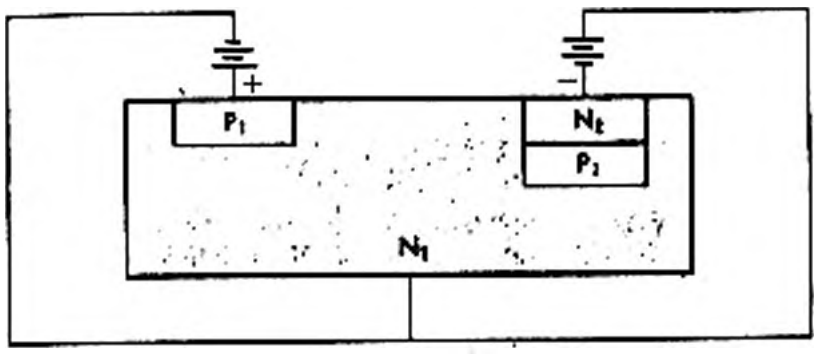


Figure 4-77. Construction of N-Type Point-Contact Transistor

N-Type Point-Contact Transistors

The illustration of an N-type point-contact transistor, in figure 4-77, shows that under the emitter point there is a thin layer of N-type germanium, N_2, followed by a thin layer of P-type germanium, P_2.

You can see in this drawing that the emitter-base section is biased in the forward direction. Current flow, consisting mainly of holes in the P_1-type germanium and of electrons in the N_1-type germanium, will occur. Under these conditions, an increase in emitter current causes a large increase in collector current, and a decrease in emitter current causes a large decrease in collector current.

To help understand the current amplification of a point-contact transistor, consider the collector-base circuit of the transistor. An enlarged view of the collector-base circuit is illustrated in figure 4-78.

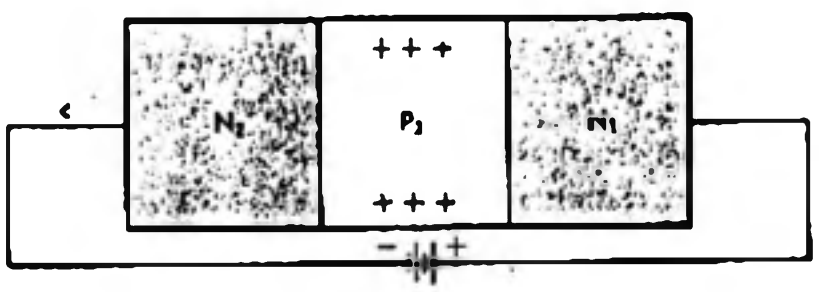


Figure 4-78. Collector-Base Circuit of N-Type Point-Contact Transistor

The potential energy of electrons in the crystal portion of this circuit is shown in figure 4-79. Under steady-state conditions (no signal applied), few electrons climb the potential-energy hill between N_2 and P_2. Those that do climb the hill do not combine with the hole in the P_2 region because it is a thin region. Instead, these electrons fall quickly down the potential-energy hill between P_2 and N_1.

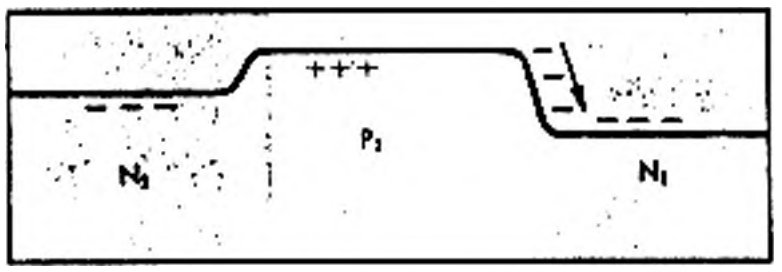


Figure 4-79. Potential Energy of Electrons Without Holes from Emitter

When the emitter-base region conducts, some of the holes which leave the N_1 region of the emitter drift into the P_2 region of the collector. With holes in the P_2 region from the emitter, the potential-energy diagram for electrons between the collector and base conforms to that shown in figure 4-80.

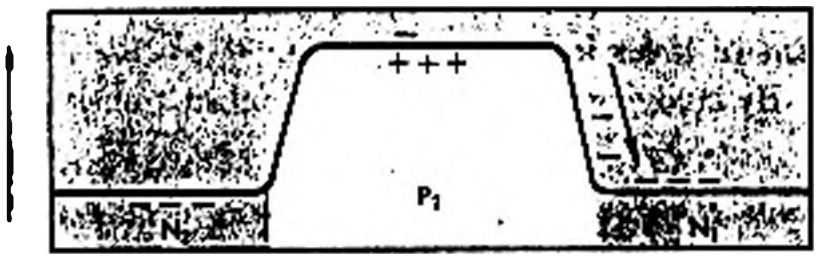


Figure 4-80. Potential Energy of Electrons With Holes from Emitter

Note that the potential-energy hill between N_2 and P_2 has been substantially reduced, and electrons in the N_2 region can climb readily into the P_2 region.

Most of these electrons, instead of combining with holes in this region, fall rapidly into the N_1 base region. For every hole that enters the P_2 region from the N_1 region, many electrons flow from the N_2 region through the P_2 region down into the N_1 region. It is evident that large amplification of current results.

This amplification of current is expressed as α (alpha) and is equivalent to the voltage amplification factor of a vacuum tube (μ). Mathematically,

$$\alpha = \frac{\Delta I_c}{\Delta I_e}$$

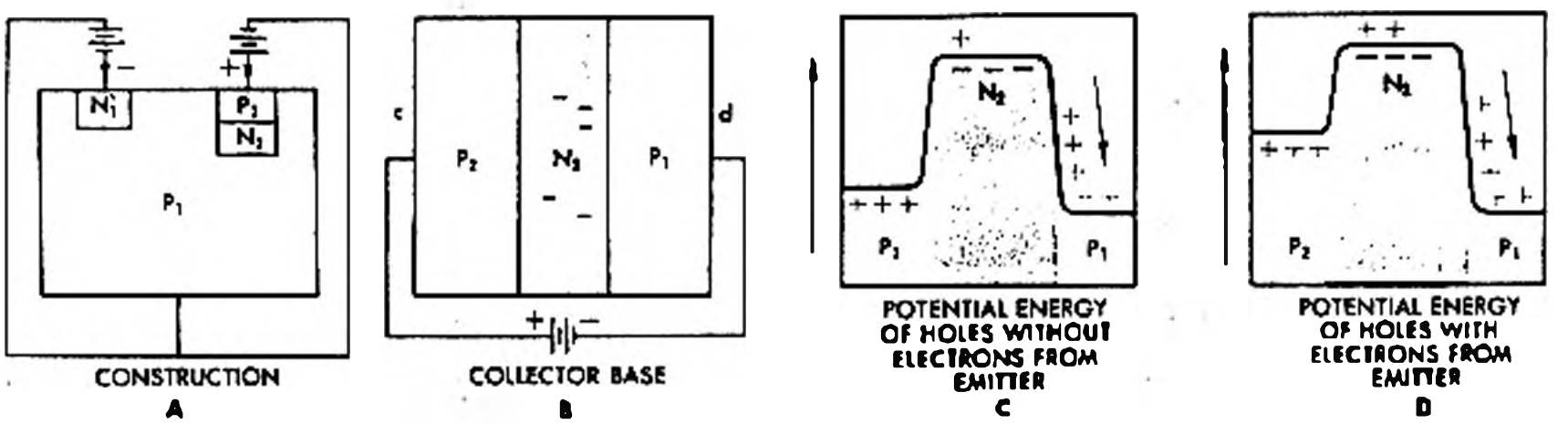


Figure 4-81. P-Type Point-Contact Transistor

where ΔI_c is a small change in collector current, and ΔI_e a small change in emitter current.

P-Type Point-Contact Transistors

Point-contact transistors have been made with P-type germanium as the main body. The theory of the operation of this type of transistor is similar to that of the operation of the N-type. The difference is that the main current-carrier consists of holes instead of electrons.

By studying the P-type point-contact transistor illustrated in figure 4-81 and reviewing the preceding paragraphs, you can understand the operation of the P-type point-contact transistor. Note that the polarities for the emitter with respect to the base and the collector are changed.

Comparison of Point-Contact Transistor to Triode Vacuum Tube

The operation of a point-contact transistor can be compared to the operation of a triode vacuum tube. The emitter is equivalent to the grid, the base to the cathode, and the collector to the plate. The base current of the transistor is relatively

Table 4-3. Comparison of Vacuum Tube and Point-Contact Transistor

VACUUM TUBE	POINT CONTACT TRANSISTOR
Cathode	Base
Grid	Emitter
Plate	Collector
Voltage amp	Current amp
$\mu = \Delta E_p / \Delta E_g$	$\alpha = \Delta I_c / \Delta I_e$
High input impedance	High output impedance
I_p	E_c
E_{in}	I_e
I_g	E_o
E_g	I_o
Constant E supply	Constant I supply
Capacitance	Inductance
Large in size	Small in size
High power consumption	Low power consumption
Fragile to shock	Rugged
Reverse bias	Forward bias

Table 4-4. Comparison of Vacuum Tube and Junction Transistor

VACUUM TUBE	JUNCTION TRANSISTOR
Cathode	Emitter
Plate	Collector
Grid	Base
$\mu = \Delta E_p / \Delta E_g$	$A = \Delta I_c / \Delta I_e$
Voltage amp	Current amp
High input impedance (grounded cathode)	High input impedance (grounded emitter)
Low output impedance	Low output impedance
High power consumption	Low power consumption
Fragile	Rugged
Large in size	Small in size

large, and the major portion of it goes to the collector. Unlike the grid of a vacuum tube, the emitter draws a continuous current.

In the point-contact transistor, the input (emitter) impedance is low and the output (collector) impedance is high. In a vacuum tube the input (grid) impedance is high and the output (plate) impedance is low. A comparison of point-contact type transistors and vacuum tubes is made in table 4-3.

A junction-type transistor has already been compared to a vacuum tube. This comparison is given in table 4-4.

BIAS VOLTAGES

In a vacuum tube the grid is usually biased by means of a voltage which is negative with respect to the cathode. In the case of transistors, the collector may be positive or negative with respect to the base, and the emitter may be positive or negative with respect to the base. Figure 4-82 illustrates the bias polarities required for the various types of transistors.

The bias polarities depend on whether the transistor is an NPN or PNP junction type, or a P or N point-contact type.

If you remember that the N-type point-contact transistor is, by stretching the imagination a bit, actually a PNP-type transistor and the P-type transistor is an NPN transistor, bias polarities can be remembered easily by the following simple rule:

"The emitter is biased in a forward direction with respect to the base, and the collector is biased in the reverse direction with respect to the base, in any transistor."

Since the transistor was developed, many improvements in techniques have been made. Also many special adaptations of semiconductors have been developed and transistors are used as rectifiers, amplifiers, oscillators, multivibrators, sawtooth generators, etc.

NEW TECHNIQUES

The field of electronics seems to change almost daily. Processes that were unheard of a few years ago, even in the laboratory, are now commonplace. Since your job in the missile field may involve some of these processes, a few of them are discussed. The presentation is limited to three important areas of solid state electronics—the laser, the printed circuit, and the integrated circuit.

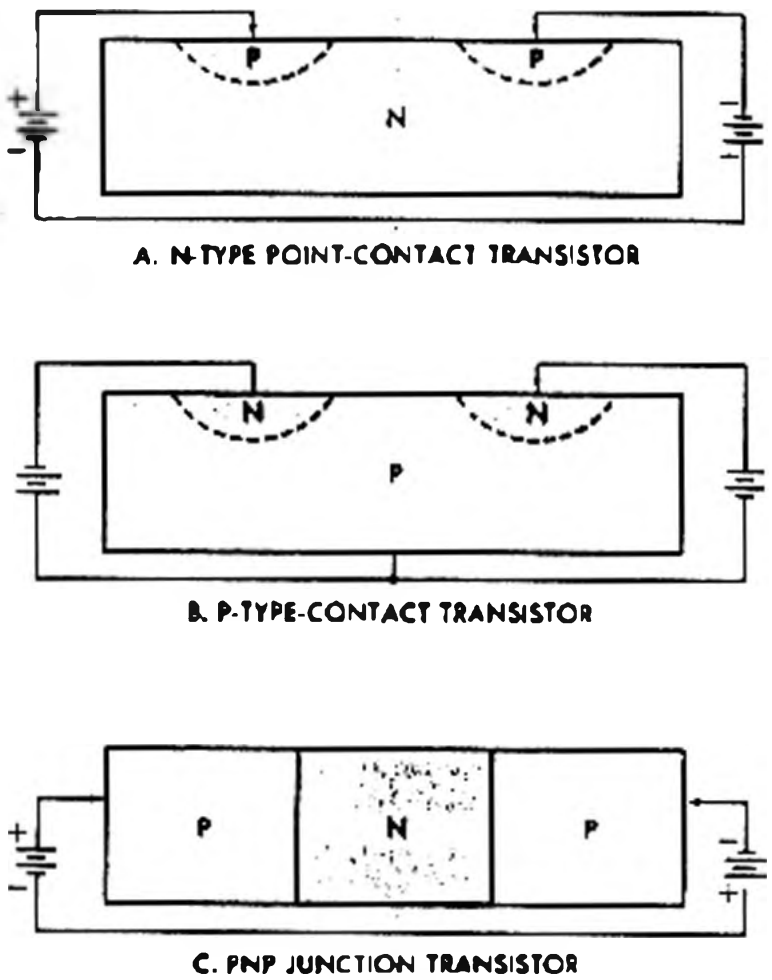


Figure 4-82. Bias Polarities

Laser

One of the fastest growing techniques in the electronic field today is the laser. "Laser" comes from *Light Amplification by Simulated Emission of Radiation*. Atoms exist in more than one energy state. In a laser, atoms are brought up to a high energy level and then the energy is suddenly released. The sudden release of energy results in a beam of light having a very precise frequency. In addition, the beam is coherent because all the energy is sent out in the same phase.

Because the laser light is coherent, it has a very high concentration of energy. By using lenses, the light can be concentrated to a very small spot of light. This concentrated beam can be used in a number of ways. It will machine small holes in diamonds for drawing extremely fine (0.0001 in. diameter and smaller) wires. In other applications this energy is used to cut steel, brick, and other hard materials. In the medical field the laser can be used to perform "bloodless" surgery and to "weld" detached retinas in the eyes.

Another big use for the laser is in the field of communications. Large amounts of information can be carried over a long distance on a laser beam. To get a small idea of this application, imagine all of the long distance telephone traffic in the United States plus four or five television channels being transmitted on one laser beam.

More detailed information about the laser is presented in a later chapter of this manual.

Printed Circuits

Transistors are used extensively with printed circuits. One thing that has made extensive automation possible is the printed circuit. The mechanization of electronics was deterred by the maze of wires and components commonly characteristic of the point-to-point wiring used in electron tube circuits. In a printed circuit, connections are made without wires. At the present time, plated circuits, etched circuits, ceramic-based printed circuits, and transfer plating are used. These are only a few of the many different methods of applying metals to insulators to make printed circuits.

The two types of board generally used for construction of printed boards are laminated plastics and fiberglass. Copper foil approximately 0.002 inch in thickness is laminated to the plastic or fiberglass sheet when the sheet is originally made.

The first step is to produce a master drawing. This is photographed to obtain a negative the proper size for the circuit. The negative is used to produce a silk screen stencil. Ink is forced through the stencil onto the copper foil. The board is then placed in an etching reagent to chemically remove all the copper except that covered by the ink. After etching, the board is rinsed to remove all the acid and a solvent is used to remove the ink.

The final process involves placing the necessary holes, notches, or slots in the circuit for attaching components such as transistors, capacitors, resistors, etc. Remember, this is only one method used to construct a printed circuit.

Integrated Circuits

The last new technique presented here is the integrated circuit. This text does not explain any one particular integrated circuit because each one is different. Integrated circuits are actually nothing more than normal circuits reduced in size and put on one small piece of material. In an integrated circuit any one particular component cannot be changed.

The making of integrated circuits involves many different processes. The circuit is first designed and drawn to devise the correct layout. After the layout is arranged, it is photographed. By the use of photographic and various methods of depositing materials on the base substance, an operating integrated circuit is built up. Leads are welded to the proper places in the circuit, and the whole assembly is encapsulated. After testing, the circuit is ready for use.

The design and manufacture of integrated circuitry has been aimed at reducing the physical size of the complete assembly to the point where many of these assemblies can be placed on your thumbnail. These miniaturized circuits have revolutionized the design of small compact computers and other missile systems.

SECTION D

Modulation of Carrier Waves

Modulation is a process of impressing intelligence upon a carrier by altering its amplitude, frequency, or phase in accordance with the variations of the speech or signal being transmitted. The carrier may be a direct current, an alternating current, or a series of uniform pulses repeated at a constant rate. An unmodulated carrier in itself conveys no intelligence except that the transmitter is in operation.

When some characteristic of the carrier is caused to vary as a function of the instantaneous

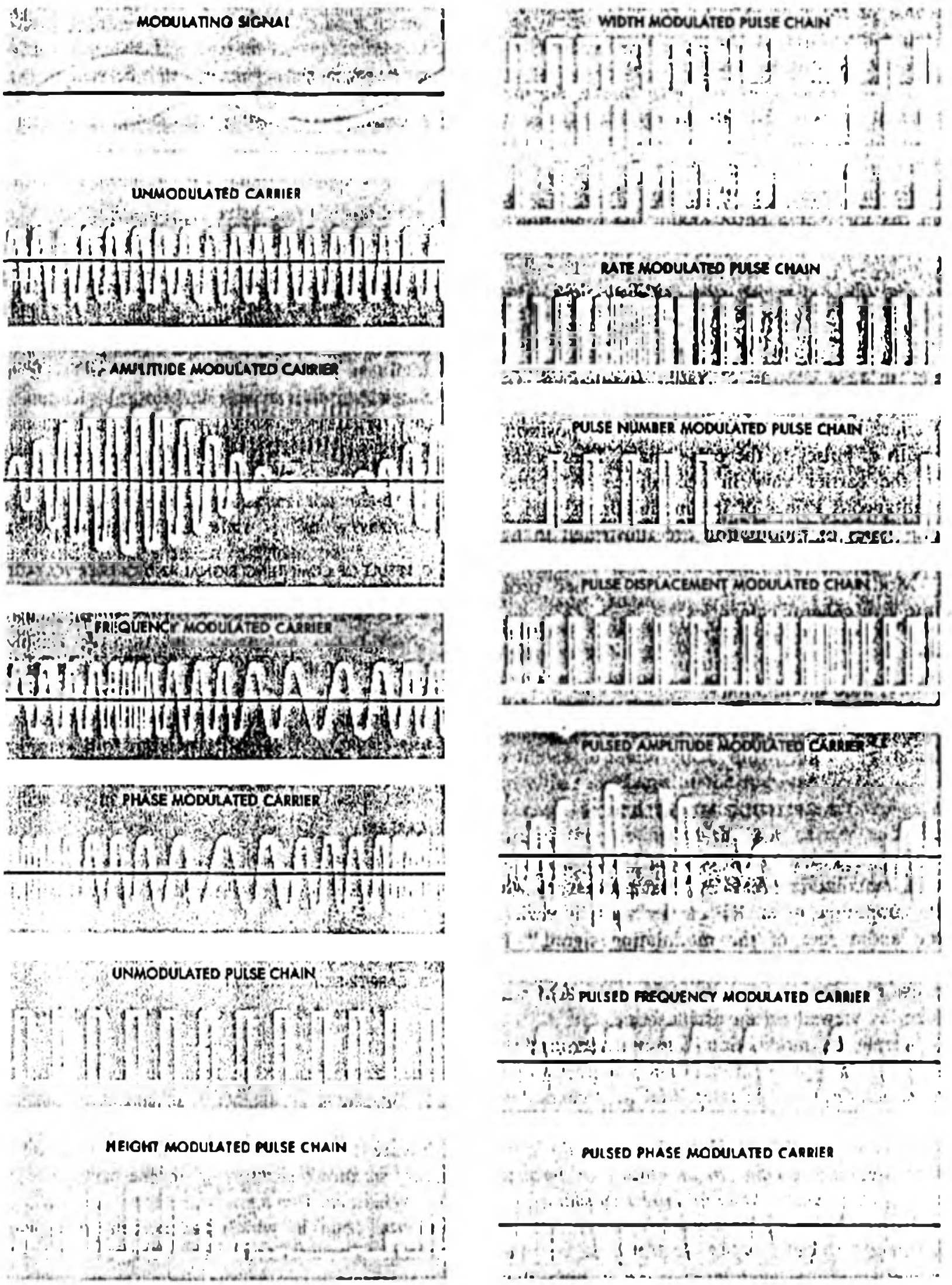


Figure 4-83. Types of Modulation

value of the modulating signal, the receiver can detect the variation (demodulate the carrier) and translate it into an intelligible form.

If the carrier is in the form of an alternating current or wave, its frequency must be greater than the highest frequency present in the modulating signal to include all of the intelligence or variations present in the signal. If the carrier is in the form of a pulse chain, the repetition rate of the pulses must be at least double the highest frequency in the modulating signal.

The frequency of the modulating signal itself is not present in the carrier at any time, but the intelligence produced by the modulation is present in new groups of radio frequencies (sidebands) above and below the unmodulated carrier frequency, or in changes in the time, position, or width of pulses in the pulse chain. These changes in the carrier vary in accordance with the instantaneous values of the modulating signal. Several types of modulation are illustrated in figure 4-83.

Modulation methods can be classified broadly into five general categories:

1. Amplitude modulation.
2. Frequency modulation.
3. Phase modulation.
4. Pulse modulation.
5. Doppler principle.

AMPLITUDE MODULATION

The definition of the popular concept of amplitude modulation (AM) is "that process by which the amplitude of an RF carrier wave is varied at the audio rate of the modulating signal." This definition has been accepted for many years, because it easily describes the composite AM waveform as viewed on an oscilloscope.

Amplitude modulation is more accurately defined as "a type of modulation that results when a modulating voltage is applied to a carrier voltage." It produces amplitude and frequency changes in the RF carrier. However, *the carrier itself does not change in amplitude or frequency during the process of amplitude modulation.*

The typical amplitude-modulated carrier is illustrated in figure 4-84A and B. The amplitude of an amplitude-modulated wave is high during the positive peaks of the signal voltage, but it is near zero during the negative peaks. The shape of this wave is the result of applying the carrier and signal voltages to a nonlinear device. A nonlinear device is one in which the current does not vary in direct proportion to the voltage.

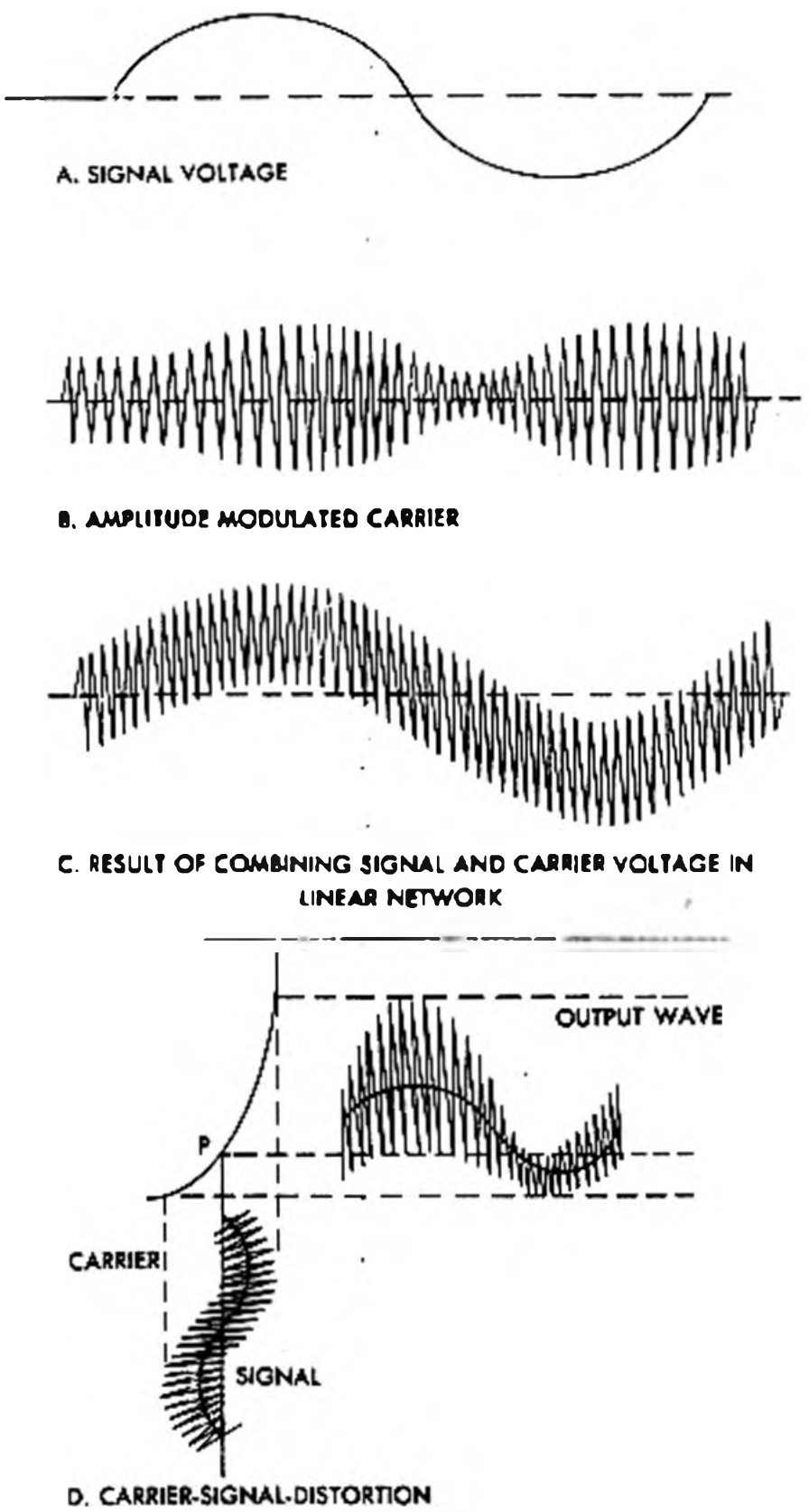


Figure 4-84. Basic Amplitude Modulation

Whenever two signals are impressed on a linear device (one in which the current does vary in direct proportion to the voltage), the output waveshape contains the same frequencies as the input and looks like the waveshape shown in fig-

ure 4-84C. In this case, the RF amplifiers in the transmitter amplify the carrier but discriminate sharply against the modulating frequency. This causes only the carrier to be transmitted. Such an arrangement would defeat the purpose of modulation, as the signal arriving at the receiver would not contain a modulation component at all.

When you impress the carrier and the modulating signal frequencies on a nonlinear network, the output waveshape, shown in figure 4-84D, is produced. The network used in producing this waveshape is a vacuum tube biased to operate on the nonlinear portion of its characteristic curve (point P). The input signal at D is the result of simultaneously impressing the modulating signal and the carrier signal onto the input circuit. The instantaneous input voltage varies around the bias value, as indicated by the input signal. The distorted output voltage produces the distorted output waveshape. This latter is the waveshape which would be seen on an oscilloscope.

The output of a wave impressed on a nonlinear network contains both the carrier and the signal frequencies. As you can see in the output waveshape shown in figure 4-84D, the negative half-cycles are altered differently from the positive half-cycles. A sine wave altered in this manner is a combination of the fundamental frequencies and additional frequencies caused by distortion. The distortion is introduced by the nonlinear resistance.

When a carrier frequency and a signal frequency are applied to a nonlinear device, the output wave contains the following:

- The carrier frequency.
- The upper side frequency, which is equal to the carrier plus the signal frequency.
- The lower side frequency, which is equal to the carrier minus the signal frequency.
- The signal frequency.

When an output wave containing these frequencies is present in a transmitter, the tuned circuits respond to the carrier, the upper side frequency, and the lower side frequency, but not to the modulating signal. The result of this frequency response is a composite waveshape, like the amplitude-modulated carrier shown in figure 4-84B.

Look at the composite AM waveform from a frequency spectrum standpoint. In figure 4-85, (1) represents the carriers; (2) the lower sideband (LSB); and (3) the upper sideband (USB). Assume the constant amplitude 100-kHz RF carrier is modulated by a voice-frequency signal that is varying in frequency from 100 to 6000 Hz and is also changing in amplitude. The output of this device will contain difference frequencies ranging from 94 to 99.0 kHz (LSB) and sum frequencies ranging from 100.1 to 106 kHz (USB). Notice that each sideband has the same frequency separation from the carrier. They contain the same intelligence, which can be extracted from either sideband at the receiver.

These sideband frequencies are generated as a result of the mixing of the RF carrier and the modulating signal in a nonlinear device. The radiated signal consists of three frequency com-

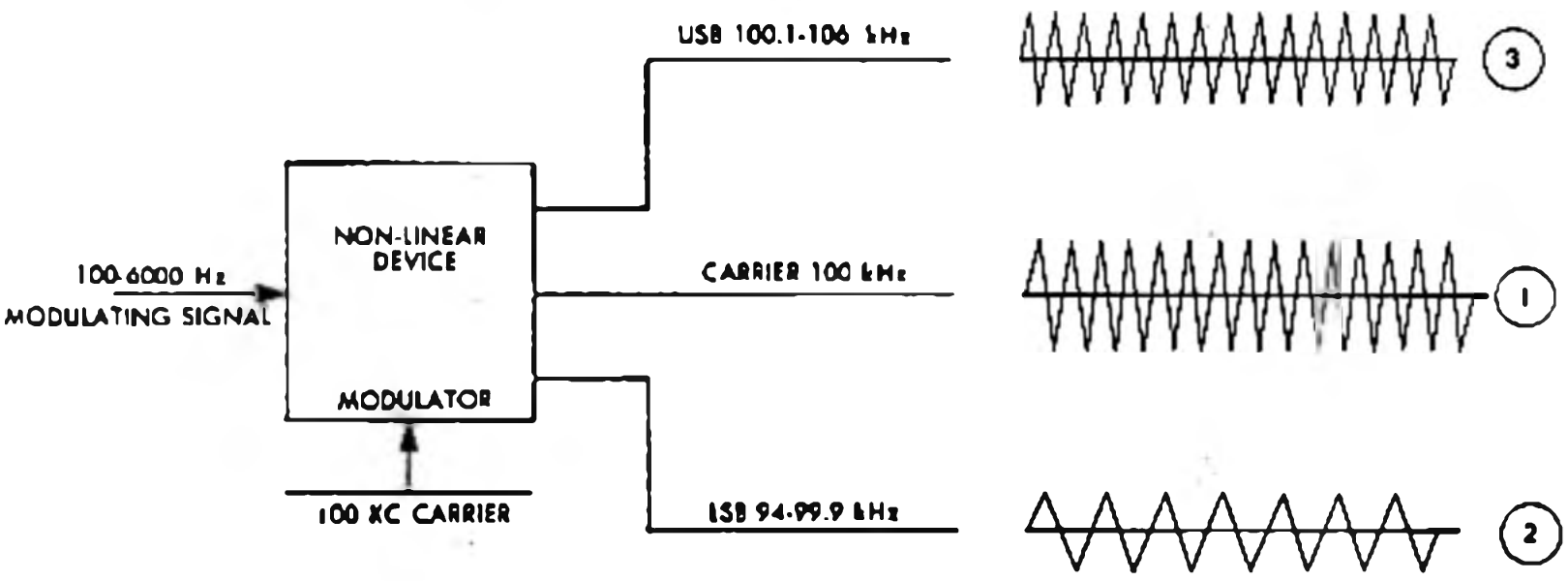


Figure 4-85. Amplitude Modulation

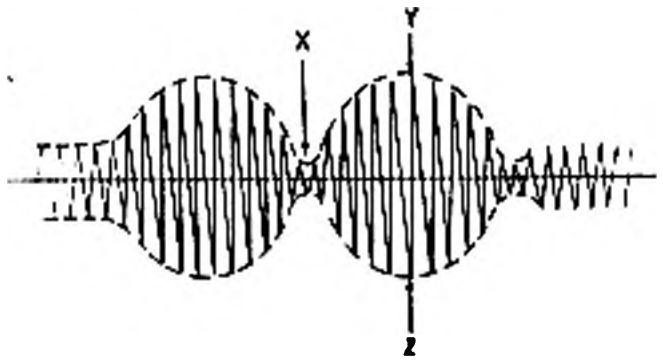


Figure 4-86. Composite AM Waveform

ponents (carrier, USB, and LSB) of a definite amplitude and frequency relationship.

When you view the radiated signal on an oscilloscope, you see a pattern referred to as a composite AM waveform. The outline of this pattern is called the modulation envelope. It will always resemble the original modulating signal if no distortion is introduced. A 100-kHz carrier modulated by a 6 kHz tone will appear as shown in figure 4-86. The composite AM waveform appears to be a radiofrequency carrier varying in amplitude. This waveform is the pattern of the instantaneous frequency-voltage relationship of the carrier and the sideband frequencies, or the vector sum of their three frequency components. It can also be identified as the overall resultant signal voltage of the phase relationship of the three frequency components at one given instant. The carrier, upper sideband, and lower sideband power and frequency are distributed throughout this waveform. That is, there is as much carrier, upper sideband and lower sideband power and frequency, at point X as there is at points Y and Z.

Percentage of Modulation

The amplitude of the sideband generated is always proportional to the amplitude of the audio or modulating signal. The amplitude of this modulating signal also determines the *percentage of modulation*. This term is used to describe the power distribution of the carrier and the sidebands; it determines the usability of the transmitted signal. One hundred percent modulation takes place when the system is delivering maximum undistorted power output. With less than 100 per cent modulation, the intelligence in the envelope is weaker. With more than 100 percent modulation, excessive distortion results. The percentage of modulation, therefore, is quite important. Just as sideband generation depends upon the audio or modulating signal, so also the percentage of modulation depends upon this signal.

When two equal voltages of different frequencies are applied to a nonlinear device, the voltage at the output, or plate, of the stage varies from twice its normal value to zero, depending upon whether the two voltages are in phase or out of phase. Because of the frequency difference between the two signals, it is impossible for them to be completely in or out of phase at all times. This explains the oscilloscope presentation of the composite AM waveform, where the carrier appears to be varying at an audio rate. The generated sidebands form a definite phase relationship with respect to the carrier. Once this relationship is formed, it must be maintained all the way to the detector in the receiver. Otherwise,

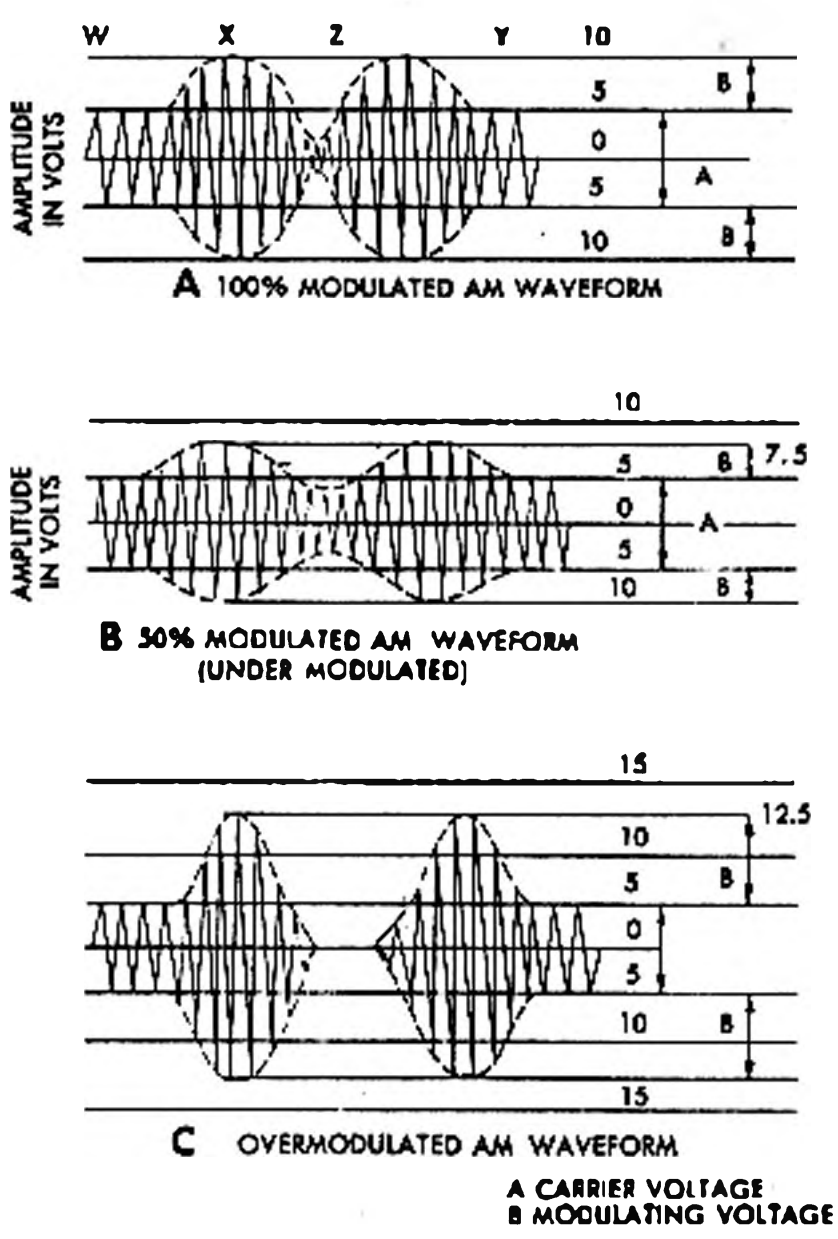


Figure 4-87. Percentage of Modulation

severe distortion will occur; and in some cases, complete loss of carrier or signal intelligence will occur.

Figure 4-87 shows examples of modulated AM signals. Figure 4-87A represents a 100 percent modulated wave. Notice the section labeled "W" is of constant amplitude. This is the carrier before it has been modulated. Since the carrier and modulating voltage are of equal amplitude, the carrier appears to increase in amplitude until at point X the amplitude of the composite waveshape is twice the unmodulated carrier. At Z the amplitude has decreased to zero. After one and one-half cycles of the modulating signal, the carrier is again unmodulated and of constant amplitude.

In figure 4-87B, the modulating voltage amplitude is less than that of the carrier. The composite waveshape does not increase as much as in the first example, nor does the amplitude ever drop to zero. Figure 4-87C shows a carrier modulated by a signal of greater amplitude. The peak voltage is greater than twice the unmodulated carrier, while the output drops to zero at an earlier point than in example A and remains there for a period of time. The envelope in figure 4-87C does not resemble the original modulating signal.

The percentage of modulation can be figured by using the following formula:

$$\text{Percentage of modulation} = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \times 100 \quad (1)$$

when E_{max} is the maximum voltage of the modulated wave and E_{min} is the minimum voltage of the modulator wave.

Another way of finding the percentage of modulation is this:

$$\text{Percentage of modulation} = \frac{\text{peak amplitude of modulating voltage}}{\text{peak amplitude of carrier voltage}} \times 100 \quad (2)$$

Examples:

Find the percentage of modulation for the waveshape in figure 4-87 using both formulas where applicable.

- For those in figure 4-87A:

$$\frac{E_{max} - E_{min}}{E_{max} + E_{min}} \times 100 = \quad (1)$$

$$\frac{10-0}{10+0} \times 100 = 100 \text{ percent}$$

or

$$\frac{\text{peak amplitude of modulating voltage}}{\text{peak amplitude of carrier voltage}} \times 100$$

$$= \frac{5}{5} \times 100 = 100 \text{ percent} \quad (2)$$

- For those in figure 4-87B:

Using formula (1).

$$\frac{7.5-2.5}{7.5+2.5} \times 100 = \frac{5}{10} \times 100 = 50 \text{ percent}$$

Using formula (2).

$$\frac{2.5}{5.0} \times 100 = 50 \text{ percent}$$

- For those in figure 4-87C:

Formula (1) cannot be used for overmodulation (greater than 100 percent modulation).

Using formula (2),

$$\frac{7.5}{5.0} \times 100 = 150 \text{ percent}$$

Power Distribution

In the composite AM waveform, there are two values of power to be considered—the carrier power and the peak power. *Carrier power* is the average power in the unmodulated carrier. This is point W in figure 4-87. *Peak power* is the power developed when the carrier and both sidebands are in phase. If the voltages are known, power in any of the components of the AM waveform can be calculated.

If you modulate a carrier 100 percent by a single sine wave of voltage of constant amplitude, the computations are not too difficult. Assume the root mean square (RMS) voltage of the unmodulated carrier is 100 volts and that the radiation resistance is 50 ohms. The power in the unmodulated carrier can be found from the power formula,

$$P = \frac{E^2}{R} = \frac{(100)^2}{50} = \frac{10,000}{50} = 200 \text{ watts}$$

Since the amplitude of the carrier remains constant, modulated or unmodulated, the power in the carrier is 200 watts at all times.

For 100 percent modulation, the modulating signal voltage is equal to the carrier voltage, in

this case 100 volts. Then the RMS value of the sideband voltage is 50 volts. The power in either sideband can be found from the power formula,

$$P = \frac{(50)^2}{50} = 50 \text{ watts}$$

The power in each sideband is one-fourth the power in the carrier at 100 percent modulation. The total sideband is 100 watts. With a carrier power of 200 watts, this gives a total power of 300 watts. This means that to attain 100 percent modulation of an RF carrier with a sine wave modulating signal, a modulating power equal to one-half the carrier power is required. The additional modulating power is divided equally between the upper and lower sidebands.

The peak power for this 100 percent modulated wave can be found as follows: The RMS value of E_{max} is 200 volts. Using the power formula,

$$P = \frac{(200)^2}{50} = \frac{40{,}000}{50} = 800 \text{ watts}$$

Single-Sideband Modulation

While all of the intelligence is contained in the sidebands, two-thirds of the total power is in the carrier. It would appear that a great amount of power is wasted during transmission. If the carrier and one of the sidebands were eliminated, the one remaining sideband could be used to transmit the intelligence. Since the carrier amplitude is unchanged, it contains no intelligence. Also, both sidebands could be transmitted, each containing different intelligence, with a suppressed or completely eliminated carrier. Transmission with one sideband eliminated is referred to as single-sideband transmission (SSB).

POWER COMPARISON. The total power output of a conventional AM transmitter is equal to the carrier power plus the power output of the modulator. Conventional AM transmitters are rated in carrier power output. In single-sideband operation, there is either a greatly attenuated carrier or a completely eliminated carrier. For this reason, the SSB transmitter is rated in peak envelope power. In the example of the 200-watt transmitter previously used, the peak power is 800 watts. If this same transmitter is converted to SSB operation, it is rated as an 800-watt single-sideband transmitter; and a 6-decibel gain in power would be realized. (Any power doubled equals a 3-decibel gain.) This gain results from the elimination of the carrier and one of the sidebands.

The bandwidth of the 800-watt SSB transmitter would be only one-half that required for a conventional AM transmitter. This is an advantage that results in a 3-decibel gain at the receiver. This gain is due to the better signal-to-noise ratio obtained. Another advantage realized is that the sideband signal is less subject to the effects of selective fading, which is caused by the out-of-phase relationship between the carrier and sidebands which commonly occurs in conventional AM transmission. In this type of single-sideband system, a theoretical 9-decibel gain is realized over the conventional AM system.

TYPES OF SIDEBAND TRANSMISSION. Several types of single-sideband systems have been developed. It is necessary to clarify the various applications of the transmission principle involved. The term "single-sideband modulation" is defined as a form of amplitude modulation in which one sideband and the carrier are suppressed.

In one system, the carrier and one of the sidebands are eliminated at the transmitter; only one sideband is emitted. This is called single-sideband suppressed carrier (SSSC). It is quite popular with amateur radio operators. The chief advantage of this system is maximum transmitted signal range with minimum transmitter power.

Another system eliminates one sideband and suppresses the carrier to a desired level. The suppressed carrier is then used at the receiver for a reference; automatic volume control; automatic frequency control; and, in some cases, demodulation of the intelligence-bearing sideband.

The type of system most commonly used in Air Force communications is referred to as twin-sideband suppressed carrier or independent sideband transmission. This system involves the transmission of two independent sidebands, each containing different intelligence, with the carrier suppressed to a desired level.

SIDEBAND GENERATION. The techniques of generating sidebands vary with different designs. In an SSB system, the sidebands are generated before the final stage. In conventional AM systems, the sidebands are generated before the final power amplifier stage. Also in the AM system,

the fundamental or carrier frequency is produced by an oscillator and then applied to a series of multiplier stages until the final operating frequency is obtained. At this point, the RF carrier is modulated by the audiofrequency intelligence and applied to the antenna.

The basic principle of SSB requires the suppression or elimination of the RF carrier and one of the sidebands. Therefore, specially designed circuits perform these functions at a low-powered level. Modulation at the low level necessitates the use of linear power amplifiers rather than frequency multipliers to keep distortion to a minimum. The most common method used to generate sidebands is frequency conversion coupled with selective filtering.

Since the primary objective is to convert the audio intelligence into a sideband frequency, a means must be devised to eliminate the carrier and one of the sidebands. This is easily accomplished by using a balanced modulator circuit. The main characteristic of a balanced modulator is that the carrier or conversion frequency is canceled in the output of the modulator. The output contains only the upper and lower sidebands of intelligence. The undesired sideband may be eliminated by using a selective filter, leaving only the desired sideband for transmission. This signal is applied to linear amplifiers and further converted until the final operating frequency is obtained.

Broadcasting audio frequencies up to nearly 10 kHz is possible by special methods such as single-sideband transmission, but certain difficulties are centered in such systems. These difficulties generally restrict their application in commercial broadcasting. If the carrier frequency is not raised and lowered simultaneously and inversely with modulating frequency, the modulated carrier overlaps into the adjacent channels when the modulating frequency exceeds 5 kHz.

The undesirable effect of swinging that occurs in amplitude modulation has led to the development of other types of modulation. Frequency modulation is one of the types developed.

FREQUENCY MODULATION

Frequency modulation is the process through which intelligence is transmitted by varying the frequency of the RF carrier in accordance with the frequency of the modulating signal.

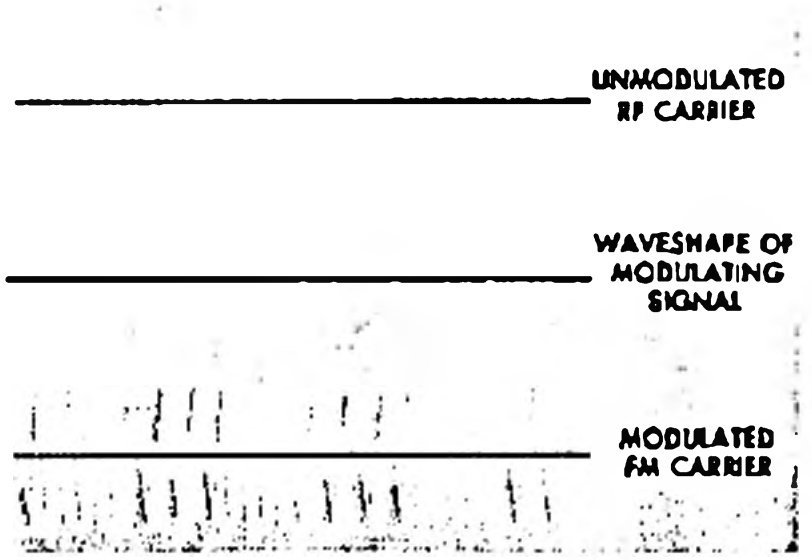


Figure 4-88. Graphic Representation of Frequency Modulation

In the frequency modulation (FM), the frequency of the RF carrier when it is not being modulated is referred to as the *center frequency.*

When the carrier is modulated by a positive signal voltage, its frequency increases in proportion to the amplitude of the positive signal voltage. And when the carrier is modulated by a negative signal voltage, its frequency decreases in proportion to the amplitude of the negative signal voltage.

Figure 4-88 is a graphical representation of frequency modulation. It illustrates an unmodulated carrier in which each RF cycle occupies the same amount of time, a sinusoidal modulating signal, and a modulated carrier.

The maximum frequency change from the center frequency is governed by the amplitude of the modulating signal and is called the *deviation.* While the modulating signal does not vary the amplitude of the carrier, it does shift the frequency and concentrate the power in new frequency sidebands.

Rate of Deviation of a Signal

A sinusoidal signal voltage causes an FM transmitter to radiate a signal which changes in frequency from the carrier center frequency to a higher frequency, back to the center frequency, then to a lower frequency, and back to the center frequency in accordance with the frequency of the

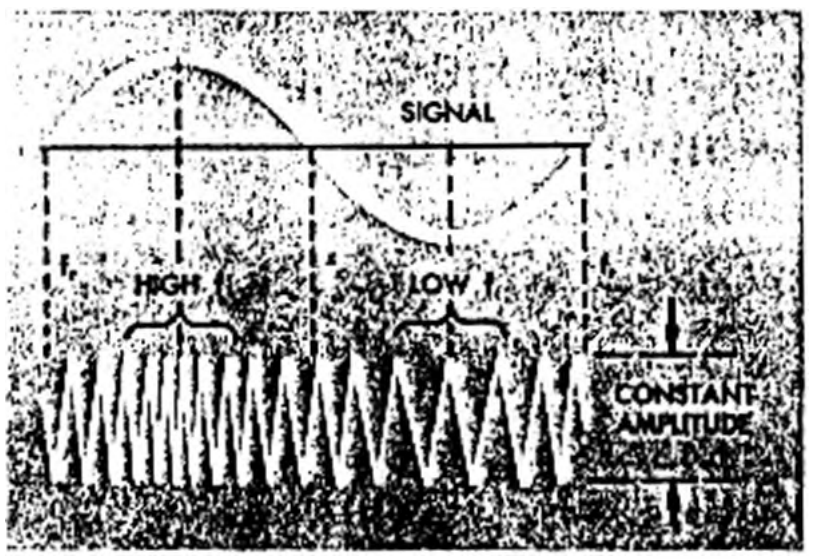


Figure 4-89. Carrier after Frequency Modulation

modulating sine wave. This process, known as the *rate of deviation*, is illustrated in figure 4-89.

An FM transmitter circuit which is designed to produce a deviation of 40 kHz can produce signals that differ from the center frequency by amounts much greater than 40 kHz. The circuit can do this because in addition to the normal upper and lower frequency sidebands which are produced by the amplitude of the modulation frequency, other sidebands differing from the center frequency by multiples of the modulating frequency also are present.

The spread of these frequencies becomes greater with the higher modulation frequencies such as are required in carrier telephone transmissions. This characteristic is inherent in the circuit employed. The strength of the emission on frequencies beyond the normal deviation, generally, does not decrease in a manner directly proportional to the frequency separation from the center frequency. The strength of the sidebands at frequencies considerably remote from the center frequency may be greater than that of frequencies close to the frequency of maximum deviation.

These sideband frequencies can cause interference on other radio channels which may be quite remote in frequency from the interfering transmitter. This interference is most likely to occur with transmitters using very high indexes of modulation.

Modulation Index

The modulation index of an FM transmission is comparable to the percentage of modulation which is applied to amplitude-modulated signals, even though the index is quite different in character. The *modulation index* often is defined as the "ratio of deviation of the carrier frequency in relation to the frequency of the modulating signal." It can be expressed by the equation:

$$\beta = \frac{\text{frequency deviation of the RF carrier}}{\text{frequency of the modulating AF}}$$

where β (beta) = modulation index

The limiting modulation index is called the

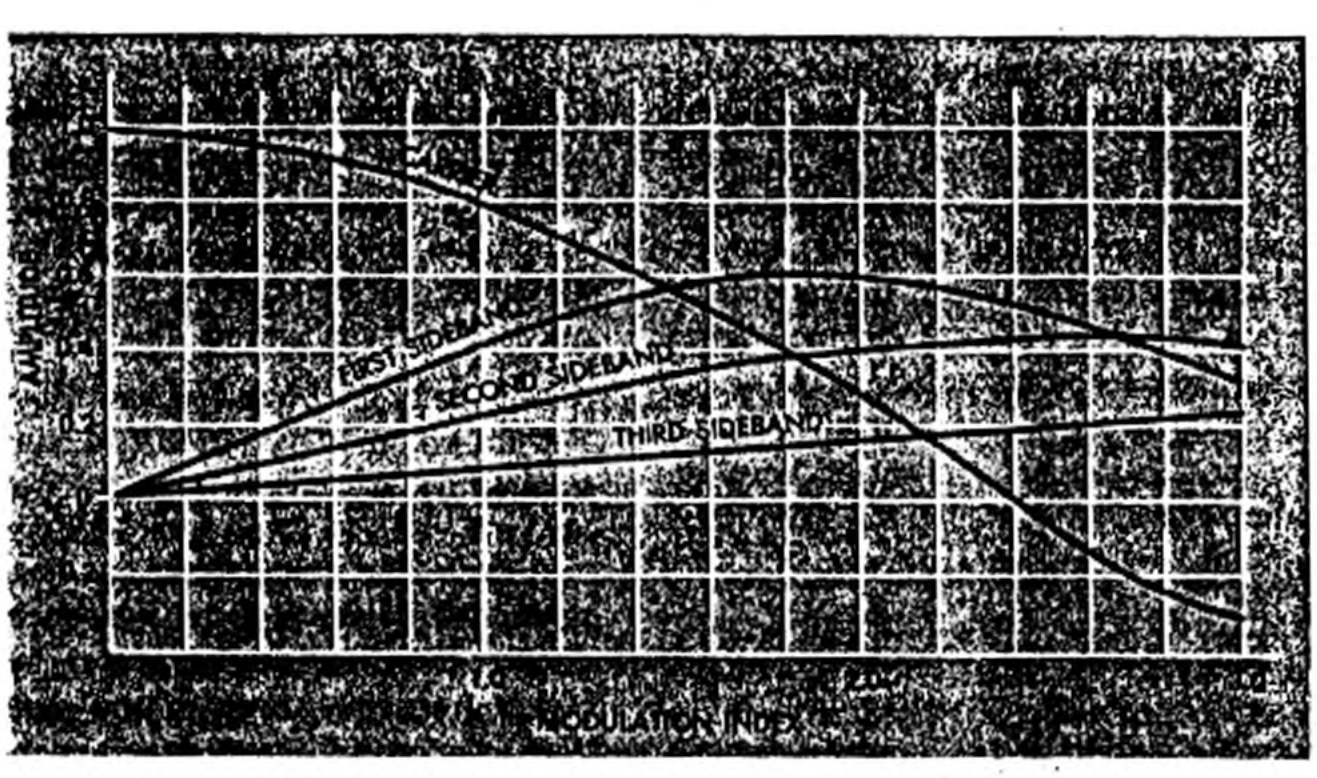


Figure 4-90. Variation of Signal Strength with Modulation Index

deviation ratio, which is the ratio of the maximum carrier frequency deviation to the highest modulating frequency used.

In amplitude modulation the carrier strength is constant and only the sideband amplitude varies, but in frequency modulation the carrier strength varies with the modulation index, as shown in figure 4-90.

At a modulation index of approximately 2.4, the carrier disappears entirely. It then reverses phase (as compared to its phase without modulation), thus becoming negative. It remains negative until the modulation index is increased to approximately 4.6, at which point it again passes through zero, reverses phase, and becomes positive once more.

If the curves shown in figure 4-90 were extended to higher indexes of modulation, additional sidebands would be developed, and the carrier would continue passing through periodic reversals in phase. The carrier would continue crossing zero points at specific indexes of modulation. The mathematical solutions for this behavior are too complex to be treated here, but they can be visualized if you consider a number of sine waves of different frequencies adding and subtracting at center points to cancel or reverse the phase of any particular wave.

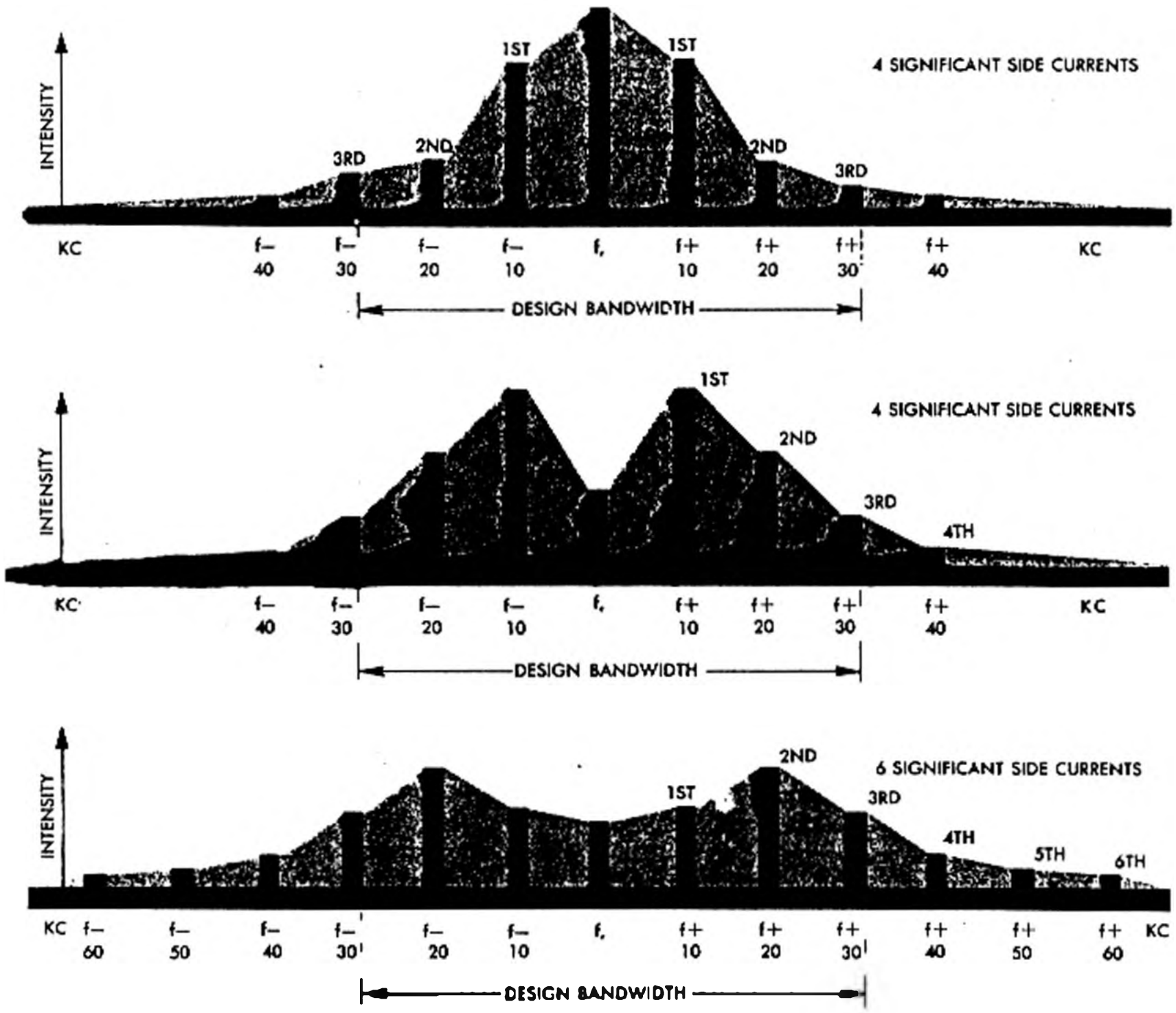


Figure 4-91. Spectrum Distribution for Modulation Index

In frequency modulation and phase modulation (PM), which is discussed next, the energy that goes into the sidebands is taken from the carrier. However, the total power in the carrier and its sidebands remains the same regardless of the modulation index. More power may be in the first sideband than in any other sideband at one value of modulation index; the greatest amount of power may be in the second or third sidebands at some other value of modulation index. Figure 4-91 shows the spectrum distribution for modulation index.

Both FM and PM systems possess the advantage of being relatively unaffected by atmospheric or manmade interferences which are essentially variable in amplitude. But both FM and PM radiation are more or less like light waves in character as a result of the high carrier frequencies employed.

FM is more generally used than PM because it is capable of handling more intelligence (producing more variations in frequency) within a channel of specific frequency limits.

The same receiver can be used for both FM and PM because the discriminator or *frequency detector* responds to change in phase in the same manner that it responds to variations in frequency.

PHASE MODULATION

Phase modulation is an indirect method for obtaining a frequency-modulated carrier. It is accomplished by passing the RF and AF signals through various phase-finding networks which tend to change the phase of the AC radio-frequency carrier voltage.

Thus, the phase of a carrier cycle shifts with respect to each preceding cycle, and the last cycle is either ahead of or behind its normal phase. In other words, the carrier cycle reaches its peak value either earlier or later than it does in the unmodulated carrier. This phenomenon is illustrated in figure 4-92 in which the normal period is shown as t_1 and the longer period as t_2. The longer period corresponds to a lower frequency, f, since $f = 1/t$.

Theoretically, the only difference between frequency-modulated waves produced by PM and FM systems is in the phase relationship between the modulating signal and the deviation. But,



Figure 4-92. Phase Modulation of a Carrier

practically, there are other differences. One difference arises from the fact that with phase modulation, the carrier frequency is much more stable than with frequency modulation. The frequency is more stable in the phase-modulation system because the oscillator can be crystal controlled and the modulating frequency applied to a later stage. In the FM system, the modulating signal must be applied directly to the oscillator circuit. Since the oscillator is designed to be resonant at the center frequency, it offers a different impedance to frequencies above or below the center frequency; therefore, oscillator output varies somewhat in amplitude at frequencies other than the center frequency.

Another basic difference between FM and PM is that in the FM system the deviation is determined only by the amplitude of the modulating signal, but in the PM system the deviation is proportional to both the frequency and the amplitude of the modulating signal.

PULSE MODULATION

The increasing demands for communication services in both military and commercial applications have led to the development of new types of radio and radar systems, which in turn have necessitated new methods of modulation for the transmission of intelligence in varied and complex form. Pulse modulation with its many possible variations has proven most practical in many of these applications.

Fundamentally, pulse modulation differs from other types of modulation in that the intelligence to be transmitted is sampled during brief periodic intervals, and these samples are used to modulate the carrier. The carrier is varied in some manner

in accordance with the instantaneous value of the modulating signal at the moment of sampling. The carrier is a chain of pulses which are uniform in nature and generated at a fixed rate. The carrier is modulated by a pulse chain from the sampled data which serves as a subcarrier.

Pulse modulation is valuable for applications which require multiplexing or simultaneous transmission of more than one data signal on a common carrier, such as telemetering, wire or microwave television cables, oceanic cables, and others. Pulse modulation permits the transmission of many types of data in short periods of time and with a minimum of equipment.

Simultaneous transmission of multiple channels is accomplished by one of two systems. One system involves using a *separate subcarrier frequency* for each channel, and the other involves transmitting data samples from each channel in a *specific time sequence*.

In the first system, known as *frequency division*, each data signal modulates the subcarrier assigned to its specific channel and is identified by the frequency of the subcarrier. In the second system, known as *time-division multiplexing*, the instantaneous amplitude of the signal is sampled from one channel at a time and transmitted in a regular sequence until all channels have been sampled. The process is then repeated in the same sequence until all desired data has been transmitted.

The nature of the data to be transmitted determines the bandwidth required for the transmission and frequency of sampling. For voice transmission, the rate of sampling must be sufficiently rapid to prevent the listener from detecting the intervals between the increments of data transmitted. A sampling rate of approximately 8000 samples per second is sufficient to produce the effect of normal continuous sound to the listener.

Since only one instantaneous value of the modulating signal is transmitted through one channel at any given instant in the time-division multiplexing system, there is no cross-talk or interchannel modulation. This interchannel modulation might occur in the frequency-division system because of the nonlinear frequency response of the modulation amplifiers.

Pulse modulation readily lends itself to multiplexing by time-division because it must employ instantaneous sampling. When pulse modulation is combined with time-division, the system is known as *pulse-time multiplex* and possesses many characteristics which are desirable in a communication system for handling complex data. Some of these characteristics are as follows:

1. High signal-to-noise ratio, which is made possible through the use of limiting and clipping circuits.

2. The "on-off" nature of pulse modulation makes it adaptable to simple repeater systems for increasing range of transmission.

3. Elimination of the need for complex filter networks.

4. Adaptability and flexibility of application.

5. Freedom from interference due to interchannel cross-modulation.

Pulses have individual characteristics (parameters) with respect to height, width, duration, repetition rate, formation time, decay time, shape, and displacement from normal occurrence. These characteristics may be used singly or in combinations to provide a wide variety of pulse-modulation methods adaptable to the transmission of intelligence in many forms and degrees of complexity.

In pure pulse modulation, both the rate and timing of the signal sampling remain constant regardless of how the pulses in the chain are modulated.

When the pulse-modulated system is used for transmitting intelligence from several sources or data of a highly variable nature, the ratio of noise to signal may become excessive (as is true in systems using conventional types of modulation). Signal-to-noise ratio may be improved through the use of clipping and limiting circuits and by increasing the bandwidth, if possible, to permit wider separation of individual data channels. The transmission bandwidth depends upon pulse shape as well as pulse rate, and the method and degree of modulation.

The fact that in pulse modulation energy is transmitted only for short intervals makes possible the use of high peak power in the carrier and the use of magnetrons and klystrons to generate carriers of very high frequency.

When high-frequency carriers are used, the bandwidth can be made much greater and more data channels can be accommodated. The customary interruption of the carrier between modu-

lation pulses is unnecessary except to effect a slight saving in power consumption.

Of the several types of pulse modulation, pulse-amplitude modulation is covered first.

Pulse-Amplitude Modulation

When the amplitude of a pulse is modulated the transmission is referred to as pulse-amplitude modulation (PAM). Although easily applied, this system is not widely used because it requires equipment with linear characteristics. Too, it can operate at an average of only 50 percent of its maximum capabilities. Figure 4-93 illustrates a pulse-amplitude modulated pulse chain.

In PAM systems, either of two types of amplitude modulation are used. One of these is known as *unidirectional PAM*. This type employs pulses of one polarity. The other type is known as *bi-directional PAM* because it employs pulses which vary in amplitude above and below a fixed reference level. The pulses are positive and negative in amplitude with an average value equal to zero. Figure 4-94 shows two types of pulse-amplitude modulation.

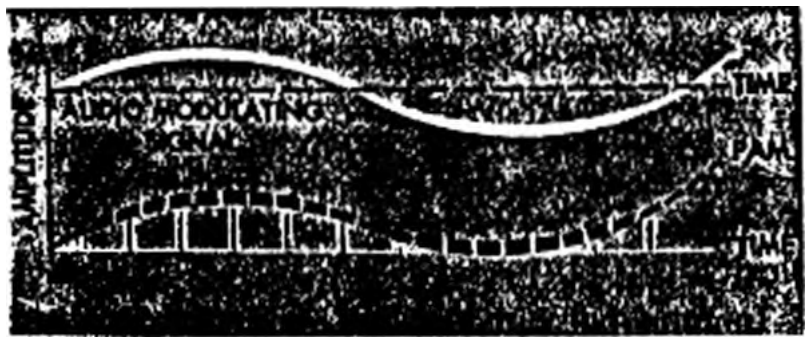


Figure 4-93. Pulse Amplitude Modulation

Pulse-Width Modulation

Pulse-width modulation (PWM) occurs when the width or duration of the pulse varies symmetrically with the modulating signal or when the leading or trailing edges of the pulses are modulated. Pulse-width modulation is commonly used on modulating and demodulating other pulse modulations. If PWM is differentiated or *gated*, *pulse-displacement modulation* is obtained. If a width-modulated pulse chain is passed through a low-pass filter, the signal is removed directly.

Pulse-width modulation is a convenient medium for transmitting intelligence, but it involves a variation of duty cycle that reduces the operating efficiency of the equipment. A combination of pulse-width modulation and pulse-rate modulation could be used to hold the duty cycle constant by

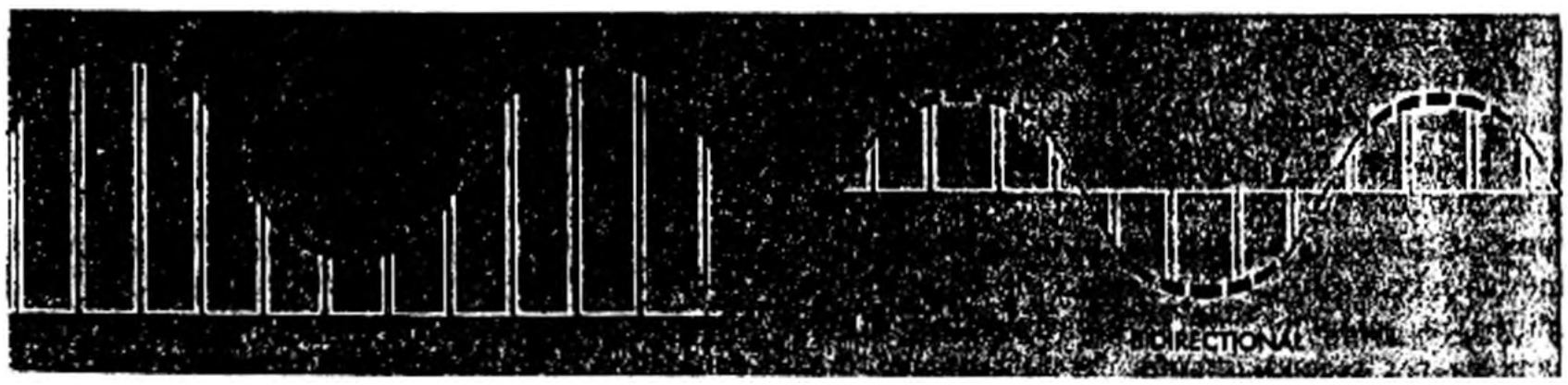


Figure 4-94. Types of Pulse Amplitude Modulation

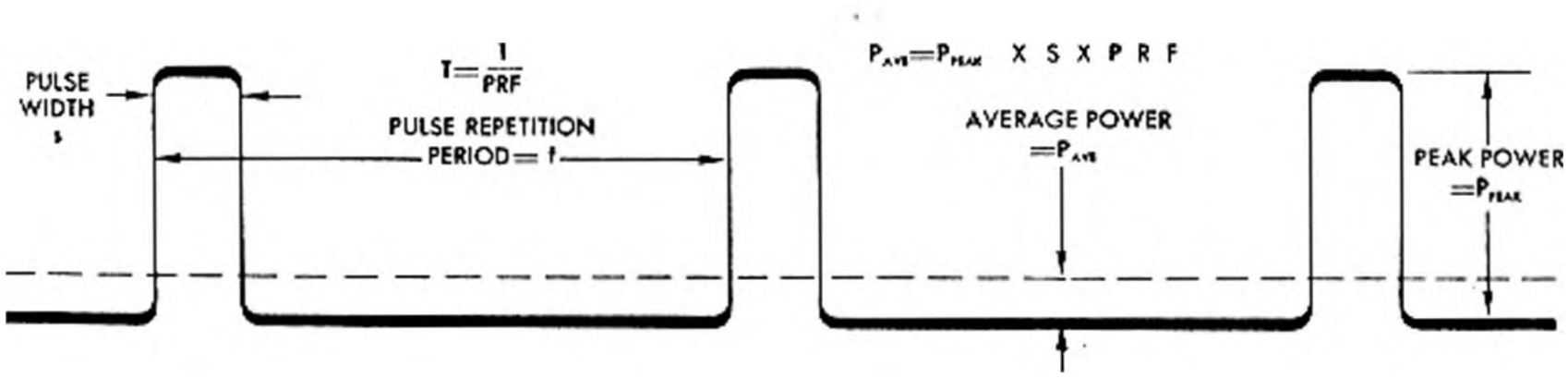


Figure 4-95. Definition of Pulse Characteristics

automatically decreasing the rate of wide pulses and increasing the rate of narrow pulses at rates inversely proportional to the width of the pulses.

The *duty cycle* is the ratio of the pulse length (duration) to the pulse-repetition period. To convert peak power into average power, peak power is multiplied by the duty cycle. These terms are illustrated in figure 4-95.

In pulse-width modulation, the variation of average power in the carrier is approximately the same as in pulse-amplitude modulation, but the amplitude of the pulses remains constant. The constant amplitude of the pulses facilitates the limiting and clipping of the pulses to eliminate extraneous noise, as shown in figures 4-96, 4-97, and 4-98.

Pulse-rate modulation also is subject to variations of duty cycle and loss of average power during the modulating cycle. It is not commonly used because of this drawback.

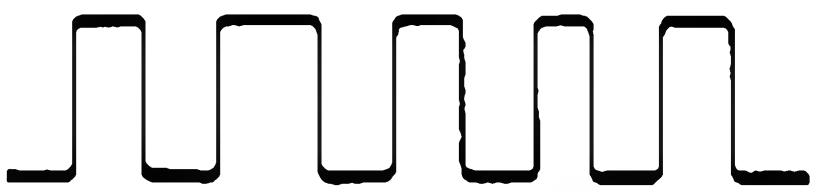


Figure 4-96. Original Train of Width-Modulated Pulses

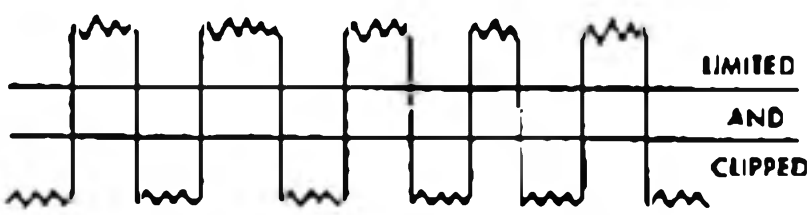


Figure 4-97. Pulses with Addition of Noise



Figure 4-98. Amplified Output of Limiter and Clipper

Pulse-Displacement Modulation

Pulse-displacement modulation, also called *pulse-time and pulse-position modulation*, is widely used in microwave time-division multiplex because it permits interlacing several pulse chains without confusion. Other pulse modulations offer similar possibilities if the extent of modulation is limited.

Pulse displacement modulation (PDM) is accomplished by varying the time between pulses or by varying the displacement of the signal pulse with reference to a *marker pulse*. A marker pulse is supplied from a separate marker generator, such as a free-running multivibrator. The marker generator modulates the carrier at uniform intervals with pulses of constant width and amplitude, or with pairs of pulses which are readily distinguished from those produced by the signal.

There are several methods by which the audio-frequency signal can be applied to produce pulse displacement modulation. One method employs a driven blocking-oscillator. The blocking-oscillator conducts when the positive-going portion of the audio signal voltage is applied to its control grid, or when the negative-going portion of the audio signal is applied to its cathode.

The positive and negative portions of the audio signal are applied successively to the grid and cathode of a single blocking oscillator through limiting or clamping circuits; or two properly biased blocking oscillators are used, one of which is biased to conduct only on the positive portion of the audio signal, while the other is biased to conduct only during the negative portion. In either case, the end result is the same. The blocking oscillator conducts and produces a modulating pulse when the amplitude of the input signal attains a predetermined positive and/or negative value. The frequency of the audio signal in this instance determines the frequency with which the modulating pulses are generated. When these pulses are superimposed upon, or used to trigger, a chain of carrier pulses which is being modulated at uniform intervals by a marker pulse (or pair of pulses), the position of the signal pulses varies with respect to the marker pulses in accordance with the frequency variations of the signal. The intelligence conveyed by the signal is represented in terms of the relative position between the signal pulses and the marker pulses, as illustrated in figure 4-99. In the figure, the audio modulating signal is represented by three sinusoidal waves each of a different frequency.

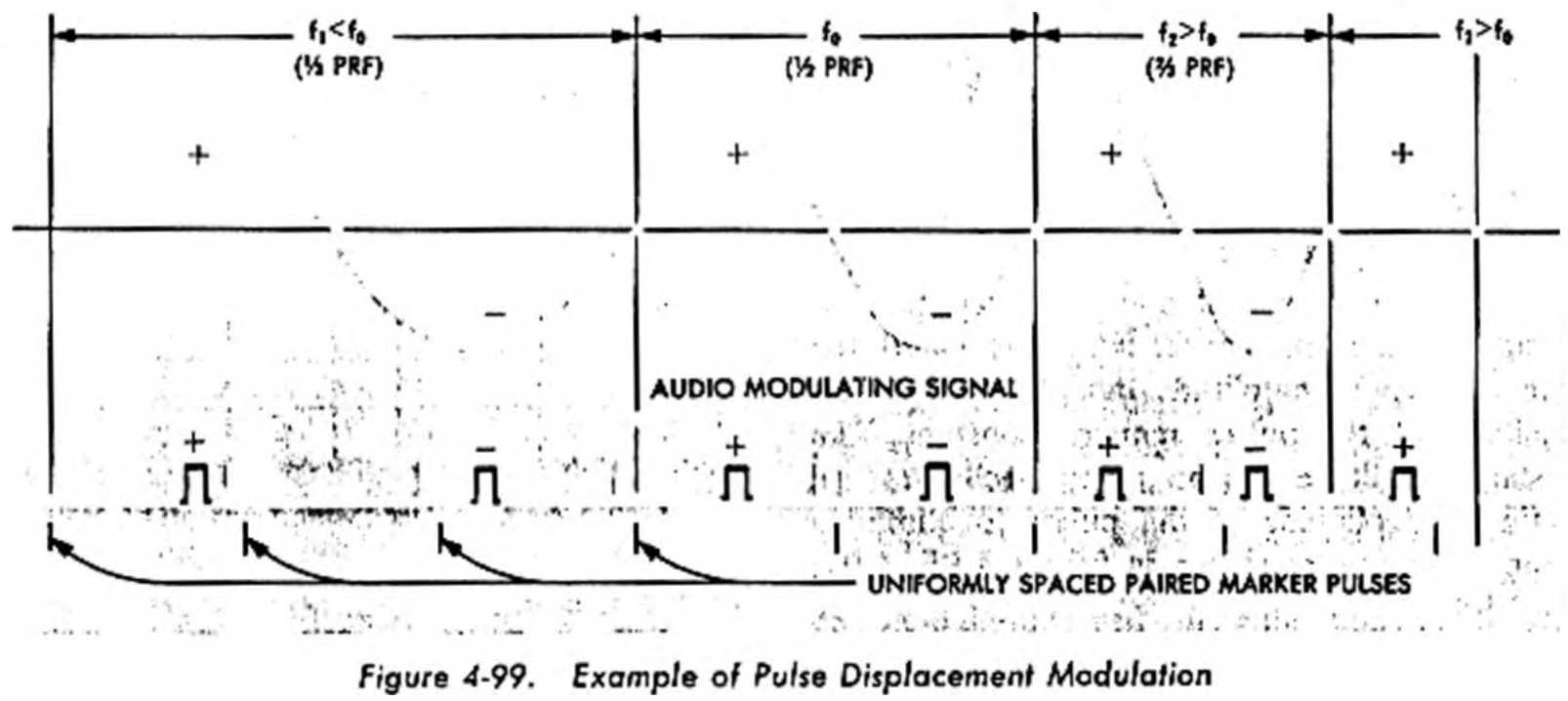


Figure 4-99. Example of Pulse Displacement Modulation

For illustrative purposes, the frequency of the middle sine wave is represented as being equal to one-half the pulse recurrence frequency(PRF) of the paired marker pulses. Each half of the audio cycle produces a pulse which is exactly midway between the adjacent paired marker pulses. The two signal pulses are separated by the same distance as the marker pulses.

The first sine wave represents a signal of lower frequency (one-third of the marker PRF in this example), and the distance between the signal pulses is correspondingly greater. No signal pulse appears between one pair of marker pulses. The third and fourth sine waves represent a signal of higher frequency (two-thirds of the marker PRF as shown). The distance between the signal pulses is correspondingly less, with two signal pulses appearing between one alternate pair of markers and one signal pulse appearing between the other alternate pair of markers. In other words, within two marker intervals three signal pulses appear.

If the audio signal were of the same frequency as the marker PRF, two signal pulses, one for each half of the audio cycle, would appear between each successive pair of markers, as shown in figure 4-100.

Pulse-displacement modulation can also be limited and clipped to reduce noise, but it gives less average carrier power output than pulse-width modulation for the same peak power.

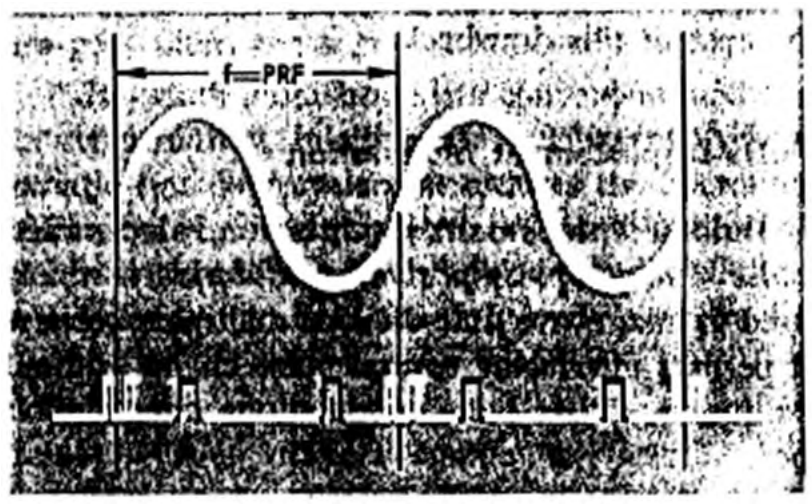


Figure 4-100. Pulse Displacement Modulation Produced When Modulating Signal Frequency is Equal to PRF

Pulse-Frequency Modulation

Pulse-frequency modulation (PFM) is a system in which the frequency of the carrier pulses is varied in accordance with variations in the amplitude or frequency, or both, of the modulating signal. The average power in the carrier for both pulse-displacement modulation and pulse-frequency modulation is fairly constant.

Pulse-frequency modulation is quite similar to pulse-displacement modulation except that no separate marker pulses are employed, and the variations of the audio signal are used to produce corresponding variations in the PRF of the carrier pulses. The development of frequency-modulated magnetrons makes the pulse-frequency modulation

system practical for many applications which require the transmission of complex data in which the signal varies greatly in frequency or amplitude within short intervals of time.

The principle of pulse-frequency modulation is illustrated in figure 4-101. Only the positive por-

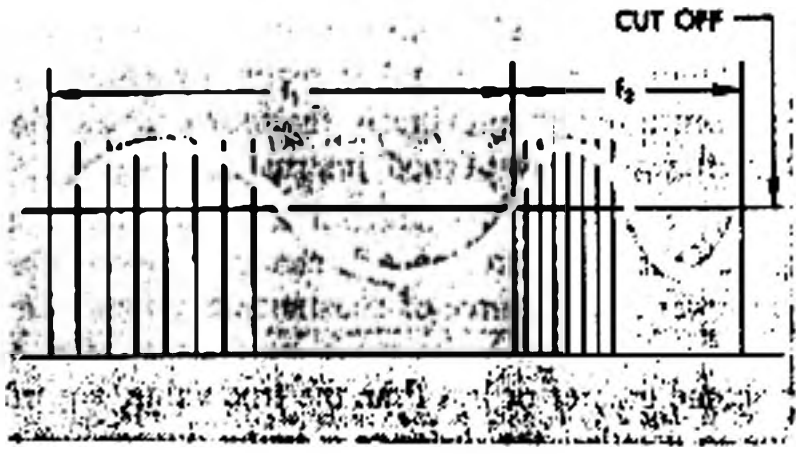


Figure 4-101. Pulse Frequency Modulation Produced by Audio Signals (f_1 and f_2, with f_2 equal to $2f_1$)

tion of the audio signal is used for producing the modulation which is in the form of groups of pulses. The frequency of the pulses contained in a group varies with the frequency of the audio signal.

In the example, only the positive half of each audio cycle is used to produce modulated pulse groups. This procedure is followed so that the groups can be separated by unmodulated intervals for ease of interpretation and decoding.

This type of pulse-frequency modulation is accomplished by applying the audio signal to a multivibrator-type modulator or to a magnetron oscillator circuit. In the case of the latter, an FM-CW-type magnetron modulation is employed. Modulation is accomplished in such a manner that the field strength of the magnetron (and consequently its frequency) is varied with the variation in frequency of the audio signal.

Multivibrator-type modulators are now most commonly used in pulse-frequency modulation systems, but are being supplanted by the FM-CW magnetrons wherever practicable because the latter are more reliable in operation and give an output with a much higher signal-to-noise ratio.

Pulse-Code Modulation

The variations in the modulating signal can be used to produce groups of pulses varying in number and concentration with respect to a *normal group* count without modulation. This process is called pulse-code modulation (PCM).

Pulse-code modulation is produced by using the signal voltage to vary the bias of a *keying circuit* which in its normal (no signal) unmodulated operation produces groups of pulses. The pulses are uniform in number and concentration. The keying circuit may contain a "1-shot" multivibrator.

The keying circuit is designed so that a positive-going signal voltage increases the number and/or concentration of the pulses in the normal group. Also, a negative-going signal voltage decreases the number and/or concentration of pulses in proportion to the variations in the amplitude or frequency, or both, of the audio signal.



Figure 4-102. Pulse Code Modulation

The principle of PCM is illustrated in figure 4-102. Pulse-code modulation is quite similar to pulse-frequency modulation except that in the PCM system the frequency of the carrier is not altered. Intelligence is conveyed by the number of pulses contained in each *signal-modulated group* as compared to the number of pulses in the *normal group*.

Pulse Modulation Thus Far

Some types of pulse modulation, such as pulse-amplitude modulation, can be detected by conventional circuits employing crystal-diode mixers and detectors. But more complex types, such as pulse-displacement, pulse-width, and pulse-code modulation, require special synchronizing and decoding circuits. These circuits contain delay and coincidence stages, integrators, and pulse-width and PRF discriminators in addition to the crystal-diode mixers and detectors. The circuitry and

principles of these special units are covered in detail under Coding and Decoding Systems, Chapter 6.

Crystal diodes have supplanted vacuum tubes as mixers and detectors in many UHF circuits because of their negligible transit time and better signal-to-noise output ratio.

In the conventional diode vacuum tube, the physical space between the cathode and anode is large compared to the wavelength of the UHF carrier, and the transit time required for electrons to pass from cathode to anode limits the frequencies at which the tube may function as a detector. Crystal diodes are relatively free from these limitations, and thermal noises generated within the crystal are negligible. A much higher signal-to-noise ratio may be obtained from the crystal, making it much more desirable for the demodulation of complex signals.

Many radar systems use 100 percent pulse amplitude modulation because it is the simplest method for modulating magnetron- and klystron-type oscillators, and it produces the highest peak power in the carrier. Also, at ultra high frequencies, the intensity of static interference becomes less appreciable, making the advantages of the frequency-modulated and phase-modulated systems less apparent.

Pulsed-phase modulation is another method for achieving the same result as obtained by pulse-frequency modulation. PPM is more suitable than PFM with equipment in which the signal pulses are applied directly to the carrier oscillator.

Although bandwidths are wide at ultra-high frequencies and selectivity normally would be high, the frequency drifts inherent in UHF equipment restrict selectivity and, consequently, limit the available bandwidth.

In whatever form pulse modulation may be used as dictated by the nature of the data to be transmitted, it provides the means for obtaining the fullest use of the bandwidths available with a minimum of interference from extraneous noises and intercarrier cross-modulation. For these reasons, pulse modulation is becoming more widely used in both military and commercial communication systems.

Before going on to the Doppler principle, look over table 4-5. It gives a compact review of three types of modulation.

DOPPLER PRINCIPLE

The Doppler principle involves the apparent change in the frequency of light, sound, or electromagnetic waves observed while the source and the observation point are in motion relative to each other. This principle is employed in radar equipment which supplies precise altitude, space-position, and velocity information.

The Doppler effect produces a frequency-shift modulation external to the transmitting equipment, as contrasted to previously discussed modulation which was accomplished internally.

In 1842, Christian Johann Doppler of Prague stated that if the distance is changing between an observer and a source of constant vibrations (such as sound or light) the wave number appears to become greater or less than the true value, according to whether the distance is being diminished or lengthened. This effect is commonly observed in the change of pitch of a train whistle as the train approaches and then continues on by.

Electromagnetic waves possess many of the characteristics of light, particularly as their frequencies approach that of light. The Doppler principle is applied to electromagnetic radiations in connection with radio and radar equipment for determining velocity and distance of moving targets, such as guided missiles.

In one system used to track a missile, the transmitter operates at a constant and accurately known frequency, f. The beam from the transmitter triggers a beacon transmitter in the missile. The beacon transmitter, in turn, operates at twice the original frequency, 2f. The latter signal is picked up by ground receiving stations. At the same time, the ground stations receive the ground transmitter's signal and double its frequency internally through a frequency-multiplying stage. The two doubled frequencies are then mixed. The resulting difference frequency is proportional to the velocity of the missile. The difference is proportional to the velocity because the motion of the missile effectively changes the transmitted frequency by an amount proportional to the velocity of the missile. The velocity vector between the missile and any point on the ground can be accurately measured by means of a Doppler radio system.

By integrating the velocity values from the instant of takeoff, position vectors can be deter-

Table 4-5. Comparison of Modulation Methods

	Amplitude Modulation	Frequency Modulation	Pulse-Displacement Modulation
SIGNAL LEVEL FROM TRANSMITTER	Varies with level of modulation.	Remains constant during modulation.	Remains constant during pulses.
AMPLITUDE OF MODULATING VOLTAGE	Determines instantaneous change in signal level. The stronger the signal, the greater the instantaneous change in carrier level.	Determines instantaneous deviation in frequency from carrier resting frequency. The stronger the audio signal, the greater the frequency deviation of the carrier.	Determines instantaneous deviation in time of channel pulse from rest position. The stronger the audio signal, the greater the time deviation.
MODULATING VOLTAGE FREQUENCY	Determines rate of change of amplitude of RF wave.	Determines rate at which carrier frequency changes between high and low values.	Determines number of samples transmitted for each cycle of modulating voltage.
SIDE BANDS TRANSMITTED	Width of transmitted side bands determined by frequency of modulating voltage. Present general limit — 5 kc each side of carrier.	Width of transmitted side bands determined by the amplitude of the modulating voltage. Present limits in military are 40 kc each side of rest frequency. In addition, a 20-kc guard band is provided for separation of adjacent channels. Divided equally on both sides of resting frequency.	Width of transmitted band is determined by pulse width of transmitted pulses. Bandwidth is several times that required by AM or FM. The band width to transmit the same amount of intelligence may be approximated by formula: Bandwidth in mc = 1 pulse length in microseconds.
MODULATOR POWER	One-half plate power input to modulated stage.	Negligible enough to supply plate power loss in modulator tube.	Negligible enough to supply plate power loss in modulator tube.
CARRIER POWER	Final amplifier must be capable of supplying four times the rated carrier power on 100 per cent modulation peaks.	Final amplifier must be able to supply rated carrier power only.	Final amplifier must be able to supply rated carrier power only.
FREQUENCY LIMITATION	None.	Normal above 20 mc, practicable at 2 mc.	Above 1000 mc.
MODULATION SAMPLING	Continuous.	Continuous.	Periodic.

mined with great accuracy. If two or more stations are used, the position of the missile with respect to a fixed coordinate system can be readily established.

If instead of transmitting a continuous wave, the Doppler system transmits a sawtooth frequency-modulated wave, both the velocity and the instantaneous distance between the missile and the ground station can be determined simultaneously.

This principle is illustrated in figure 4-103. Note that the received wave from a distant moving object is shifted both to the right and upward with respect to the original transmission. Consequently, the beat frequency will be alternately very small and very large on succeeding half cycles. The sum of the two different values of beat frequency produced is a measure of the distance of the missile from the ground station, and the difference between the two values of beat frequency is a measure of velocity of the missile with respect to the ground station.

Frequency-modulated radar determines the distance to a reflecting surface by measuring the frequency shift between transmitted and reflected waves.

While the wave is traveling to the surface and back, the transmitter frequency is changing under the influence of frequency modulation. When the reflected wave arrives back at the transmitter, its frequency is slightly different from the frequency being transmitted at that instant. The transmitted and reflected signals are combined in a mixer-detector circuit. The frequency difference between them is developed as a *beat note*.

The frequency difference becomes greater as the distance between the transmitter and the reflecting object increases. Altitude and range distances may be obtained by interpretation of the *beat frequency* between the transmitted signal and the received signal at any given instant of time. Velocity data may likewise be obtained.

This FM Doppler system is capable of measuring short distances accurately, while a pulse radar system is limited by the width of the pulse. With a pulse width of only 0.2 microsecond duration, the minimum range is about 100 feet. The FM Doppler system is more practical for such applications as aircraft altimeters and telemetering where relatively short distances and velocities must be measured accurately. The pulsed radar system is suitable for long-range or high-altitude measurements, for which purposes it is commonly used.

To employ the FM Doppler system for long-range measurements would require lowering the rate of deviation of the frequency modulation. Lowering the rate would reduce its accuracy for shorter range measurements.

Time and Frequency Relationship in Modulation Envelopes

Figure 4-104 illustrates the time and frequency relationships existing when a triangular modulation envelope is used in an FM radar system. The transmitted and received signals combine to produce beat notes at audio frequencies.

The black sawtooth wave represents the transmitted signal, and the white sawtooth wave rep-

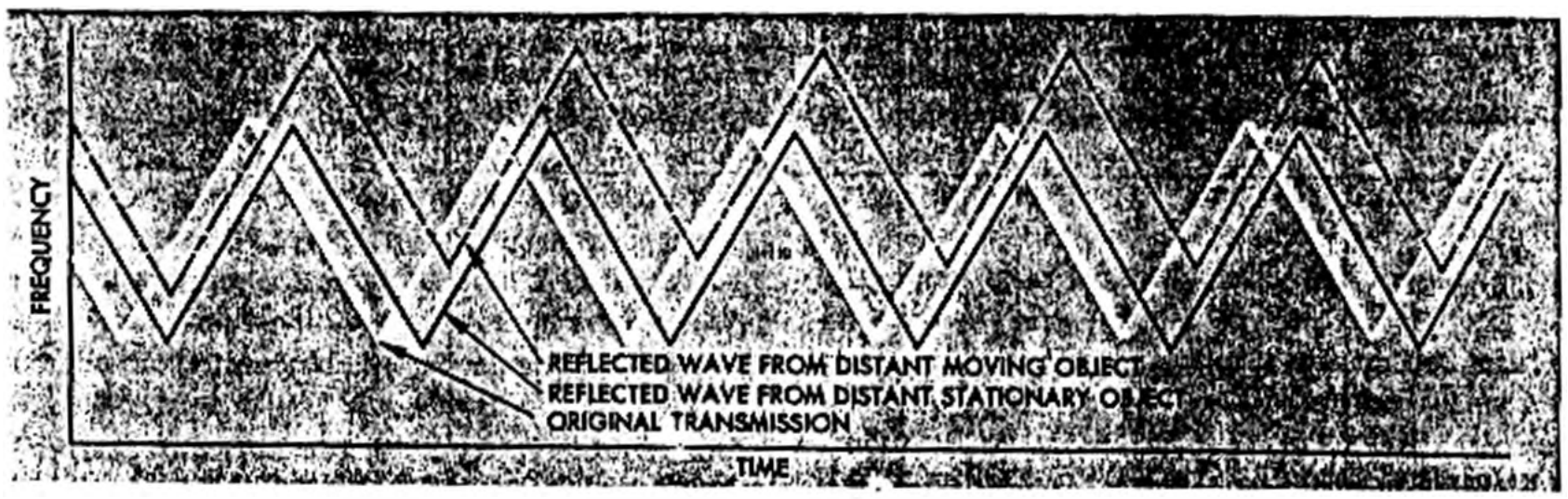


Figure 4-103. Doppler Effect on Frequency Modulation (sawtooth wave)

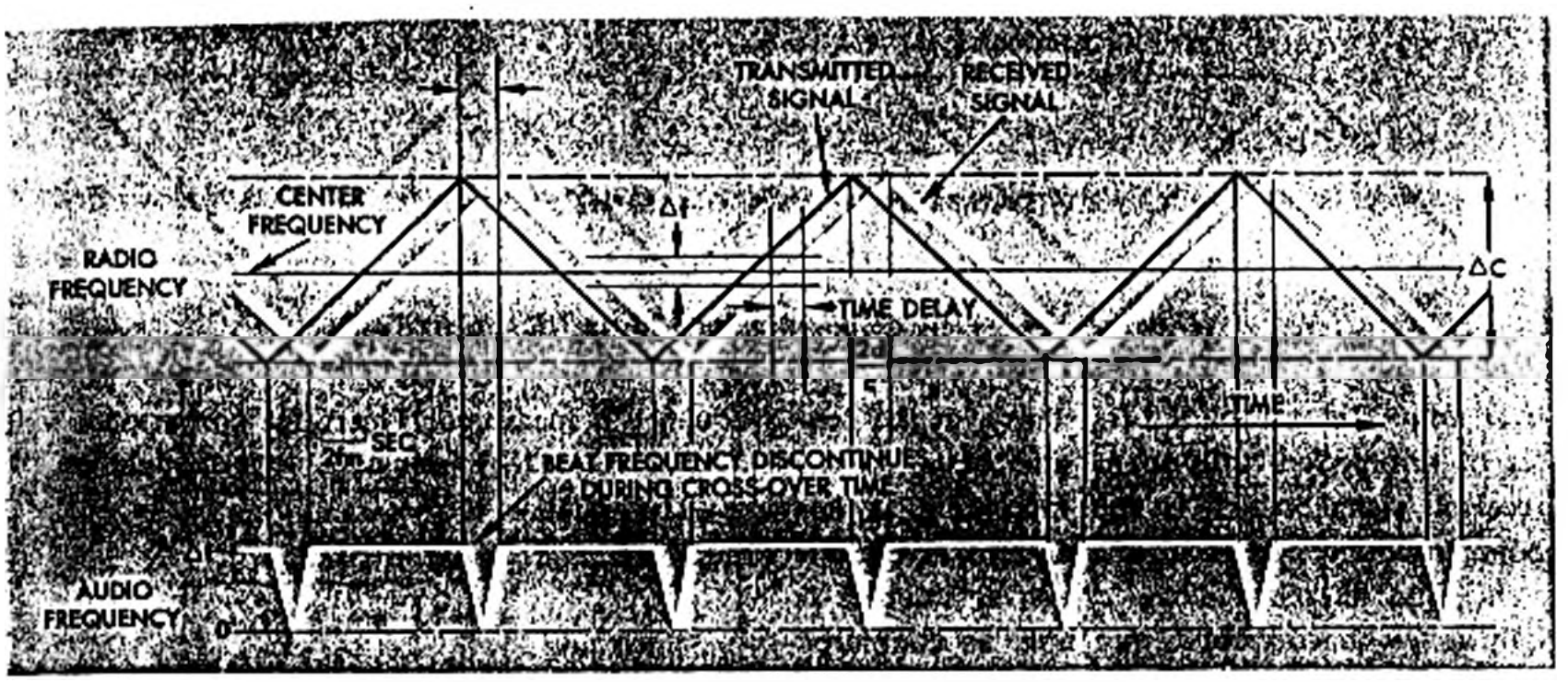


Figure 4-104. Graphic Representation of Time and Frequency Relationships

resents the received signal, both plotted as functions of time.

The total deviation of the frequency modulation (the peak-to-peak value of the modulation envelope) is indicated as Δ C megaHertz. If the center-frequency is 220 megaHertz and the deviation is plus or minus 400 kiloHertz (Δ C = 0.8 megaHertz), the carrier frequency then deviates from 219.6 to 220.4 megaHertz.

The received signal is frequency modulated by the same envelope as the transmitted signal because the signal preserves its form during reflection, but the received signal is displaced in time by the reflected interval. The reflected interval is equal to 2 d/c, where d is the distance in feet from the transmitter to the reflecting surface and c is the velocity of radio wave propagation in feet per second (about 984,000,000 ft/sec or 186,000 miles/sec).

As the time displacement occurs, a corresponding frequency displacement results. This is indicated by the vertical separation, Δ f, between the solid lines and the dashed lines on the graph.

The relationship between the distance, d, and the frequency difference, Δ f, can be determined by comparing similar triangles in the figure. The height of the triangular waveform is Δ C and half its base is ½fm, where fm is the frequency of the modulation envelope. The ratio of the height to half the base is 2 Δ Cfm, and this is equal to the frequency difference divided by the time difference:

$$2\Delta Cfm = \frac{\Delta f}{2d/c}$$

where Δ C, fm, and Δ f are all measured in Hertz per second. The relationship between frequency difference and distance is then found to be

$$\frac{\Delta f}{d} = \frac{4\Delta Cfm}{c}$$

Note that the sensitivity of the indication, in Hertz per second frequency-difference per foot, can be increased by using a wide frequency deviation or a high modulation frequency, or both.

If Δ C is 0.8 megaHertz, fm is 120 Hertz per second, and c is 984,000,000 feet per second, the ratio of frequency difference to distance Δ f/d is approximately 0.39 Hertz per second per foot of distance. The equation reads

$$\frac{4 \times 0.8 \times 10^6 \times 120}{984 \times 10^6} = 0.39$$

Thus, the maximum beat frequency developed from a reflecting surface 5280 ft. distant is 5280 X 0.39 or 2059 Hertz per second.

A small error occurs during the period between reversal of the transmitted frequency deviation and the corresponding reversal in the received wave. The period is shown in the graph as *cross-*



Figure 4-105. Time Relationship of 120 Hz Triangular Wave

over time. The transmitted frequency, at the center of this period, falls to a value equal to the received signal, which is still increasing at this point, resulting in a beat note of zero at this instant. The duration of the crossover interval is only a few microseconds per mile as compared with 4166 microseconds, the half-period of the modulation envelope at 120 Hertz per second; thus, the error of crossover effect is negligible. These time relationships are shown in figure 4-105.

If the modulation is in the form of a sinusoidal envelope as in figure 4-106, instead of a triangular envelope, the frequency difference between crossovers is no longer a constant. Instead, the difference varies from zero at the crossover to a maximum which occurs when the transmitted frequency is passing through its center value. By means of suitable circuits, the average beat frequency is measured. This value corresponds with that produced by triangular-wave modulation.

The equation

$$\frac{\Delta f}{d} \quad \frac{4\Delta Cfm}{c}$$

is used to predict the average beat frequency, Δf, when sinusoidal modulation is used.

Employing Doppler Principle in FM Radar Altimeter

The functional block diagram in figure 4-107 shows a typical airborne FM radar altimeter using the Doppler principle for determining altitude of aircraft.

A typical double-diode counter circuit, which can be used in the indicator counter listed in the preceding block diagram, develops a positive DC

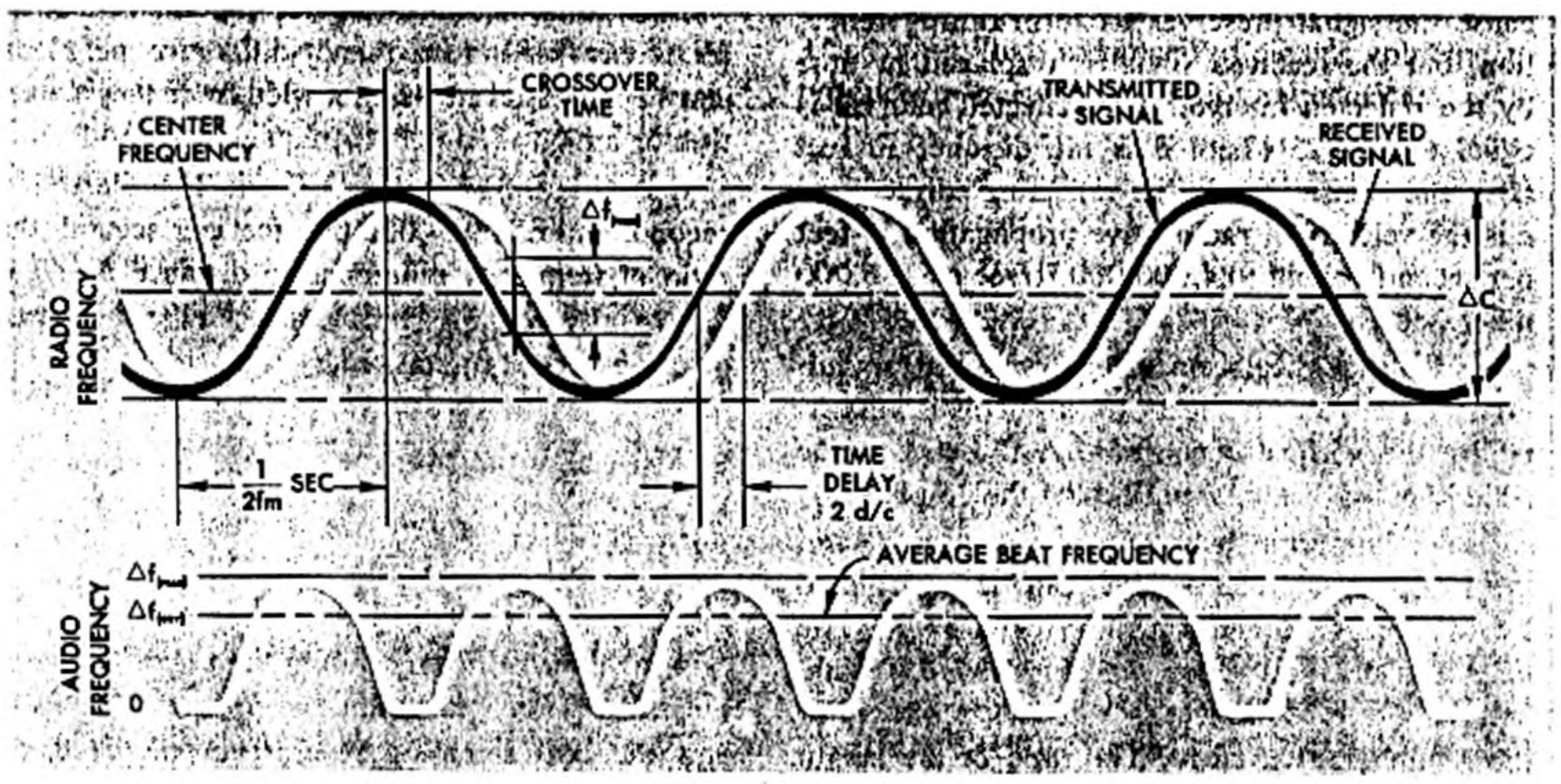


Figure 4-106. Beat Frequency Produced by Crossover

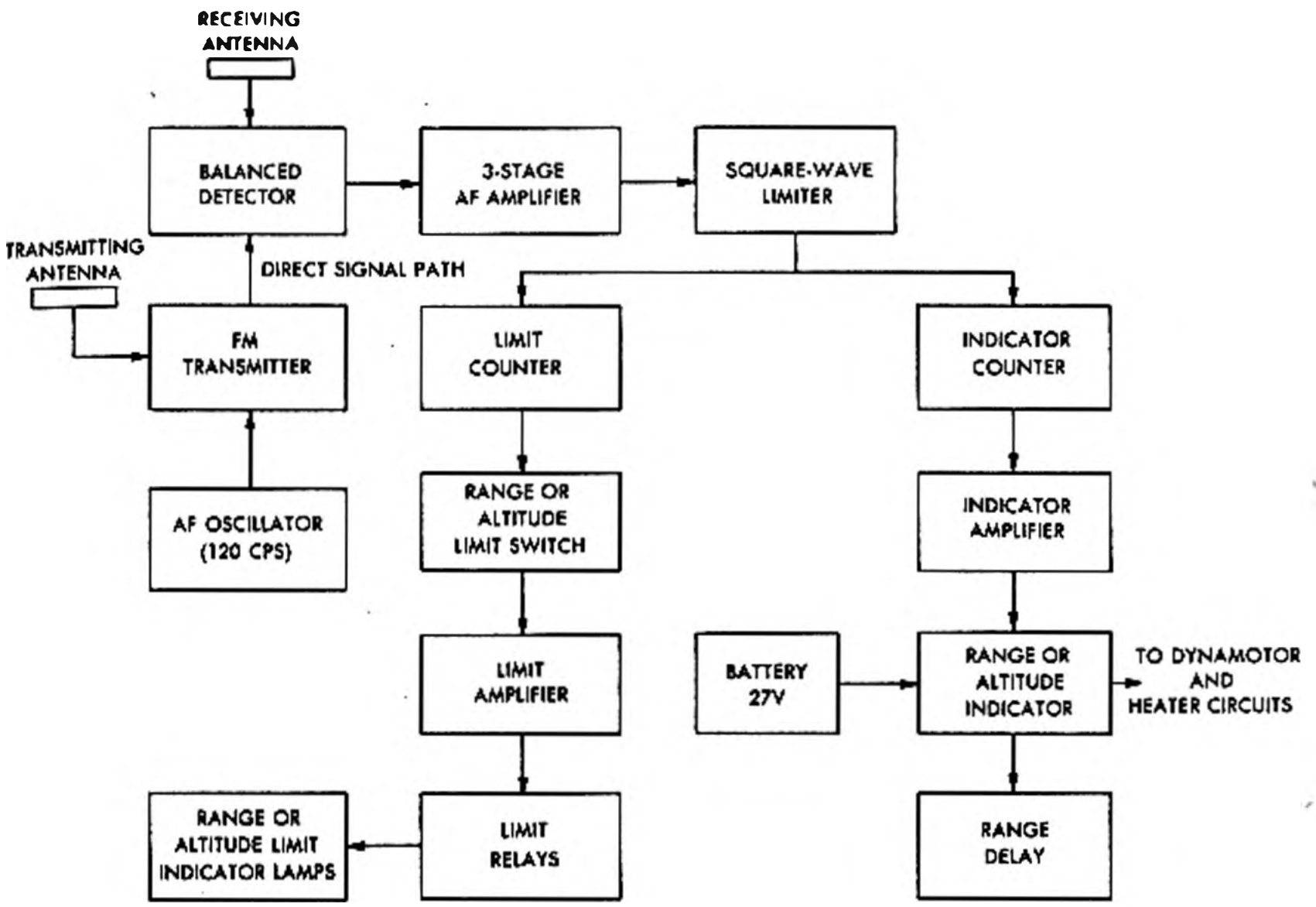


Figure 4-107. Typical Airborne FM Radar Altimeter

voltage proportional to the beat note frequency. This circuit shown in figure 4-108 includes a double-diode (12H6), one section of which passes the positive half-cycles of the limiter output, charging the 0.125 μf capacitor and load resistor in shunt. The values of capacitance and resistance are so chosen that the charge on the capacitor leaks off through the resistor at a rate approximately equal to the conduction of charge through the diode.

When the positive rectangular waves from the limiter arrive at an increasing rate (higher beat-note frequency), the direct voltage across the capacitor tends to increase. The voltage decreases when the waves arrive at a decreasing rate.

This voltage is passed through a low-pass RC filter which averages the direct voltage and applies it to the grid of the output amplifier tube (12SH7).

A milliammeter in the cathode circuit of the output amplifier tube registers, over a range of 5 milliamperes (1.5-6.5 ma), the average value of the direct voltage on the grid. This meter is calibrated directly in feet. The HI-LO scale shown is switched synchronously with the switch which controls the total deviation of the transmitted signal so that the meter scale corresponding to wide or narrow deviation range is always correctly selected.

Part of the voltage across the cathode resistance is fed back to the other section of the dual-diode (12H6) in the counter circuit. This diode passes the negative halves of the limiter output to ground. This discharges the coupling capacitor, leaving it ready to receive the next positive half of the limiter output wave. The feedback from the indicator amplifier also biases the diode and makes the indication more linear.

You can now recognize the importance of the Doppler principle to the missile field. As this section has brought out, the Doppler effect is widely used in electronic equipment.

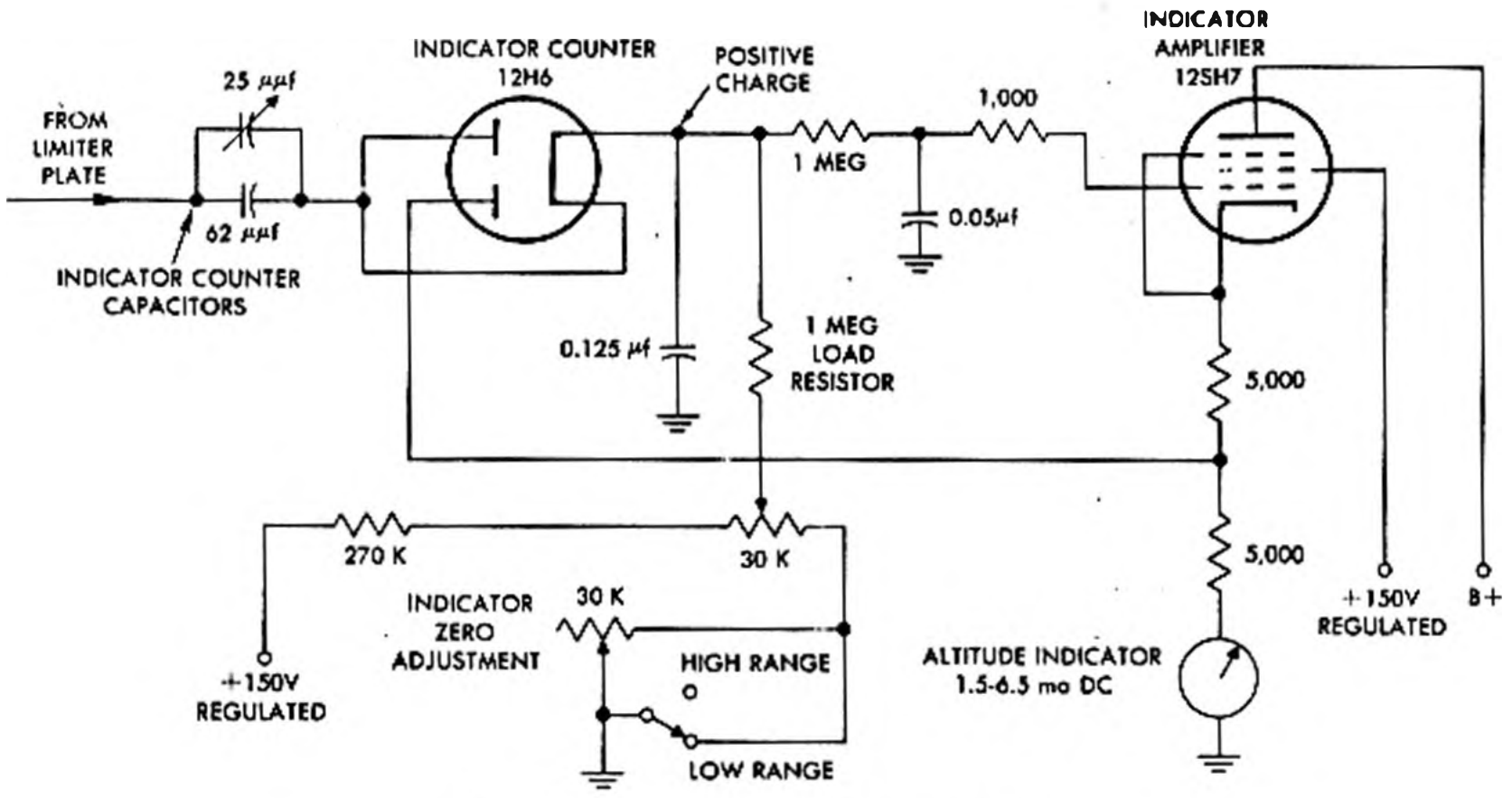


Figure 4-108. Double Diode Counter Circuit

A REASON FOR UNDERSTANDING MODULATION

Since a carrier must be altered in amplitude, frequency, or phase before it can transmit intelligence, such altering becomes vital to the effective use of electronics equipment.

You should be fully acquainted with the types and functions of modulation. You should understand the phenomenon of modulation before going into the next chapters which take up the detailed study of the components of a missile system.

CHAPTER 5

Control System Components

Devices which make up the various types of guided missile control systems are presented individually in this chapter. To help you visualize the relationship of these components in the control systems, a brief explanation of overall system operation is included. A later chapter presents a more detailed explanation of system operation.

Guided missile control system usually refers to a system that automatically controls the flight of the missile. The system is similar to the automatic pilot system installed in some piloted aircraft. The system moves controlling devices the proper amount at the proper time to keep the missile from erroneously pitching, changing heading, or rolling.

The principles of the missile control system are the same as any automatic control system. Such a system constantly makes corrections of some controllable item and then checks the results as a basis for further corrections. This forms a cycle of interdependent actions which is called a closed-cycle control system. Any block diagram of the actions is connected to form a closed loop.

There are four requirements of any automatic control system.

- The first requirement of a control system is the presence of something which is controllable. For example, it would not be feasible to attempt to control the temperature of the outside atmosphere.
- The second important requirement is a means of detecting or sensing a deviation of the controllable item from the desired condition or reference. This deviation is usually called error.
- A third function is to convert the error information into a form which can be used to regulate the controlling device.
- The controlling or correcting device is what actually effects the correction of the controlled item and is the fourth requirement of any system.

One of the most common automatic control systems is the one used to maintain a certain temperature in a room. A heating system consists of the four basic system requirements as shown in figure 5-1. The thermostat is adjusted for a certain room temperature. This adjustment represents the only command to the system. The temperature for which the thermostat is set then becomes a reference. If the temperature deviates from this reference temperature, the thermostat detects the deviation and sends an error signal to the valve controlling the fuel supply. Heat is produced by the equipment which obeys the sig-

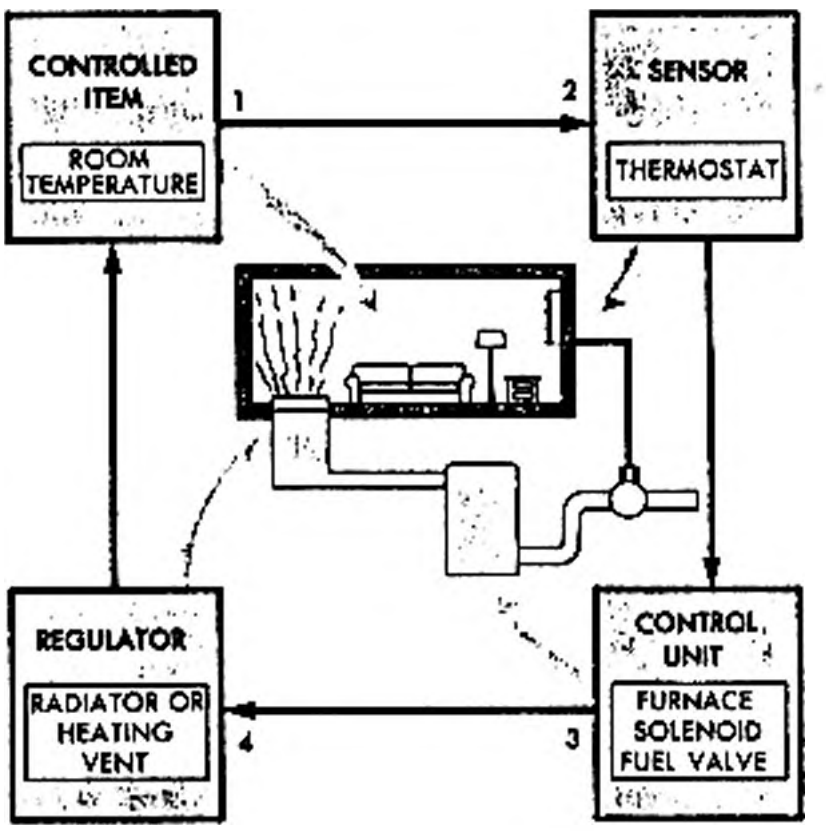


Figure 5-1. Basic Parts of Automatic Control System

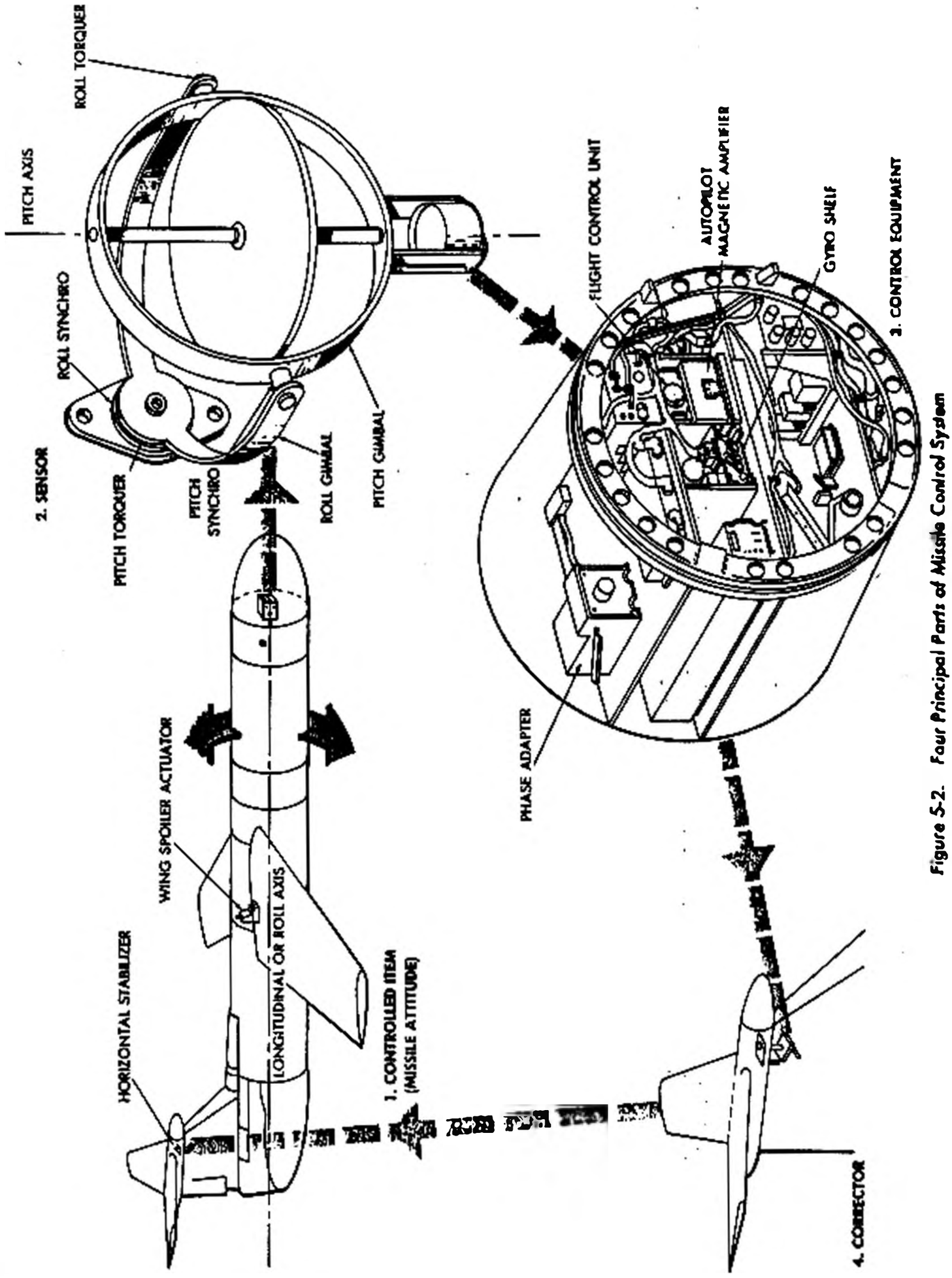


Figure 5-2. Four Principal Parts of Missile Control System

nals of the thermostat. The heat enters the room by means of a vent or radiator. Meanwhile, the thermostat, still measuring temperature, insures that the fuel supply is cut off at the proper time to maintain the reference temperature.

Figure 5-2 shows the same four requirements of an automatic system, applied this time to a guided missile flight control system. The missile attitude is a controllable item. Deviations may be sensed by a gyro. The controlling equipment changes this signal to a force that can move the control surface the proper amount. The device which actually imparts correction is the control surface (or movable jets, etc).

The series of events has not stopped. When the missile moves as a result of corrective action, this motion is fed back as information to the sensor, since it is continually detecting attitude. This feedback information completes the major loop of the system. This major loop is sometimes called the dynamic loop since it includes the motion of the guided missile.

Figure 5-2 also lists major functions of a guided missile system. Each component forms all or part of the four requirements of a missile system. These major functions are arranged in block diagram form in figure 5-3.

Basic control system units are all applied to one of the eight functions of the basic missile control system.

These functions are not necessarily fixed as shown in the block diagram. The diagram provides a basis for understanding component operations.

The control system used in a missile consists of a number of components which are installed as a part of the missile equipment and are interconnected to operate as a complete control system. The control system is designed primarily to provide stabilization of a missile during its flight from the launch site to the target site. There are also secondary control-system design considerations which are needed because of the extensive research and testing phase a missile must go through before it becomes an operational weapon. Test range safety requirements also play a part in the control system design of any specific missile.

The control system must be capable of moving the craft's control surface to maintain a given flight attitude and a fixed course or directional heading. The system must be so designed that an operator in a director aircraft or at a ground station can maneuver the missile, through remote control, to accomplish dives, climbs, or various other functions necessary to establish the aerodynamic characteristics of the craft. To accomplish this, the control system must be capable of doing the following:

1. It must set up fixed reference lines in space, from which deviations in course or change in flight attitude can be measured.

2. It must provide a mechanical or electrical and mechanical means of operating the control surfaces of the missile as required. Such a device is commonly called a servo.

3. It must provide a means for measuring the magnitude and direction of the angular deviations which the missile makes from these reference lines and which the servos make in the position of the control surfaces.

4. It must provide a means for translating the amount and direction of missile deviation and the amount and direction of control surface move-

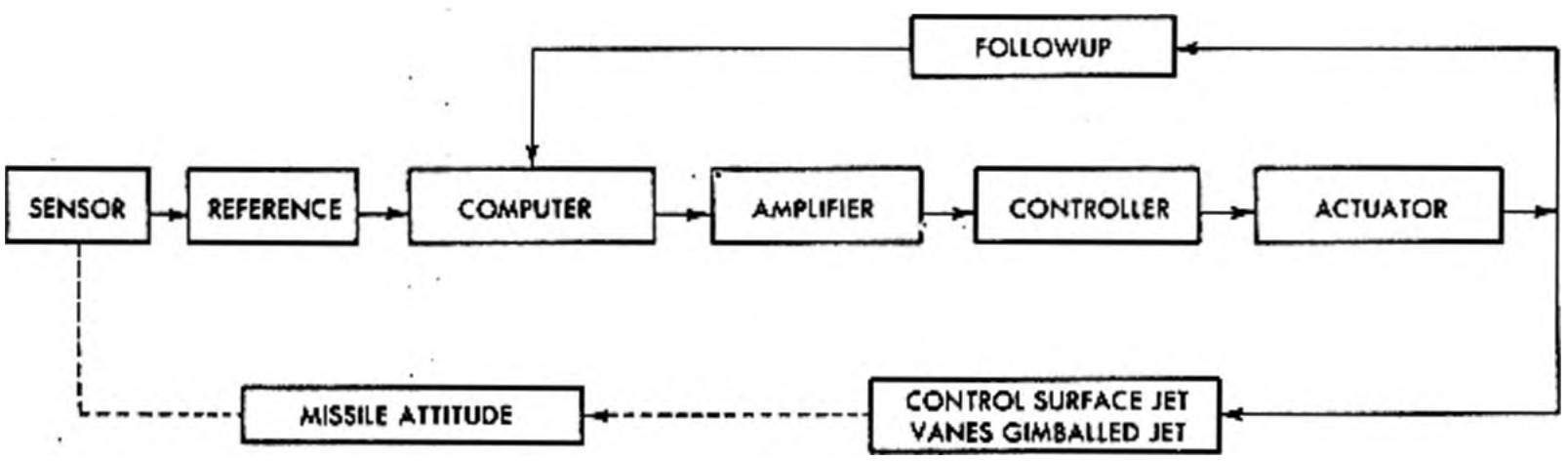


Figure 5-3. Block Diagram of Missile Control System

ment into the operation of a complete system so that the missile will perform properly.

The way in which a missile control system meets these requirements is covered in subsequent chapters. In this chapter the operation of the individual components and subsystems which comprise the makeup of a control system are explained.

SECTION A

Sensor Units

A sensor unit in a guided missile control system is a device capable of detecting deviation of the missile from a desired flight condition. Sensor units discussed in this chapter include gyroscopes, altimeters, and transducers. Gyroscopes generally are considered to be the basic sensor unit in any missile control system. Altimeters and transducers are used as sensors in secondary or auxiliary servo loops in the control system. Pickoffs, devices that make the detected intelligence useful, are also presented in this chapter.

GYROSCOPES: THE BASIC SENSOR UNIT

Before discussing the application of gyroscopes in a control system, first consider some basic gyroscopic terms and definitions.

A gyroscope is a mechanical device containing an accurately balanced rotor. The rotor spins about its central or spin axis, which passes through the center of gravity. A free gyroscope is so mounted that it can tilt or turn in any direction about this center of gravity. Figure 5-4 shows such a gyroscope. When the rotor is rotated at a high speed, it assumes the characteristics of a gyroscope, that is rigidity in space.

The characteristic of rigidity in space makes the gyroscope useful as a reference or sensor unit in controlling the flight of a missile.

Rigidity of Gyroscopes

Rigidity, gyroscopic inertia, or angular momentum, is that property of a gyroscope which resists any force tending to displace the rotor from its plane of rotation. Three factors determine a gyro-



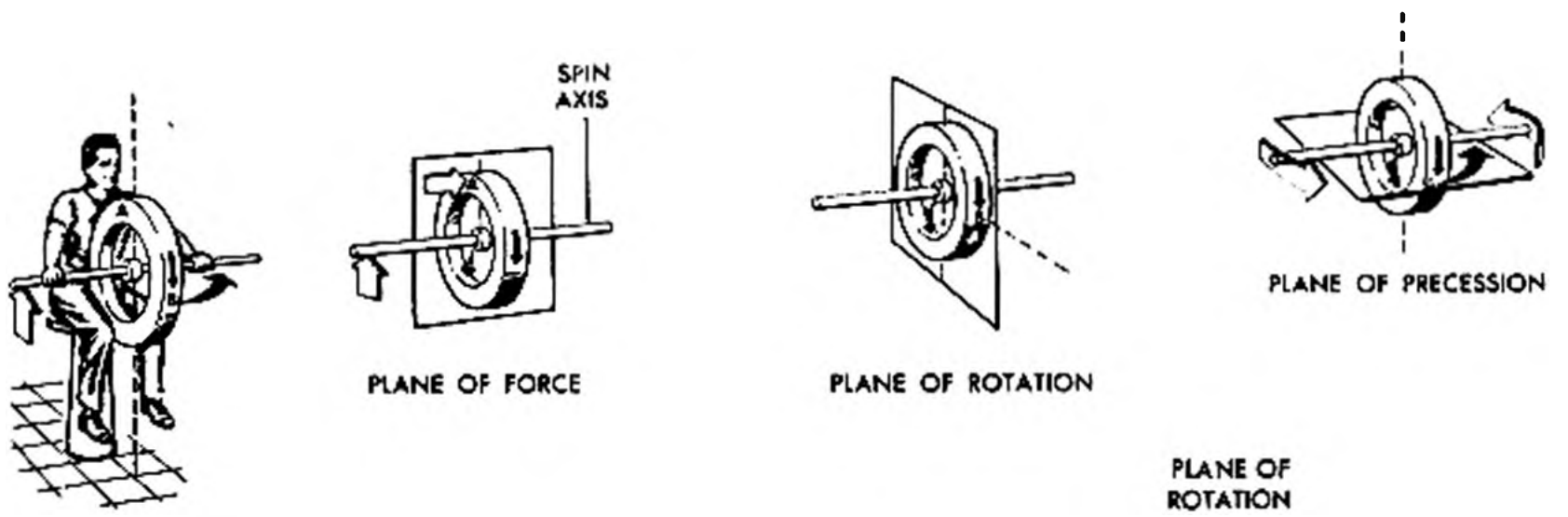
Figure 5-4. Free Gyroscope

scope's strength or amount of rigidity. These factors are (1) the weight of the rotor, (2) the distribution of this weight, and (3) the speed at which the rotor spins. Rigidity may be increased by adding to the weight of the rotor. A gyro with a heavy rotor has more rigidity than one with a light rotor if the speed of rotation is the same for both. Rigidity can also be increased if the weight of the gyro is distributed to the outer rim of the rotor, as far from the spin axis as possible, even though there is no increase in the weight of the rotor. Rigidity is also increased when the speed of rotation is increased. A slowly spinning rotor gives the gyro little or no rigidity.

Gyroscopic Precession

There are two types of gyro precession: real or induced precession and apparent precession.

REAL PRECESSION. Real precession is movement of the gyro spin axis from its original alignment in space. This precession is caused by a force applied to the spin axis and can be predicted if the point to which the force was aplied is known. Spin axis precession occurs 90° from the point of applied force, in the direction of rotor rotation,



By applying an upward pressure on the GYRO SPIN AXIS, a deflective force is applied to the rim of gyro at point A (PLANE OF FORCE). The resultant force is 90° ahead in the direction of rotation to point B (PLANE OF ROTATION), which causes gyro to precess (PLANE OF PRECESSION).

Figure 5-5. Real Precession

and in the same direction as the applied force (refer to figure 5-5).

An easy way to remember the direction in which a gyro will precess when an external force is applied, is by the hand rule shown in figure 5-6. Using either hand, place the fingers in the direction of rotation and extend the index finger in the direction of the applied force. The thumb will then extend in the direction of precession.

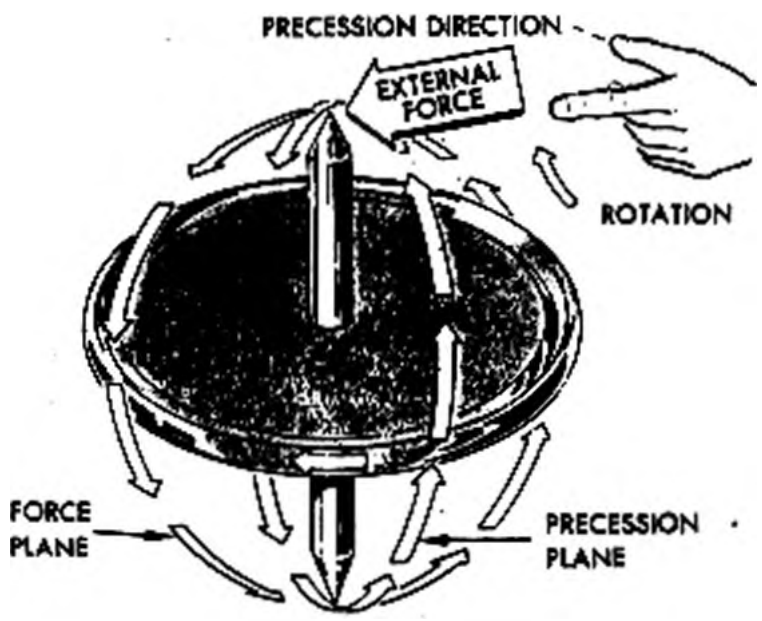


Figure 5-6. Hand Rule for Determining Direction of Gyro Precession

It is apparent from the preceding explanation of gyro precession that if a force were applied to the gyro at the center of gravity, it would not act to tip the gyro spin axis from its established position. No precession would take place. A spinning gyroscope can be moved in any direction, provided its axis remains parallel to its original position in space. The gyro provides stability only against tipping its spin axis. A spinning gyroscope can be used to provide stabilization only in planes containing its spin axis. For complete stabilization in aircraft, two gyroscopes, which have their spin axes at right angles to each other, are required. For this reason, both a vertical gyro and a horizontal gyro are needed to give complete stabilization and to set up the necessary reference lines from which deviation can be measured.

APPARENT PRECESSION. Because rigidity fixes the spin axis of a gyro in space, the axis points in a fixed direction. The earth, which is rotating, turns under the gyro. The axis of the gyro appears to tilt. Imagine a gyroscope at the equator with

the spin axis parallel to the equator. Apparent precession may be explained by reference to figure 5-7. Start at 0000 hours. The gyro at the equator has its spin axis horizontal to the earth's surface and parallel to the equator. As the earth rotates at 15° hour, the gyro spin axis changes its relationship to the earth's surface at the same rate. After six hours, it appears to the observer to have tilted 90°; at 12 hours, 180°; and at the end of 24 hours, it is back in its original relationship to the earth's surface. Throughout this time, the gyro has maintained its rigidity in space, so the precession was not real, but apparent.

Apparent precession of a gyro makes it unfit for use as a reference over an extended period of time unless some sort of compensating or erecting mechanism is used to keep the gyro in a fixed position in relation to the earth. Over a relatively short period of time a gyro can be used to estab-

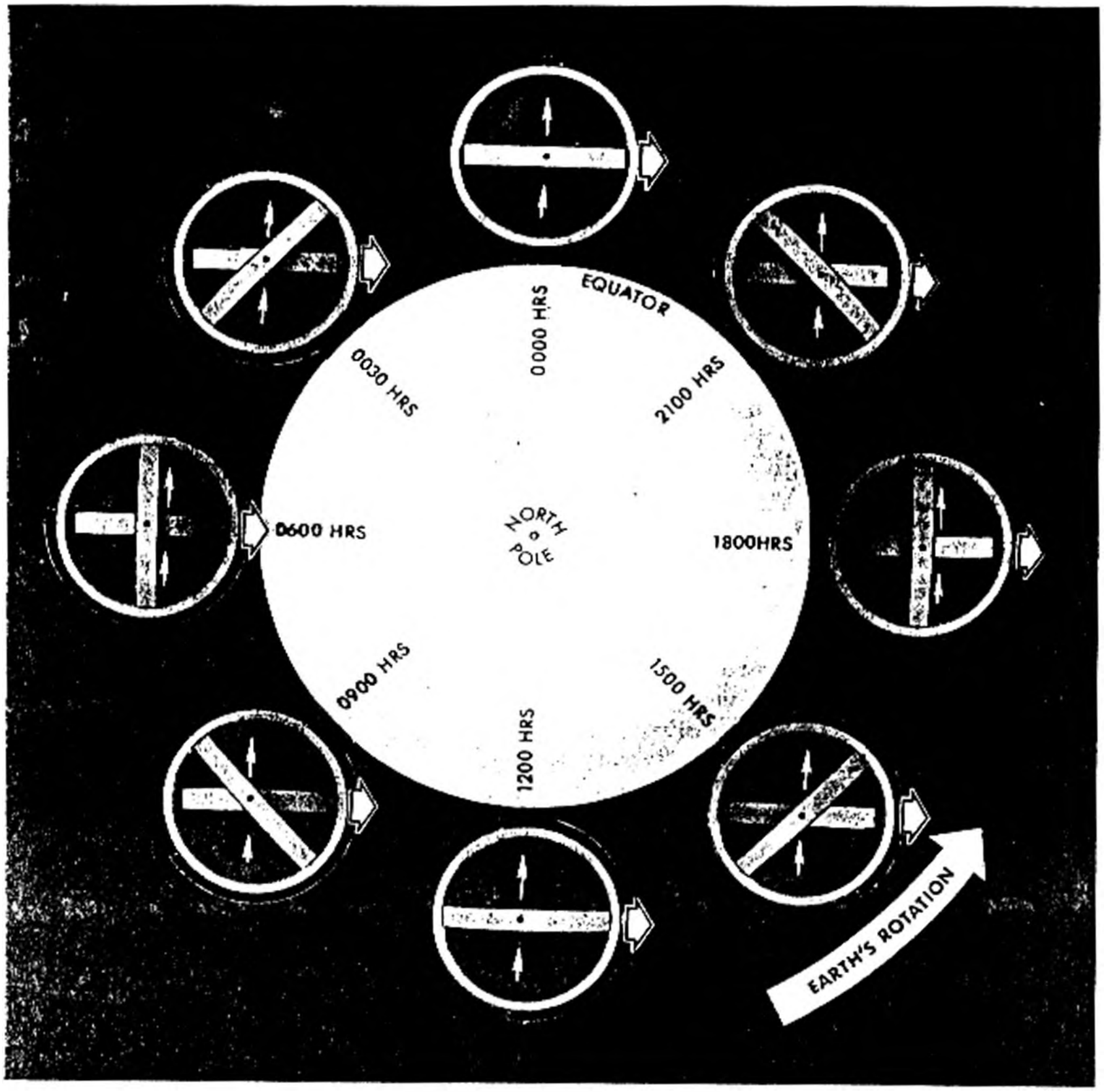


Figure 5-7. Apparent Precession

lish a satisfactory reference without the use of an erecting mechanism.

Gyro Drift

The line of direction of a gyro is not always in the direction in which it theoretically should

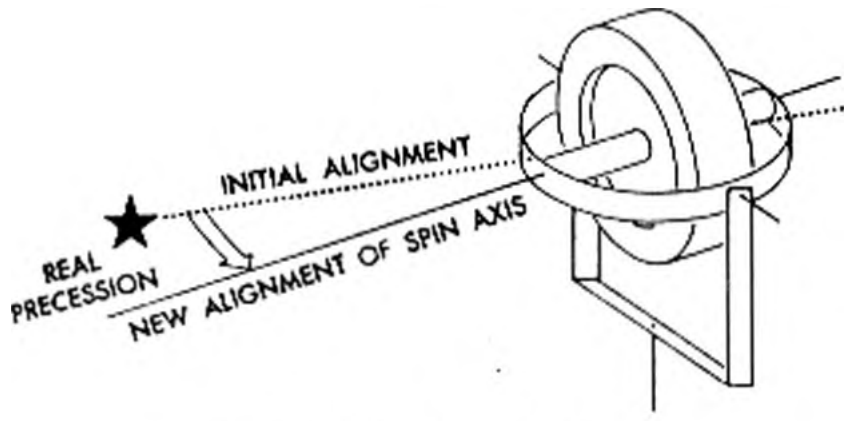


Figure 5-8. Gyro Drift

point as shown in figure 5-8. This error in the gyro is produced by random inaccuracies in the system. The resulting change in position of the spin axis is called *drift*. There are three general causes of drift:

UNBALANCE. A gyroscope often becomes dynamically unbalanced when operating at a speed or temperature other than that for which it was designed. Some unbalance exists in any gyro since manufacturing processes do not give perfect symmetry.

BEARING FRICTION. Friction in the bearings of the gimbals results in lost energy and incorrect gimbal positions. Friction in the spin axis bearing causes drift only if the friction is not symmetrical.

INERTIA OF GIMBALS. Energy is lost whenever a gimbal rotates because of the inertia of the gimbal. The greater the mass of the gimbal, the greater the drift.

The complete elimination of drift in common types of gyroscopes appears to be an impossibility. However, great strides have been made in recent years toward reducing the amount of drift. The methods being used are explained briefly later in this section.

Gimbal Lock

If two gimbals of a gyro are positioned in the same plane, the gyro is not free to precess. As a result, the forces of precession lock the gyro in a rigid position when a torque is applied.

The two basic properties of a gyro, rigidity in space and precession, are used in gyroscopic instruments. Rigidity is used to establish a reference in space unaffected by movement of the supporting body; and precession is used to control the effects of the earth's rotation, bearing friction, and unbalance, thus maintaining the reference in the required position.

Rate Determination

A gyro having two gimbals may be used to determine the rate of deviation of a missile from its established reference. By restraining one of the gimbals by springs or a similar means, the force exerted on the restraint by the precession of the gyro may be measured, giving an indication of rate.

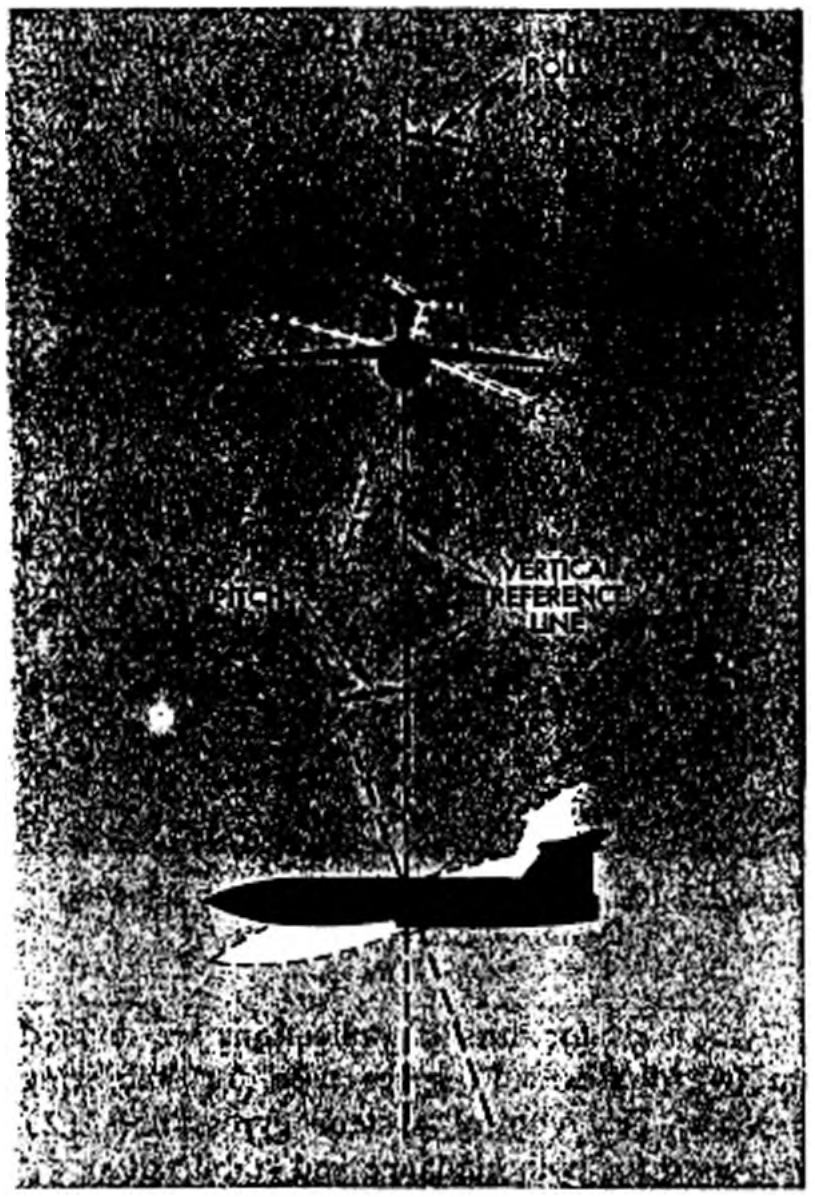


Figure 5-9. Vertical Reference Line showing Deviation about Roll and Pitch Axes

Use of Gyroscopes in Guided Missiles

A minimum of two displacement sensing gyros is necessary for missile stabilization. Each of the two gyros is used to establish a fixed reference line from which missile deviations in directional heading or flight attitude are measured. One of these lines is vertical and is established perpendicular to the earth's surface. This line passes through the spin axis of a vertical gyroscope. From this line, deviations can be measured from flight attitude about the roll axis or about the pitch axis of the missile, as shown in figure 5-9.

Deviations from directional headings cannot be measured by a vertical reference line. A second reference line fixed in space horizontally must be established so deviations in directional heading can be measured. This reference line passes through the spin axis of a horizontal gyro and is, therefore, stablized in space. Since deviations from course are measured from this reference line, it is established parallel to the horizontal axis of the missile as shown in figure 5-10.

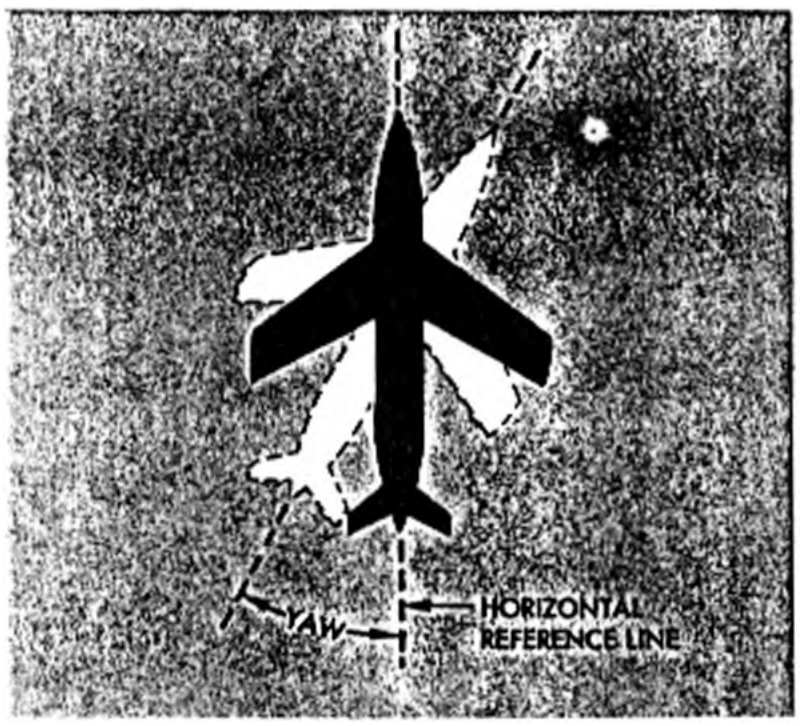


Figure 5-10. Horizontal Reference Line Showing Deviation about Yaw Axis

Gyros used for missile applications are divided into two classes: (1) gyros used for stabilizing purposes; i.e., control system gyros, and (2) gyros used for both guidance and stabilization.

Gyros used for stabilizing purposes alone are usually adaptations of existing gyro structures. A typical displacement sensing vertical or horizontal gyro to be used with external guidance signals might have the following characteristics:

- Power—alternating current, 2- or 3-phase, 400 to 2000 cycles per second.
- Sensing element—magnetic or capacitive pick-off, or precision potentiometer.
- Size— 5- to 6-inch cube, in hermetically sealed housing.
- Weight—3 to 5 pounds.

Such a gyro provides adequate missile stability when supplied with guidance signals at a precession rate greater than 1° per minute.

As stated previously, a minimum of two displacement sensing gyros are necessary for missile stabilization. When turns or other maneuvers are indicated, it often becomes necessary to use a 3-gyro system with but one sensing axis per gyro. Usually 3-gyro systems can use the same gyro design in all three positions, reducing production and maintenance problems.

Gyros used for both guidance and stability are free gyros which depend upon design skill and accurate manufacturing techniques to provide the necessary space reference. As a class, such gyros are usually larger and heavier than non-guidance types. In general a guidance gyro is built around a comparatively large, high-speed rotor of tremendous angular momentum, supported on micro-friction gimbal bearings. A "perfect" free gyro of this type, balanced in neutral equilibrium would maintain its spin axis in a constant angular position in space. Two or three such reference axes, usually mutually perpendicular, can be used as the basis of a guidance system.

Rate Gyro

In addition to the control signals supplied by the vertical and horizontal gyros, which are proportional to the deviation of the missile, a signal proportional to the rate of deviation is required for accurate control. This rate-of-deviation signal is supplied by a rate gyro. A rate gyro has a restricted gimbal, free to rotate about one axis only. It is a true gyroscope, conforming otherwise to the same basic principles as other gyros. The spin axis of a yaw rate gyro is mounted with its spin axis parallel to the missile line of flight. A roll rate gyro is mounted with its spin axis parallel to

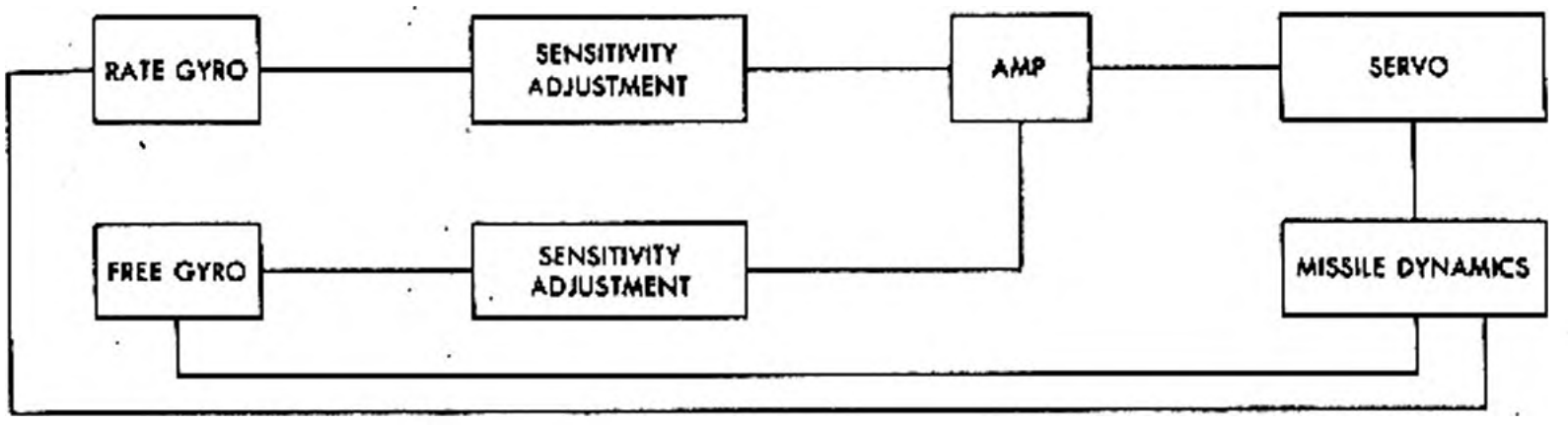


Figure 5-11. Missile Control Channel with Rate Gyro and Free Gyro

the missile pitch axis, at right angles to the line of flight. A pitch rate gyro is mounted with its spin axis parallel to the yaw axis of the missile, also at right angles to the line of flight.

The use of displacement signals alone as a means of applying corrective control in a flight control system results in a tendency of the craft to yaw or pitch about its desired course. When displacement signals are used alone, the application of control to bring the craft back on course results in overcorrection because of the momentum of the missile. The control system must have not only a means of recognizing the position of the craft in space, but also a means of detecting the rate of position change. By adding the rate of position change signal to the displacement signal, the tendency of the craft to overcorrect is minimized and a better degree of stability is obtained.

Present systems of missile stabilization employ both free and rate gyros, or a differentiating network in conjunction with a free gyro, to provide both position change and rate of position change signals to the control servos. The introduction of the rate signals results in good aerodynamic damping. Figure 5-11 shows a functional diagram of a system used to sense motion of a missile about its control axis. The system is for a missile control channel using both a rate gyro and free gyro. The sensitivity adjustments set the ratio of the rate signal from the rate gyro to the position signal from the free gyro. These signals are fed to an amplifier which adds the two signals and gives an output voltage proportional to their sum. This voltage is applied to the servo which positions the control surface in the proper position to drive the angle of change and the rate of change to zero.

Figure 5-12 is a functional block diagram of a control channel employing a free gyro and a computer network. The voltage signal from the free gyro goes to the proportional and derivative channels of the computer. The proportional channel consists merely of gain control. In the derivative channel the signal is differentiated by an RC network and amplified. These two signals are then

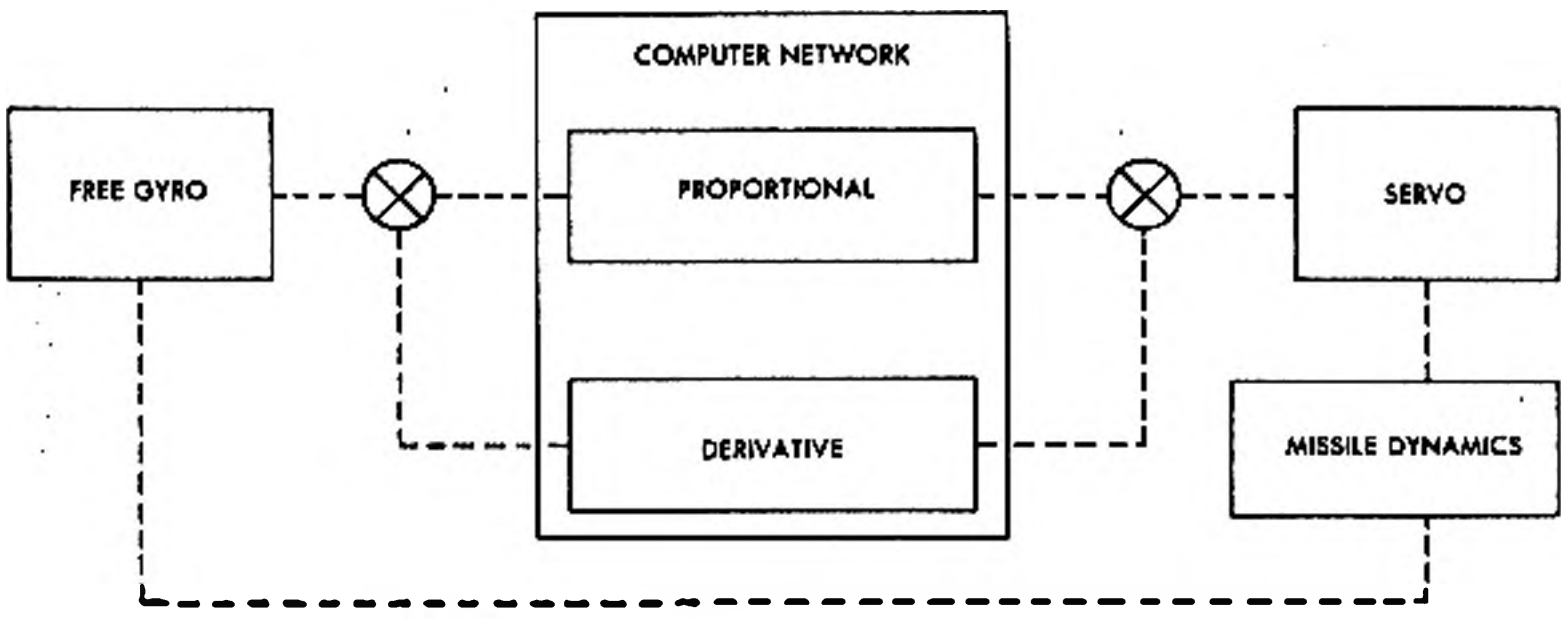


Figure 5-12. Missile Control Channel with Free Gyro and Computer Network

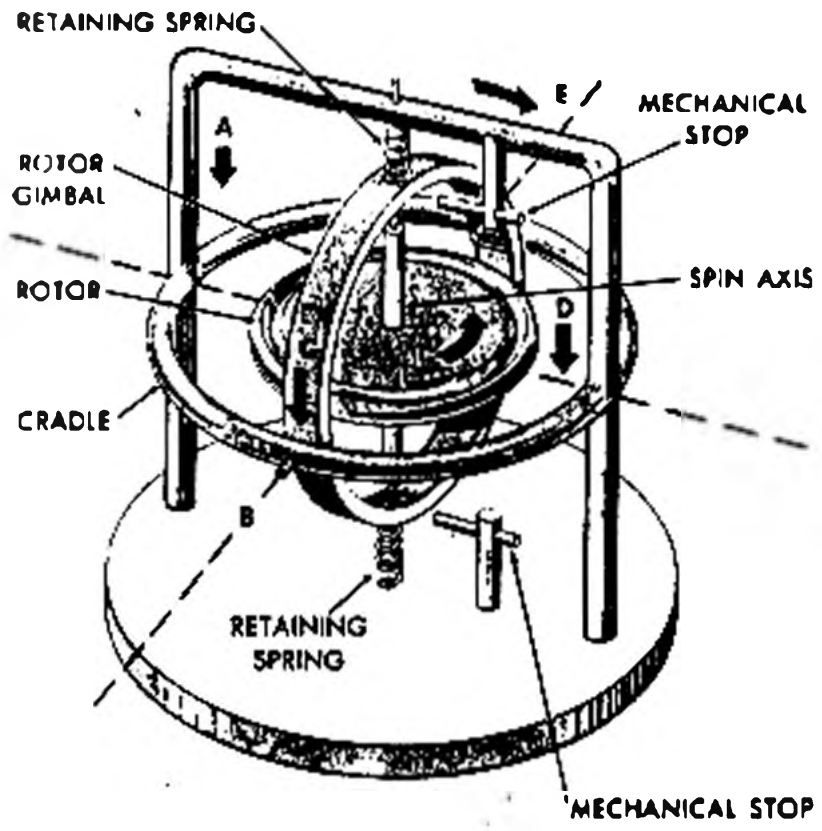


Figure 5-13. Action of Rate Gyro

combined and fed to the servo. The servo positions the control surface, and the angle of deviation is driven toward zero by the response of the missile to control surface deflection.

The rate gyro used in the first application is similar to a free gyro in that it has a high gyroscopic momentum while the rotor is spinning. It differs in that it is free to rotate about one axis only and in that it is sensitive only to angular rates of movement; that is, angular movement of its cradle in excess of a certain rate causes its spin axis to change its position in relation to space. Such changes occur 90° in the direction of rotation from the angular movement applied, and in the direction of movement. Figure 5-13 illustrates the action of a rate gyro when used in a missile control system.

An angular force applied to the cradle at A is transmitted through pivot B as a downward force at C. This force travels 90° in the direction of rotation and causes downward movement at D with subsequent precession at E. As soon as the angular rate applied to the cradle ceases, restraining springs restore the gyro to its neutral position. Stops are provided to limit the precession of the gyro to a few degrees in either direction.

Rate gyros have proved to be of considerable value in high-performance missile applications. Three rate gyros, combined with compact electronic integrators, offer an excellent source of displacement reference whenever guidance is not a function of the gyro system. Control systems of this type are within the scope of gyro and electronic experience.

Gyroscope Error

As explained previously, random drift in a gyroscope is that drift caused by bearing friction and dynamic unbalance. This type of drift results in an unpredictable precession of the gyro. Apparent drift in a gyro is that drift which occurs because of the rotation of the earth. It is toward the elimination of random drift that designers and manufacturers of gyros concentrate their efforts.

Many improvements have been made in the design of free and rate gyros in the past years. These units still have uncertain characteristics caused by friction, especially when the gyro case is subjected to accelerations. Gyros are delicate instruments and cannot stand rough handling. Their fine bearing surfaces may deteriorate in storage. Nevertheless, under certain conditions of missile launching and for the simplest stabilization system, the use of the free gyro is often required.

One requirement for an ideal missile guidance system is that the system be invulnerable to jamming signals. Two systems that meet this requirement are the celestial system and the inertial system. Both of these guidance systems require gyros for basic references. The celestial system must have a gyro stabilized platform on which a telescope is mounted, and the inertial system must have a stabilized platform for mounting accelerometers.

Random errors in the order of 1° of arc per hour will result in a course error of 60 nautical miles per hour of flight. Since the main cause of random drift in gyros is gimbal bearing friction, the problem of improving gyro stability is centered around the reduction of this friction.

FLOATED GYRO UNIT. A floated gyro unit is a good example of the progress that has been made toward the development of more accurate instruments for use as guidance gyros in guided missiles. This unit, also called the Draper gyro and HIG gyro, is a viscous-damped, single-degree-of-freedom gyro with a microsyn torque generator

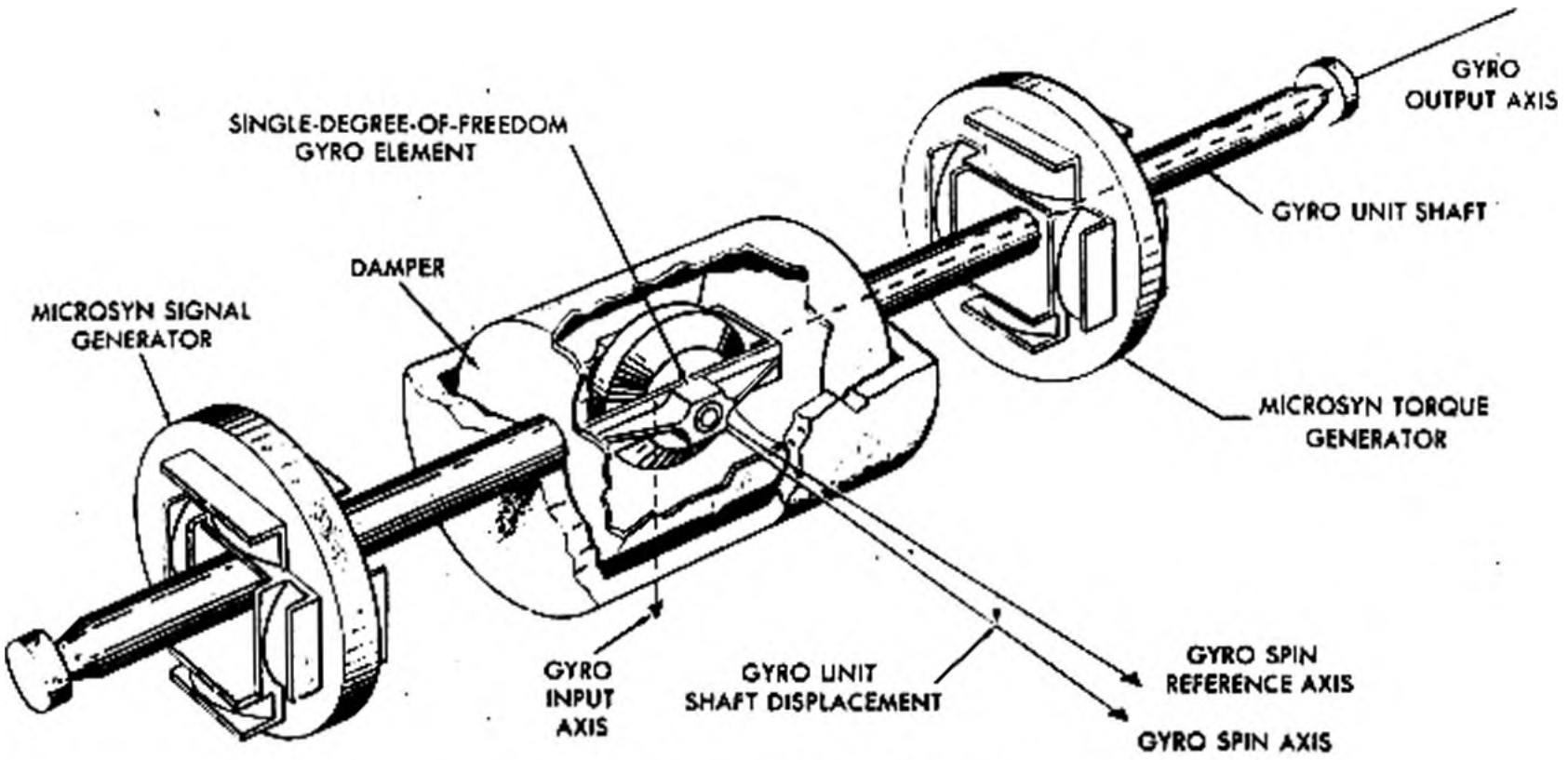


Figure 5-14. Integrating Gyro Unit

and a microsyn signal generator mounted on its output shaft. The microsyn torque generator places a torque on the gyro gimbal.

A single-degree-of-freedom gyro is so called because its gimbal is free to rotate with respect to the gyro case about a single axis. This axis, called the output axis, is perpendicular to the gyro spin (reference) axis. The third axis, the gyro input axis, is perpendicular to both the spin axis and the output axis. If the gyro case is subjected to an angular velocity with a component about the input axis, a precessional torque develops about the output axis.

Figure 5-14 shows a cutaway view of the integrating gyro unit. The gyro wheel is contained within the damper housing. The microsyn signal generator and torque generator units are mounted on the gyro shaft as shown. The space between the damper housing and the gyro case is filled with the viscous damping fluid of high specific gravity. Because of the high specific gravity of the fluid, it serves to float the gyro damper housing and gyro gimbal shaft, thus reducing the gimbal bearing friction. In this way, the random drift resulting from friction is greatly reduced. Because the gyro is supported by the viscous damping fluid, the unit is relatively free from the undesirable acceleration effects experienced with the usual gyro construction. This is a decided advantage of the integrating gyro for application to guided missiles.

The sensitivity of the integrating damper for angular velocity input and torque output is a function of the viscosity of the damping fluid, which in turn depends on the temperature of the fluid. Heaters with suitable thermostatic controls are placed in the gyro case around the space containing the damping fluid to keep the fluid at the desired operating temperature, maintaining the desired damping characteristics.

If this gyro is mounted in a missile so its input axis is parallel to the pitch or yaw axis of the missile, the torque applied to the output shaft would be proportional to the difference between the desired angular velocity of the missile and its actual angular velocity about the respective axis. Since the liquid has the function of integrating the torque, the unit can be considered as a combination angular velocity error measuring instrument and an integrator in cascade.

AIR BEARING GYRO. Another method by which random gyro error caused by bearing friction has been greatly reduced is through the use of air bearings. With this type of bearing, friction is reduced to a degree so negligible it might be considered a zero component. An air bearing works on the same principle as a vacuum cleaner display in which a large rubber ball is held suspended in a

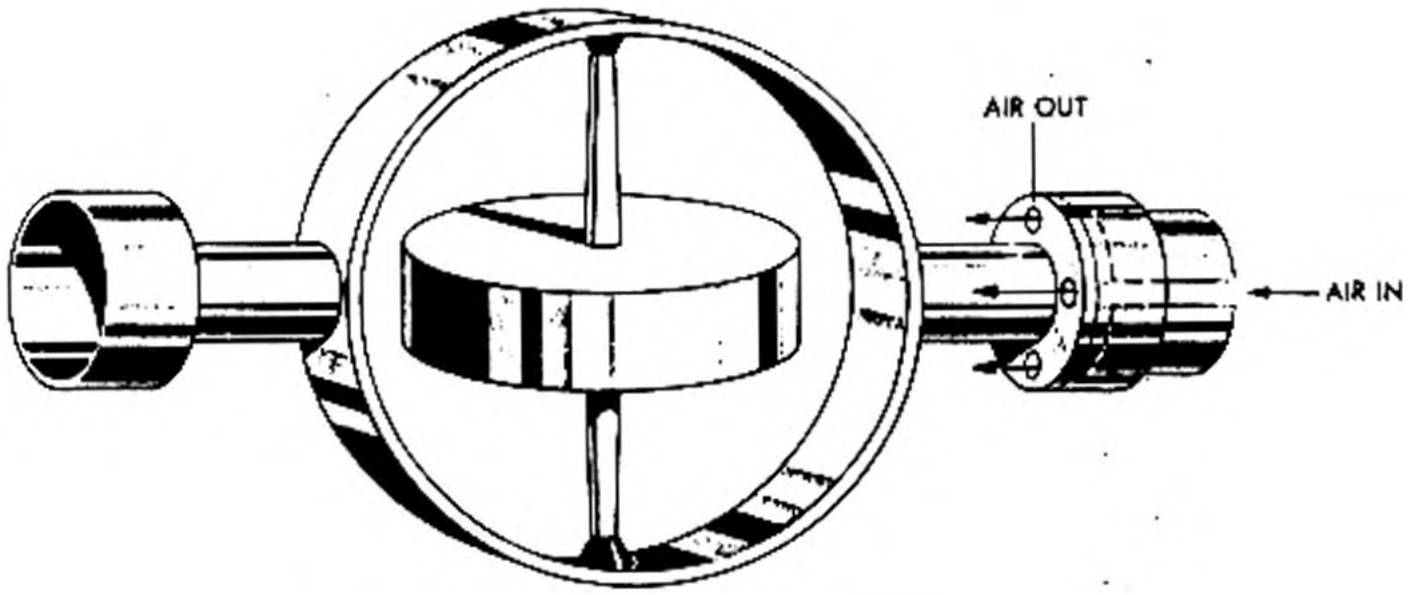


Figure 5-15. Air Bearing Gyro

conical air stream. The basic principle of an air bearing gyro is illustrated in figure 5-15.

Another innovation is an air suspension gyro in which the entire rotor is supported by pressurized gas. Electromatic suspension as well as the use of a nucleus of an atom to furnish stabilization are also being studied.

Perhaps the ultimate in gyros will result from experiments being conducted to develop a LASER (Light Amplification by Stimulated Emission of Radiation) gyro. More detailed information is presented about LASER in a later chapter of this manual. Basically, the LASER gyro consists of counter-rotating coherent light beams. Rotation rate can be sensed by the different frequency of the two beams. Further development of this principle will result in a gyro that will have no random drift and will be extremely accurate.

Gyros: Heart of the Missile Control System

The gyro system is the heart of the missile control system. The function of a gyro system is the ultimate actuation of the missile controls to maintain the missile in a specified attitude.

Accuracy requirements for control system gyros are not so rigorous as for guidance system gyros. On first thought, one might not readily see why the displacement gyro used in a missile autopilot need not be as accurate as a gyro used in a missile guidance system. Closer examination reveals that entirely different principles are involved. In the case of a guidance gyro, the instrument must give an accurate reference of the missile relative to its desired trajectory over the earth. In the case of the displacement gyro, the instrument must furnish a reference to the missile control system relative to its flight attitude only.

A vertical displacement gyro can be equipped with an erection system to keep it in a relatively level position. The erection system must be capable of maintaining the displacement gyro within allowable limits necessary for flight attitude reference.

GYRO ERECTION AND SLAVING CIRCUITS

Displacement gyros are controlled or kept erect by means of an erection and/or slaving system. Vertical gyros are controlled by an erection system, while horizontal gyros are controlled by both

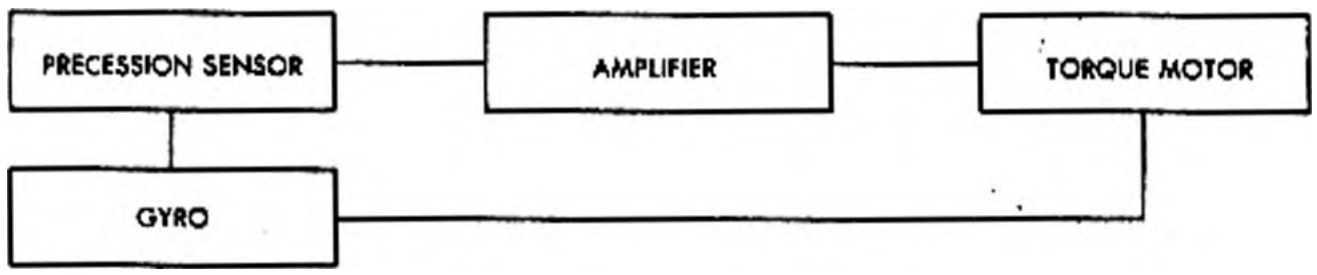


Figure 5-16. Vertical Gyro Erection System

a slaving system and an erection or leveling system.

Vertical Gyro Erection

A vertical gyro is used to stabilize a missile about the pitch and roll axes. Two torque motors are provided to maintain the gyro axis erect in both planes. A basic vertical gyro erection system is represented in block diagram form in figure 5-16.

The gyro precession sensing device (pickoff) may be any one of several types. As an example, the pendulous weight type shown in figure 5-17 is explained here in detail.

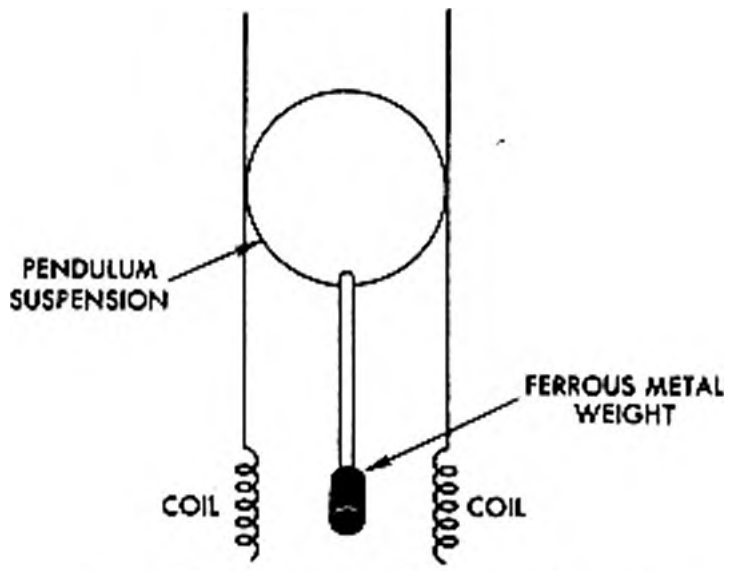


Figure 5-17. Gyro Precession Pickoff

Four secondary windings are connected in pairs, each coil in a pair being phase-opposed. The pairs are located beneath a pendulous, ferrous metal weight which acts as the magnetic coupling between the primary winding and the four secondary windings. When the gyro axis is vertical, the coupling to each pair of secondary windings is equal; but since the coils of each pair are phase-opposed, the signal output is zero.

The coils of this transformer-like device are wound on the gyro rotor, and the pendulous weight is suspended in the shaft of the rotor. When the spin axis of the gyro moves away from the vertical, the pendulous weight increases the coupling to one coil of a pair and decreases the coupling to the other. The amplitude of the signal developed is determined by the amount of gyro precession, and the phase of the signal depends upon the direction of precession. The signal, which may appear across one or both pairs of secondaries, is then amplified and sent to the appropriate torque motor, which applies a force to cause the gyro spin axis to precess back to the vertical position. The pendulous weight is liquid damped to prevent excessive oscillations. Figure 5-18 shows the gyro and torque-motor assembly.

The erection amplifier is a conventional amplifier capable of increasing the precession signal to a voltage large enough to drive the torque motor.

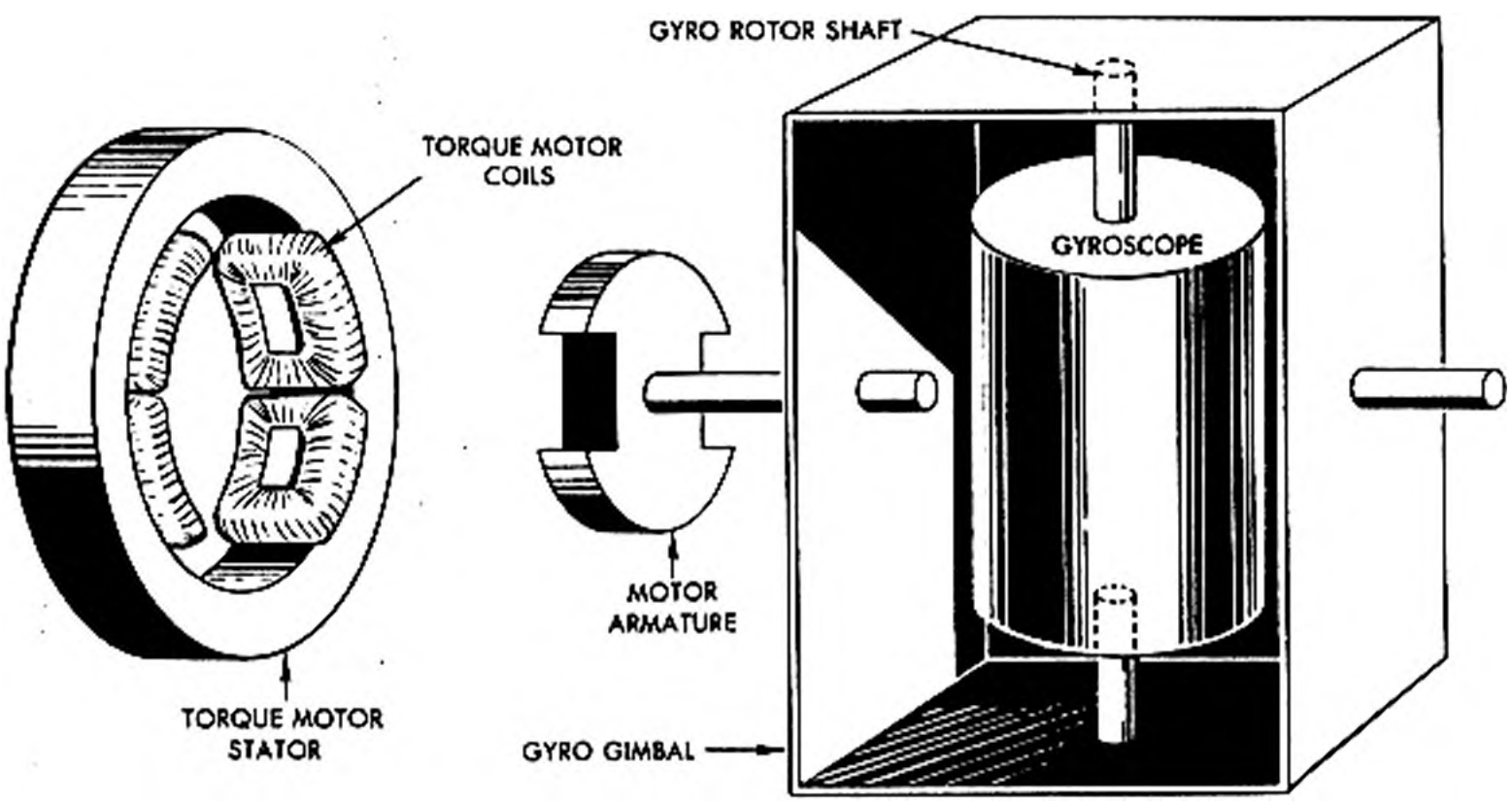


Figure 5-18. Gyro and Torque-Motor Assembly

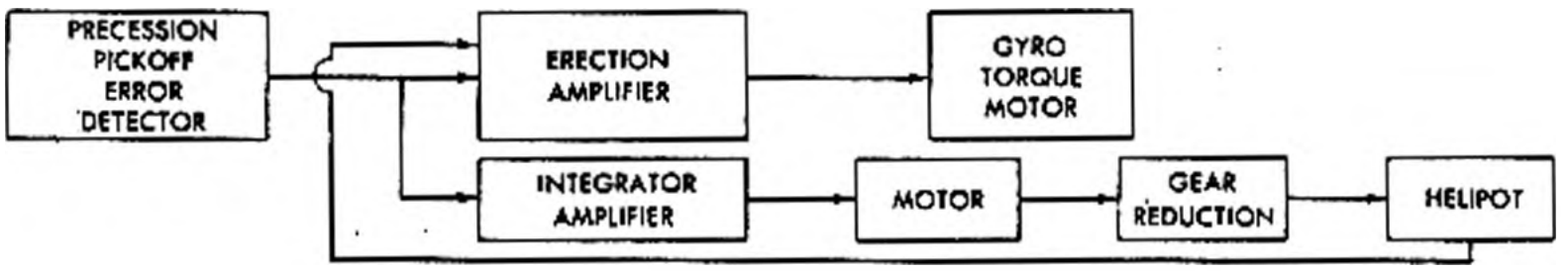


Figure 5-19. Gyro Erection System with Integrator

The torque motor is generally a 2-phase, squirrel-cage type induction motor. Some erection systems may include an integrator loop to correct for any dynamic unbalance in the gyroscope. A block diagram of such a system is shown in figure 5-19. The precession pickoff feeds a signal to the erection amplifier which amplifies it to a large enough voltage to drive the gyro torque motor. The integrator amplifier also receives the signal from the precession pickoff, and its output is applied to a motor which drives the wiper arm on a helipot through a large gear reduction. For occasional small errors, the helipot output is very small. For a constant error, which would result if the gyro were dynamically unbalanced, the helipot is driven long enough to compensate completely for the gyro unbalance. The helipot output is fed back to the erection amplifier input and appears as a constant signal to the gyro torque motor.

Gyroscope Slaving and Erection Systems

Horizontal gyros use a leveling system and a slaving system. The leveling system serves to maintain the gyro on a level plane, while the slaving system permits the gyro to give a directional indication.

The best known reference for slaving a horizontal gyro is the earth's magnetic field. This is accomplished by use of a flux valve which is a unit that senses the direction of the earth's magnetic field. It consists of a primary coil and three secondary coils wound on a metal core. Power to energize its primary coil and the signals originating in its secondary coils are transmitted through terminals located under a compensator which serves as a cover. The flux valve unit is pendulously suspended on a universal joint within the bowl and is weighted so that, within limits, it continuously responds to gravity. To prevent excessive swinging in flight, the bowl is filled with a damping fluid. The fundamental flux valve unit

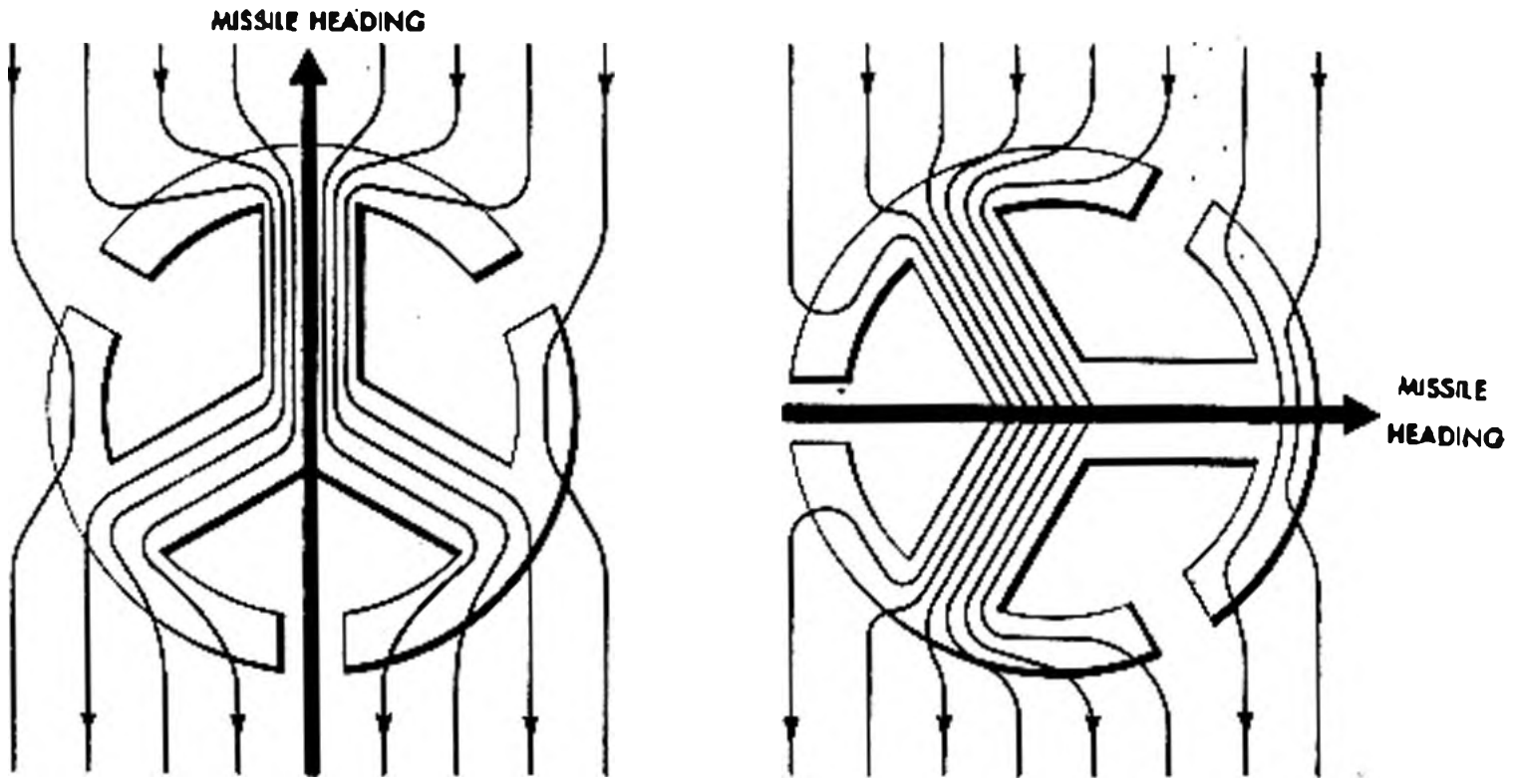


Figure 5-20. Changes in Flux Through Flux Valve as Heading Changes

is mainly a spider-like core made of laminated metal of high permeability. The core resembles a 3-spoke wheel, split through the rim between the spokes. Note figure 5-20. The hub is widened to receive a core about which the primary (exciter) coil is wound. The coil is energized by a 400-Hertz power source. Each spider leg is encircled by a pickup (secondary) coil and terminates in arcs of metal which serve as flux collectors.

The flux valve unit functions in this manner: As the positive half-cycle of the exciting current builds up, it sets up a magnetic field which expels the earth's flux from the spider legs. The expelled flux cuts the pickup coils on each leg, inducing current in each. As the positive half-cycle of the exciting current decays, the earth's flux, because of the higher permeability of the legs compared to the surrounding air, is drawn into the legs, again cutting the pickup coils and again inducing current but of opposite polarity. A half-cycle of alternating current produces one full cycle of induced current. The negative half-cycle of the exciting alternating current reproduces the action of the positive half-cycle. Since the exciting current has a frequency of 400 Hertz, the induced current is doubled in frequency to 800 Hertz.

The magnitude of the induced current in the pickup coils varies according to the number of magnetic lines cutting them. The number of lines cutting the coils varies according to the position of the individual spider leg with regard to the magnetic north. Each pickup coil produces a voltage of a given magnitude, which is transmitted to the corresponding "Y" coils of the flux valve synchro in the directional gyro control. The magnitude of the transmitted voltage depends upon the pickup coil's position relative to the magnetic meridian. This transmission creates a magnetic vector. The vector varies in its position across the stator in the same relationship as the spider legs of the flux valve unit vary in their orientation to the earth's magnetic lines. The rotor of the flux valve synchro, if positioned at right angles to this vector, produces no voltage. For the purposes of this discussion, such a no-voltage condition is referred to as rotor alignment with the vector.

The compensator counteracts constant distortions originating from magnetic aircraft parts and electrical apparatus located in the vicinity of the flux valve. The compensator is a flat, compact unit which contains four permanent magnets to oppose the deflecting forces. The magnets are rotatable by means of two slotted shafts. When the compensator is attached to the transmitter (flux valve), the shafts are clearly identified. One shaft has "NS" stamped adjacent to it and the other "EW." Alignment of the dots on the shafts with those on the compensator cover indicates that the magnets are in neutral position.

Type S-1 Directional Gyro Control

A type S-1 directional gyro control houses an electrically driven gyro. The gyro's spin is not only maintained tangent to the earth's surface, but is also slaved to the earth's magnetic field for the purpose of furnishing a basic magnetic reference to which the magnetic heading of the missile can be slaved.

Most of the internal assemblies are supported by a frame which is mounted on a cast base. The cover, which is transparent to permit reading of a compass card, fits over all of the internal assemblies and fastens onto the base.

The leveling torque motor is of the 2-phase induction type and has a vertical torque axis. The motor consists of a stator attached to the top of the vertical ring and a squirrel cage (rotor) secured to the frame of the unit. When actuated by a liquid level, which is mounted on the gyro housing, the motor returns the spin axis of the gyro to the horizontal plane.

The slaving torque motor is likewise of the 2-phase induction type. It has a horizontal torque axis and consists of a stator mounted on the gyro housing and a squirrel cage fastened to the side of the vertical ring. It is actuated by an amplified and phase-detected voltage which aligns the gyro to the magnetic heading sensed by the flux valve. The voltage originates in the rotor of the flux valve synchro in the directional gyro control.

The rotor of the flux valve synchro of the directional gyro control is fixed to the lower pivot of the vertical ring and turns relative to its stator, which is fastened to the frame of the unit. When the system is energized, the transmitter instantly establishes a magnetic vector across the stator of the flux valve synchro. This magnetic vector is

the resultant of the earth's magnetic lines of force as detected by the three transmitter spider legs. The flux valve synchro rotor, fixed to the vertical ring or outer gimbal of the gyro, may not be aligned with that vector. Any misalignment between the rotor and the vector is itself the corrective factor, for under this condition voltage is continuously generated in the rotor. This voltage, after amplification and phase detection, energizes the slaving torque motor until perfect alignment between the synchro rotor and the vector is attained. The gyro rotates the vertical ring as it responds to the precessing force of the torque motor. This brings the rotor of the flux valve synchro into alignment with the vector, stopping generation of voltage in the rotor and halting the action of the torque motor. Should the spin axis of the gyro drift in azimuth during flight, a similar action takes place. Conversely, effective alignment between the synchro rotor and the vector is maintained during turns by the rotation of the directional gyro control case, with its fixed synchro stator, about the gyro-stabilized vertical ring and its synchro rotor. Since the magnetic vector sensed by the remote compass transmitter remains fixed in space, both the transmitter and directional gyro control case effectively rotate about it, maintaining the alignment between the synchro rotor and the vector.

During turns, centrifugal forces tend to swing the pendulous flux valve unit into the vertical component of the earth's magnetic field. This slightly distorts the signals of the flux valve unit. Because of the slow precessing response of the gyro to these signals, the indications of heading remain effectively accurate.

The heading synchro of the directional gyro control is located vertically adjacent to the flux valve synchro, and the rotors of both are affixed to the same vertical ring pivot. The heading synchro stator is attached to the frame above the flux valve synchro stator. The rotors or stators of both synchros move simultaneously, depending on whether the vertical ring or the case produces the rotation. The heading synchro rotor is energized by a 400-Hertz power supply to produce a magnetic field which induces voltages in the "Y" connected coils of its stator.

When the missile turns in either direction, the flux valve, being fixed to the missile, turns with it. As it turns, the flux valve spider legs and coils also turn and continuously change positions and relationships relative to the magnetic meridian. Because of these changing relationships, the earth's flux penetrates each spider leg in differing quantities and the magnitude of the induced voltages in each spider pickup coil continuously varies. Correspondingly, it varies in the "Y" connected coils of the stator of the flux valve synchro in the directional gyro control. New vectors are being continuously formed across both the transmitter and flux valve synchro stator coils which, in effect, rotate relative to these coils as a pointer would move over the face of a dial. The directional gyro control case, to which the synchro stator is fixed, also rotates with the missile and angularly displaces the stator. These magnetic vectors, although rotating with respect to the stator coils, remain constantly aligned with the flux valve synchro rotor which is held rigidly fixed in position by the gyro. Under such conditions, misalignment can occur only if the gyro drifts.

The slaving amplifier used performs two functions: (1) it amplifies the signal of the rotor of the flux valve synchro in the directional gyro control; (2) it detects its phase, controlling the direction and amount of torque in the slaving torque motor. Figure 5-21 shows a typical slaving system.

The second type of sensor to be presented in this chapter is the altitude sensor.

ALTIMETERS: SENSOR UNITS FOR MEASURING ALTITUDES

Instruments used to measure altitude are called altimeters. Two main types of altimeters are pressure altimeters, which gave an approximate true altitude from which a more accurate value can be calculated, and absolute altimeters or radio altimeters, which give absolute altitude directly.

Altitude refers to the height of a vehicle in the air. Altitude is defined as vertical distance above some point or plane used as a reference. There may be as many kinds of altitude as there are reference planes from which to measure.

Pressure Altimeter

A pressure altimeter is simply a mechanical aneroid barometer, registering atmospheric pres-

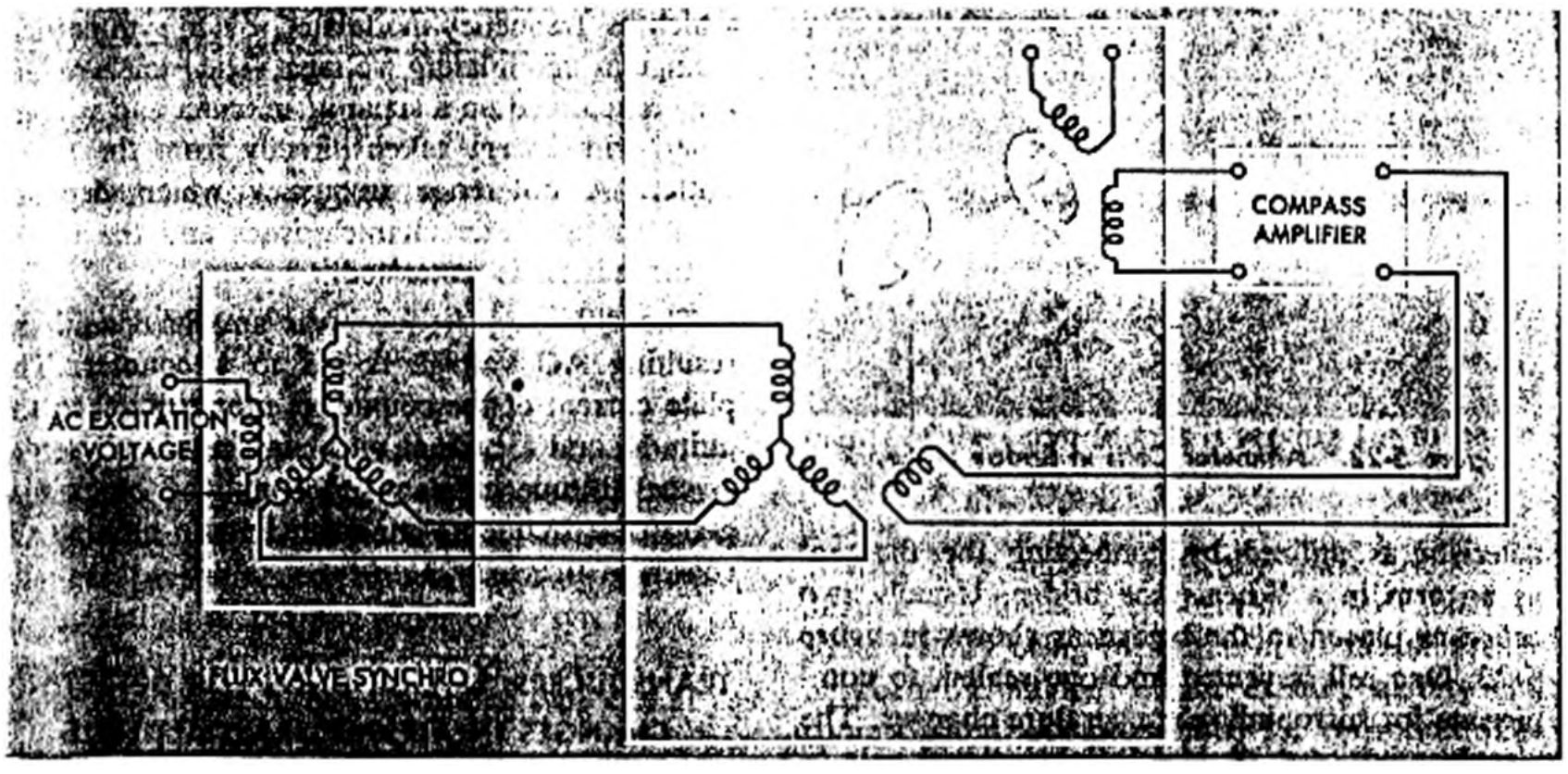


Figure 5-21. Basic Functional Diagram of Slaving System

sure on a scale calibrated in terms of altitude instead of inches of mercury. It consists of a small, airtight chamber from which most of the air has been removed. The pressure or weight of the outside air tends to collapse the chamber, but this tendency is resisted by a spring. As the atmospheric pressure increases, the chamber is compressed; as the pressure decreases, the chamber is expanded by the spring. This slight motion is magnified mechanically, and is registered in terms of the altitude that would produce a corresponding change in pressure under standard conditions.

The pressure altimeter, although calibrated in feet, actually measures atmospheric pressure at flight level and interprets this value in terms of feet above a certain pressure level. If the pressure were constant for each level of altitude, an altimeter could be designed to indicate the true altitude corresponding to each pressure. But since the pressure does not remain exactly constant at any one level, the altimeter cannot indicate true altitude directly.

As a compromise, the altimeter is designed to indicate changes of altitude as the pressure varies according to an arbitrary rate. This arbitrary rate is called the standard pressure lapse rate. For any change of pressure, the altimeter indicates the corresponding change of altitude according to the standard pressure lapse rate.

The pressure altimeter is designed to measure altitude from 28.00 inches to 31.00 inches of mercury above any one pressure level. Whatever pressure is set in the window on the dial face is the pressure level above which the altimeter measures standard altitude.

Accurate absolute altitude is an important requisite for good navigation and bombing. Absolute altitude can be computed from the pressure altimeter readings, but the results are often unreliable. Under changing atmospheric conditions, corrections applied to obtain true altitude are only approximate. Besides, any error made in determining the terrain altitude results in a corresponding error in the absolute altitude. Such an error in absolute altitude makes ground-speed by timing unsatisfactory for dead reckoning at low altitudes.

An altimeter cell is an electrical application of a barometer for determining altitude. An altimeter cell detects changes in altitude and converts this information into an electrical signal. The unit consists of a filament of fine platinum wire, heated by an electric current and enclosed in a vented envelope. The vent is always connected to a static pressure line. When the altitude or static pressure changes, the rate at which the filament can release its heat changes, and the temperature and resistance of the filament also change. This char-

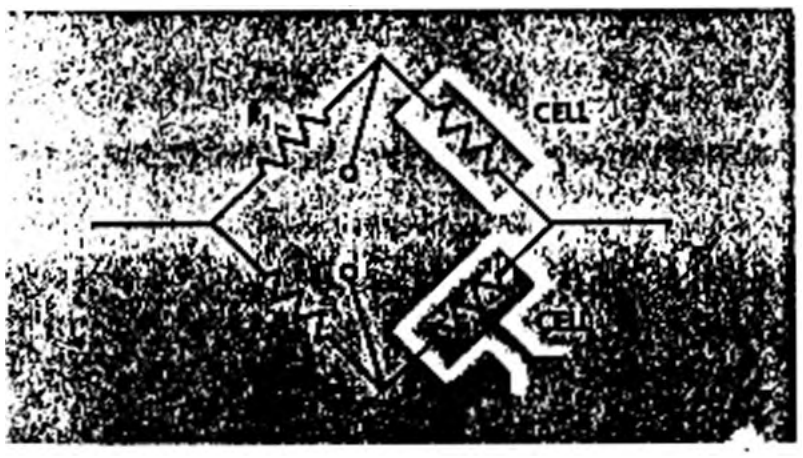


Figure 5-22. Altimeter Cells in Bridge Network

acteristic is utilized by connecting the filament as an arm in a Wheatstone bridge. Usually two cells are placed in the bridge as shown in figure 5-22. One cell is vented and one sealed, to compensate for surrounding temperature changes. The signal output of the bridge is proportional only to pressure changes since both cells change with temperature while only one changes with pressure.

An altimeter cell measures altitudes as high as 500,000 feet. This wide range gives the cell an advantage over a mechanical aneroid. Another advantage of altimeter cells over a mechanical aneroid is that a unit consisting of two cells is quite compact, the height of each individual cell being only 1½ inches.

Radio Altimeters

Frequency-modulated radio altimeters have been in common use on many types of aircraft for some time. These altimeters can be used to keep a missile at a preset altitude automatically. It is also possible to use the radio altimeter as a proximity detonating device for missiles.

In frequency-modulated radio altimeters, a transmitter radiates toward the earth a wave which is frequency-modulated by a symmetrical triangular modulating voltage. The reflected energy is received on a separate antenna and is combined with energy taken directly from the transmitter. A difference frequency which depends upon the altimeter characteristics and the height of the missile is generated.

The difference frequency is amplified, and the resulting AC voltage is fed to a counter. The plate current of the counter is proportional to the altitude, and a current meter in this circuit serves as the altitude indicator. A simplified block diagram of a frequency-modulated radio altimeter is shown in figure 5-23.

TRANSDUCERS: SENSOR UNITS FOR DETECTING CHANGES IN AIRSPEED AND ALTITUDE

A transducer is defined by Webster as "a device actuated by power from one system and supplying power in the same or another form to a second system." In most missile applications, the transducer is used to change a mechanical action caused by a pressure change into a reference voltage proportional to the mechanical change. This is accomplished by linking the mechanical action to a wiper arm on a linear potentiometer. In the missile field, a transducer is used as an end instrument of the telemetry system. Transducers are the components used to translate changes in airspeed and altitude into electrical control signals.

An airspeed circuit is used to maintain desired constant indicated airspeed of a missile. Changes in airspeed may be accomplished either by changing the throttle setting or by changing the pitch attitude of the missile. The airspeed transducer may consist of a bellows mechanically attached to the wiper on a linear potentiometer (pot) or

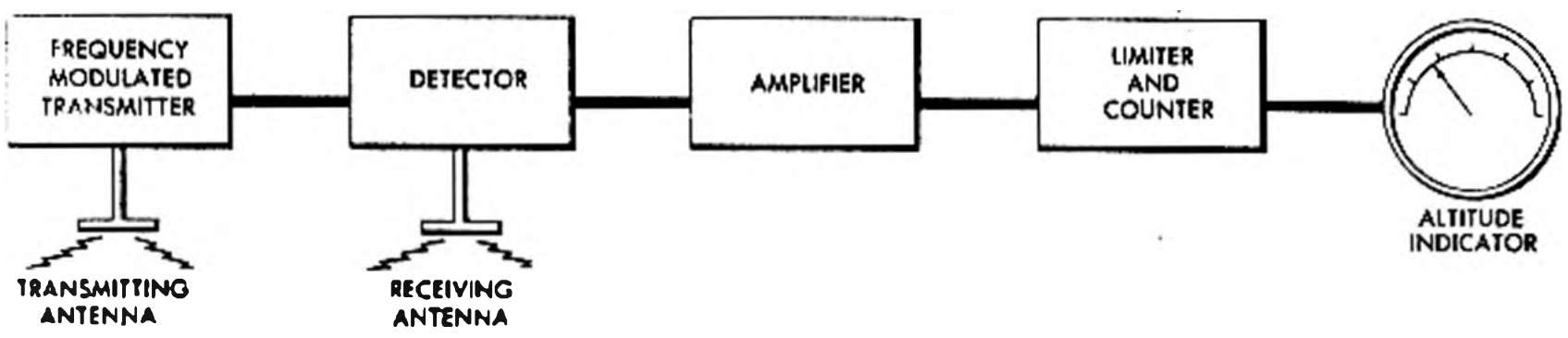


Figure 5-23. FM Radio Altimeter

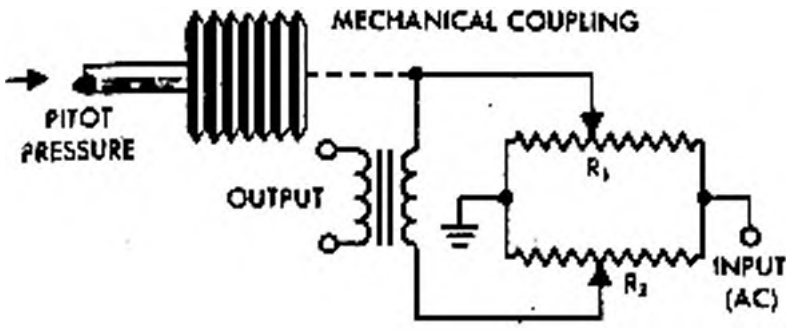


Figure 5-24. Resistor Bridge-Type Airspeed Transducer

a bellows mechanically attached to the rotor or stator of a selsyn. The bellows is actuated by ram air pressure. As the airspeed changes, the bellows changes (expands or contracts) causing a change in the setting of the pot wiper or selsyn rotor, or stator. The potentiometer type of airspeed transducer is generally used in a bridge type circuit as illustrated in figure 5-24.

The wiper of potentiometer R_1 is attached to the bellows and is repositioned on the pot whenever air pressure changes occur. The wiper of potentiometer R_2 is adjusted at a point on the pot in such a way that the bridge will be balanced at the desired reference airspeed.

The bridge balancing operation is performed before the missile is launched by applying air pressure equivalent to the air pressure of the reference airspeed. The wiper on R_2 is fixed at this particular point by means of a set screw or similar device.

Figure 5-25 shows a selsyn-type airspeed transducer. The stator of the selsyn is mechanically connected in such a way that expansion or contraction of the bellows causes a rotation of the selsyn stator.

Once a sensor unit detects a change in the attitude of a missile, the newly acquired intelligence must be transmitted in a useful form. Pickoffs serve this purpose.

PICKOFFS: DEVICES THAT MAKE INTELLIGENCE OF SENSOR UNITS USEFUL

A pickoff is a device which produces a useful signal from the intelligence developed by a sensor. This signal must meet the particular requirements of the servo loop it is serving (as to phase, amplitude, loading effect, etc.).

The sensing devices pertinent to the attitude control of a missile generally make use of an angular or linear displacement proportional to some quality being measured or to the difference between the amount of the existing quality and some desired amount needed for operation. The physical dimensions of the displacement will vary; each variation requires different types of pickoffs to meet different needs in the servo loop.

The pickoff must first have an output sense. That is, it must be able to distinguish the direction of the displacement and produce a signal indicative of the direction. In electrical pickoffs this is frequently done by using a phase or polarity difference. The ideal pickoff should also have a maximum output for a minimum movement of the pickoff. It should have a linear output. It should also have a minimum torque or friction loss which would be reflected back to the sensor element.

Small size and lightweight construction are additional requirements imposed on pickoffs to be used in missiles. Pickoffs also should have a fine null range, and electrical pickoffs should have no phase shift through a varying displacement.

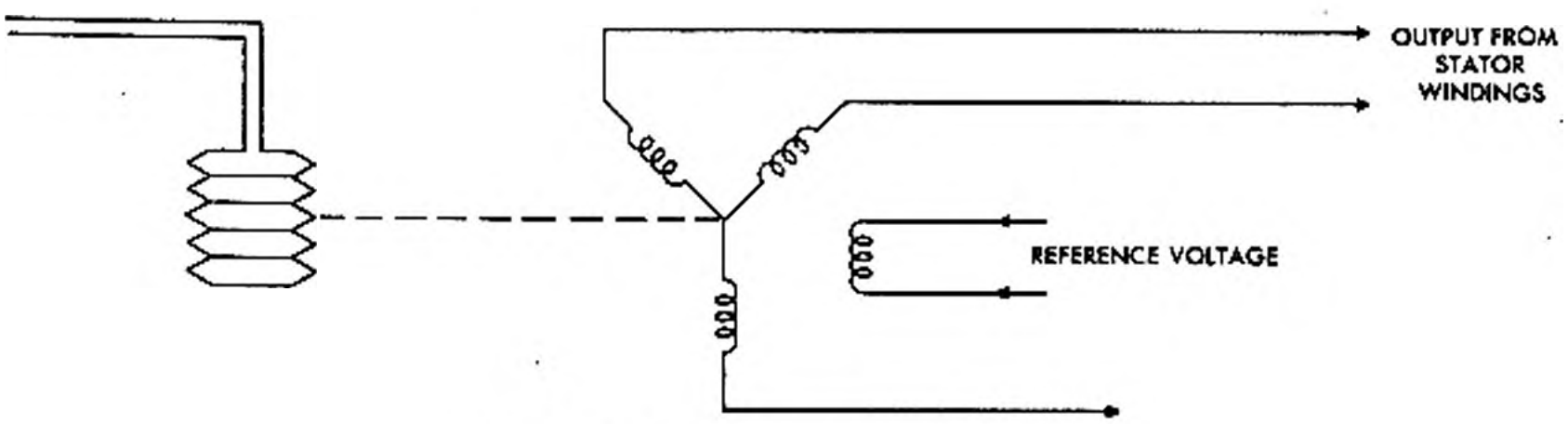


Figure 5-25. Selsyn-Type Airspeed Transducer

A number of pickoffs have been used or experimented with in missile control systems. The majority of these are electrical. Pneumatic devices have been used in some systems in which it was desired to keep electrical components to a minimum, but they have not proven too satisfactory and are not considered here.

Of the electrical pickoffs generally in use, the following are the most common:

- Selsyn.
- Potentiometer.
- Reluctance.
- Capacitance.

Each of these is found with variations to make it suitable to specific applications in which its particular characteristics are needed.

Selsyn Pickoffs

A selsyn pickoff (also called synchro, autosyn, microsyn, etc) usually consists of a pair of selsyns wired as a generator and a control transformer. They give an accurate electrical identification of angular movement, can be made small and light, and have a constant phase for a varying displacement. They lack sensitivity for minute variations in displacement.

Two typical selsyn pickoff systems are shown in figures 5-26 and 5-27. When a selsyn generator and a selsyn control transformer are properly connected, they form a complete pickoff system. As shown in figure 5-27, a differential selsyn may be included between the generator and control transformer. The application of the differential selsyn is explained in detail later.

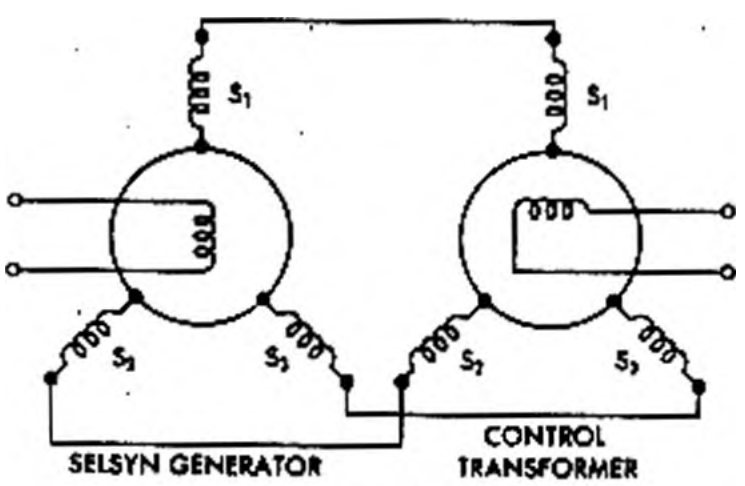


Figure 5-26. Selsyn Hookup

Both schematics represent the electrical zero position. For a control transformer, electrical zero is defined as that position of the rotor in which no voltage is induced in the rotor windings from the stator windings. This condition occurs when the axis of the rotor is perpendicular to the axis of the S_1 winding.

OPERATION OF A CONTROL TRANSFORMER. A control transformer and a selsyn generator are connected in the electrical zero position. Note the relative positions of the two rotors in the accompanying diagrams. No voltage is induced by the S_1 coil since the rotor and coil are perpendicular. The currents of the S_2 and S_3 coils are equal and opposite and induce voltages in the rotor which are equal and opposite. The net effect is that no voltage is induced in the rotor in this position.

When the shaft of the control transformer is turned 90° clockwise or counterclockwise, the position of S_1 and the rotor is such that maximum voltage is induced in the control transformer

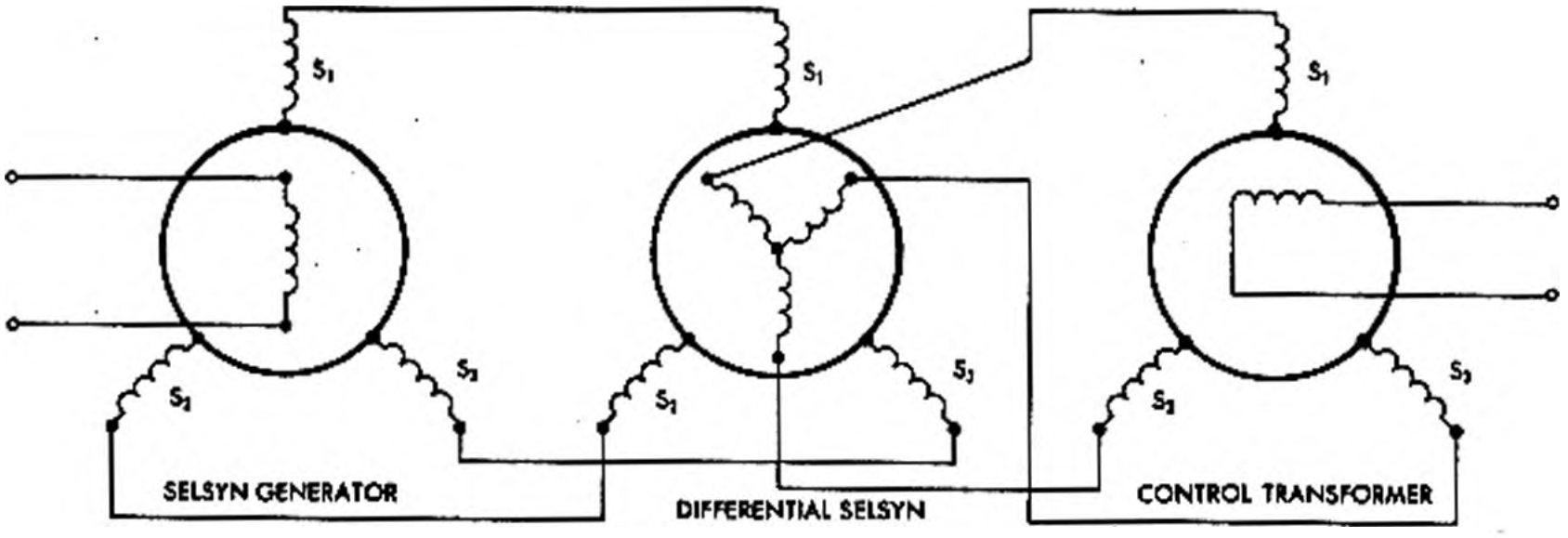


Figure 5-27. Selsyn Hookup with Differential

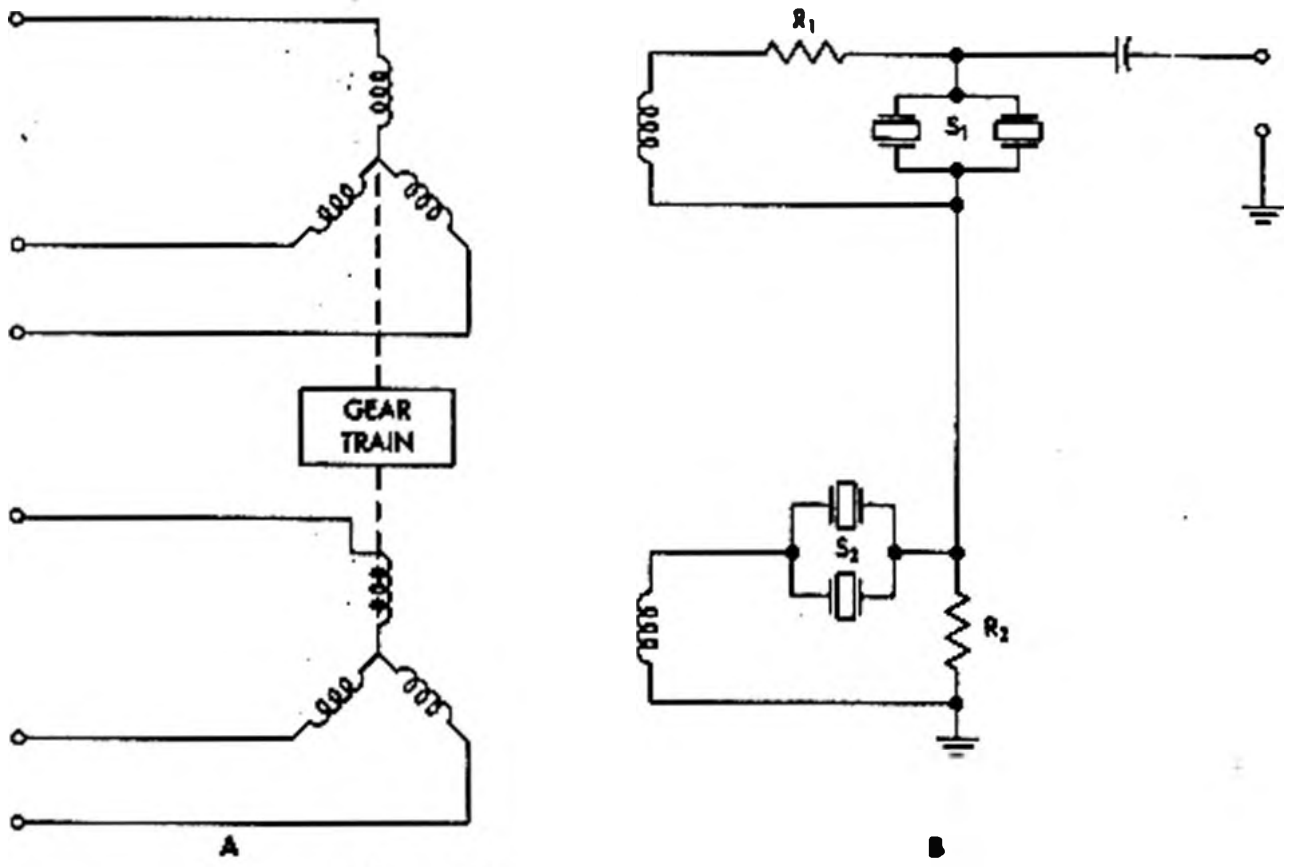


Figure 5-28. Two-Speed Selsyn System

rotor. The currents in the S_2 and S_3 coils are equal and in the same direction, and they induce voltages in the rotor which are equal and of the same polarity. The net result is that maximum voltage is induced in the rotor in these positions, and the induced voltage is either in phase or 180° out of phase with the voltage of the rotor of the selsyn generator.

The output voltage obtained from a control transformer is called an error signal, because the magnitude and phase of this voltage are an indication of how much and in what direction the two rotors are out of correspondence. Turning the rotor of the control transformer so that its position is perpendicular with that of the generator rotor cancels or nulls the error signal. This position is sometimes referred to as the null position. The ordinary selsyn pickoff system has a wide null position; that is, it lacks sensitivity for very small displacements. For this reason 2-speed selsyn systems are often employed to increase their pickoff sensitivity.

TWO-SPEED SELSYNS. In a 2-speed selsyn system, the fast selsyn of the pair is mechanically attached to the slow selsyn through a stepdown gear train as shown in figure 5-28A. With an appropriate relaying arrangement, the error signal can be picked off the fast speed selsyn during small displacements of the rotor shafts. When the error signal exceeds a certain preselected value, the relay positions change, and the error voltage then is taken from the slow selsyn rotor. Sensitivity is thus increased since the fast selsyn measures an appreciable angular displacement for small displacements of the slow selsyn.

Another type of 2-speed selsyn system does not require any switching or relaying between selsyns. This system, shown in figure 5-28B, uses the characteristic curve of a selenium rectifier for forward resistance versus voltage applied. As shown in figure 5-29, the resistance of the rectifier is high for small voltages and decreases as voltage is increased.

When the angular displacement of the slow

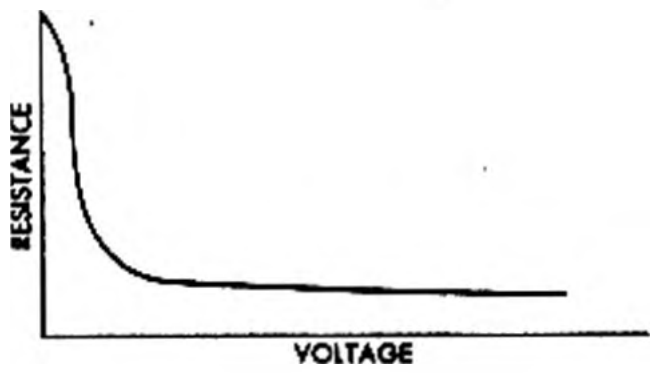


Figure 5-29. Resistance vs Voltage Curve of Selenium Rectifier

selsyn rotor is large with respect to the null position, the slow selsyn error voltage is large. This voltage is divided between the selenium rectifier and the resistor. Since this voltage is large, the resistance of the selenium rectifier is small in comparison to that of the resistor. Most of the voltage drop is across the resistor. The fast selsyn output can be large or small at this time since it cycles several times for each revolution of the slow selsyn. The fast selsyn output voltage is limited by the selenium rectifier S_1 and resistor R_1 to 2 or 3 volts. The output voltage to the succeeding circuit is the sum of the voltage across R_2 and S_1. As long as the slow selsyn error output voltage is greater than 2 or 3 volts, it predominates over the total servo output voltage. As the system comes closer to its null position, most of the slow selsyn output is dropped across S_2.

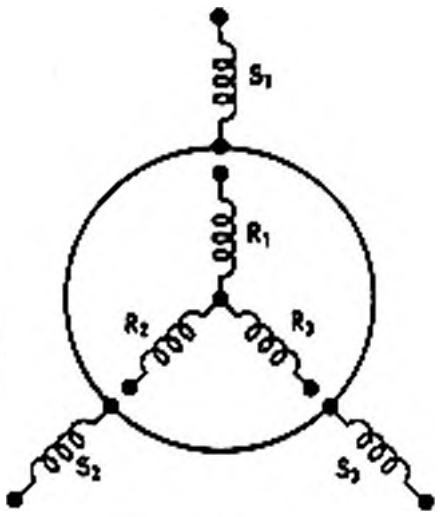


Figure 5-30. Differential Selsyn

When the system is 3° or 4° from null position, the slow selsyn voltage across R_2 is so small that it contributes little to the error signal. The error signal is essentially that of the fast selsyn alone.

DIFFERENTIAL SELSYNS. A differential selsyn performs the same function electrically that a mechanical differential performs in a mechanical system. The differential in a mechanical system connects three shafts so one shaft turns an amount which is equal to the difference between the amounts that the other two turn. A selsyn differential generator subtracts two inputs, and a selsyn control transformer indicates the difference between the two inputs. The action of the differential selsyn can be reversed so that two inputs are added by a control transformer.

The stator of a differential selsyn is similar to that of an ordinary selsyn. It consists of three sets of Y-connected coils wound in slots. The slots are spaced 120° apart around the inside of the field structure. The rotor of the differential is completely different from that of the ordinary selsyn as shown in figure 5-30. It is cylindrical in shape and has three sets of Y-connected coils wound in slots which are spaced 120° apart around the circumference. Three slip rings connect the rotor leads to the external circuit as in the ordinary selsyn. In the electrical zero position, the rotor coils R_1, R_2, and R_3 are turned so that they are aligned with the stator coils S_1, S_2, and S_3, respectively.

Like the ordinary selsyns, the differential operates on the transformer principle. The stator

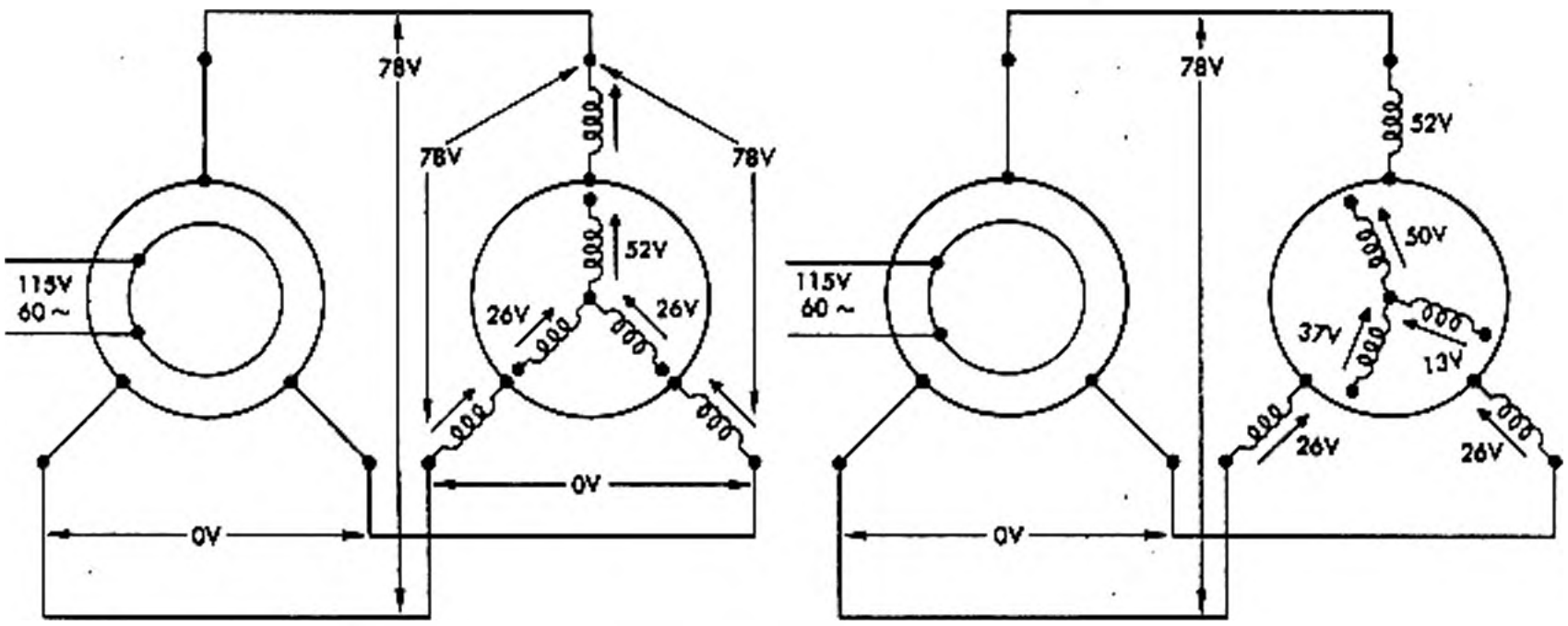


Figure 5-31. Transformer Action in Differential Selsyn

acts as the primary; the rotor, as the secondary of a 1 to 1 transformer. Because of the air gap, more turns are wound on the rotor coils than on the stator coils to achieve the 1 to 1 ratio. The stator must always be used as the primary and the rotor as the secondary of a differential selsyn.

The transformer action of a differential selsyn is illustrated in figure 5-31. Assume that both rotors are in the electrical zero position, the stators are connected as shown, and the rotor of the differential selsyn is open. Since the stators are connected together and are in parallel, the voltages in both sets of stators are equal to each other and are in phase. The stator voltages of the differential, through transformer action, cause similar voltages to be induced in each of the rotor windings.

If the rotor of the differential were turned to the 15° position and the rotor of the selsyn generator remained fixed, the stator voltages would be equal as before. The voltages induced in the rotor coils of the differential are less because all of the flux from each stator winding no longer cuts its corresponding rotor winding.

A differential selsyn receives two inputs. One is electrical, and the other is mechanical. It subtracts one input from the other and transmits the difference.

Potentiometer Pickoffs

A potentiometer is a tapped resistor in which the position of the tap can be altered by some type of mechanical control. The standard high resistance potentiometer used in communications circuits uses a resistance element consisting of a thin film of carbon deposited on some insulating material. This element is not practical for servo use because the resistance of such a carbon film changes with wear and with variation in temperature and humidity. Since a more rugged type is required, the potentiometers used in servos generally use resistance elements consisting of a number of turns of resistance wire.

A potentiometer pickoff can be mounted in any position and works well in bridge-type control networks. It usually has a large loading effect upon the network. It also lacks sensitivity for minute variations in displacement which limits its application to certain sensors.

The degree of complexity of the bridge network depends upon factors other than the pickoff. All are variations of the same basic bridge circuit shown in figure 5-32.

An output voltage is not developed when the potentiometer (AB) is arranged in a bridge network, and the wiper is at position C with the resistance equally divided on each side of the wiper. An output voltage is developed when the wiper is displaced some distance (X) to position C_1.

Since the potentiometer is attached to the sensor so its displacement is the distance "X," the output voltage varies directly with the displacement.

CONVENTIONAL POTENTIOMETERS. Most potentiometers used in servos are essentially the same, consisting of wire resistors and movable sliding contacts. In the conventional type wire-wound potentiometer shown in figure 5-33, the resistance wire is wound on an insulating strip. The strip, together with the wire, is called a resistance card. The card is bent so that it forms an arc of nearly 360° and is mounted so that the slider makes contact with the winding along one edge of the strip. The slider usually can rotate

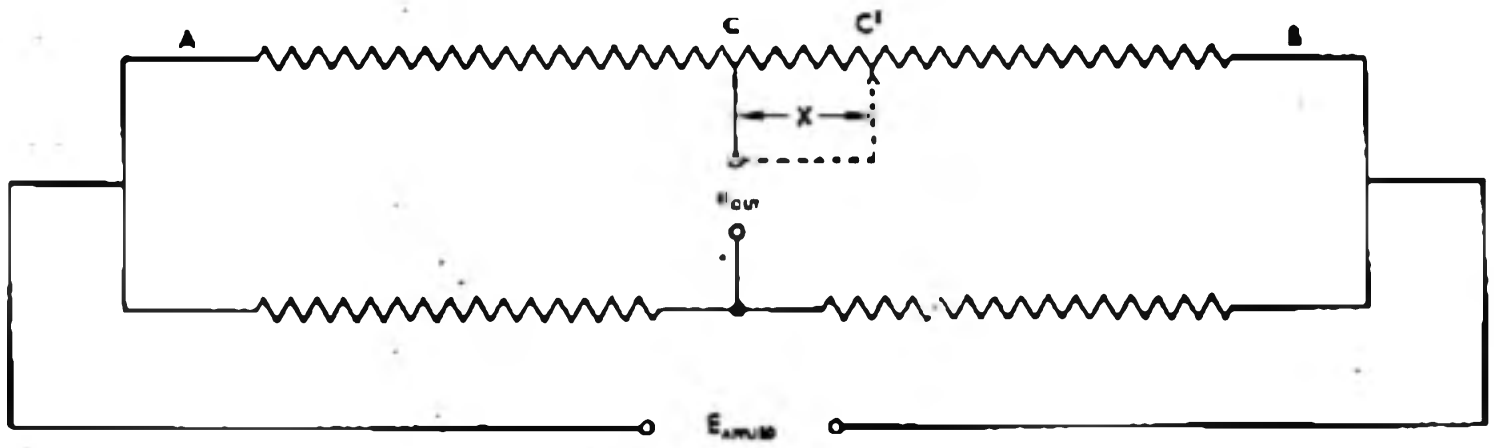


Figure 5-32. Basic Bridge Network using Potentiometer

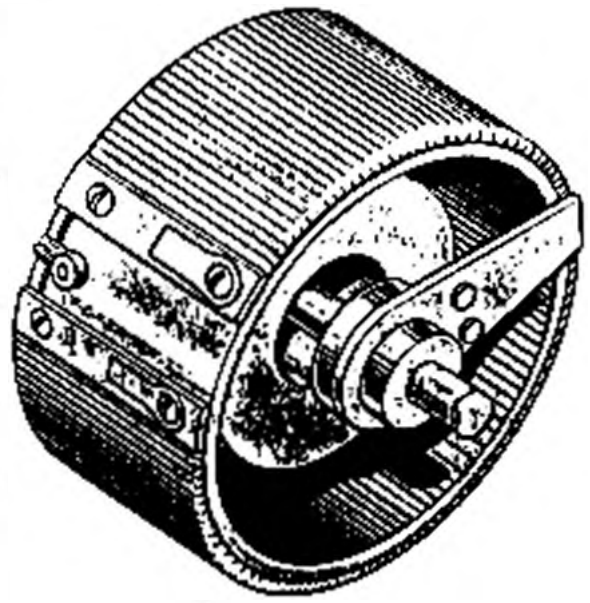


Figure 5-33. Wire-Wound Potentiometer

continuously, but it loses contact with the resistor over a small arc.

In servo usage, a potentiometer should be thought of as a variable voltage source rather than as a variable resistor since an input voltage is applied to the resistance element and a fraction of this voltage appears as the output between the slider and one end of the resistance. Unfortunately the output of a potentiometer does not change smoothly as the slider is moved. Instead, the output voltage changes in jumps, each jump being equal to the voltage difference existing between adjacent turns of wire. A potentiometer having 1000 turns of wire on the resistance element is said to have a resolution of 1 part in 1000 or a resolution of 0.1 percent. A resolution of 0.1 percent means that the smallest change in the output voltage is 1/1000 of the input voltage.

To improve the resolution, the resistance element is sometimes wound in a helix as shown in the helipot illustrated in figure 5-34. The slide may make as many as 10 turns in covering the whole element, and although this construction allows many more turns on the resistance element, it has the disadvantage that the slider cannot rotate continuously. Since the change in voltage along the resistance element is proportional to the change in resistance, it is simple to make a nonlinear potentiometer by using a nonuniform resistance card.

INDUCTION POTENTIOMETER. An induction potentiometer is similar to a selsyn but has only a single rotor winding and a single stator winding. By shaping the pole pieces, it is possible to make the output voltage of such a unit proportional to the shaft angle rather than the sine of the shaft angle over a limited range. The induction potentiometer is more complicated than a conventional potentiometer and can be used only with alternating current. It has higher resolution than most wire-wound potentiometers.

MICROSYN. A microsyn, as shown in figure 5-35, is used occasionally as an induction potentiometer. There are no electrical connections to the rotor; consequently, there are no brushes. The magnetic field of the input windings magnetize the iron rotor. The magnetic field of the iron rotor then induces voltage in the output windings unless the rotor is in its zero position. Over a limited range, the output voltage is proportional to the displacement of the rotor. The microsyn, like

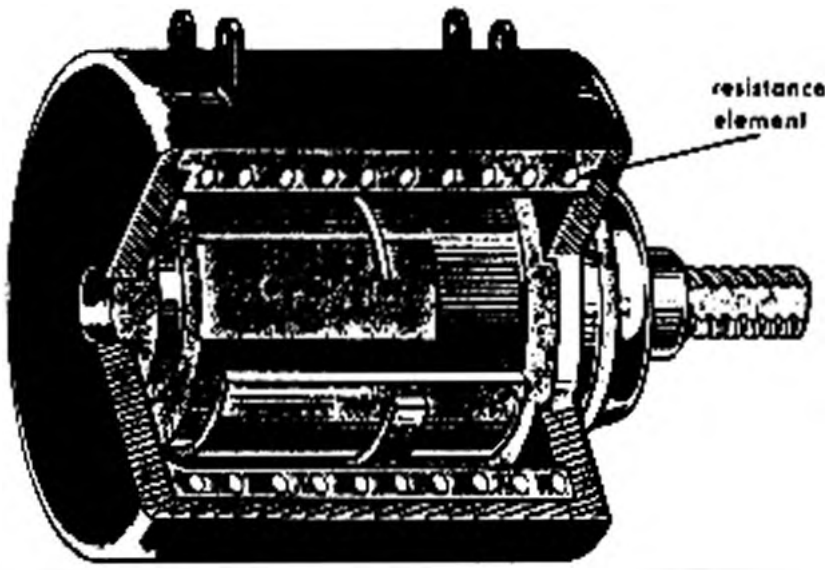


Figure 5-34. Helipot (showing resistance element wound in helix)

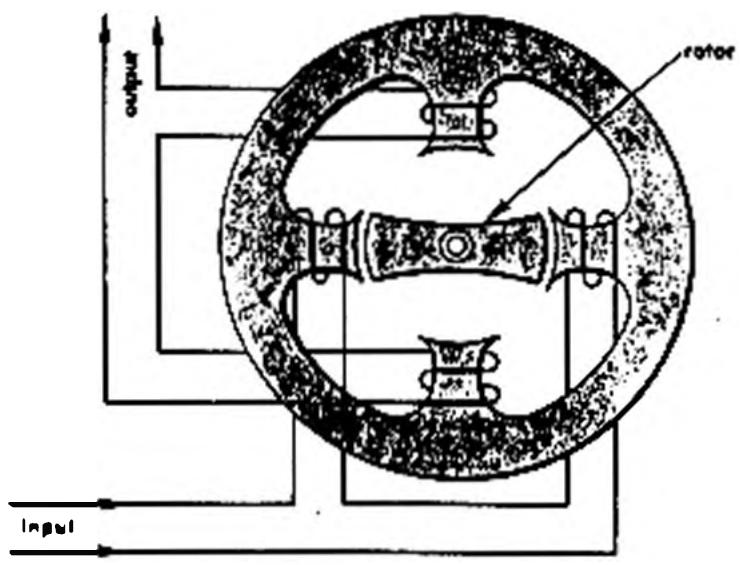


Figure 5-35. A Microsyn

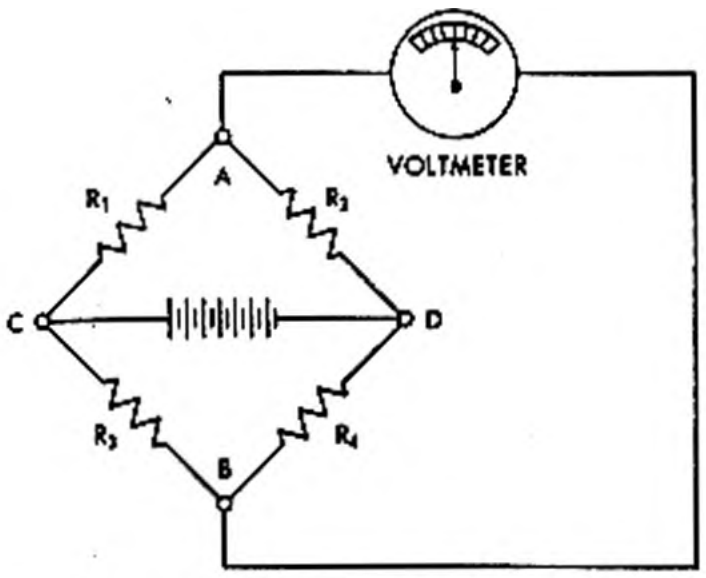


Figure 5-36. Simple DC Resistance Bridge Circuit

other magnetic elements, requires alternating current.

POTENTIOMETERS IN BRIDGE CIRCUITS. Potentiometers are sometimes used in bridge circuits to derive a signal proportional to direction and magnitude of deviation in all three control axes. For a brief analysis of resistance bridges, refer to the bridge in figure 5-36. Equal resistances R_1 and R_2 are connected in series across a DC voltage source. Also connected in series across this voltage source are equal resistances R_3 and R_4. R_3 and R_4 are in parallel with R_1 and R_2. Since R_1 and R_2 are of equal value, applied voltage is divided equally across them. In a similar manner, the voltage drop across R_3 is equal to that across R_4. In this circuit then there is no voltage difference between points A and B.

The bridge circuit is essentially the same if the resistances are replaced by two potentiometers, one in each branch of the circuit as shown in figure 5-37. When the potentiometer wipers are placed on the exact electrical centers of the resistances, the voltmeter shows no voltage difference between points A and B. When wiper A is moved to the left, as shown in figure 5-38A, the bridge becomes unbalanced, and voltage difference exists between wiper A and wiper B. The amount of voltage recorded is in proportion to the distance which the wiper has moved across the winding, provided the resistance of the winding is distributed uniformly from one end to the other. When a balanced bridge network is used for control purposes, the voltage which is developed at the wiper when the bridge is unbalanced is called a control signal.

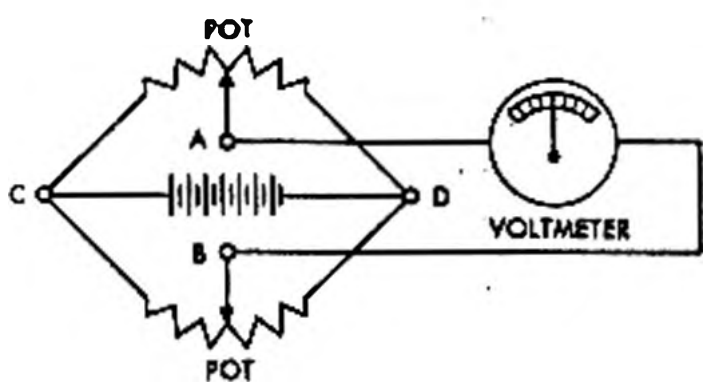


Figure 5-37. DC Bridge Circuit with Potentiometers

If the potentiometer wiper B is moved across the winding in the proper direction and the proper amount, the bridge circuit can be rebalanced without returning wiper A to its original position. This fact is demonstrated in figure 5-38B. Potentiometer B has produced a control signal in the opposite direction, and this signal has balanced out the signal supplied by potentiometer A. The polarity shown by the voltmeter when the bridge is unbalanced depends on the direction of movement of potentiometer wiper A.

The simple bridge circuit described in the previous paragraphs is energized by direct current. The direction of the control signal is indicated by

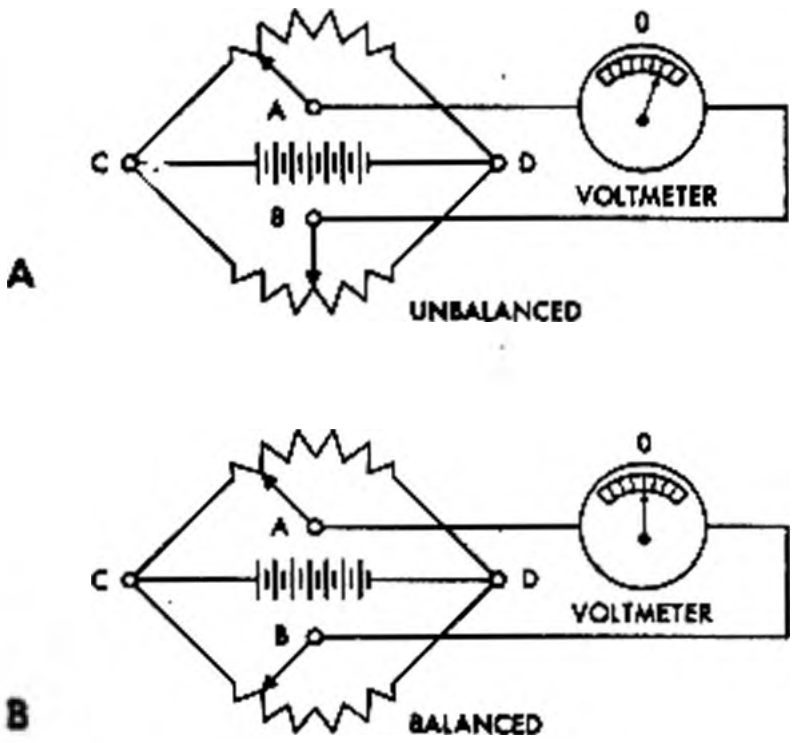


Figure 5-38. Unbalanced Bridge Supplying Control Signal and Rebalanced Bridge

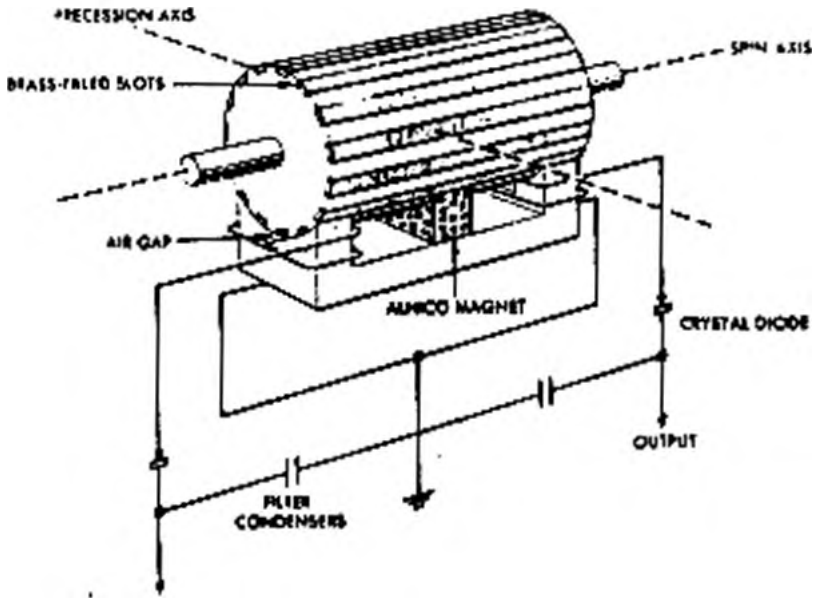


Figure 5-39. Internally Operated Reluctance Pickoff

the polarity at the voltmeter. In actual practice, it is more practical to energize the bridge with alternating current. The direction of the control signal in an AC bridge circuit must be determined by the phase of the alternating current.

Variable Reluctance Pickoffs

Variable reluctance pickoffs have been used extensively. Such pickoffs have been used with sensors which have small displacement values or oscillatory movements. They give the largest output per unit displacement and with a minimum loading effect, but they also require stable oscillators, making space and weight demands on the missile.

A number of types of reluctance pickoffs have been developed. Since reluctance pickoffs require an oscillating voltage for their operation, they can be classed by the method used to supply this voltage. An internally oscillated pickoff is illustrated in figure 5-39. The figure shows a rate gyro rotor and its reluctance pickoff. The pickoff consists of an E-shaped metal mass with coils wound around its extremities and a permanent magnet located in the center. The gyro rotor is a ferrous material which has been slotted and the slots refilled with brass to restore the lost weight.

The magnetic force set up by the magnet causes a flux to flow through each end of the pickoff mass, through the gyro rotor, and back into the magnet. The gyro rotation causes regular variations in the reluctance of the flux paths as the brass or ferrous metals pass over the end pieces. This causes a regular variation in flux density to be established, inducing an AC voltage in the coils. The voltages in the coils cancel each other out.

When the gyro precesses, the air gaps at each end vary oppositely in proportion to the acceleration and cause different voltages to be induced in the coils. The difference in these two voltages is then proportional to acceleration, and after being rectified the difference is the signal voltage. The output phase shows up here as a difference in polarity.

Another type of reluctance pickoff, shown in figure 5-40, consists of a stator containing two pairs of coils mounted in space quadrature. One pair of coils is supplied with a constant amplitude AC voltage from a reference oscillator. Voltage is induced in the second pair of coils through a split iron core rotor coupled to the gyroscope gimbal. As the angle between the rotor and the pickoff coils changes because of variations in missile attitude, the amplitude of the AC voltage induced in

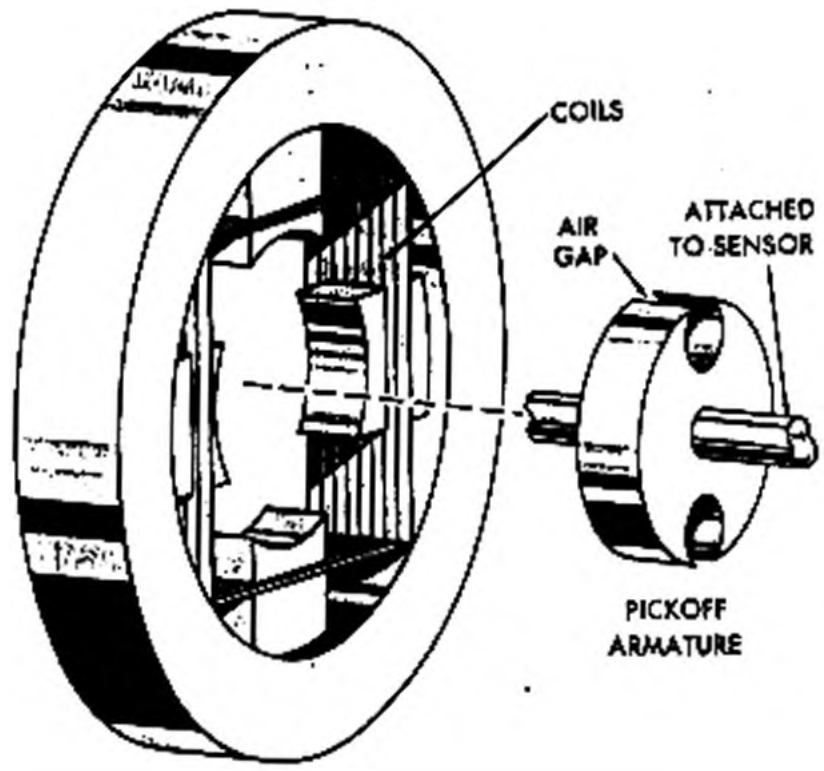


Figure 5-40. Externally Operated Reluctance Pickoff

the pickoff changes proportionally. The phase of the voltage induced differs by 180°, depending upon the direction of the missile attitude change from the zero position.

Capacitance Pickoffs

This presentation of pickoffs is concluded with a brief mention of capacitance pickoffs. A variable capacitance pickoff, shown in figure 5-41, makes

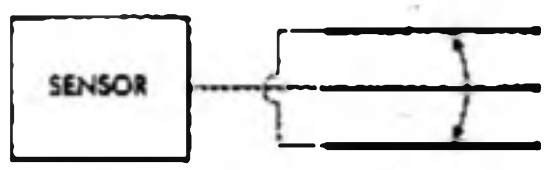


Figure 5-41. Capacitance Pickoff

use of a movable capacitor plate located between two fixed plates. Upon movement of the center plate by the sensor unit, the capacities of the two condensers are varied and their output, when wired into the appropriate circuit, is useful. This pickoff is extremely sensitive.

SENSOR UNITS: THE SECOND REQUIREMENT OF MISSILE CONTROL SYSTEMS

This section has covered the second requirement of a missile control system — sensor units. Gyroscopes, altimeters, and transducers provide a means of detecting error information, any departure of a missile from the desired reference.

The next section is concerned with computer units which help to satisfy the third requirement of an automatic control system. Computers, along with other types of units, convert error information into a form that can be used to regulate the controlling devices of a missile.

SECTION B

Computer Units

Up to this point, methods of detecting various missile deviations from desired conditions have been presented. These deviations are detected by means of the sensor block operating with the reference block. The output of a sensor is referred to as an error, which represents an existing missile deviation from desired conditions. The most important deviations are changes in missile attitude. Other conditions — such as changes in missile airspeed, altitude, angular acceleration, or forward and sideward acceleration — also can be detected. Deviations from a desired condition may be caused by changing winds, changing air pressure, and/or changing engine thrust.

Although the output of a sensor represents a situation which must be corrected, this output seldom goes directly to the actuator which effects a missile change. The error signal must first be "operated" on and amplified. This operation consists of changing the signal to represent the additional information required for proper control functioning. The operation is represented by the "computer" block which follows the sensor block in figure 5-3.

A computer in a control system is comprised of three general types of components, which are mixers, integrators, and rate components.

MIXERS

A mixer is any circuit or device which combines information from two or more sources. Every missile control system, which operates as shown in the basic block diagram in figure 5-3, must combine several signals. A followup signal is combined with the sensor error signal. A rate or integral signal, which is produced within the computer, is often combined with the sensor error signal. In some missiles several inputs are mixed together. This mixing occurs when the required aerodynamic reaction depends on more than one variable. For example, a system may be controlling the pitch of a missile. Pitch can be influenced by missile attitude, altitude, and airspeed. A separate sensor and reference is used to detect each of the three variables. All of these signals are mixed together so each variable has the required influence on the pitch actuator.

Requirements of Mixers

To function properly, a mixer must combine signals in the correct proportion, phase, and amplitude. If a followup signal is combined with an attitude signal, the followup signal must be of a certain strength as compared to the error signal. This proportion can be indicated by means of weighting factors. Such a means determines the

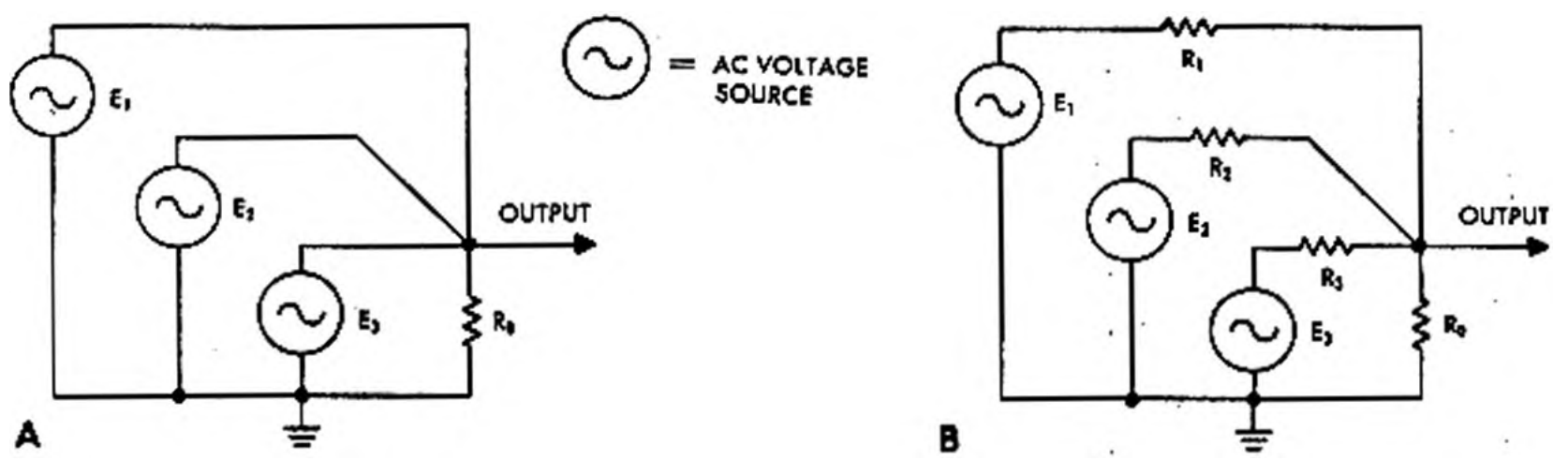


Figure 5-42. Parallel Mixer Circuits

comparative weight or influence of a signal. Suppose that three signals — X, Y, and Z are to be summed and that only one-third of Y and one-half of Z should be used as compared to the strength of signal X. One-third of Y and one-half of Z are weighting factors.

A sensor signal must produce action in the desired direction. The phase of a followup signal must be such as to have a counteracting effect on the output. The total output must be of proper amplitude for the right amount of control. Normally the signals either add or subtract, although other functions can be performed. These processes which are necessary for proper mixing can be considered a computer function.

The type of mixer used depends mostly on the type of control system. For example, most systems are basically electronic; therefore, most mixers are electronic. Mixers also may be mechanical, pneumatic, or hydraulic. These types are explained on the following pages. A knowledge of different types of mixers will help you to recognize a mixer function in missile circuitry. Such knowledge also will help you to analyze mixer malfunctions which may occur.

Electronic Mixers

Fundamentally, the principles of any electrical mixer reduce to basic network and vacuum tube theory. Mixers consist either of an impedance network of resistors, inductors, and capacitors or of several vacuum tube stages. The information to be combined is represented by the amplitude and phase of each voltage.

Mixers combine voltages from such sources as the output from pickoffs, rate components, integrators, followup generators, or guidance circuits. Whatever the sources, they are indicated in the following diagrams by the standard symbol of an AC generator.

Electric signals in a control system are usually AC. These signals are then changed to DC before they are applied to the actuators which move the control devices.

The phase of the AC signals determines the sense. If a certain two signals must make the same effect on a rudder, they must be in phase in the mixer. Occasionally, capacitors or inductors are necessary to alter the phase of signals so that they will mix in exactly the correct phase. If DC signals are used, they must be mixed with the proper polarity.

PARALLEL MIXING. The simplest method to combine voltages is to apply them across a common impedance by means of a junction. The voltage sources are effectively in parallel as shown in figure 5-42A.

This circuit is normally altered in actual practice. The change consists of an added impedance in each branch as shown in figure 5-42B. These added resistors reduce the voltages applied to R_e. In analyzing a circuit, such resistors usually are called attenuators.

Actually, there are three possible reasons for these resistors.

- First, they reduce the current flow from the sources when an increase in the resistance of R_e is not feasible. This reduction of load current decreases the power consumption and sometimes

makes the output of the pickoffs (voltage sources) more linear.

• Second, these resistors allow the adjustment of the amount of signal which should be applied to R_o. Such adjusting makes possible the mixing of the signals in the proper ratio (or with the proper weighting factors), a condition which is important. To alter a pickoff output by using a series resistor of a certain value is often easier than to acquire a pickoff with a certain output.

• Third, these added resistors reduce errors caused by coupling between the signal sources. This coupling is produced by current from one source changing the voltages which would normally be produced by the other source. Such an error occurs from a source in which the internal impedance varies, such as a potentiometer pickoff or an inductance pickoff in which the inductance field builds up with current flow. The cause of this error can best be understood after a review of the superposition theorem.

The operation of a circuit with many voltage sources generally is much harder to visualize than the operation of one with just a single generator. One of the best ways to visualize the former type is by the use of a standard electrical law of operation, the superposition theorem: If a network of linear impedances is energized by two (or more) generators, the current or voltage at any specified point in the network is expressed as the sum of the voltages or currents that each generator would produce if it alone were connected to the network and if the other generator were replaced by its internal impedance. This theorem can be applied in analyzing the circuit in figure 5-43.

The circuits in figure 5-44 are a breakdown of the circuit in figure 5-43. The breakdown will help

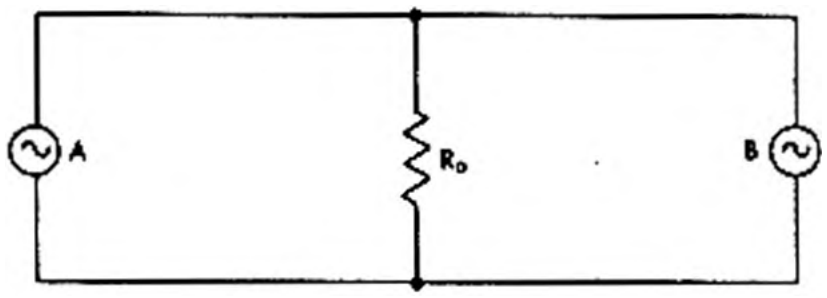


Figure 5-43. Simple Network with Two Voltage Sources

you in considering separately currents and voltages from each source. The electron flow indicated by the arrows results from assuming an instant when the polarity of the AC generator is as shown. In each case the flow represents current due to a single voltage. When the circuit is considered with both voltage sources, the actual current in each path is the vector sum of the components that would result from the sources considered separately. The vector sum is the result produced when voltages combine in a manner that depends upon the amplitudes and phase differences.

The application of the superposition theorem proves that the circuit in figure 5-43 can be used as a mixer and that the series resistors can be used to adjust the ratio of the signals. The total voltage across R_o depends on the total current. Since the total current is the vector sum of the branch currents, the output voltage depends not only on the strength of the voltage sources but also on the impedance in each branch.

As mentioned before, changes in internal impedance of pickoffs which are interconnected may result in signals of incorrect amplitude. Suppose a potentiometer pickoff signal is combined with a signal from source A as shown in figure 5-45. The current developed from source A depends on the total impedance in the circuit. When this impedance is varied by movement of the potentiom-

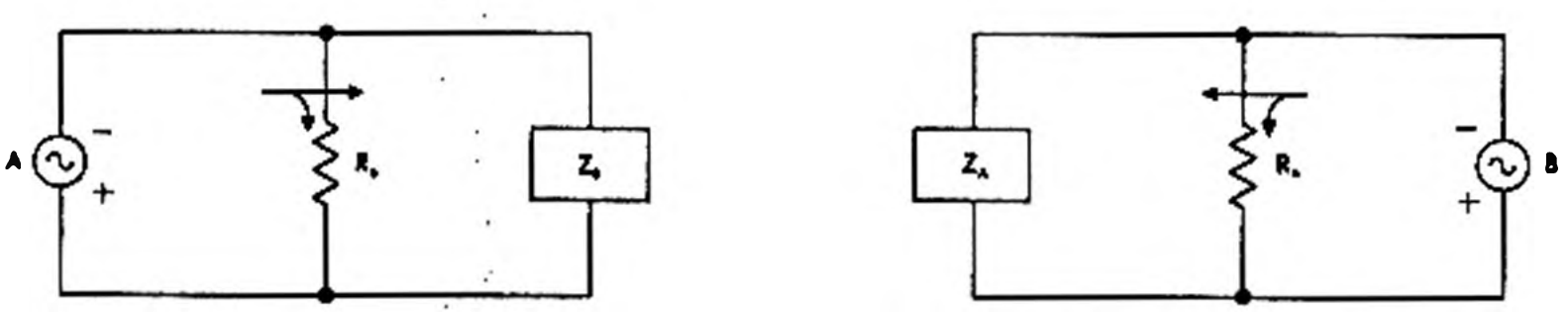


Figure 5-44. Circuit Breakdown (considering each applied voltage separately)

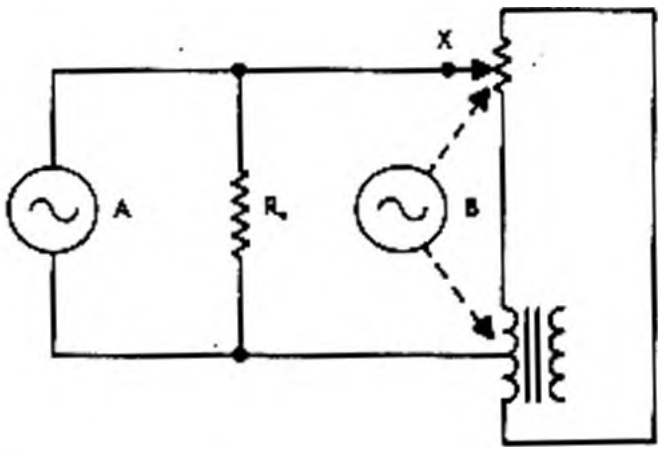


Figure 5-45. Circuit with Variable Impedance Input

eter, the current from source A varies. The varying of the current changes the source A voltage applied to R_o, even though the cause of the change does not represent a change of the information represented by voltage A.

If a resistor were inserted at point X, the current variation due to movement of the potentiometer would be reduced. There would be less effect on the signal voltage at source A.

SERIES MIXING. Voltages also are combined by applying them in series with a common load, as shown in figure 5-46A, basic series-type signal

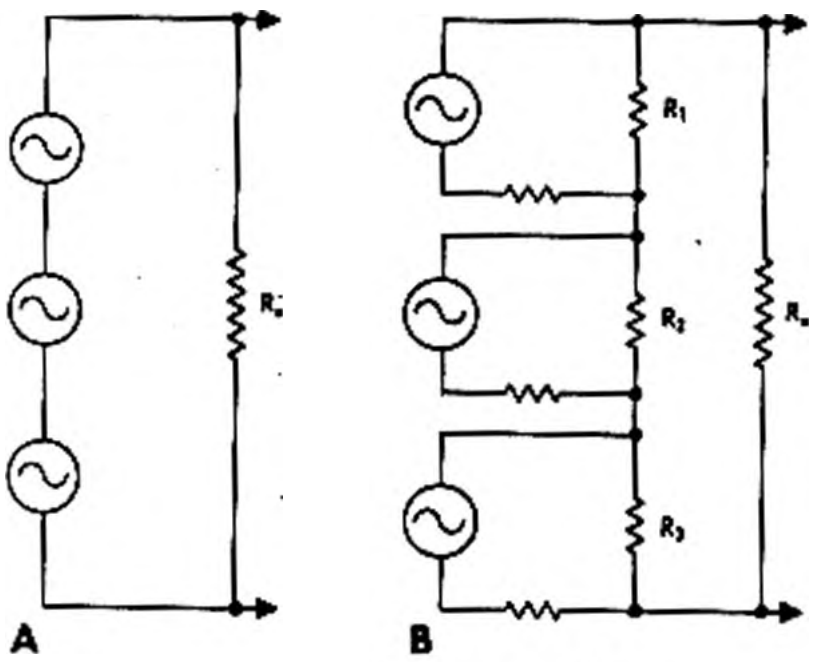


Figure 5-46. Series Mixer Circuit

mixer circuit. A disadvantage of this method is that all the load current must flow through each source. This fundamental circuit is normally altered as shown in figure 5-46B. The resistors are included for the same three reasons as in the case of parallel-type mixers.

The sum of all voltage drops and rises around a closed loop must equal zero. In the basic circuit, the voltage across R_o is the sum of the generator terminal voltages. In the actual circuit, voltage across R_o is the sum of the voltages across the resistors labeled R.

RESISTANCE BRIDGE MIXER. A resistance bridge network mixes position information from two sources. Suppose that position information of a gyro and a control surface must be detected and combined. Two sliding contacts are used in the bridge, as shown in figure 5-47. One contact is

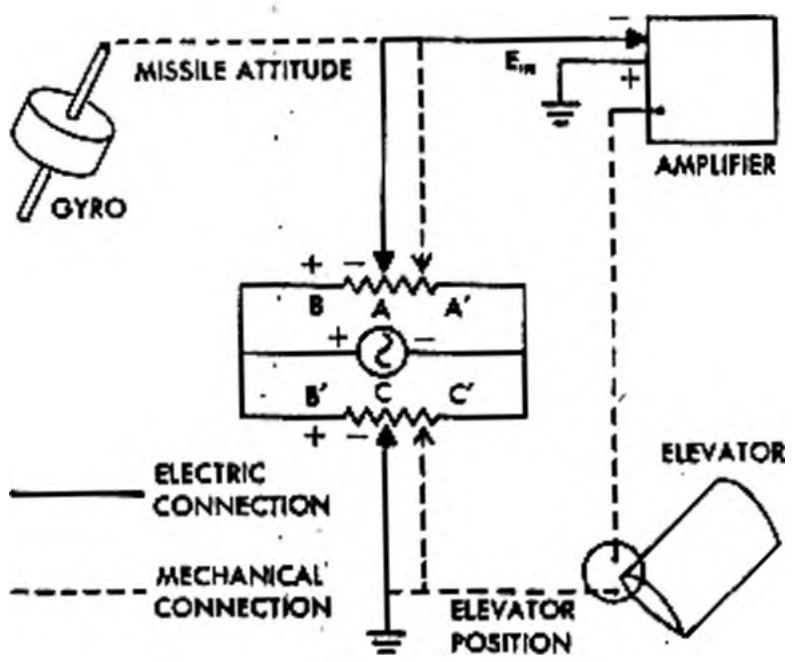


Figure 5-47. Mixing Two Signals using Resistance Bridge

mechanically connected to the gyro gimbal and one to the control surface. Physically, the potentiometers may be located quite a distance apart. The bridge serves the purpose of two electrical pickoffs and a mixer.

The purpose of the circuit is to mix a voltage proportional to gyro displacement with a voltage proportional to control surface position. These two particular signals are often combined in systems since one signal represents the input error signal, and the other is the followup signal. The potentiometers are adjusted so that the bridge is balanced if the missile is in the proper attitude and the control surface is streamlined. No signal is sent to the amplifier to cause a correction. Assuming an instant when the polarities of the AC voltages are as shown, the voltage across AB is exactly cancelled by BC, so E_{in} is zero.

If the missile deviates, a voltage proportional to the distance the upper slider moves (AA') is applied to the amplifier (assuming a linear pot and no load current flowing to the amplifier). At that instant, the voltage is negative. If the missile is in the correct position but the control surface is not yet streamlined, a voltage proportional to the distance the lower slider moves (CC') is applied to the amplifier. It is positive at that instant. Normally, both actions occur at the same time, and the voltage applied to the amplifier is the resultant of the two voltages. In this application, the mixer performs a subtraction since the signals caused by the displacements tend to cancel out. The resultant E_{in} is negative since the deflection of the upper pot is shown greater than that of the lower pot. (The control surface would then be deviating further from streamline.) That the resultant would be negative can be shown by the voltage equation for a closed loop. Starting at the grounded input terminal, $E_{in}-E_{ab}+E_{bc}=0$. The weighting factors depend upon the mechanical linkages and on any resistors which are added to the bridge circuit.

Impedance networks often are used to divide a signal among more than one load. The same problems and network analysis apply to the separation of signals as apply to the combination of signals. The problem becomes more involved when these loads are fed by other signals since a network must be designed to "decouple" the unwanted voltages from each load.

TRANSFORMER MIXING. Transformers may be used to mix two AC signal voltages. If the primary of a transformer is placed in series with two AC sources, as shown in figure 5-48, the output of

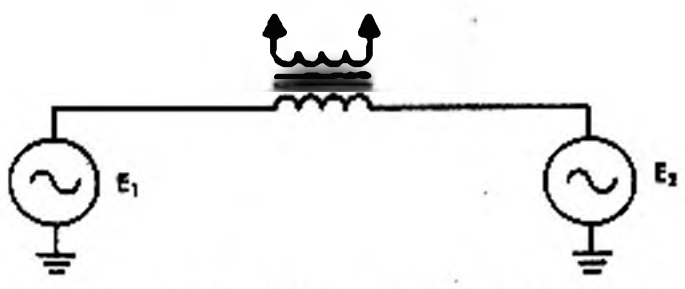


Figure 5-48. Series Mixer using Transformer Output

the secondary will be proportional to the vector sum of the source voltages. Assuming a perfect transformer (no losses) and a 1-to-1 turn ratio, the secondary voltage will be equal to the vector sum of the two input voltages. If a higher secondary voltage is desired, a step-up transformer may be used and voltage gain as well as mixing may be accomplished.

Suppose, in figure 5-49, a phase check is made

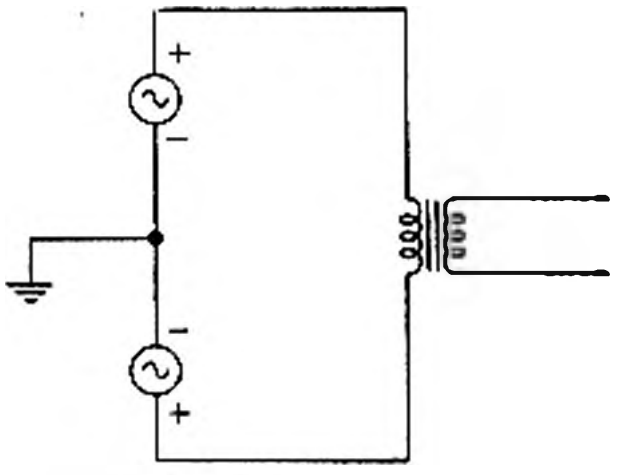


Figure 5-49. Series Mixer with Opposing Inputs

and the instantaneous polarities are in phase with respect to ground. Would the voltages add or subtract across the primary? Examining the circuit you will find current from the upper source will tend to flow counterclockwise around the circuit and the current from the lower source will tend to flow clockwise. The two voltages are 180° out of phase in respect to the primary winding, and the voltage and polarity will depend upon the difference of the two voltages and upon which one is the larger. If the polarity of one of the sources is changed by 180°, the primary voltage will be the sum of the two sources.

Mixing action may be accomplished by applying one voltage to the primary and another voltage to the secondary. This is especially useful when a DC component is present in one of the signals. The signal with the DC component is applied to the secondary winding. The secondary voltage is proportional to the vector sum of the primary and secondary voltages.

A transformer does not isolate signals as well as some other mixing devices because of mutual coupling between primary and secondary windings. A current flow in the primary induces voltage in the secondary, but current flow in the secondary will also induce a voltage in the primary which affects the source. Secondary loading is an important consideration when using a transformer mixer.

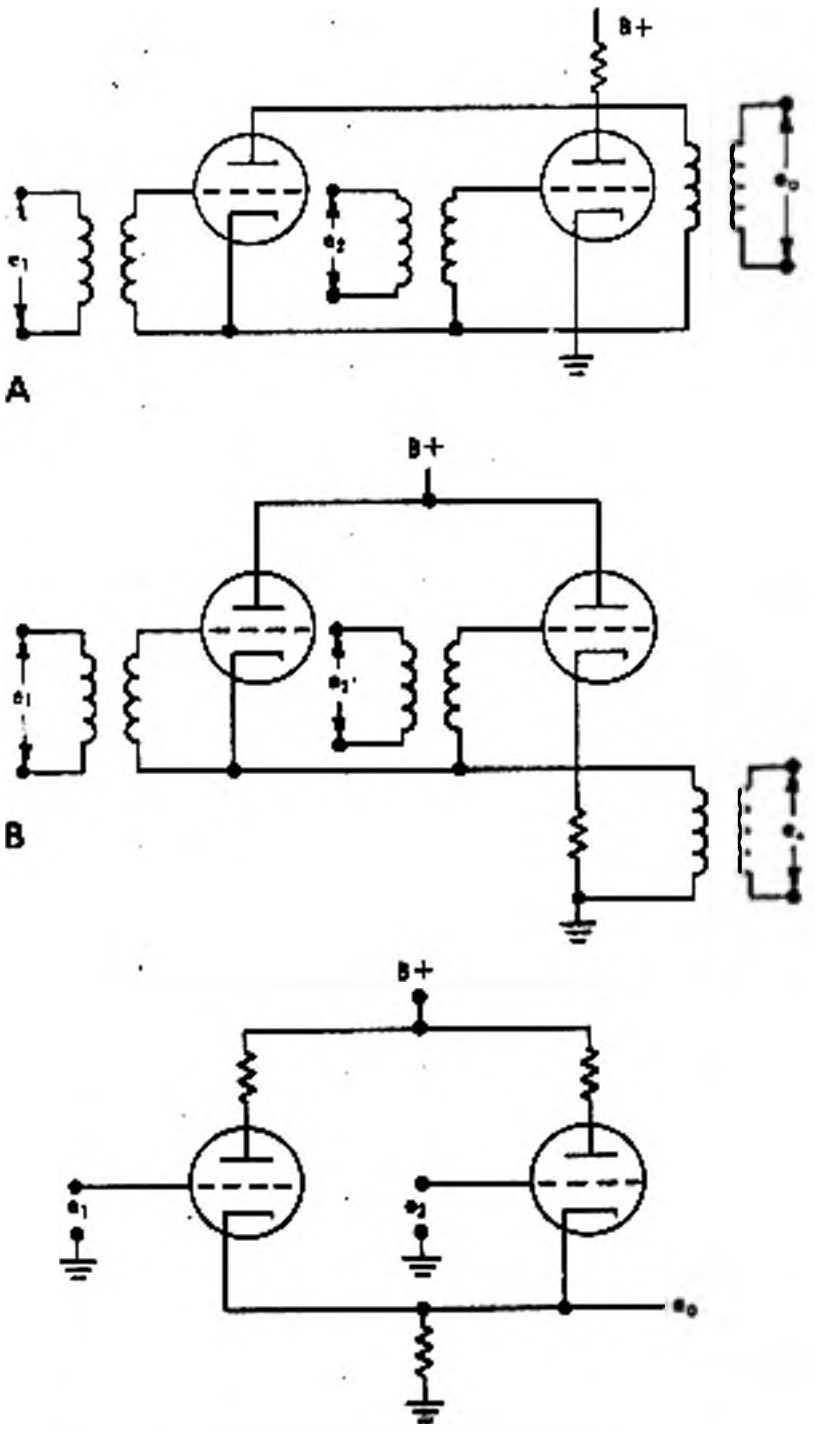


Figure 5-50. Parallel-Type Vacuum Tube Adders

VACUUM TUBE MIXERS. A method used to eliminate coupling between voltage sources is to apply the signals to vacuum tubes within the amplifier circuit. If the input impedance is high, very little current will flow from the source. There is little loading effect. Because feedback impedance of the tubes is very high at the low frequencies used in flight control systems, little or no feedback voltage is felt at the source. As a result, the sources are well isolated from each other and from the output.

Figure 5-50A shows a typical vacuum tube adder for mixing signals from two sources. If the tubes have the same characteristics, each signal will have equal influence on the output because of the common plate load resistor. Weighting factors can be introduced by using potentiometers at the input. Amplification as well as mixing is accomplished in this circuit.

When voltage amplification is not necessary, the cathode follower circuit shown in figure 5-50B is used. This circuit will have an even higher input impedance and a much lower output impedance than the circuit shown in figure 5-50A. It may also be used to match the impedance of the next stage to transfer a maximum amount of power. A potentiometer at the input or separate plate resistors may be used if weighting factors are necessary.

The parallel amplifier circuit can be altered to produce a sum and difference output. In the differential amplifier mixer in figure 5-50C, the output voltage, e_o, at the cathodes is proportional to the sum of the input voltages. The voltage between the plates is proportional to the vector difference of the input voltages.

Another method of vacuum tube mixing is the combining of signals within the tube itself. Several signal voltages can be impressed upon different electrodes in the tube to vary the plate current, as shown in figure 5-51. The effect of the plate current variations resulting from one signal has little effect on the other voltage sources because the impedances between the electrodes of any multigrid tube are very high.

The number of inputs that may be applied to a tube is limited by the number of electrodes. For a triode tube, the maximum would be two. The pentode shown in figure 5-51 could have three inputs. A pentagrid tube might have as many as

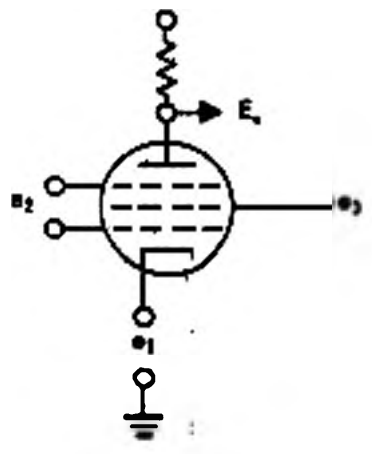


Figure 5-51. Multigrid Mixer

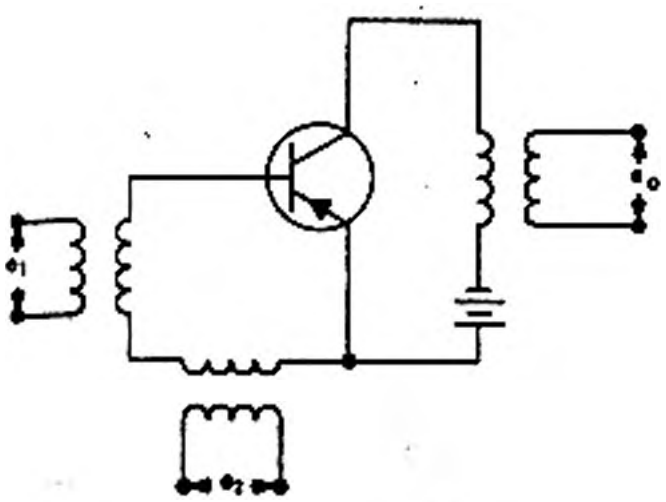


Figure 5-52. Transistorized Mixer

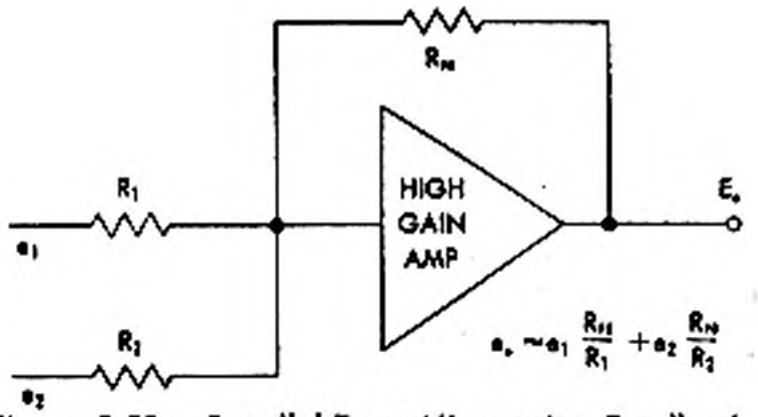


Figure 5-53. Parallel-Type Mixer using Feedback Amplifier

five inputs. The weighting factors depend on the characteristics of the tube and the electrodes used.

TRANSISTORIZED MIXERS. The transistor has several significant advantages over the electron tube. The power efficiency is greater than that of the vacuum tube because the transistor operates with small DC voltage and current and, in addition, does not require heater power. Other advantages of the transistor are its low noise level, small size, and rugged construction.

The transistor is much smaller than even the miniature electron tube and requires less power. The transistor is ideally adaptable to the design of flight control and guidance systems for missiles. This circuitry is especially advantageous for air-interceptor missiles such as the AIM-4 Falcon, AIM-7 Sparrow, and AIM-9 Sidewinder, in which space is so limited. Because the components (resistors, capacitors, coils, etc.) can be so small the entire guidance system for the AIM-4 can be in the space previously taken up by only a power supply chassis. Figure 5-52 is an illustration of a simple transistor mixer circuit. Compare this with the mixers in figure 5-50, and you can see that one transistor can be used to replace two vacuum tubes.

The parallel-type resistance mixer previously mentioned is often fed into a high gain amplifier which is made stable by the use of negative feedback. The feedback amplifier maintains almost a constant ratio between input and output because gain variations are reduced by feedback. The relationship depends almost entirely on the values of the resistors as shown in figure 5-53. The output depends entirely on the ratio of resistance, if the amplifier is assumed to have infinite gain.

Differential Selsyns

So far, only the mixing of information represented by the amplitude of electrical AC or DC

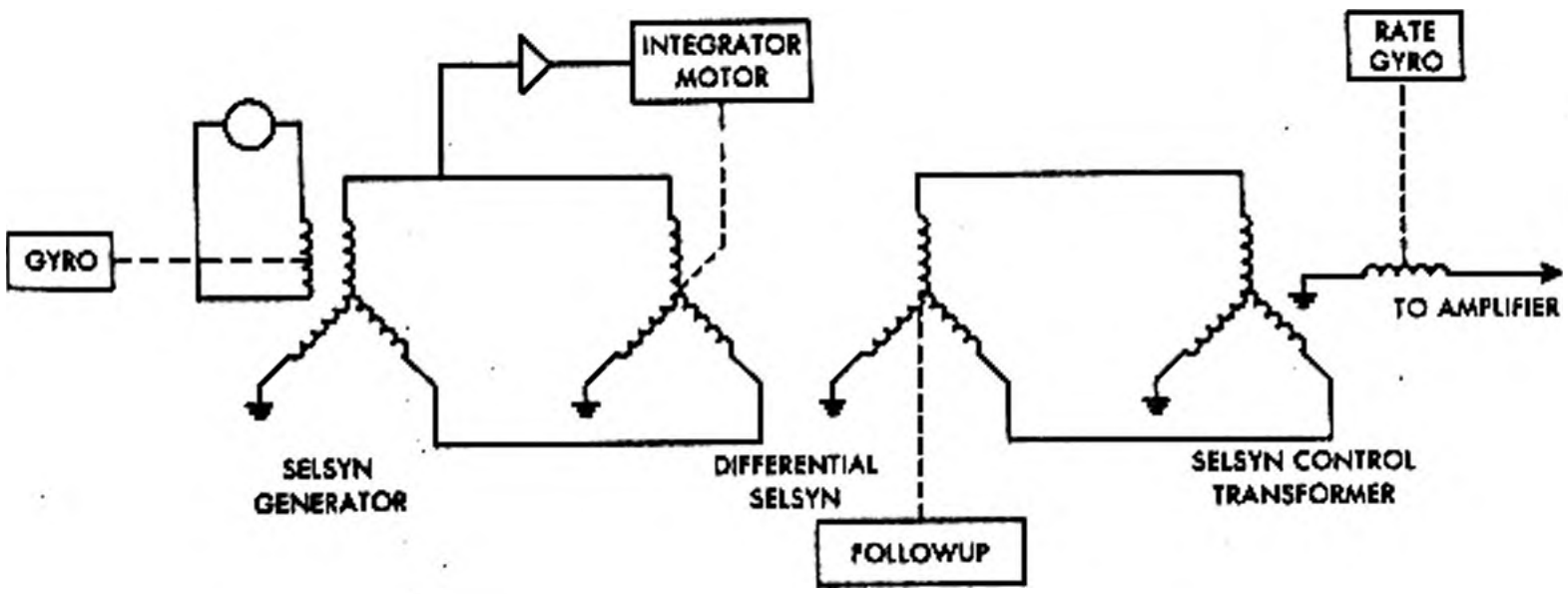


Figure 5-54. Using Selsyns to Combine Information

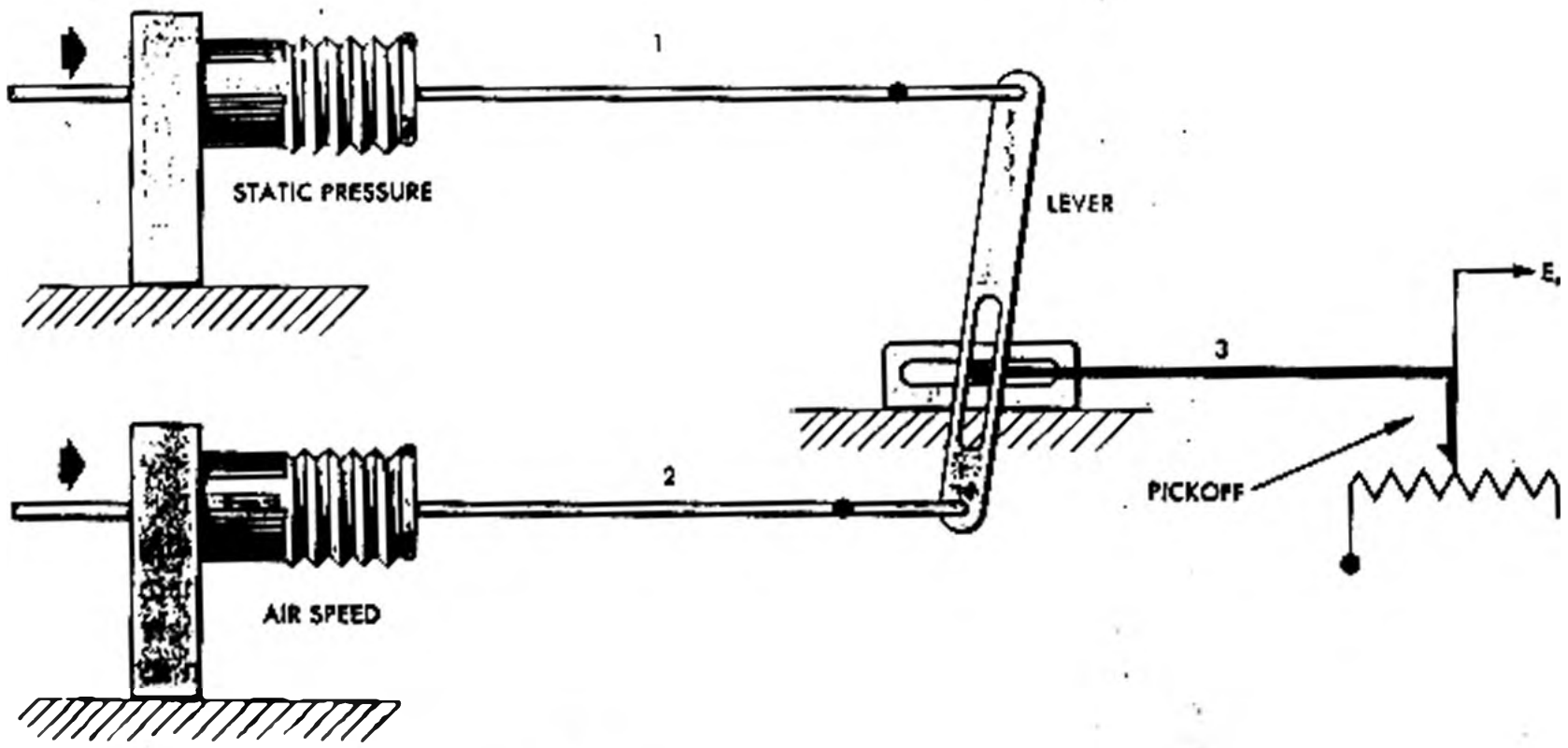


Figure 5-55. Mechanical Mixer

signals has been considered. Information represented by the angular position of shafts can be transferred and mixed by the use of differential selsyns.

A sample part of a control system using differential selsyns for mixing position information is shown in figure 5-54. The gyro locates the rotor of a standard selsyn. Since an excitation voltage is applied to the rotor, a magnetic field is created in a direction depending upon the rotor position. This magnetic field is converted to stator voltages so that the field can be transferred to a differential selsyn. The direction of the magnetic field in the final selsyn depends on the original selsyn position and the position of both differential selsyn windings in relation to the final selsyn. The position of four devices has contributed to the input to the amplifier. The amount of influence from each device is determined by the mechanical linkages. The phase of the position information can be established by the mechanical and electrical connections to the selsyns and differential selsyns. Any reasonable number of differential selsyn windings can be included in a circuit since the weight and power losses are the only limitations. In figure 5-54, the selsyns are positioned to provide a null (zero) output to the amplifier.

Mechanical Mixers

Information can be combined by means of mechanical mixers which are made up of shafts and gears. Lateral positions can be combined by using plain levers as illustrated in figure 5-55. Assume that shafts number one and two operate independently and that the positions represent information which must be combined. The three connections pivot freely. The position of shaft number three represents a weighted average of the other two shafts, because vertical lever arms from shaft number three are not the same length. The sense is obtained by the direction of shaft movement.

The output from shaft number three could operate an electrical pickoff. The lever mixer would then eliminate the need for two pickoffs for converting the shaft positions to an electrical signal before mixing. If the inputs to shafts one and two provide enough power, the output could operate a pneumatic or hydraulic valve directly.

Gears can be used to combine position or angular velocity information by means of a standard gear differential of the same type as used in the rear axle of an automobile. If the input shafts contain position information, they will move slowly and maintain approximately the same average position. The position of the output shaft

constantly indicates the difference between the two shaft positions. If the information is represented by the speed of rotation of the shafts, the angular velocity of the output shaft represents the difference between the input velocities.

The input shafts can be selected so that the output represents the sum of the inputs rather than the difference. This summation of inputs applies to both position and angular velocity information. The weighting factors can be controlled by changing the gear ratios.

Pneumatic and Hydraulic Mixers

Information sometimes is transferred by means of air or hydraulic tubes. Signals are created by the varying pressure within the tube. Two such signals can be combined by a union of the two tubes into one. An air pickoff controls the amount of airflow. The ratio of signals depends on the adjustment of the pickoff or the valve adjustment of the air pressure prior to the pickoff.

With this brief mention of pneumatic and hydraulic mixers, considerations of mixers is concluded and we move on to integrators.

INTEGRATORS

An integrator is part of the computer section of the control system since it performs a mathematical operation on an input signal. The integrator changes the input to a form that represents new information which is desired for one of several reasons presented later.

Simply stated, the integral of a constant signal is proportional to the amplitude multiplied by the time this signal exists. For example, suppose the output of an integrator is 4 volts, and it is produced by a constant input signal which lasts for 1 minute. If the same input signal had lasted for only one-half minute, the output would have been 2 volts. This is assuming the integrator is a perfect one. The control integrator is a device for computing the time that an input signal exists. In contrast, the output of a sensor which provides continuous control is normally instantaneously proportional to some missile error. This proportional signal provides one fundamental means of continuous control. Integral control is another method of providing continuous control.

Normally, an actual missile error is not constant. Instead, the amplitude and phase of the error change, depending on missile flight conditions. Even so, the correct integration is proportional to the product of the operating time and the average error during that time. If the phase of the error should change during the integration period, the signal of opposite phase would decrease the final output of the integrator. The integrator can be considered to be a continuous computer since it is always producing a voltage which is proportional to the product of the average input and the time.

The integration of an error with respect to time represents a summation or accumulation of error over a specified period of time. Actually, one definition for the term integrate is: "To indicate the whole of, to give the sum or total of."

Any integrator possesses a lag effect. Assume that a constant error is applied. Although the input signal suddenly reaches a certain value, no time has elapsed, so the output is zero at that instant (assuming the output originally nulled). As the input signal continues, the output signal will steadily increase as in the previous example. The reaction is considered to be delayed.

The graphs for constant error inputs and the output, assuming the integrator itself operates

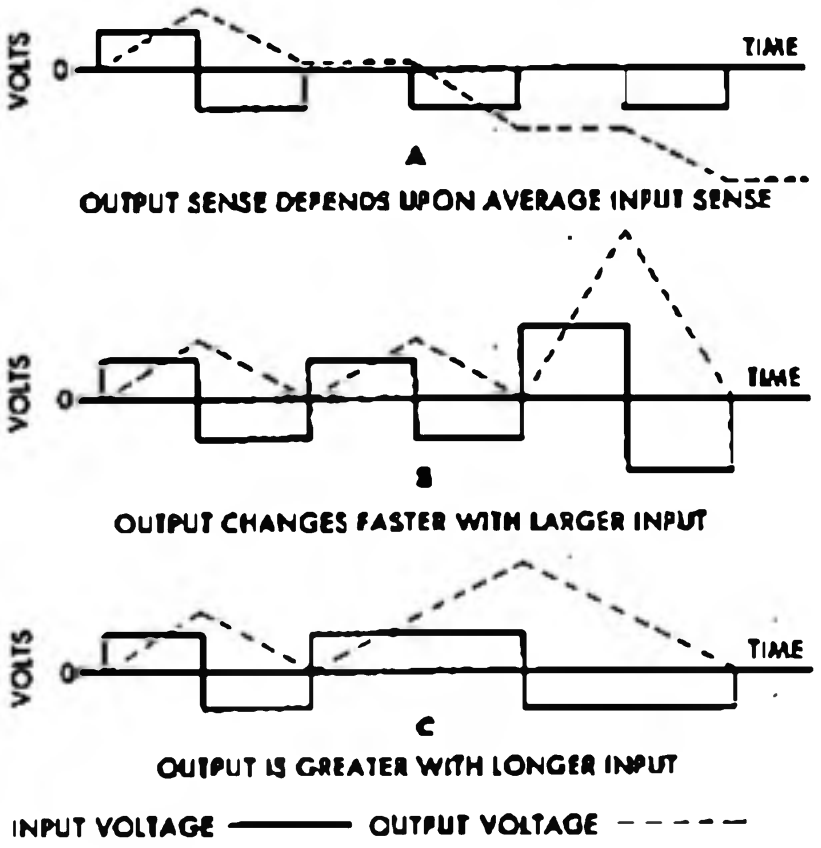


Figure 5-56. Integrator Outputs

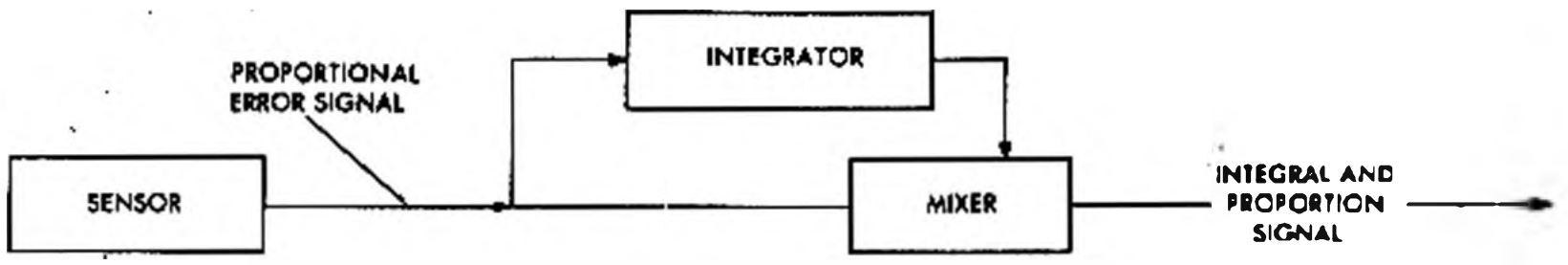


Figure 5-57. Inclusion of Integrator Control with Proportional Systems

without error, are shown in figure 5-56. The lag effect can be observed since time is involved before the output presents an appreciable signal. Time is also required for an appreciable reduction in the output after the input is removed.

The outputs from actual integrators may deviate from the outputs shown in the figure. Such deviation would result from characteristic errors in the particular integrating devices.

An integrator also has a filtering effect on high frequency signals. Suppose a rapidly varying signal with the same amplitude is applied to the input. The output would indicate practically nothing since the period of the input signal is so short. The device filters high frequency signals since it is insensitive to rapid variations in the error signal.

Some types of integrators are more accurate than others by virtue of the principle of operation. Several problems are involved in producing an integrator which will deliver an output exactly proportional to the integral of the input. First, the integrator must be able to react linearly to any signal level which may be applied to it. If the input signal doubles, the output of the integrator should increase at twice the previous rate. Second, this reaction should occur rapidly. Third, the integrator must have the ability to "remember." Suppose a certain signal is developed at the integrator output and there is no further input signal. For perfect results this same signal should remain at the output. The above three qualities are present to different degrees in integrators, depending mostly upon the type used.

Uses of Missile Integrators

There are situations in missile systems in which a signal proportional to an accumulation of error over a period of time is important. Normally this integrator control signal is combined with a proportional signal by means of a mixer. This process is shown in figure 5-57. The integrator signal either supports the error signal or subtracts from it, depending upon the purpose.

An integrator signal is used to support the proportional error signal to guarantee that sufficient correction will always be made by the system. Such an application appears later in the explanation of the hydraulic-electric system. The degree of control that a pure proportional signal can exert is limited. Overcontrol and undercontrol cause excessive movement of the missile. There are instances when proportional control alone is not sufficient to overcome a strong, steady force (such as a crosswind) deviating a missile. In such a case, the proportional error signal has a steady component which affects the integrator. Since the error is of one phase (determined by the crosswind direction), the integrator output increases with time and augments the proportional signal until correction ultimately takes place. The integrator output remains to supply the necessary correction signal component to overcome the relatively constant influence of a steady crosswind. In this case, the integrator supplies a necessary constant signal.

An integrator signal subtracts from a proportional signal to eliminate an undesirable signal component. The general connections are the same as shown in figure 5-57, except that the phase of the integrator output is reversed. As stated earlier, the output responds to long-lasting errors but is relatively insensitive to rapid variations. When the proportional and integral signals counteract, the result is to reduce or eliminate an undesirable steady component of the proportional signal.

A need for the latter characteristic exists in self-balancing an amplifier output. A signal caused

by unbalance appears as a steady component of the total output and is opposed by the integral of the input. Similarly, any unbalance which arises from a sensor, such as autosyn misalignment or gyro unbalance, can be counteracted. A difficulty arises from the possibility that a desirable steady signal also will be eliminated. In such an application, sufficient feedback is allowed from the integrator output to the input to limit the amplitude of the integrator output.

The slowly varying component of voltage produced by the output can be considered as a reference for the more rapidly varying signals. Since this reference voltage is not restricted to a particular value, integral control is sometimes referred to as floating control.

A further important application of integrators is in inertial-type guidance systems. They are used to compute range information from acceleration or velocity data.

Variable-Speed Motor-Type Integrator

The first type of integrator presented is a variable-speed motor type, shown in figure 5-58. Briefly, the signal to be integrated is supplied to a motor. The speed of the motor is proportional to the input signal. The motor slowly drives a pickoff. The distance the pickoff moves is proportional to the integral of the input signal. Assuming that the pickoff was initially zeroed, the voltage output from it will also be proportional to the integral of the input signal. The diagram shows such a circuit using a linear potentiometer (pot) to detect the integration.

The error signal is first fed to a power amplifier so that the signal will be strong enough to drive the motor at varying speeds. A large gear reduction follows the motor to prevent it from driving the contact to either end of the pot during normal integrator operation. The signal slowly increases with time as the potentiometer is driven further from the midpoint or balanced position.

The motor rotates in a direction dependent on the polarity or phase of the input signal. If the motor is AC, the integration is performed on the envelope of the input signal. In actual use, the error amplitude is not constant but varies and may even reverse phase. Physically, this reversal would produce opposite rotation of the motor.

For example, suppose the integrator suddenly receives a signal of amplitude B. The speed of rotation of the motor is proportional to the input signal,

$$\text{or } V_M = K_1 B.$$

The contact moves to the right or left of the zero position at a speed proportional to the speed of the motor,

$$\text{or } V_C = K_1 K_2 B.$$

The distance the contact moves is the speed times the time:

$$d = V_C t, \text{ or}$$
$$d = K_1 K_2 B t,$$

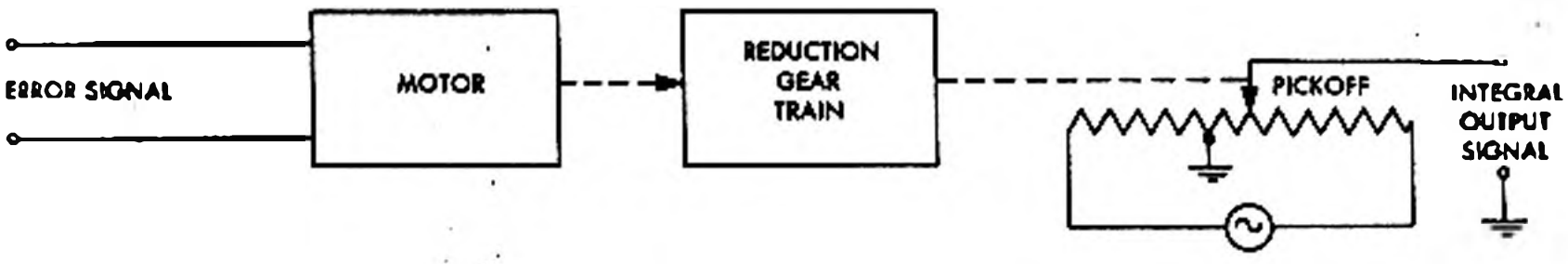


Figure 5-58. Variable-Speed Integrator

where t is the time the constant error exists. The output voltage is proportional to the distance the potentiometer contact moves from the zero or balanced position,

$$\text{or } E_o = K_1 K_2 K_3 B t.$$

Each of the three constants depends on certain independent factors of the device itself such as motor rpm, gear ratio, and voltage applied to the bridge circuit. These can be combined as:

$$E_o = KBt,$$
$$\text{where } K = K_1 K_2 K_3.$$

This shows that the output voltage is proportional

to the input voltage multiplied by the time. The constant of proportionality, K, depends on the machine and is considered in the design of the system.

For this type of integrator to operate perfectly, two conditions must be met. The speed of motor rotation must be directly proportional to the input error voltage, and the output from the potentiometer must always be proportional to the distance the contact has moved from the zero position.

These two ideal conditions are difficult to meet in actual practice. The curve of input power versus motor speed of rotation is never exactly a straight line. An inverse feedback circuit helps to remedy this problem. The second problem requires the use of linear potentiometers. If a selsyn is used as the pickoff, error will be introduced for large angles of deflection. The extremes of the particular pickoff establish the operating range of the integrator.

An integrator need not be perfect for it to operate satisfactorily in a missile control system since it needs to operate only within certain limits. These limits are established during the designing of the control system. They depend on missile aerodynamics and on the control system itself.

Normally, the pickoff which provides the output is zeroed before the beginning of integration. Sometimes a beginning position other than null is required. The integrator then begins with a constant output, and any changes depend on the input. Such a beginning bias signal is important since it represents a condition which should exist at the beginning of integration. It is called the initial condition.

RC Integrator

A simple and commonly used integrator consists of two circuit elements: a resistor and capacitor. The fundamental circuit of an RC integrator is shown in figure 5-59.

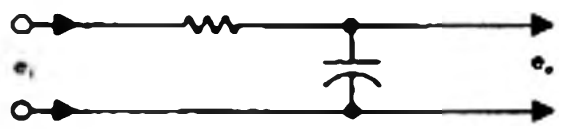


Figure 5-59. Fundamental RC Integrator Circuit

The voltage across a capacitor is proportional to the integral of the charging current. It can be explained by considering that the voltage across a capacitor is

$$E = \frac{Q}{C}$$

For any given capacitor (C), the voltage depends directly on the charge (Q) which is the unbalance of electrons on the two plates. The amount of this charge depends on the current flow and the time which this flow exists. These conditions are comparable to the definition of integration given previously.

The fact that the voltage is proportional to the integral of the charging current allows the RC circuit to be used as an integrator. The capacitor voltage is the integrator output. Provi-

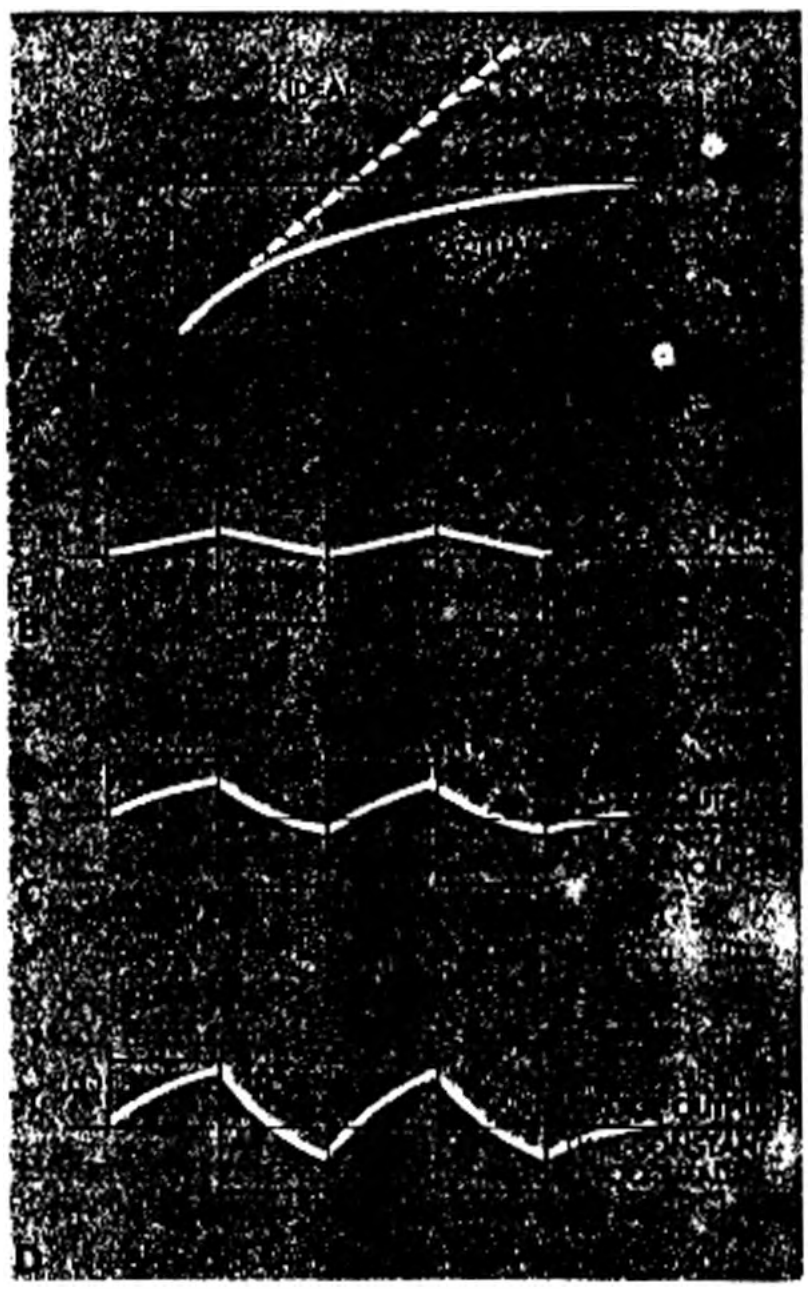


Figure 5-60. Ideal and Actual Outputs from RC Integrator

sion must be made to supply a charging current that is proportional to the input information. The purpose of the resistor is to produce this proportional current from an input signal voltage e_i. At the instant this voltage is applied, the charging current becomes

$$i = \frac{e_i}{R}$$

Unfortunately this proportionality does not continue to exist. As the capacitor becomes charged, the capacitor voltage opposes the charging current, and the charging current becomes less proportional to the input signal. The resulting error in the output is shown by the curve of actual e_o in figure 5-60A. The ideal output, for a constant input signal, would be steadily increasing as shown on the graph. This steady increase is attained only when the signal voltage is first applied and the capacitor has not become appreciably charged.

A remedy to this error in the RC integrator is to use a circuit with a large time constant. Such a circuit slows the charging of the capacitor. The result is a more accurate integration of an input signal as shown in figure 5-60B and 5-60C. The ideal output would be a perfect triangular wave. Although a long time constant produces more accurate results, it also provides a much lower output for the same input signal. Better integration is possible by the use of a high gain, feedback amplifier.

Amplifier RC Integrator

The common integrator shown diagrammatically in figure 5-61 uses a high gain amplifier, with a capacitor forming a feedback path for output variations. This is the Miller integrator. The amplifier produces an output which is not limited by the input signal as it is in the simple RC integrator. The amplifier also supplies any energy

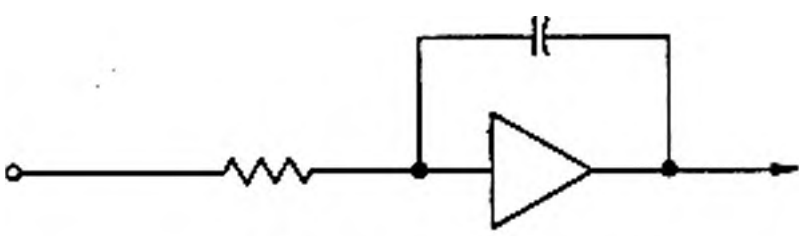


Figure 5-61. Amplifier Integrator

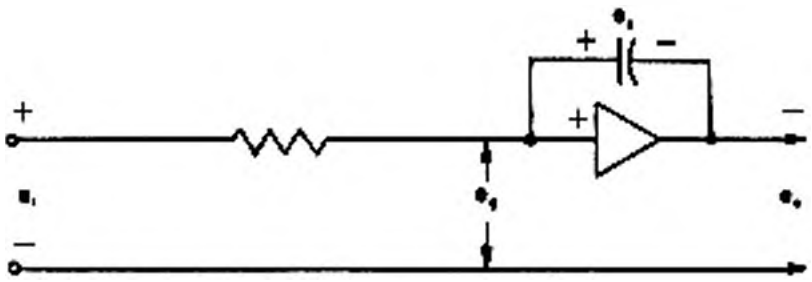


Figure 5-62. Amplifier Integrator with Positive Input

which is required in the output. The function of the input signal is to control the charging current.

The operation can be understood by assuming a constant input as shown in figure 5-62. At the start, assume the initial condition is zero;

$$e_i = e_g = e_o = 0.$$

Also assume that the capacitor is discharged. The positive voltage to be integrated, e_i, is then applied. The capacitor charges with a polarity as shown, since electrons are attracted from the left plate. The charging path is shown in figure 5-63.

A voltage measured at the amplifer input, e_g, tends to rise in the positive direction since this point is directly coupled to e_i. This rise tends to be opposed by the degenerative feedback voltage from the output. The output will be $-Ae_g$. The letter A stands for the amplifier gain. The minus sign indicates that the output polarity or phase is opposite to the input. The output changes A times faster or steeper than e_g. The output voltage will be negative and will aid the charging of the condensers.

For a certain input voltage, the charging current is limited to a particular value which tends to keep e_g practically zero. If the current exceeds this value, e_g decreases a small amount due to the increased voltage drop across R. Then e_o ($e_o = -Ae_g$) decreases and the charging current de-

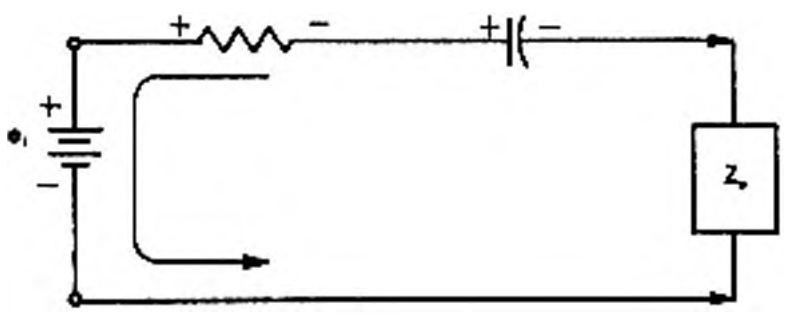


Figure 5-63. Electron Path for Capacitor Charging

creases to the original value. If the initial charging current should decrease, the opposite action would occur. The value of the charging current is therefore stabilized to a specific value proportional to the input voltage. This eliminates the error caused by e_i and the charging current not remaining proportional in the fundamental RC integrator.

This constant charging current must be produced by e_o despite the fact that the steadily increasing capacitor voltage opposes the charging current. To do this, e_o must also steadily increase. This steady increase in e_o is exactly the integrator output voltage desired for a constant signal input.

Similar action would be produced for a condition in which the input signal suddenly became negative. Polarities then would be in reverse to those shown in the example given. Remember that simple examples are used for explanation on the assumption that the desired result also will be produced for a more complicated signal input. Removal of e_i would produce little effect upon the output which existed at that instant, since the amplifier output would oppose the tendency for C to discharge.

The limits for e_o are determined by the amplifier and not by e_i or the range of e_g. The output range is designed to produce an increasing output for any probable input amplitude and period of application. The exception to this is an integrator which is designed to function also as a limiter.

RL Integrator

A resistor and inductor combination can also be used as an integrator. Notice in figure 5-64



Figure 5-64. Simple RL Integrator

that the components of the RL integrator are connected in reverse to the RC integrator. This circuit produces the same response as the RC circuit. It also possesses the same defects. A further error exists because of the resistance in the coil. The RC circuit is used more often than the RL circuit.

Thermal Integrator

An integrator can operate on the principle that the temperature of a heating element is approximately proportional to the integral of the current input for small intervals of time. A heat detector provides the integrator output. Such an integrator is limited in "memory."

Integration from Viscous Damped Integrating Gyro

A method of integration is made possible by using a viscous damped integrating gyro, such as was explained in the section on gyros. This gyro is not considered a free gyro since it has only two degrees of freedom. The integration function is produced as the gyro is allowed to precess through the restrictive force imposed by the coil. The force causing precession of the gyro is produced by angular deviation of the missile. This force depends on the rate of change of missile deviation. Precession of the integrating gyro is not limited to a few degrees as in a rate gyro. The amount of precession depends upon the force causing precession and the duration of precession. The missile deviation also depends upon the rate and time of angular motion. A definite ratio exists between the angle of gyro precession and the angle of missile displacement. Although the force on the gyro results from the rate of displacement, this gyro produces displacement information as a result of the integration which takes place within the gyro unit.

The last of the three general types of components of a computer unit are rate systems.

RATE COMPONENTS

A rate component is part of the computer section of the control system since it, like the integrator, performs a mathematical operation on an input signal. This operation differs from that of the integrator. The input signal is changed to a new form which represents new information. This information improves the control system stability. The input signal conveys some type of missile information such as position error, angular devia-

tion, altitude or airspeed error, or control surface position.

Briefly, a rate component is a device which produces an output that is proportional to the rate of change of the input signal amplitude. This mathematical operation that determines the speed at which a signal is changing is called *differentiation*. If an input signal is increasing or decreasing slowly, the output of the rate component is small. If a signal is changing rapidly, its output is proportionally larger. Even when an input signal is large, the output of a rate component should be zero as long as the input signal remains steady. The output of most rate components reverses phase when the input changes from an increasing to a decreasing signal. Figure 5-65 shows rate output from sample inputs.

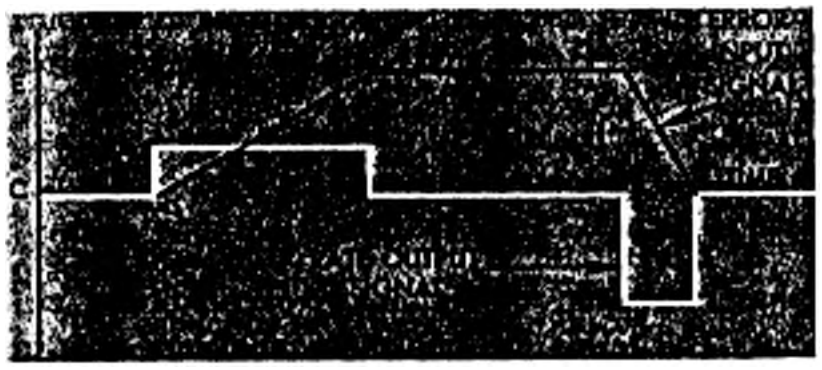


Figure 5-65. Rate Output Depends on Rate of Change of Input Signal

You probably have more contact with rate components (differentiators) than you realize. Any device which indicates the speed of some object directly is a rate computer. The automobile speedometer is an example of this. The input of the speedometer is a rotating shaft. While the mileage section is simply an extremely large gear reduction, some sort of rate sensing device is necessary to detect and indicate speed.

This rotational velocity can be detected by centrifugal force or by magnetism. Suppose a weight is fastened to a rotating member. The weight is fastened so that it can swing outward against spring tension. The amount of centrifugal force tending to pull the weight from the shaft depends on the speed of rotation. The displacement of the weight is a measure of speed.

A special generator can be used to indicate rotational speed since the rate at which magnetic lines of force are cut determines the output voltage. A voltmeter can then be calibrated in rpm, miles per hour, or comparable units.

Purposes of Rate Control

Rate systems are necessary because a definite amount of time is required to perform any type of control operation. In the case of a missile, if the correction of missile deviation occurs at the instant of deviation, no deviation whatsoever would result. Such a result would be ideal, but it is impossible since a system must detect some error in order to operate. The degree to which this ideal situation is approached depends on the design of the system and the requirements of the load such as size of airframe and missile aerodynamics.

Since the ideal control system described would have zero time lag and would result in a missile which did not deviate whatsoever, efforts are made to reduce this time lag to a minimum. The airframe and control surfaces are designed to correct missile attitude rapidly. The control surfaces are moved by powerful, fast-acting actuators using an amplified error signal to produce fast response. The reduction of lag by such measures approaches a limit.

OSCILLATION FROM OVERCONTROL. At this point, you may be thinking, "Why not increase the amplification of the error signal from the sensor so that even a small missile deviation will produce a signal to correct the missile almost instantaneously." This idea works to a certain extent. A signal which is too large returns the missile beyond the desired point to an opposite error. The resulting opposite error signal causes the missile to deviate back to the first direction. The end result is a serious oscillation about the desired attitude.

DAMPING UNWANTED OSCILLATION. The addition of a rate signal has the effect of damping the oscillation. The term "damp" means to restrain or retard. The amount of damping is classified as critical damping, underdamping, or overdamping.

Rate Component Creates Lead

If the signal which commands system action could be advanced in time, the effect would be an apparent reduction in the overall lag of the system. A rate signal does this. It is combined with the proportional signal to produce a resultant signal which leads the original proportional signal.

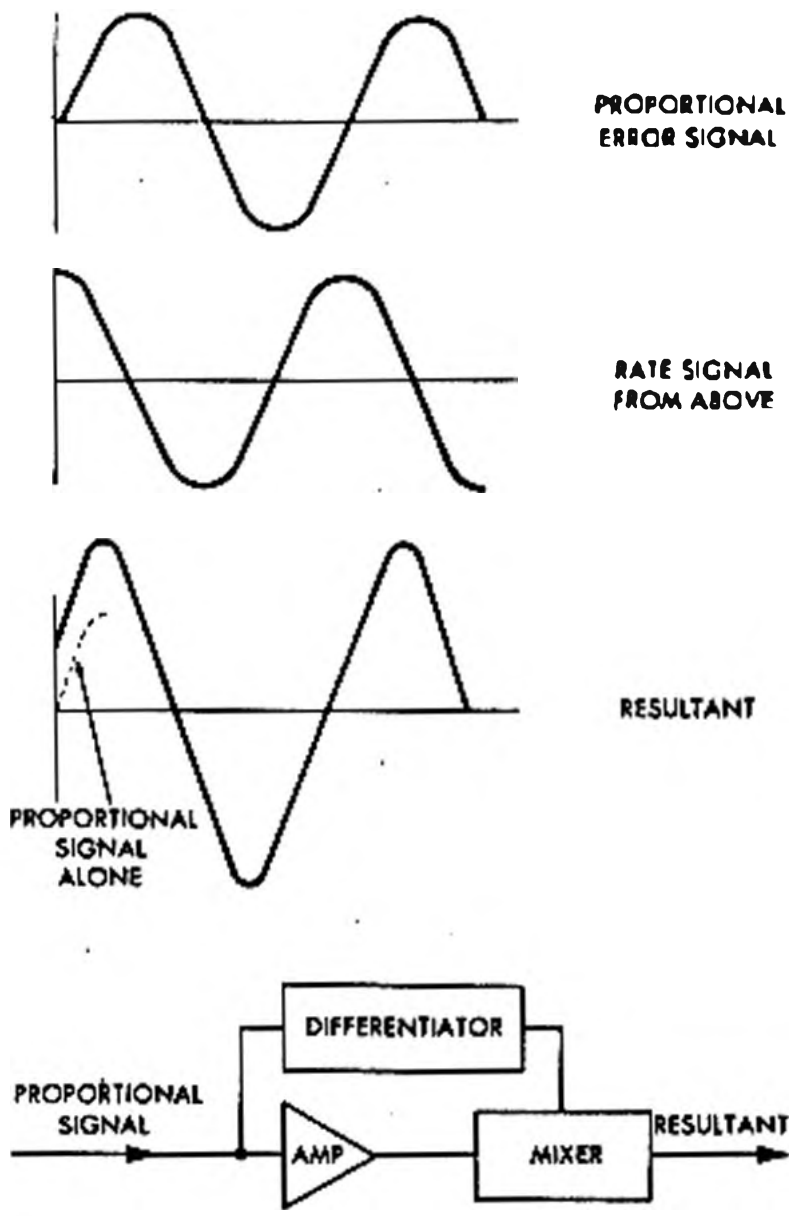


Figure 5-66. Method of Producing Error Signal Lead by Differentiation

The combination of the error signal with a rate signal to produce a leading signal is shown in figure 5-66. The proportional signal is assumed to be a sine wave as shown at the top. If it represents yaw error, this indicates that the missile is "fishtailing" to either side of the desired heading. A sine wave is used since it is most suitable for an example, despite the fact that it actually represents an undesirable flight condition.

The second waveform shows the output of a rate device when using a part of the proportion signal for an input. Notice that the output of the rate device is zero at the instant the input signal is at a crest and that the output has no rate of change. The error curve has the steepest slope when the input error signal is changing at a maximum rate. This phenomenon occurs when the error voltage is zero. At this point, the rate signal is a maximum, indicating a maximum rate of change.

The vector addition of two sine waves (rate and proportion) which are out of phase produces a sine wave with a phase somewhere in between. This result is shown in the third line of figure 5-66.

The two signals are mixed in a ratio which produces the best response from the system. This optimum ratio is determined by the other system components and the aerodynamics of the missile; it may vary with any change in missile configuration, weight, or altitude. Of course, the amount of rate signal need only to be within a certain range for acceptable missile operation.

Remember, there is a rate output only when the missile deviation is changing. If the rate signal were not combined with the error signal there would be no way for the control system to respond to a constant error. It is also possible to combine an attitude rate signal with a guidance error signal.

Predictor Circuit

A rate circuit is often called a lead circuit, phase-advancer, differentiator, or prediction circuit. The terms are analogous and vary only in the concept of the subject. The term "rate" refers to the fact that the signal is proportional to the rate of change of the input signal. The term "differentiator" refers to the mathematical operation which is performed on the equation of the input voltage to yield the proper equation for the output voltage. Figure 5-66 shows the origin of the terms "lead circuit" or "phase advance."

But, why call a rate device a "predictor"? Consider the dashed line portion of the proportional signal in figure 5-66. The error is increasing during this period. The resultant during this time is much greater than the value of the proportional signal alone. If the missile slowly veers from the desired heading, the resultant signal will be due mainly to the proportional signal.

Without a rate circuit, a rapid rate of missile deviation would normally produce a fairly large deviation before correction takes place. The rate circuit "predicts" this possible future large deviation from the existence of a high rate of deviation, and it increases the corrective effect. The result

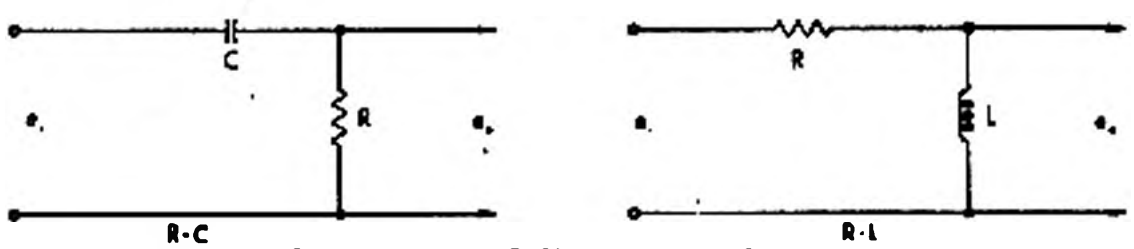


Figure 5-67. Differentiation Circuit

is that the missile does not deviate as far with a rate unit in its control system.

Electronic Rate Circuits

The RL and RC combinations which are used as electronic integrators also serve as effective differentiators when the components are connected as shown in figure 5-67.

The inductor (RL) circuit produces a differentiation because the voltage across the coil is proportional to the rate of change of current. This is assuming a perfect coil with no resistance. The IR drop across the coil causes an error in the output voltage, with the result that the RC circuit is used more often.

The capacitor (RC) circuit produces a differentiation because the charging current is proportional to the rate of change of applied voltage. The output voltage dropped across the resistor in the circuit depends upon the charging current. The basic capacitor theory states that the voltage across a capacitor is determined by the unbalanced electrons on the two plates. In figure 5-68, the battery transfers electrons from the lower plate to the upper one, so E_c equals the applied voltage. Now, suppose the voltage from the battery could be slowly increased at a steady rate. The charging current would be proportional to the rate of change of the voltage since E_c also changes. A capacitor forms the basis of a rate circuit.

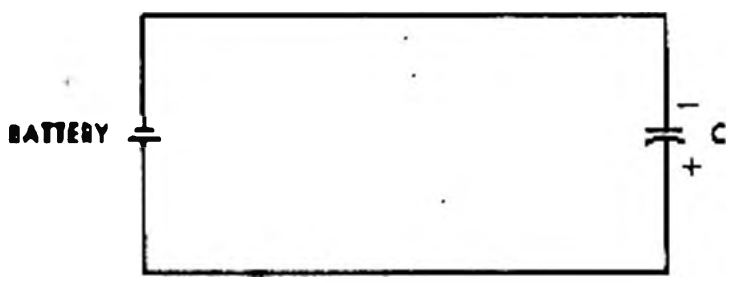


Figure 5-68. Theoretical Differentiation Circuit

The theoretical circuit in figure 5-68 should operate as a perfect differentiator since it contains no resistance or other impedance. Such a differentiator is not practical because the voltage source and wiring possess impedance. Furthermore, some method is necessary to detect the charging current. Since the current is normally small, a resistance is placed in series with the capacitor, and the voltage across it is used as the output. This impedance creates an error caused by the dissipation of energy in the resistance and the delay in the response of the output.

In most cases a control system error signal is modulated by 400-Hertz AC. This signal has to be demodulated to a varying DC voltage before it is applied to a capacitor rate circuit. This demodulation is necessary because the rate circuit is intended to detect rate of change of amplitude represented by the envelope of the input signal rather than by the 400-Hertz variations. The signal may be modulated again after it passes the rate circuit, as shown in figure 5-69. Modulation and demodulation also can be considered a computing function.

Figure 5-70 shows the rate output from a

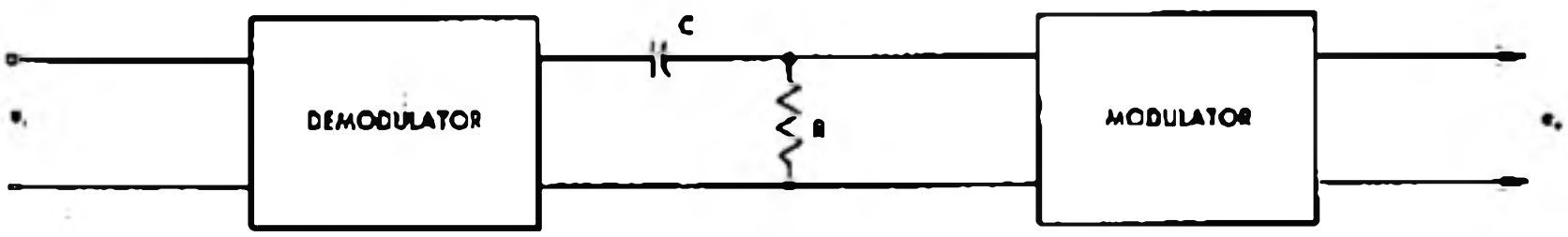
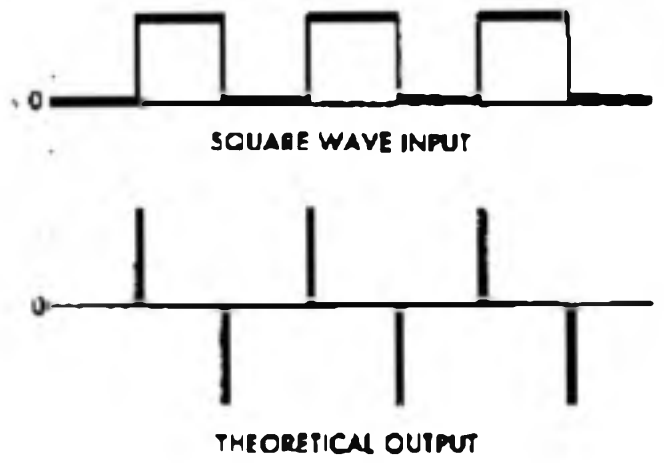


Figure 5-69. Rate Circuit for Modulated Signal



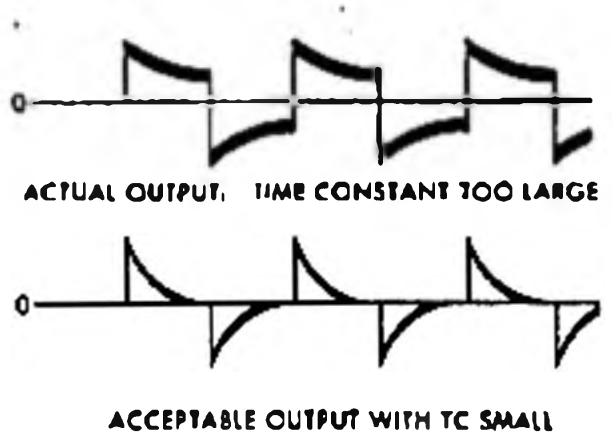


Figure 5-70. RC Differentiator Outputs

square wave input. An accurate output is obtained if the time constant of the circuit does not exceed four times the period of the input signal. The perfect output is theoretical. It would be infinitely large when the input signal changes instantaneously. In actual practice this output is limited by the resistance in the circuit.

A solution to the errors in the basic RC differentiator is to use an amplifier as was done in the case of the integrator. The amplifier supplies most of the dissipated energy and permits the use of a short time constant by amplification of the output.

RC Amplifier Differentiator

As shown in figure 5-71, an RC amplifier differentiator is just opposite to an electronic integrator. This differentiator circuit has a capacitor input and resistive feedback. The amplifier can be a modulator type so as to reduce drift and noise.

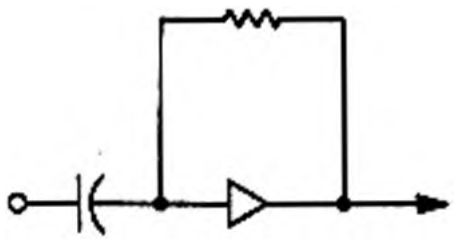


Figure 5-71. RC Amplifier Differentiator

In analyzing such a circuit, it is normal to consider the amplifier to have infinite gain. On this basis, e_g can be considered to be zero. Although infinite gain is impossible, the amplifier does have a very large gain. If e_i should change, e_g tends to change, but the inverse feedback voltage counteracts the change in e_g.

The output voltage depends on the tendency of e_g to change, and it is proportional to the rate of change of e_i.

This same analysis can be given in terms of polarities. Figure 5-72 shows the same circuit

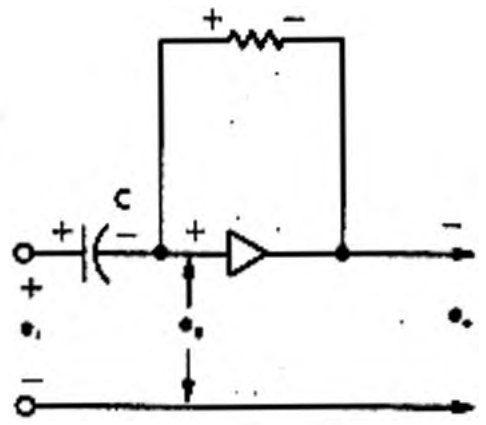


Figure 5-72. RC Amplifier Differentiator with Positive Going Output

assuming a steady, positive input signal. C is charged to the steady value of the input signal, therefore, $e_g=0$ and $e_o=0$.

Now, assume the error signal suddenly increases. C begins to lose electrons on the left and gain on the right. The charging of C tends to make e_g go positive. This action is reflected in the output as a negative voltage which increases at a rate A times faster than the input. Current produced by the output rapidly charges C to the new e_i. The output is a large pulse of voltage which indicates the rapid change in the input voltage e_i. This is the output desired. The amplitude of e_o is not limited by e_i, as was the case in the simple RC circuit, because of the energy supplied by the amplifier. If e_i should decrease, an output pulse of opposite polarity would be produced.

Rate Gyro

A common method of producing a rate signal is by means of a separate sensor. A special type gyro, called rate gyro, has been designed for this purpose. A rate gyro has a restricted gimbal, free to rotate about one axis only. It is allowed to precess only a few degrees as the gimbal is restrained by a spring which tends to return the gimbal to a neutral position. Any precession in this plane is a result of force on the gimbals. This force is caused by angular movement of the airframe about a certain axis and is proportional to the rate of deviation. The amount of small gyro deviation depends on the precessing force which must overcome the restraining force.

This gyro deviation can be detected with a pickoff, and then becomes a rate signal. A separate rate gyro is required for every plane of missile movement in which a rate signal is desired. These rate signals are normally combined with the output from a displacement gyro. Rate gyros are covered in the section on gyros.

REVIEW OF COMPUTER FUNCTIONS

Rate devices produce a method of continuous control in a system. They comprise the third method of continuous control that we have explained. The three methods are (1) proportional, (2) integral, and (3) rate.

Proportional control operates the load by means of an error signal which is proportional to the amount of deviation of the controlled item from the desired attitude. This signal is produced by a sensor. Integral control operates the load by means of a signal which represents an accumulation of error. The integral signal is produced by devices covered in the presentation on integrators. Rate control operates the load by means of a signal proportional to the speed at which the deviation is changing. When rate or integral control is used, the output is usually combined with a proportional signal to produce the desired effects on the missile. Although rate and integral control have a canceling effect, they sometimes appear in the same system. Practically every system contains a rate device since it enables the missile to react rapidly to any error which develops.

The computing operations do not necessarily occur in the position which is indicated by the basic block diagram. Regardless of where the operations take place, the mixing of signals and the development of integral and rate signals comprise the computing functions of a control system.

SECTION C

Reference Units

The mediums for control system reference can be divided into three categories. These categories are voltage references, time references, and physical references.

Before taking up voltage references, the first category, look back at the basic missile control block diagram shown in figure 5-3, so you can see where the "reference" block fits into the system.

VOLTAGE REFERENCES

If there is no reference in a control system, the system does not know how to correct for an error. Consequently, an electronic control system could not function properly without some type of voltage reference.

AC Voltage Reference

In many of the existing control systems, the error signals are in the form of AC voltages. These error signals contain two characteristics which are necessary for control systems to make proper corrections. These characteristics are (1) amount of deviation or magnitude of the error, and (2) direction of deviation or the phase of the error.

The *amount of deviation* is carried by the error signal in the form of amplitude. As the amount of deviation increases, the amplitude of the error signal also increases. If the amount of deviation should decrease because of a movement of the control surfaces, then the error signal would decrease in amplitude. When the deviation decreases to zero, the amplitude of the error signal also decreases to zero. There is no error signal when there is no deviation in missile attitude.

The *direction of deviation* is carried by the AC error signal as a function of phase. As has already

been pointed out, this phasing by itself cannot tell the control system which way to correct unless there is something to which this phasing can be compared, such as a reference voltage.

Around each axis of control there can be only two directions of deviation. Only two phases are required to tell the control system which way to move the control surfaces for any one axis. Usually one of these phases is in phase with an AC reference voltage, called zero phase, and the other is 180° out of phase with the AC reference voltage, generall called π (pi) phase.

By using phase sensitive circuits which compare the phase of the error signal with the AC reference voltage, the autopilot can detect the "direction of deviation" information from the error signal and move the control surfaces in the right direction to correct for the attitude error.

This AC reference voltage is usually the AC power supply for the control system. It also serves as the excitation voltage for the sensor unit which originates the error signal.

DC Voltage Reference

Most of the mechanisms used in the controller unit require a DC error signal, so it is necessary to show how a DC voltage can carry both of the characteristics that existed in the AC error signal. The amplitude of the DC error signal indicates *amount of deviation*. The *direction of deviation* is shown by the polarity of the DC error signal.

The phase-sensitive circuits which are used to remove the two characteristics from an AC error signal usually produce a DC error signal that carries the two characteristics in the manner described in the above paragraph.

In some control systems limiter circuits are used to prevent the error signal from becoming too large and causing overcontrol. These limiter circuits require DC voltage as their reference. This DC reference voltage can be obtained in many different ways, but regardless of the method used, the circuit functions as a part of the reference unit of the control system.

TIME REFERENCES

Time is another medium by which an autopilot can obtain a reference. Primarily, this reference takes the form of fixed periods of time during which certain autopilot functions are allowed (or not allowed) to take place. A definition of time reference as it applies to control systems is: A time-delay device used to control a specific autopilot function during a predetermined period of time.

Characteristics of Timers

Timers vary greatly in their physical characteristics and their operation. Some of these physical characteristics need to be mentioned at this point to facilitate a better understanding of specific timers and their application. These characteristics are covered again as the various timing devices are presented.

All control timers require some method of being initiated or triggered. Because all of the timing devices used in one system are not triggered at the same time, each must have its own trigger.

The method most used to trigger timing devices employs some form of electrical signal. In some cases this electrical signal is fed to a solenoid that has a core which moves when the solenoid is energized. The movement of this core, in these cases, is used to trigger the timing device mechanically. A special application of this method is in the case of the solenoid actually being a part of the timing device, rather than the triggering device. In some special cases the normal electrical signal to a time-delay device is a voltage input, and the triggering takes place by removing the voltage.

There are two other electrical triggering methods. One is accomplished by applying the electrical signal to a heater coil which heats a thermosensitive bimetallic strip. The other electrical method is to apply the electrical signal to an electric motor. This signal operates the motor which is a part of the timing device. The voltage can be either AC or DC depending on the type of motor used.

Another method of triggering timing devices which has been used in missiles is to use an inertia switch. This method is explained later with the piston-type pneumatic timer.

The degree of accuracy needed greatly affects the selection of a timer for a specific application. Some of the uses of timing devices permit use of timers having comparatively poor accuracy. Other

uses might require the highest degree of accuracy that can be obtained from timing devices.

When using timers in guided missiles, the effect that large changes in temperature and pressure have on their accuracy must be considered. A varying input voltage can affect the accuracy of some timers that make use of electric motors. These accuracy considerations on specific timers are explained in greater detail when the timing devices are described.

The simplicity or the complexity of a timer is influenced by the specific application of the device. If the timer is to be used only once during a missile flight, the mechanism may be simple in operation. However, if the timing device is to be used many times during a missile flight and has to recycle itself each time it performs its function, the mechanism will be more complex in operation. The one exception to this is the thermal timing device which recycles itself without any additional mechanism.

In most timers the output produces electrical contact switching. The use of the timing device determines whether the contacts are normally open or normally closed.

If the timer performs more than one function, there are two basic ways to accomplish the circuit changes. The first way is to employ a multicontact switch on the timing device. This method can close or open many circuits. The other way is to use a single-pole, single-throw (SPST) switch on the timer to energize or deenergize a relay that has the required number of contacts.

Application of Timers

In the following paragraphs, some of the uses described might not be applied strictly as control system functions, but the applications are related enough to justify their mention at this point in the text. You might consider some of the applications as preset guidance functions and some as launch or powerplant functions. The reason for mentioning such uses is to show the wide variety of applications of control timing devices.

One of the first uses of timers in guided missiles was to keep the gyros caged during the period of great acceleration at launch. The timer used for this purpose was triggered by a voltage either from the ground equipment during the launch preparation or from an inertia switch which closed after the missile left the launcher.

Timers can be used to trigger other timers. In some missiles, a mechanical timer is used to trigger another mechanical timer at a preset time after launch. The second timer may be used to precess the directional gyro and make a preset turn. The length of time set into the second timer controls the amount of turn. Usually these turns are around the yaw axis, but in some of the terminal guidance systems, the gyro is precessed about the pitch axis to put the missile into a dive. The timers used to precess gyros must be accurate.

Timing devices are used extensively in programming takeoff. Some timers will cut off the powerplant if the boosters haven't fired within a given time after the powerplant has developed some specified speed or thrust. This prevents releasing a missile without adequate power for a safe launch. The sequence of firing the boosters is frequently controlled by timers. Launch programmers which control the climb of the missile often use timing devices.

Another possible use of timers is their application to command guidance. A carrier relay, which remains energized at all times during which the command carrier frequency is present, controls a timer. When the carrier frequency is lost, the carrier relay triggers the timer. After a certain time delay, if the carrier has not been received, the timer completes its function and allows some preset command guidance function to take place. This type of timer is ordinarily used as a safety device and may cause the missile to remain on a predetermined course until the command signal is again received, or it may cause self-destruction of the missile.

One other specific use of timers is in connection with gyro erection and slaving systems. During a time delay set up by a timer, the gyro is precessed at a much faster rate than under normal operating conditions. After this period of time has elapsed, the circuitry is set up by the timer to allow the gyro to precess at its normal rate.

Mechanical Timers

One type of mechanical timer which has been used in missile systems is a clock mechanism. In operation it is similar to a mechanical alarm clock.



Figure 5-73. Simple Motor-Timer Operation

The ordinary alarm clock can be set for a time delay up to 12 hours, while the clock mechanism used as a timer usually can be set only for relatively short periods such as seconds or minutes. The alarm clock gets its movement from energy stored up in the mainspring by turning a knob on the back of the clock. The clock mechanism timer gets its movement in much the same way. To set in a time delay, the knob on the face of the timer must be turned so that the arrow points to the desired time on the dial. This puts energy into the mainspring of the timer and corresponds to winding the alarm clock.

A bell rings at the time set on the alarm clock. When the time has elapsed on the clock timer, a switch is closed (or opened according to the application of the timer).

The time delay set into a clock mechanism may not be used until the missile is in flight, so some sort of triggering linkage is necessary. This linkage usually consists of a catch which can be released by energizing a solenoid. This type of timing device can be used only once during a missile flight. Its accuracy is considered good.

Electrical Timers

Two types of timers presented as electrical timers are motor timers and thermal timers. In all cases the triggering is accomplished by an electrical signal. When the voltage is applied to the timers, the time interval or period of delay begins.

MOTOR TIMERS. The simplest motor timer, shown in figure 5-73, possesses the operation principles of all motor timers.

A motor is turned by the triggering voltage. The speed of the motor shaft is reduced by a gear box. The speed is such that the output shaft from the gear box does not make a complete revolution during the period of delay required. For example, if a 30-second delay is needed, the speed of the output of the gear box is about one or two revolutions per minute. There is an arm connected to

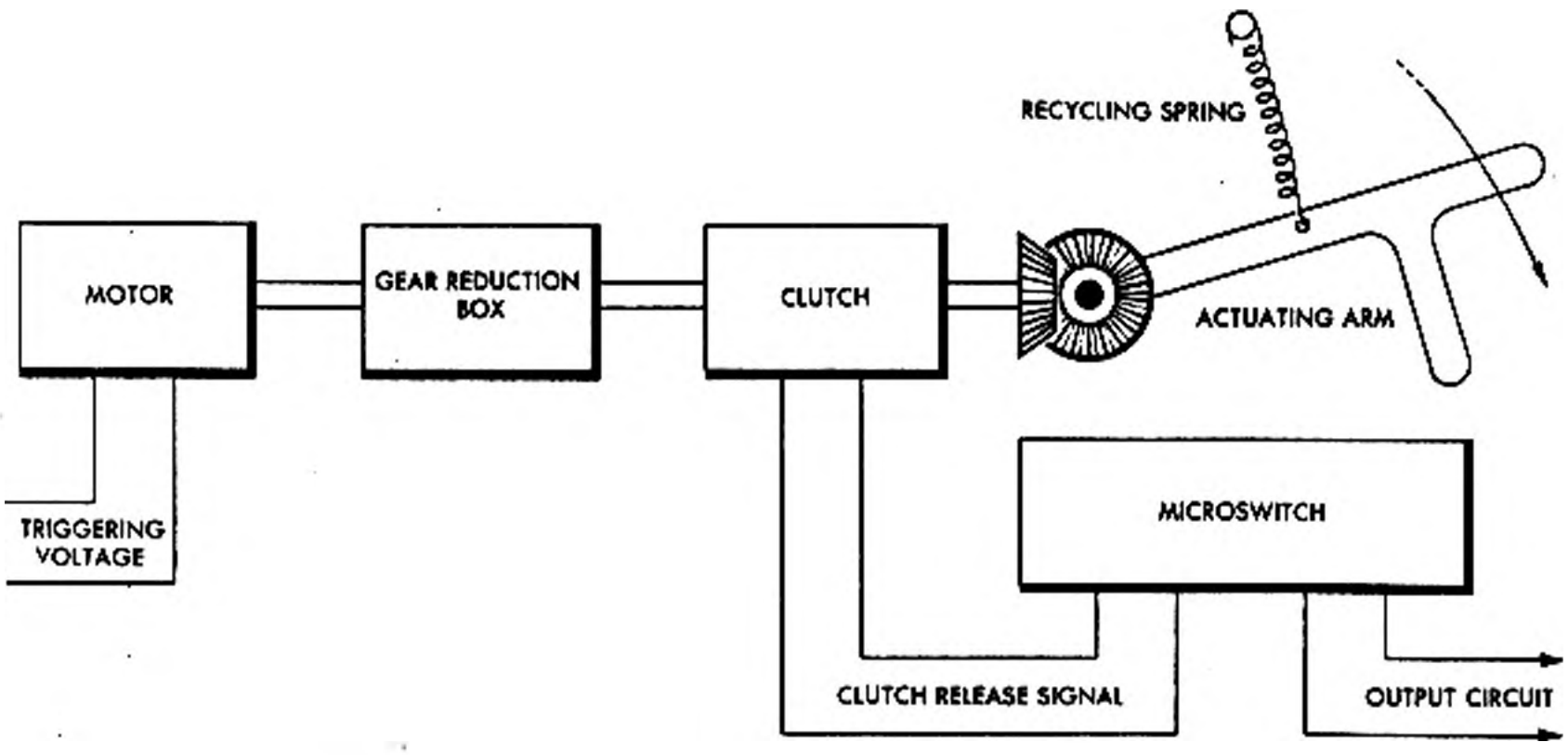


Figure 5-74. Operation of Recycling Motor Timer

the output shaft of the gear box. The time it takes this arm to go from the starting position to a point which closes a switch is the time delay of the timing device. This simple mechanism can be used only once during a missile flight.

If recycling is required during flight, the mechanism is more complex, as illustrated in figure 5-74. Again, the basic operational components are present: the triggering circuit, the motor, the gear box, and the switch for an output. But some additional stages are used to make the timer automatically recycle itself. A clutch mechanism breaks the linkage between the actuating arm and the gear box. The arm then returns to the starting position without the motor or gear box turning. The recycling spring pulls the actuating arm back to its original setting when the clutch disengages.

The microswitch, shown in the diagram, has two sets of contacts. One set is in the output circuit of the timer. The other set opens or closes the clutch circuit. In a case such as shown in the diagram, the clutch would probably be controlled by a solenoid. The actual details of the solenoid circuit depend on the construction of the clutch.

A different method of controlling the clutch may be used. In this case, the triggering voltage also energizes a solenoid which keeps the clutch engaged as the motor turns. When the triggering voltage is released, the solenoid in the clutch de-energizes, and the timing mechanism recycles.

Another method of releasing the clutch uses a mechanical means. A lever releases the clutch when the actuating arm applies pressure to the lever. The actual method used to release the clutch is determined by the application of the timing mechanism.

If accuracy is of prime importance when using motor timing devices, other additions must be made to the mechanisms. Most motors vary in speed when the input voltage varies. This is not true of a synchronous motor. When such a motor is used in timing devices, no regulation of voltage is necessary. But to make the timer accurate when other than a synchronous motor is used, either the speed or the input voltage must be regulated. Circuits already exist which can regulate input voltage. If the speed must be regulated to have still greater accuracy, a special clock mechanism

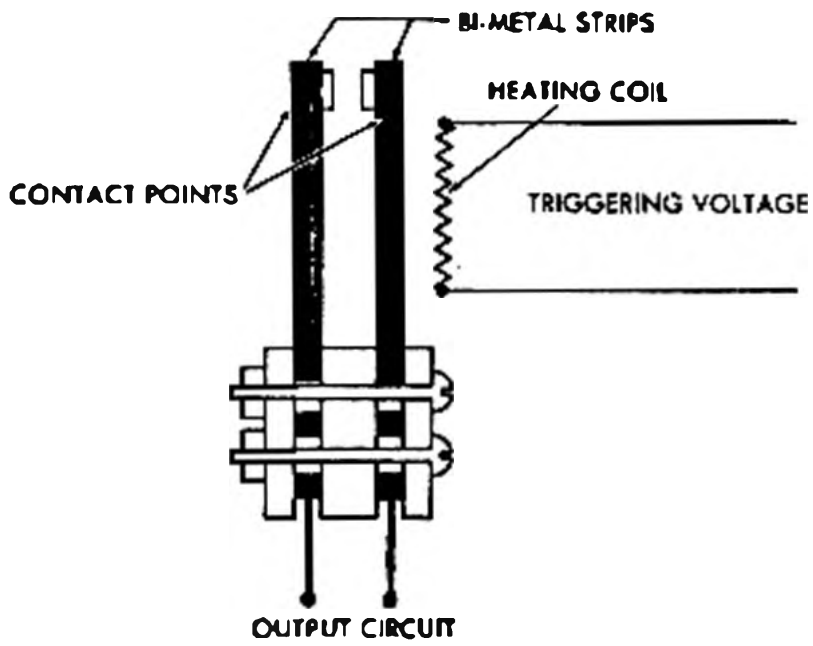


Figure 5-75. Thermal Delay Tube

is used as a governor. The torque of the motor provides the actuating force for the speed-regulating clock mechanism.

You can easily see that the more accurate the requirements of a timing device the more complex that timer becomes.

THERMOELECTRIC TIMERS. Thermal delay tubes and thermal delay relays are used extensively to perform time delay functions. The advantage of timers of this type over other timing devices is the ease with which they recycle themselves. To do this, they require no additional circuitry or mechanisms. However, contrasting this advantage is a major disadvantage. This disadvantage is a lack of accuracy. The accuracy of these timers is comparatively poor with respect to the accuracy of other timers presented in this section.

In figure 5-75, a thermal delay tube is shown. Its components are two bimetallic strips, a heating coil, a set of contacts, and a means for spacing the two bimetallic strips. A voltage is applied to the heating coil. The coil generates heat which heats up one of the bimetallic strips. This strip bends toward the other as the temperature rises. When the bimetallic strip has heated sufficiently, the contacts close, completing the output circuit of the timing device. The amount of delay depends on the distance between the contacts; i.e., how far the bimetallic strip has to bend to close the contacts. In this case the time delay is preset, and the elements are then put into a vacuum tube

which prevents any further adjustment of the time delay.

Accuracy is affected if the voltage applied to the heating coil varies from its rated value. Also, if the thermal element does not have sufficient time to cool before the timer is used again, the accuracy is affected because the thermal element now does not require so long a time to attain the temperature at which the contacts close.

Although ambient (surrounding) temperature affects the accuracy of many thermal delay devices, it has little effect on some of the later thermal delay tubes. The ambient temperature may have a wide range, approximately −50° to +70° C, without seriously affecting the operation of the thermal device. This partial immunity to temperature change is accomplished by incorporating two bimetallic strips and fastening one of the contacts to each strip as shown in figure 5-75. When the ambient temperature changes, the temperature of each of the bimetallic strips gradually changes also. An ambient temperature change causes the same amount of change in each strip, as shown in figure 5-76, because the two strips are constructed of the same two metals and are the same size.



Figure 5-76. Thermal Delay Tube in Extreme Ambient Temperature

Thermal time delay relays usually do not have this compensation for ambient temperature, so they have to be used in equipment that is temperature controlled. The time delay can be adjusted on thermal delay relays. Figure 5-77 shows the time adjustment screw which can change the distance between the contacts. The thermal time

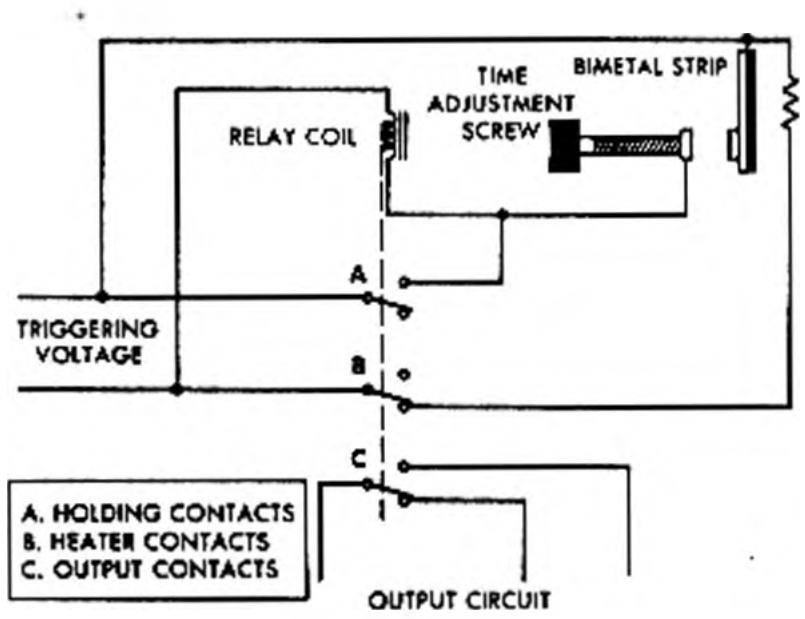


Figure 5-77. Thermal Time Delay Relay Showing Time Adjustments

delay relay shown can be considered a thermal device and a holding relay all built into one unit.

When the signal is applied from the triggering circuit, the voltage is impressed across the heater coils through the B relay contacts. As the bimetallic strip is heated, the contacts move closer together. After a period of time, the contacts close, and the voltage from the triggering circuit is applied to the relay coil. When the relay coil energizes, all of the relay contacts change. The A contacts parallel the contacts on the bimetallic strip and act as holding contacts, keeping the relay solenoid energized. The B contacts break the circuit to the heater coil, and the coil cools off. Once the relay coil has been energized and the holding contacts have closed, the bimetallic strip has served its purpose, and the timing device will not be affected if it is allowed to cool off. The C contacts form part of the output circuit. This type of timing device is easily recycled by taking away the triggering voltage which would deenergize the relay coil. The timer then would be ready to begin its operating cycle again.

Pneumatic Timers

Pneumatic timers determine a period of time by means of a small orifice through which the air must pass.

They are not very accurate in comparison with the other types that have been presented. The effects of atmospheric pressure and temperature changes on the air stored in the timer influence

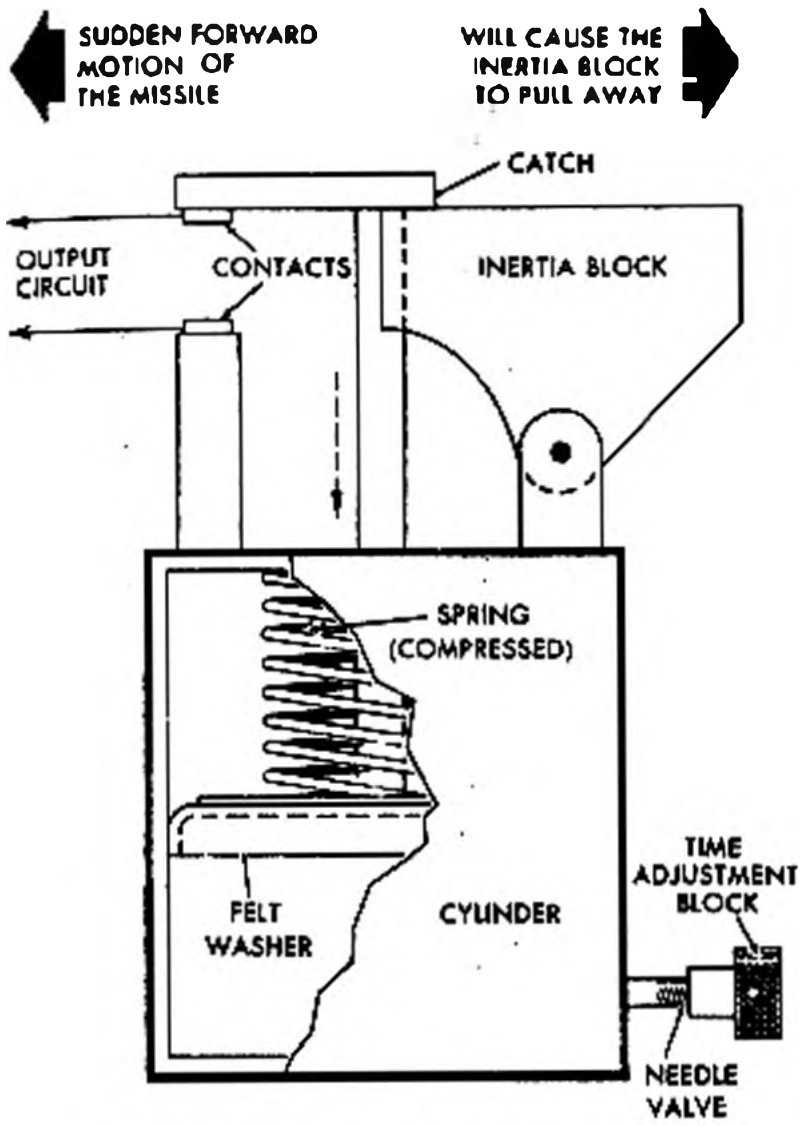


Figure 5-78. Piston-Type Pneumatic Timer

their accuracy. As mentioned, these timers get their time delay by forcing air through a small opening. If the density of air should change or if the pressure between the air stored in the timer and the atmosphere should differ, which of course would happen in a traveling missile, the accuracy would be affected. This type of timer is more suited for launch functions because it performs its function under the same atmospheric conditions as existed when the time delay was adjusted.

Pneumatic timers can be divided into two general types, piston and diaphragm.

PISTON-TYPE TIMER. A piston pneumatic timer is shown in figure 5-78. The action of this timer can be compared to that of a tire pump. As you pull up the handle of a tire pump, the air goes around the leather washer. The same is true of this timer. As you apply force on the handle of the tire pump, you push air out of the valve at the end of the hose. In the timer the spring applies the force to the piston, and the air in the cylinder passes through the opening in the needle valve. The spring exerts the same force each time the timer is cocked. The timer is cocked by pulling the piston up and hooking the catch over the inertia block. Because the spring exerts the same force each time, the amount of time it takes the timer to close the contacts of the output circuit depends on the size of the opening in the needle valve and the relative inside and outside air pressures. A sketch of the needle valve appears in figure 5-79. The needle and the tube are tapered so that the gap between them will close as the adjustment screw is turned in.

The inertia block is the trigger for this timer. It consists of a block of metal which has enough mass for it to be thrown backward when subjected to a large acceleration, thus the term *inertia block*. This type of triggering indicates that the timer has to be used during the launching phase.

DIAPHRAGM-TYPE TIMER. The diaphragm pneumatic timer is essentially the same in principle as the piston-type pneumatic timer. Air is taken in and then released through a small hole. The time it takes the air to escape through the hole determines the amount of time delay.

An example of a diaphragm pneumatic timer is shown in figure 5-80. The drawing shows the timer ready to perform a time delay.

The normal signal to this timer is a voltage which keeps the solenoid coil energized. When the coil is energized, the iron core is pulled down, depressing the spring. When the iron core is down, the leather diaphragm is stretched, allowing air to fill the air chamber through the air inlet holes.

The timer is triggered by taking away the voltage on the solenoid. When this is done, the spring

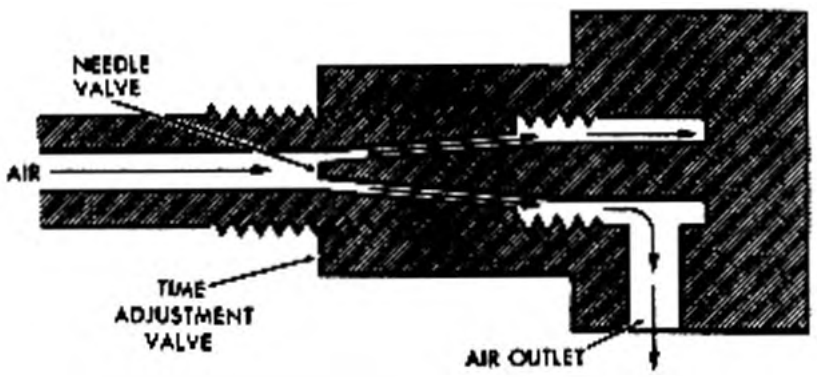


Figure 5-79. Needle Valve of Piston-Type Pneumatic Timer

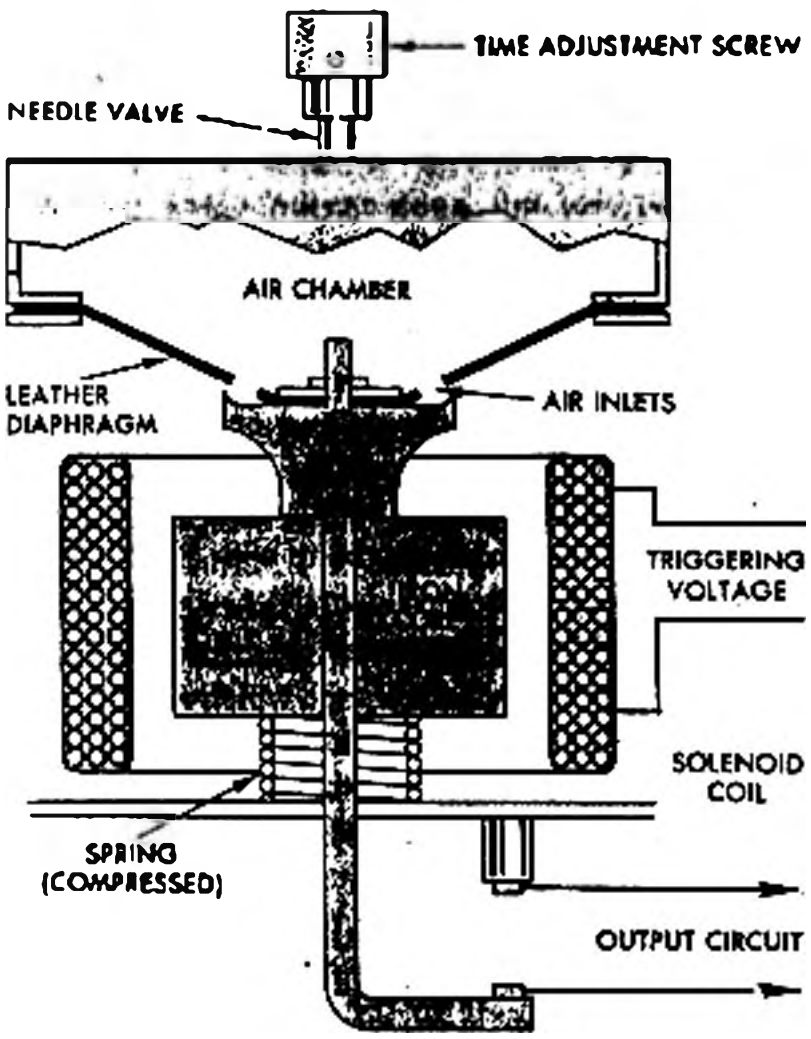


Figure 5-80. Diaphragm-Type Pneumatic Timer

pushes the pressure plate against the diaphragm, closing the air inlet holes and slowly pushing the air out of the chamber through the needle valve. Again, the adjustment screw controls the opening in the needle valve, controlling the length of time it takes the output contacts to close.

This timer is fairly accurate compared to other pneumatic timers and does not require any special mechanism to recycle itself. Reapplying the voltage to the solenoid sets up the timer for another cycle.

Timers are used in almost all control systems, and a particular timer may be used to perform a variety of functions. The requirements of a missile dictate the usage of a timing device. Some timers are strictly safety devices that must be used during the development of missiles.

PHYSICAL REFERENCES

There are many other references for missile control systems besides voltage and time. They can all be grouped under the heading of physical references. These references are space, gravity, the earth's magnetic field, barometric pressure, and missile airframe. In some measure they all contribute reference to most missile control systems. Primarily, they form the group of references that sensors use to establish error signals. A brief explanation of each physical reference is presented, sufficient only to complete the scope of the reference unit and to familiarize you with the sensors which use these references.

SPACE. Gyroscopes, because of their characteristics, are the sensors which use space for a reference. A plane of reference is established in space, and gyros sense any change from that reference.

GRAVITY. The mass of the earth sets up a strong attraction to objects on its surface. This property is called gravity. Using gravity as a reference, a pendulum senses the point where the attraction is greatest. Some gyros are precessed to a vertical reference by a pendulous pickoff and an erection system. These are generally called vertical gyros, and they usually control the pitch and roll attitudes of a missile.

EARTH'S MAGNETIC FIELD. The lines of flux of the earth's magnetic field have served as a reference for many years through the use of the compass. A compass can not be used in a missile control system, but there is an electrical device that senses the earth's magnetic field that can be used. This instrument is known as a flux valve. It primary use is to keep a directional gyro slaved to a given magnetic heading. This gyro usually controls the yaw channel of the autopilot.

BAROMETRIC PRESSURE. Pressure has been used for a long time in aeronautics to indicate altitude. A missile also may use an altimeter which senses barometric pressure and produces an error signal if the missile is not at a preset altitude.

Another instrument which uses the pressure of the atmosphere as a reference is the airspeed indicator. It compares static barometric air pressure to ram air pressure. The difference derived from this comparison is indicated airspeed.

MISSILE AIRFRAME. The airframe of a missile is also a reference. The displacement of the control surfaces can not be referenced to the vertical or to a given heading because the reference would have to change when the missile attitude changes. It is

easier to give control-surface displacement with respect (or with reference) to the missile airframe.

Selsyns are used to indicate the angular position of the control surfaces with respect to a missile airframe. Potentiometers may be used to indicate angular position when the potentiometer is fastened to the missile airframe and its wiper arm is moved by the control surface.

The autopilot sensors have been classified according to the reference they sense. For further information about sensors, refer back to the first section of this chapter.

This concludes the comments on reference units. In the next section, amplifier units are presented.

SECTION D

Amplifier Units

An amplifier is a device, usually containing one or more vacuum tubes, whose output is an enlarged reproduction of the input signal. Voltage amplifier units are designed to increase a voltage across the load in the plate circuit of the amplifier. Power amplifier units are designed to deliver large amounts of power to a plate circuit load without regard to voltage.

Both voltage and power amplifiers are used in missile control systems. Some of the less conventional amplifiers are included in this section.

VOLTAGE-TYPE AMPLIFIER UNITS

Electrical signals in a control system may be either AC or DC. Since some system components may require DC and other components in the same loop may require AC, devices which can convert electrical signals from DC to AC and vice versa are desirable. Modulators perform this function.

A *modulator* is a device which alters the amplitude, frequency, or phase of a wave (carrier) in conformity with the variations of an input signal. In a control system, modulators generally are used to convert polarized DC signals to properly phased AC signals. There are several types of modulators which perform this function in control systems. Mechanical and vacuum-tube choppers and the rectifier modulator are considered. Commutators and paraphase amplifiers also are covered under voltage amplifiers.

Mechanical Choppers

A mechanical chopper is simply a synchronized switch. Its output line is connected alternately to its input line and to ground, or zero-level DC. The explanation which follows refers to figure 5-81.

The coil is energized by AC of the appropriate

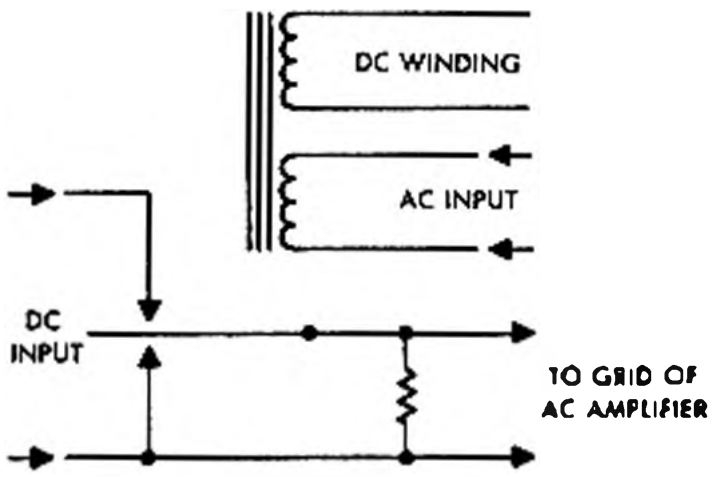


Figure 5-81. Mechanical Chopper

frequency. To prevent the vibrating arm from being attracted twice by the coil during each cycle, some direct current is passed through the coil, or through another coil wound on the same core. When the AC flows in one direction, its magnetizing force is aided by that of the DC and the vibrating arm is attracted. When the AC flows in the other direction, its magnetizing force is canceled by that of the DC, and the attraction on the arm is reduced. The arm thus vibrates at the frequency of the AC. To reduce the power required to drive the arm, the arm usually is tuned so that it vibrates naturally at the proper frequency. The output is then a square wave, having the level of the DC input for a half-cycle and having zero level for the other.

Blocking condensers in the amplifier that follows the chopper remove the DC component so that the signal becomes a square wave with an amplitude proportional to the DC signal level and a phase which reverses when the DC signal changes polarity. This amplifier also discriminates against high-frequency signals so that the square wave ultimately is converted to a sine wave at the frequency of the chopper switching.

A cutaway view of a mechanical chopper is

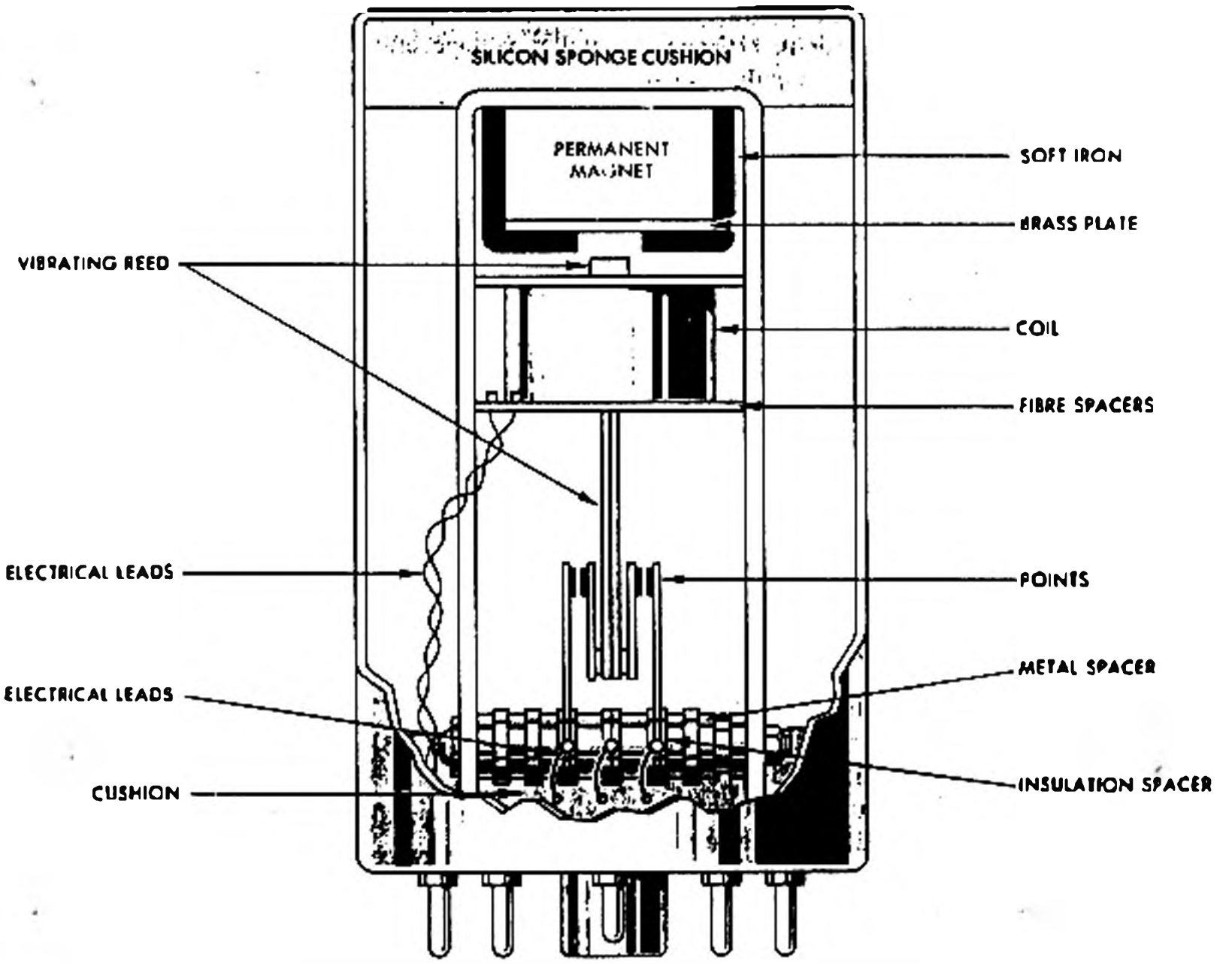


Figure 5-82. Cutaway View of Mechanical Chopper

shown in figure 5-82. In this particular chopper, a permanent magnet is used in place of the DC winding shown in figure 5-81.

Vacuum Tube Choppers

A vacuum tube chopper amplifier circuit, shown in figure 5-83, consists of two tubes, one having a pulsed DC signal on its control grid and the other a fixed voltage. Alternating current of the proper frequency is applied to the screen grids, the voltages at the two screen grids being 180° out of phase with each other. When the tubes carry the same current, the effect of raising the voltage on one screen is exactly balanced by that of lowering the voltage on the other screen. Consequently, the total current flowing in the common load resistor is unchanged, and there is no AC output. This condition is no longer true in the presence of

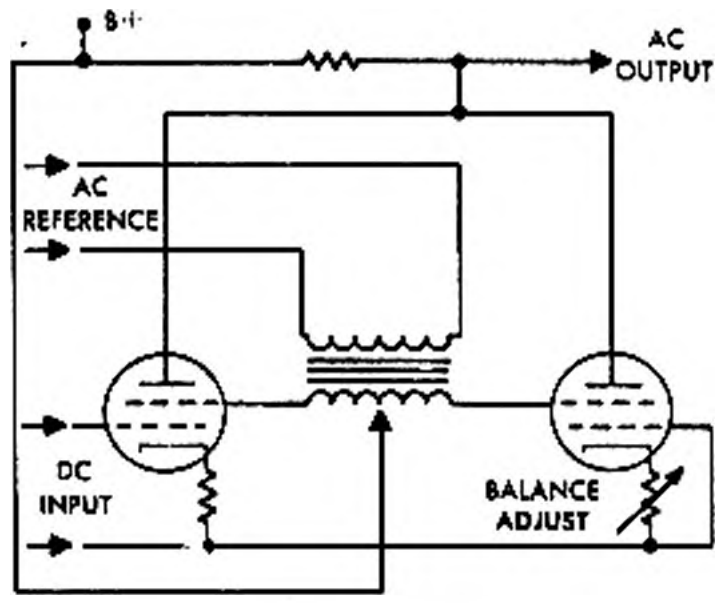


Figure 5-83. Vacuum Tube Chopper Circuit

a DC signal. Now AC appears across the load resistor with an amplitude proportional to the DC level and with a phase which reverses when the

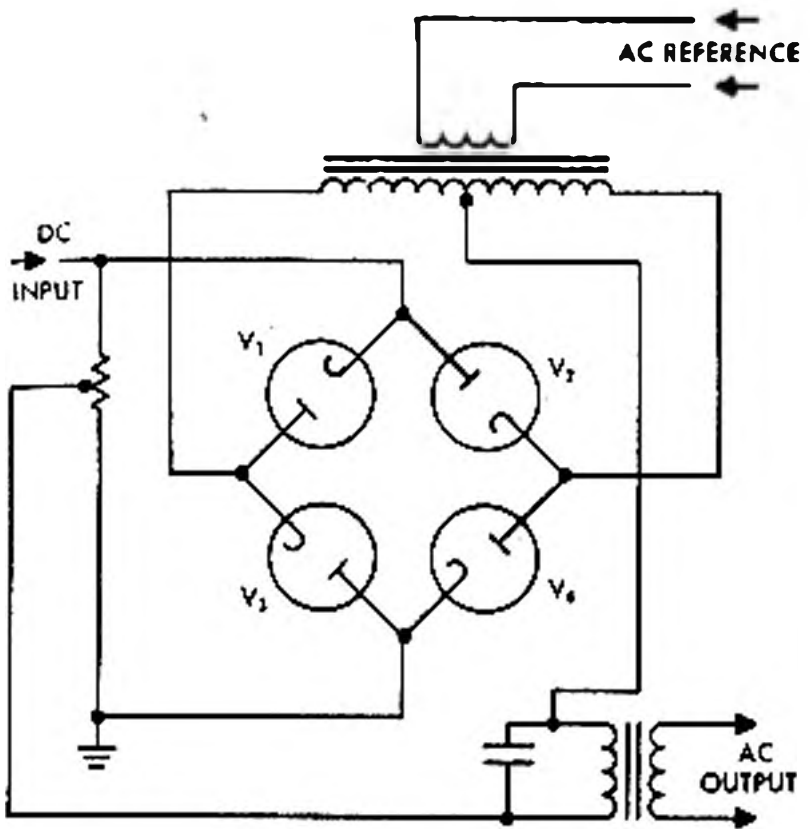


Figure 5-84. Rectifier Modulator Circuit

DC polarity is changed. The variable cathode resistor in one tube is used to balance the currents so that the output drops to zero when the input is zero.

Vacuum tubes can be employed in many other ways to produce the same result, as does the circuit just described.

Rectifier Modulators

A rectifier modulator, as illustrated in figure 5-84, is another type of synchronous switch. During a half-cycle, an AC reference signal causes current to flow through V_1 and V_2; during the other half-cycle, it flows through V_3 and V_4. The nonconducting diodes appear as open circuits, and the conducting diodes appear as relatively low resistances. One end of the input resistance is connected to the center tap of a divider across the output of the reference AC transformer. It is at the same voltage as the center tap on the transformer secondary. The output voltage is equal to that appearing across one-half of the input resistor. Because of the switching action, the output circuit is subjected alternately to the voltages across the upper and lower halves of the input resistor, both voltages being measured with respect to the center of the resistor. The resulting output voltage is equal to half of the input voltage during half of the AC cycle and is equal to half of the amount of the input voltage but opposite in polarity during the other half of the AC cycle. The action is the same as that of a mechanical chopper. Since one side of the input resistor is usually grounded, the output must go through an isolating transformer to remove the DC component.

Commutators

A commutator, sometimes known as a phase-sensitive detector, makes the same conversion as a modulator, but in the reverse order. It receives an AC input and develops a DC output. For an input of one phase, it develops a positive DC output. For an input with a phase difference of 180°, it develops a negative output. For inputs differing from these by 90°, it develops no DC output. There are several types of commutators, three of which are the synchronous mechanical switch, the vacuum tube commutator, and the rectifier commutator.

SYNCHRONOUS MECHANICAL SWITCH. A synchronous mechanical switch, such as shown in figure 5-85, depends on a vibrating arm similar to that

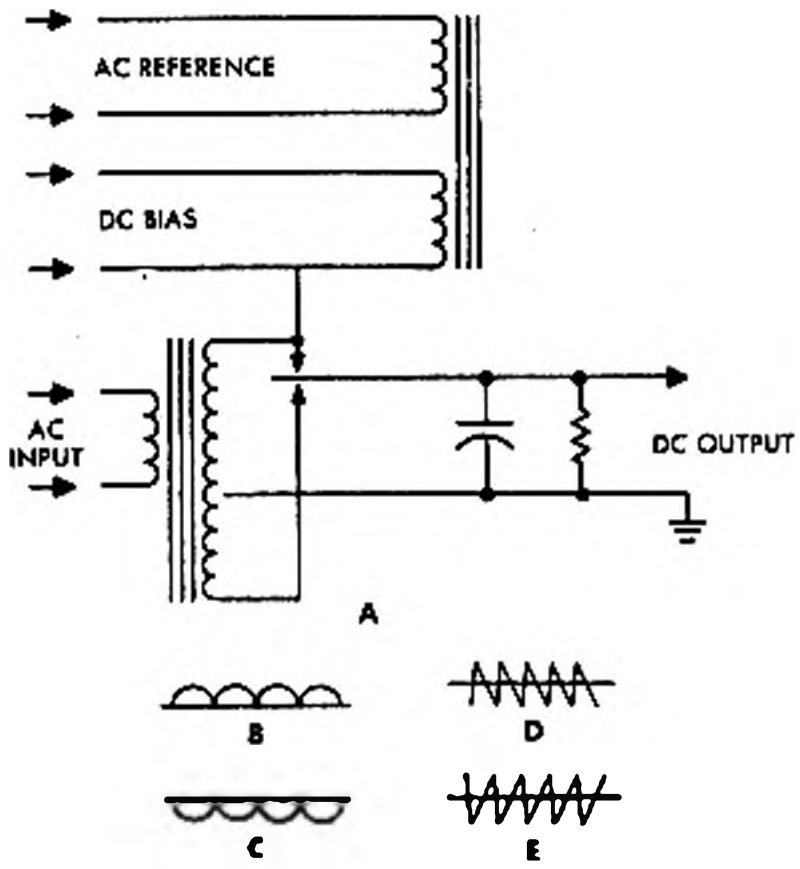


Figure 5-85. Snychronous Switch with Output Waveshapes

of a mechanical chopper. Assume that the voltage at the upper contact is positive during the half-cycle when the arm is up. During the next half-cycle, when the arm is down, the voltage at the lower contact is also positive. The resulting output is similar to that of a full-wave rectifier. After being filtered, the output has a DC component proportional to the amplitude of the AC. If the phase of the input is reversed the output is negative DC.

The signals in these two examples are shown by waveforms B and C in the illustration. For signals differing from these by 90°, the voltage at each contact reverses during the time the arm makes connection, and the average voltage is zero as shown by waveforms D and E. There is, therefore, no DC component and no output from the circuit after filtering.

VACUUM TUBE COMMUTATOR. A number of vacuum tube circuits can serve as commutators. One such circuit is shown in figure 5-86. The DC bias on the control grids is sufficient to hold both tubes at cutoff. Assume that the voltage on the control grids moves in the positive direction at the same time as that on the screen grid of tube V_1. During the half-cycle of current flow, the screen grid of V_1 always is more positive than the screen of V_2, and V_1 draws more current. The cathode of V_1, therefore, is more positive than the cathode of V_2, and the plate of V_1 is more negative than the plate of V_2.

A phase reversal of the control grid signal causes V_2 to draw more current and reverses the polarity of both the cathode-to-cathode and plate-to-plate voltages. If the grid signal differs by 90° from the screen signal, V_1 draws more current than V_2 during one-half of the conducting half-cycle. The situation is reversed during the remainder of the conducting period. The DC component of the output is therefore zero. A circuit of this sort is useful where a DC output need not be developed with respect to ground.

RECTIFIER COMMUTATOR. A rectifier commutator, as illustrated in figure 5-87, acts in much the same way as the rectifier modulator described previously. During a half-cycle, the reference AC causes V_1 and V_2 to conduct. During the other half-cycle, it causes V_3 and V_4 to conduct. The output circuit is connected alternately from the center tap of the input transformer to the junction of the upper diodes and from the center tap to the junction of the lower diodes.

Paraphase Amplifiers

Paraphase amplifiers are sometimes used in place of transformers to operate push-pull circuits. A paraphase amplifier is a combination

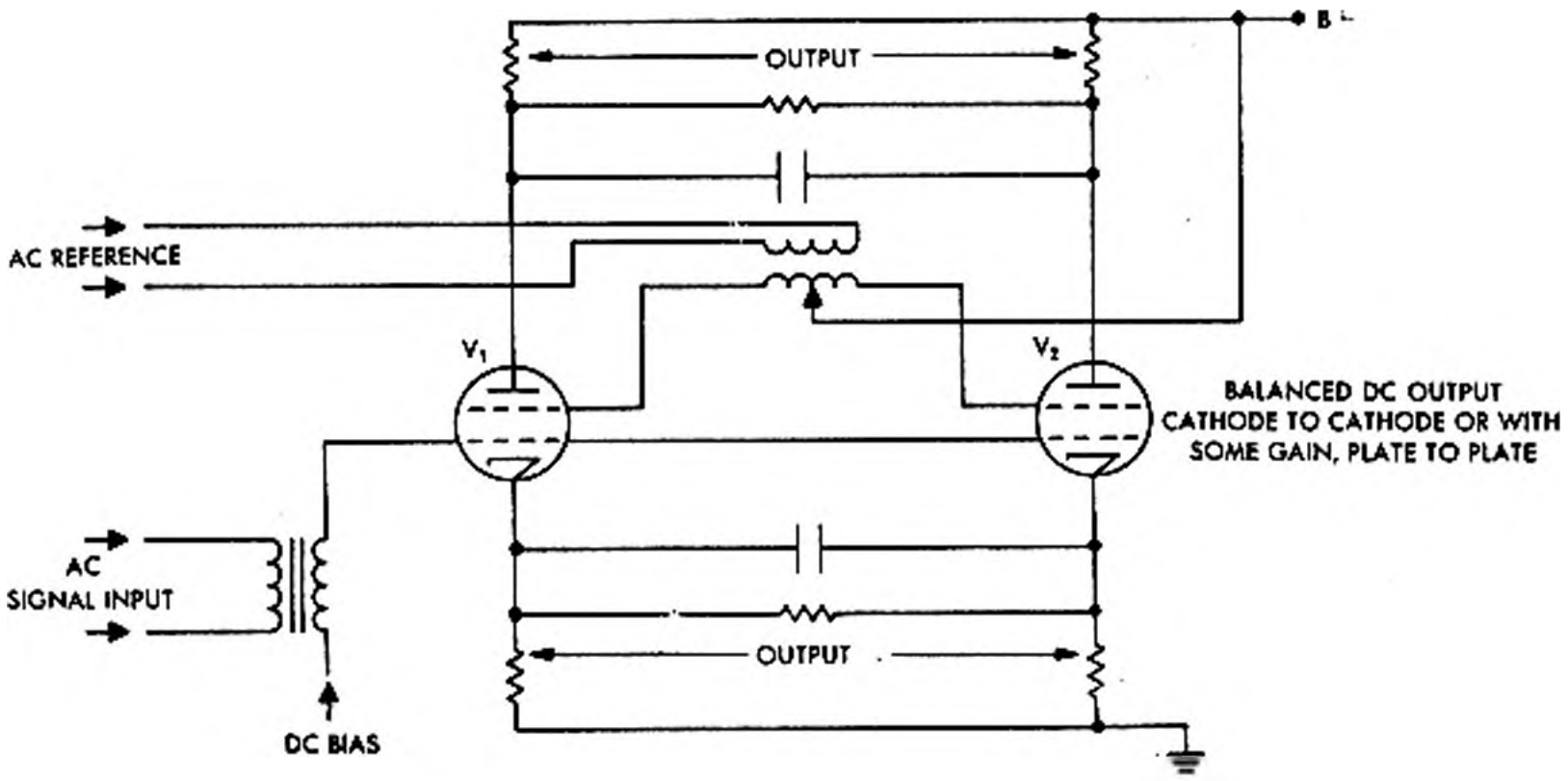


Figure 5-86. Vacuum Tube Commutator Circuit

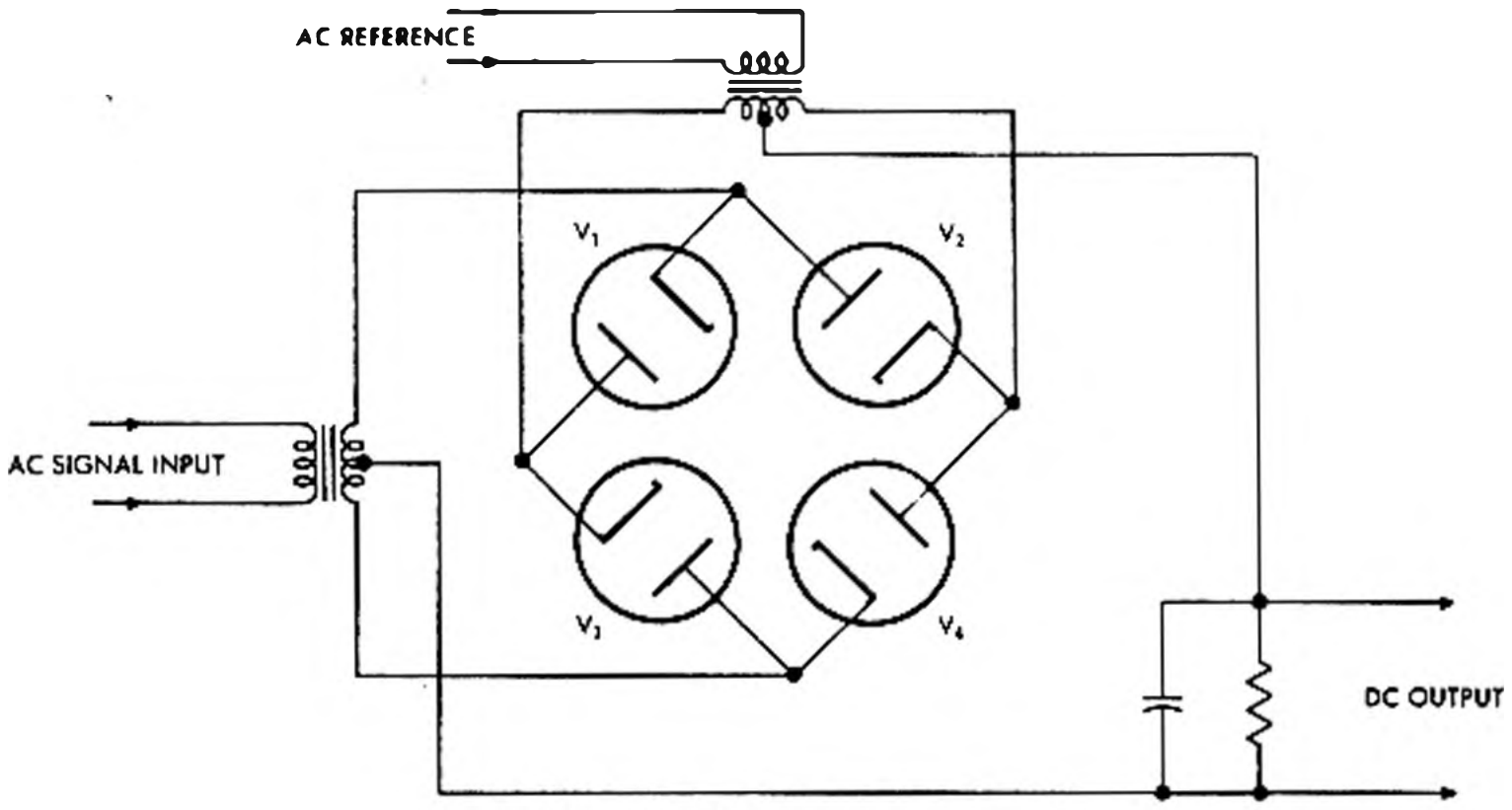


Figure 5-87. Rectifier Commutator Circuit

amplifier and phase inverter. Figure 5-88 is a circuit diagram of one type of paraphase amplifier.

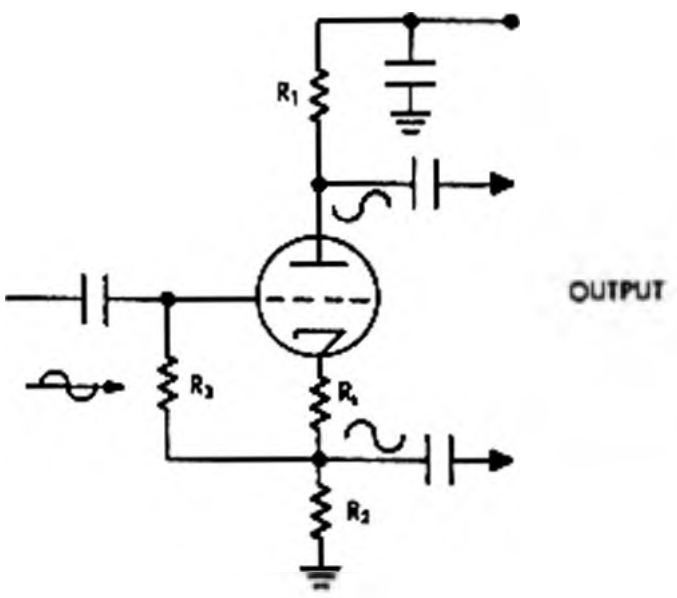


Figure 5-88. Paraphase Amplifier

In this circuit the outputs are taken across the resistors R_1 and R_2. These resistors are of equal value, and since the same current flows through both, equal voltages are developed across them. The voltages across these resistors are opposite in polarity since the output is taken from the negative end of R_1 and positive end of R_2.

Transistor Amplifiers

As in the case of mixer circuits, the vacuum tube has also been displaced to a great extent by the transistor in amplifier applications. There are several circuit arrangements possible for the transistor. Figure 5-89 shows the common-base amplifier where the signal is introduced into the emitter-base circuit and taken from the collector-base circuit. The arrows indicate the direction of current flow. The letter I represents the total emitter current. Transistors in practical use have from 92 to

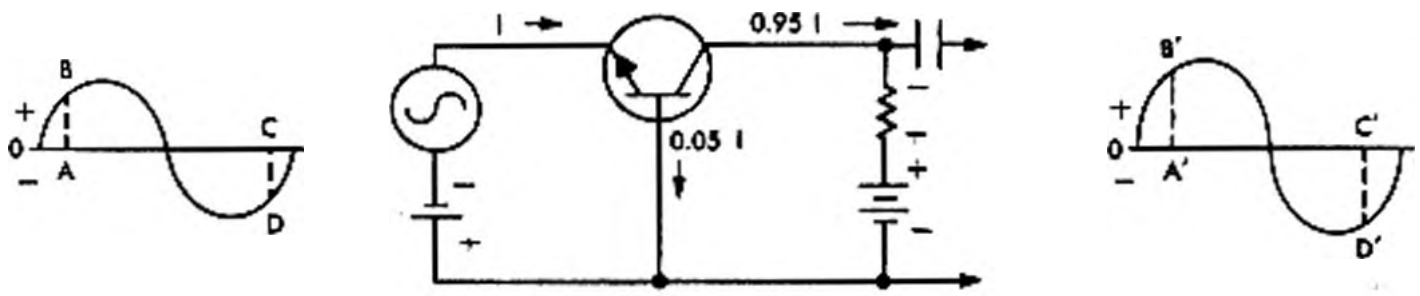


Figure 5-89. Common Base Amplifier

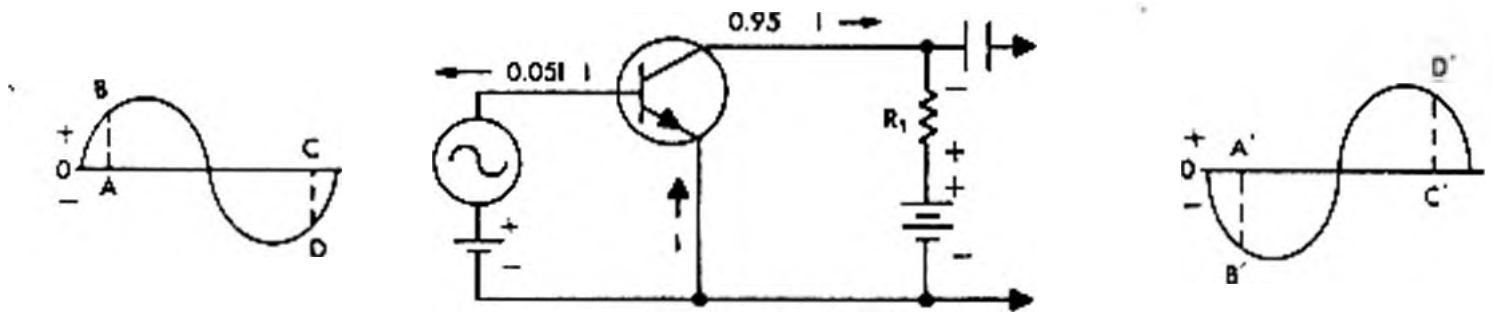


Figure 5-90. Common Emitter Amplifier

95 percent of the emitter current reaching the collector, with the remainder flowing through the base. In the example illustrated, 95 percent of the current (0.95I) reaches the collector and 5 percent (0.05I) flows to the base.

The waveforms in figure 5-89 represent voltage. The input signal is on the left of the figure and the output signal is on the right. With no signal applied, a steady current flows through R_1 producing a constant voltage drop. During the first or positive alternation of an applied signal, the signal opposes the battery voltage between the base and emitter. This decreases the forward bias. The decrease in bias reduces the total current flowing through the emitter. The collector and base currents are also reduced. The smaller current flow through R_1 results in a smaller voltage drop and the output voltage increases. The input voltage AB produces the output voltage A'B'.

During the second or negative alternation of the input, the signal aids the bias and allows more current flow through the emitter. This increases current flow through R_1 and the voltage drop increases. The output voltage decreases and you can see that input voltage CD produces output C'D'. There is no phase reversal between the input and output of a common-base amplifier.

Connecting the transistor as illustrated in figure 5-90 provides a common-emitter amplifier. Again the arrows indicate the direction of current flow.

Now the positive signal input aids the forward bias produced by the base-emitter battery and increases current flow through the emitter. This increases current flow through R_1 and increases the voltage drop. The output voltage decreases and the output is a negative going wave.

The negative portion of the input signal opposes the forward bias on the emitter. This decreases emitter current and reduces the flow through R_1 decreasing the voltage drop. Now the output voltage increases and as you see in figure 5-90 there is a 180° phase shift across this amplifier.

The common-collector amplifier is illustrated in figure 5-91. The emitter current flows through the collector and base the same as in the two previously described amplifiers.

As in the previous explanations, when the input signal aids the forward bias, the total emitter current increases. The increased current through the load resistor R_1 increases the voltage drop across R_1. During the positive alternation of the input signal, the output voltage is positive. A negative input signal decreases the forward bias. During the negative alternation of the input, the emitter current is reduced. This results in less current flow through R_1 and less voltage drop. The output voltage is negative with a negative input signal. There is no reversal in the polarity of the

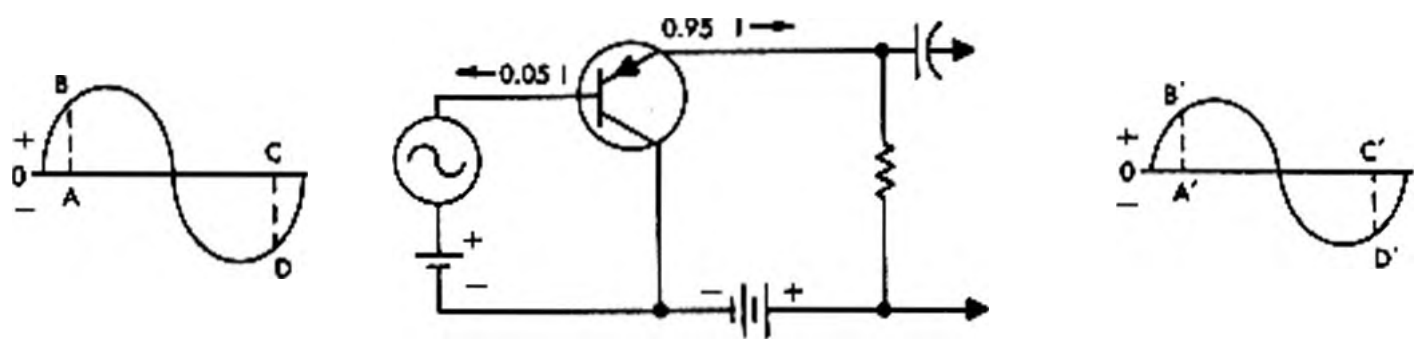


Figure 5-91. Common Collector Amplifier

signal from the input to the output in the common-collector amplifier.

Summarizing the preceding examples of transistors you can see that how each type of circuit is used depends on transistor, circuit function, and design choice. Basic choices are made based on the following facts.

- Common-emitter and common-collector amplifiers yield substantial current gain. For the common-base amplifier the gain is always less than one.
- Common-base and common-emitter amplifiers produce appreciable voltage gains. For the common-collector amplifier, it is always less than one.
- Highest power gain occurs when the product of the current-voltage gain is greatest. Since the common-emitter amplifier has both high current gains and high voltage gains, its power gains are the highest.

ADVANTAGES OF TRANSISTORS. The transistorized amplifier operates on much less power than the tube amplifier and develops more power output. The transistor requires no heater for its operation. They do not generate the heat which is one of the biggest problems encountered with vacuum tubes. Heat not only causes many failures in equipment, but much weight and space is devoted to cooling equipment in an effort to overcome the heat problem. The elimination of this equipment as well as the small size of the transistors and circuit components makes the transistorized amplifier especially advantageous to the design of guided missiles where size and weight are critical factors.

Other advantages to the use of transistors are in their operating characteristics. The transistor has a much better signal-to-noise ratio than the vacuum tube. There is less inter-electrode reaction and much less drift in a transistorized circuit.

DISADVANTAGES OF TRANSISTORS. There are some disadvantages to the use of transistorized systems. A transistor is sensitive to direct heat such as may be encountered when a circuit is designed to use vacuum tubes that are in close proximity to transistors. Transistors are easily damaged by application of a hot soldering iron. Special soldering techniques are required in repairing transistorized circuits.

POWER-TYPE AMPLIFIER UNITS

As stated before, power amplifiers deliver large amounts of power to a plate circuit load. Before investigating power amplifiers, you should become familiar with saturable reactors which are used in power amplifier units of missile control systems.

Saturable Reactor

A saturable reactor is, in effect, an inductor whose inductance can be varied by means of a control current. A magnetizing force (H) is developed when current flows through the winding of an ironcore inductor. This sets up a magnetic field (G) in the core. When the current is changed, the accompanying change in H causes a change in G. The change in G induces voltage in all of the windings that encircle the iron core. The induced voltage is proportional to the rate at which G changes. If G is proportional to H, the induced voltage is proportional to the rate at which the current changes. In some types of core material, the relation between G and H is like that shown in figure 5-92. A winding on such a core shows induced voltage when AC is passed through it. If DC is added in the same winding or in another winding, the variation in H no

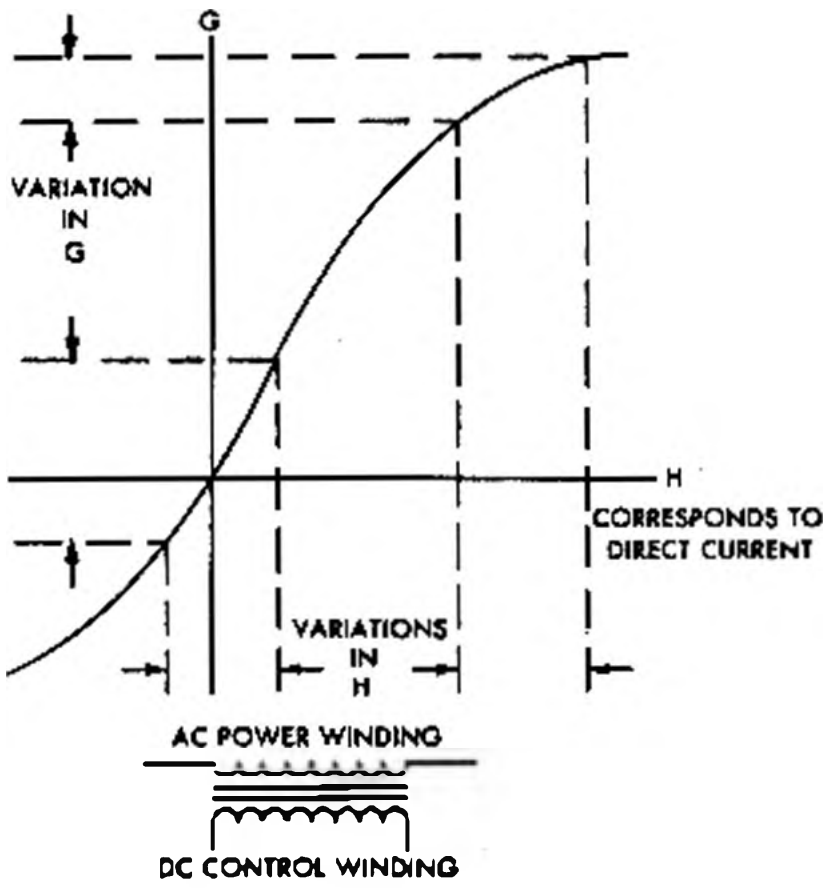


Figure 5-92. G-H Curve for Saturable Reactor

longer takes place about an average value of zero. Instead, H oscillates about a value which corresponds to the DC current. If this point is sufficiently far beyond the bend in the G-H curve, there will be only a small change in G and a small induced voltage. The result is that an inductor made of such material can be designed to have a high inductance when no DC flows in the windings and a much lower inductance when DC does flow.

Saturable-Reactor Control of AC Inductance Motor

Figure 5-93 shows a pair of saturable reactors connected so as to control an AC induction motor. The main field of the motor is connected to the line in the usual way. One end of the control field winding is connected to the center tap of a transformer fed by the line. The other end is connected through a pair of saturable reactors to both ends of the transformer. When no DC flows through either reactor, the control field is not energized. When DC flows through either reactor, the reactor's inductance is decreased so that the control field is connected to one end of the transformer through a high impedance and to the other end through a low impedance. The control field is energized in one direction or the other, and forward or reverse torque is developed by the motor. The DC through the reactors is supplied by a pair of thyratrons. These tubes receive an AC error voltage and are connected to an AC plate supply.

The AC for the motor and the reactors may be of one frequency, and the error voltage and the thyratron anode supply may be of another. Vacuum tubes can be used instead of thyratrons if the error signal is AC. If the error signal is DC, the signal might possibly be used directly to saturate the reactors, or a DC amplifier may be used when the signal is not strong enough to produce the desired degree of saturation.

Magnetic Amplifiers

A magnetic amplifier uses saturable reactors. These reactors have extra windings that carry a direct current called the bias current. Figure 5-94 shows a magnetic amplifier circuit controlling an

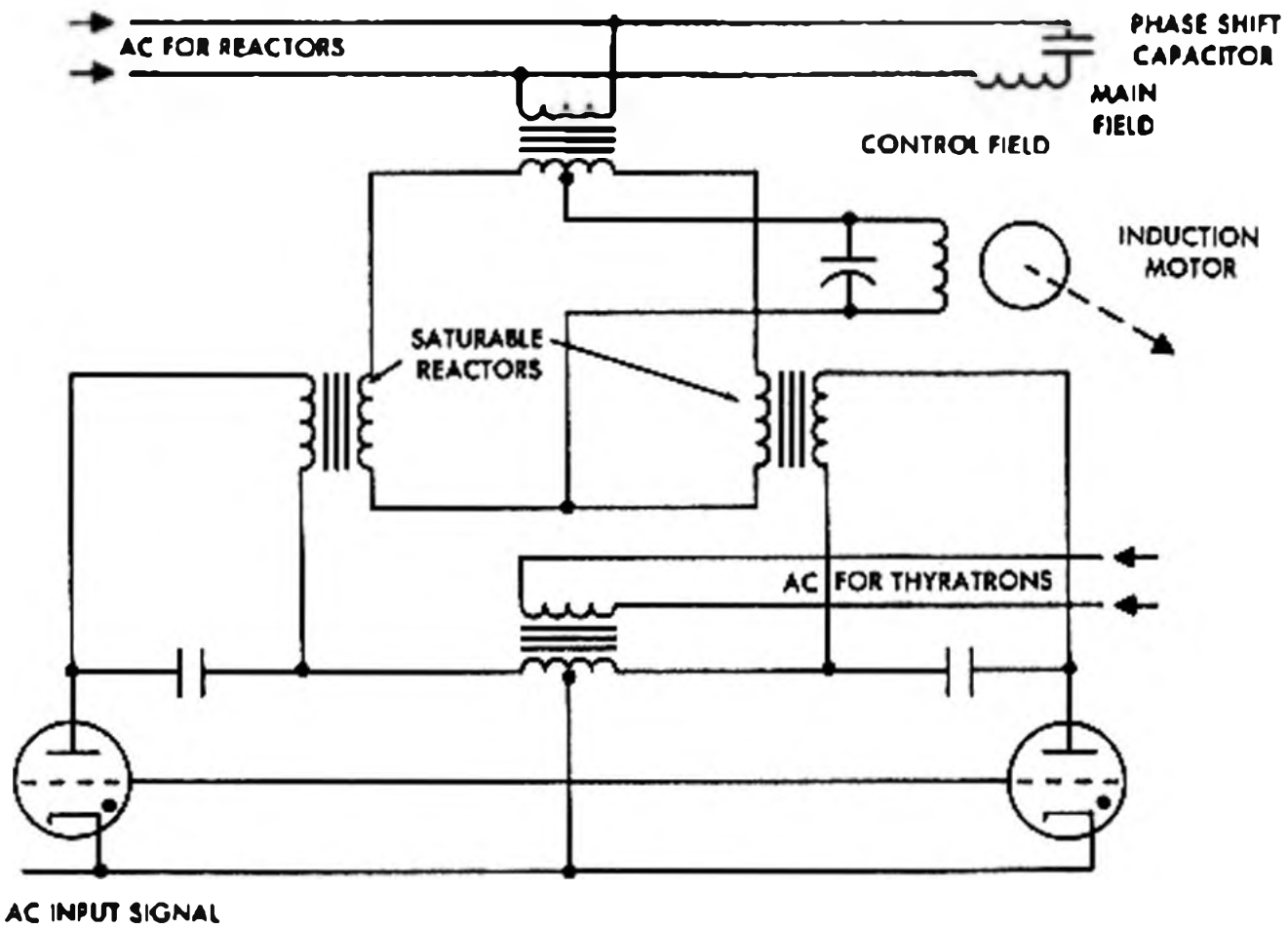


Figure 5-93. Saturable Reactor Control Circuit for AC Motor

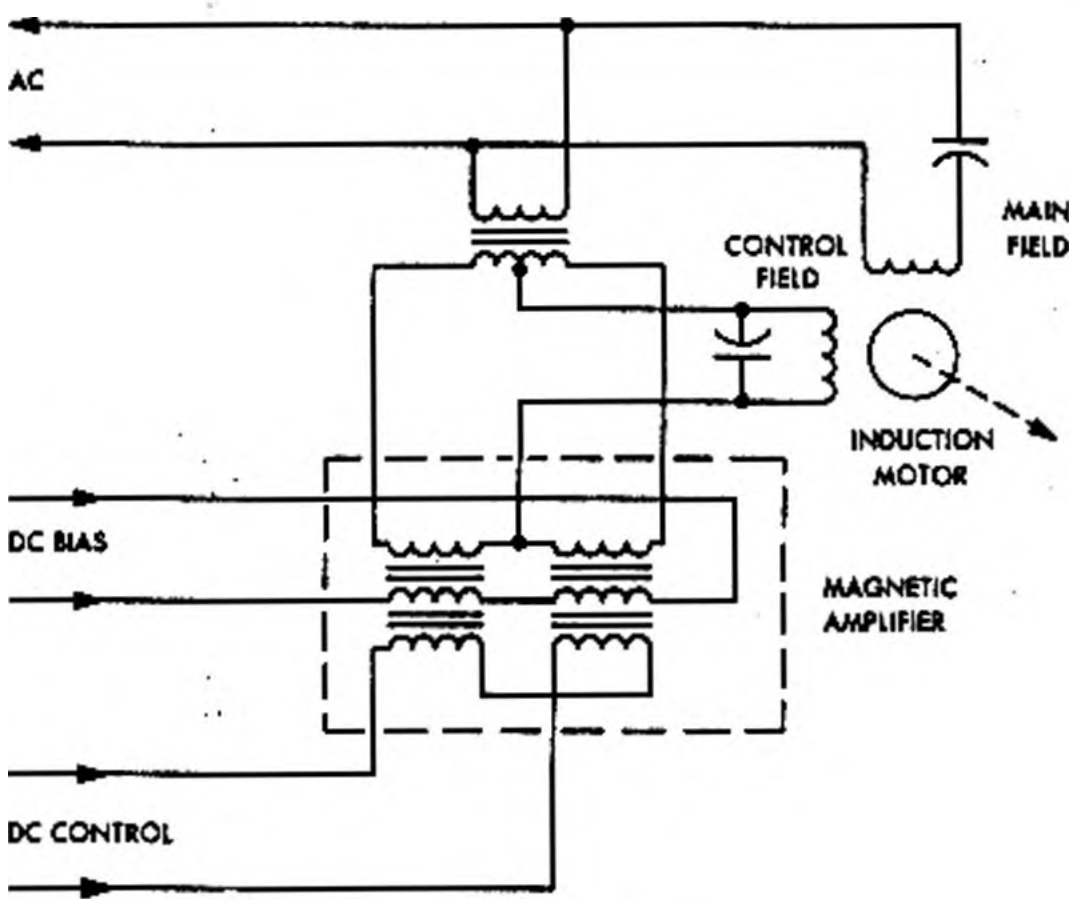


Figure 5-94. Magnetic Amplifier Control Circuit for AC Motor

AC motor. It is nearly the same as the circuit for saturable reactor control except that the two reactors have their control windings connected in series. The direction in which the bias current flows is so chosen that the magnetizing force developed by the bias current aids the force developed by the control current in one reactor and bucks the force of the other. In the absence of a control signal, the DCs are the same in each reactor, and the control field of the motor is not energized. The control current and the bias current are, in effect, added in one reactor and subtracted in the other. Both reactors are partially saturated in the beginning, and the control current saturates one even more and reduces the saturation of the other. A DC control signal is required, but this can be obtained from an AC error signal by means of a phase-sensitive detector.

Both the magnetic amplifier and the saturable reactor suffer from the fact that, because of the inductance of the control winding, a perceptible time is required to establish the control current. A magnetic amplifier is more desirable than a saturable reactor because it has neither moving parts nor vacuum tubes.

This presentation of amplifier units is not a complete explanation of all types of amplifiers used in control systems. The intent here has been to familiarize you with the more common types.

SECTION E

Controller Units

A controller unit in a missile control system controls the operation of the actuator which is responding to an error signal it receives from the sensing element. In some systems, an amplifier, whose output is being applied to a motor, is the controller unit. This section considers only those controller units, other than amplifiers, which control actuator units in a system. Note that the "controller" block follows the "amplifier" block in figure 5-3.

A description of solenoids follows because they are important components of controller units.

SOLENOIDS OPERATE VALVES AND RELAYS

A solenoid consists of a coil of wire wound around a hollow cylinder. It is used to produce a magnetic field. If a movable core of soft iron is placed inside the cylinder, the field of the coil tends to center the core into the coil when cur-

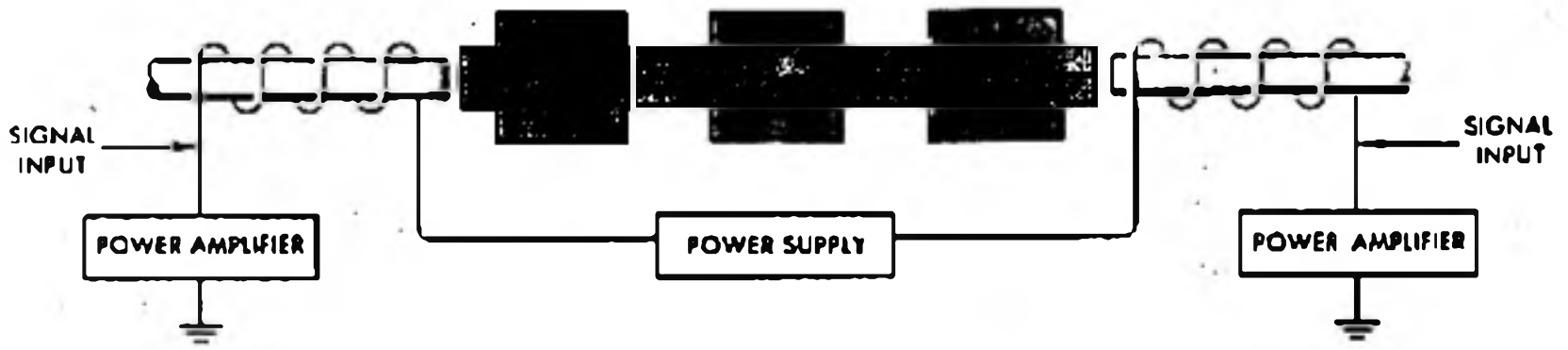


Figure 5-95. Double Acting Solenoid

rent is flowing. Solenoid coils with movable cores are used for remote control of various units such as solenoid-operated valves and relays.

If two solenoids are arranged as shown in figure 5-95, they can control movement of a needle valve in a hydraulic system, pneumatic system, etc.

TRANSFER VALVE CONTROLS FLUID TO ACTUATING DEVICE

Transfer valves are used in control systems having hydraulic actuators. The transfer valve unit illustrated in figure 5-96 consists of a double-acting solenoid which controls the position of a spool in the transfer valve unit. Figure 5-96A shows the spool in the closed position. The position of the spool determines how much fluid is allowed to flow to the hydraulic actuating device. In figure 5-96B the spool valve is shown displaced from its center position to allow fluid to flow under pressure to the actuating unit. Solenoid S_1 has a heavier current flow through its winding than solenoid S_2, thus S_1 attracts the spool. This action opens the ports. Now the hydraulic fluid is able to flow through the transfer valve to one side of the actuator. The fluid on the opposite side of the actuator is returned back through the exhaust port as shown. A more complete drawing of the transfer valve and actuator is shown in figure 5-97.

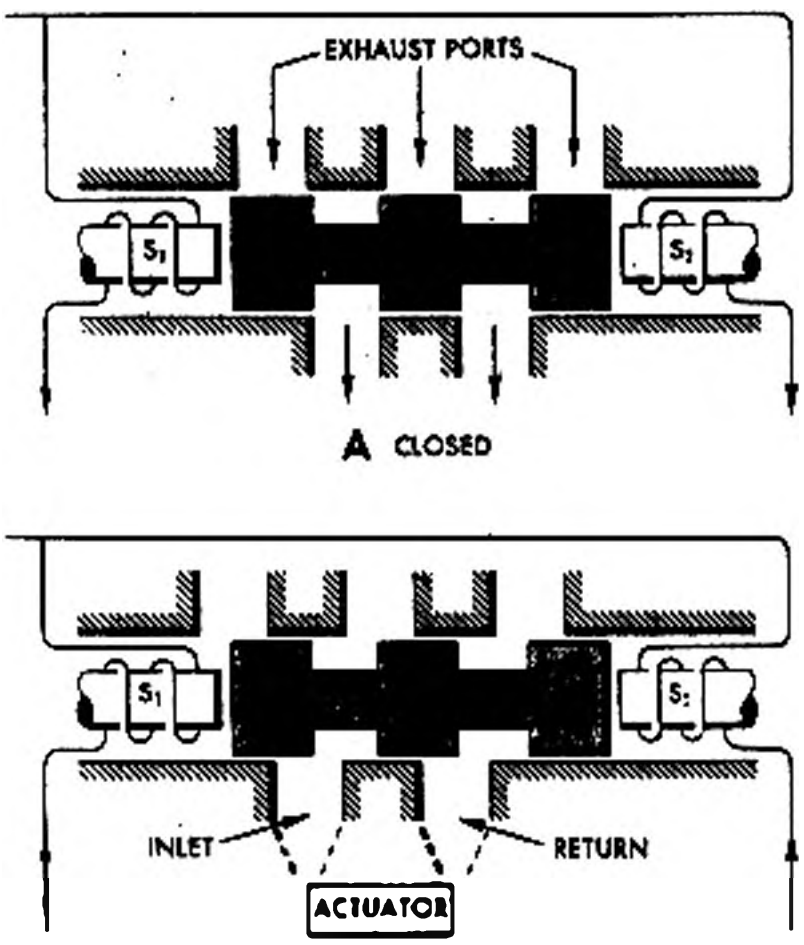


Figure 5-96. Transfer Valve

RELAY SWITCHES CONTROL HEAVY-CURRENT CIRCUITS

Relay switches are used for remote control of heavy-current circuits. They are placed directly between the source of power and the controlled unit so that the cables carrying heavy current will be as short as possible. A relay switch consists of a coil or solenoid, an iron core, and fixed and movable contacts. Small wires connect the solenoid coil terminals with the source of power which is the control signal. When a control signal is present, an electromagnetic field is set up around the coil.

In the relay switch shown in figure 5-98, the iron core is fixed. When the control signal is present, the core is magnetized by the field set up around the coil. The pull of the core on the piece of soft iron overcomes the force of a spring, thus closing the contacts. This action completes the

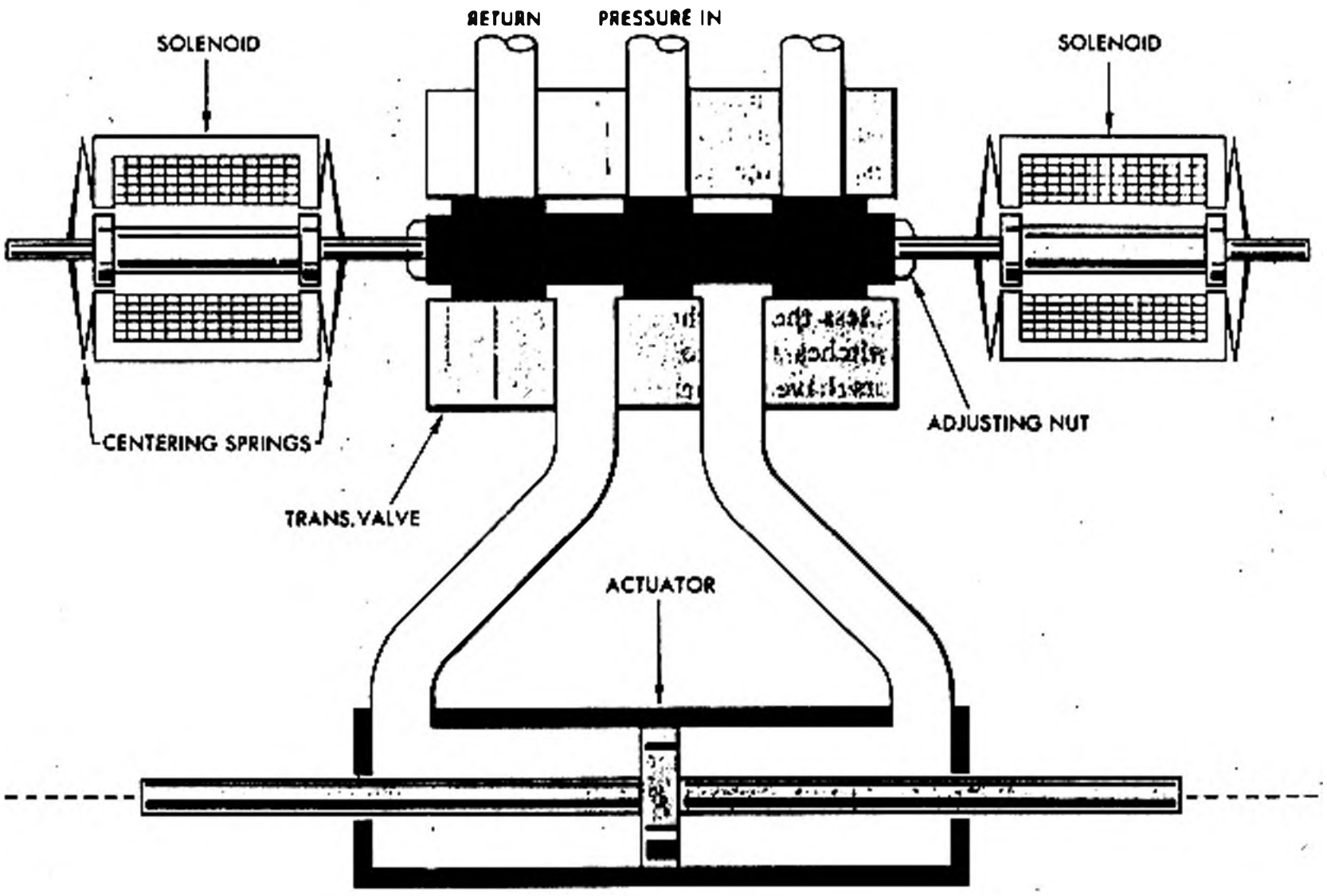


Figure 5-97. Hydraulic Transfer Valve and Actuator

heavy-current circuit. When the control signal is removed, the field around the coil collapses. Then the spring separates the contacts, breaking the heavy-current circuit.

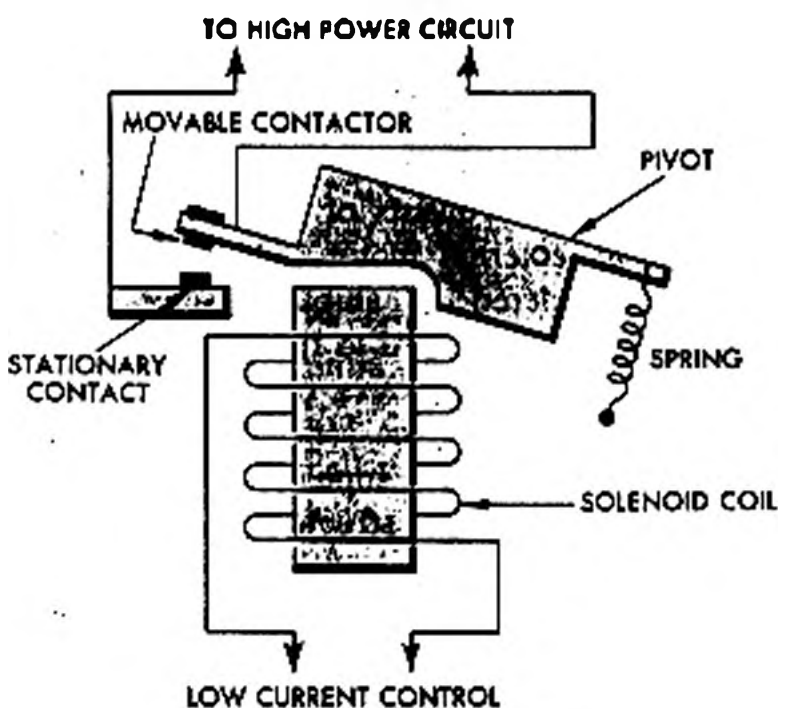


Figure 5-98. Relay Switch with Fixed Iron Core

In another type of relay switch, part of the core is movable, as indicated in figure 5-99. Contacts are attached to the coil mounting but insulated from it. When the control switch is

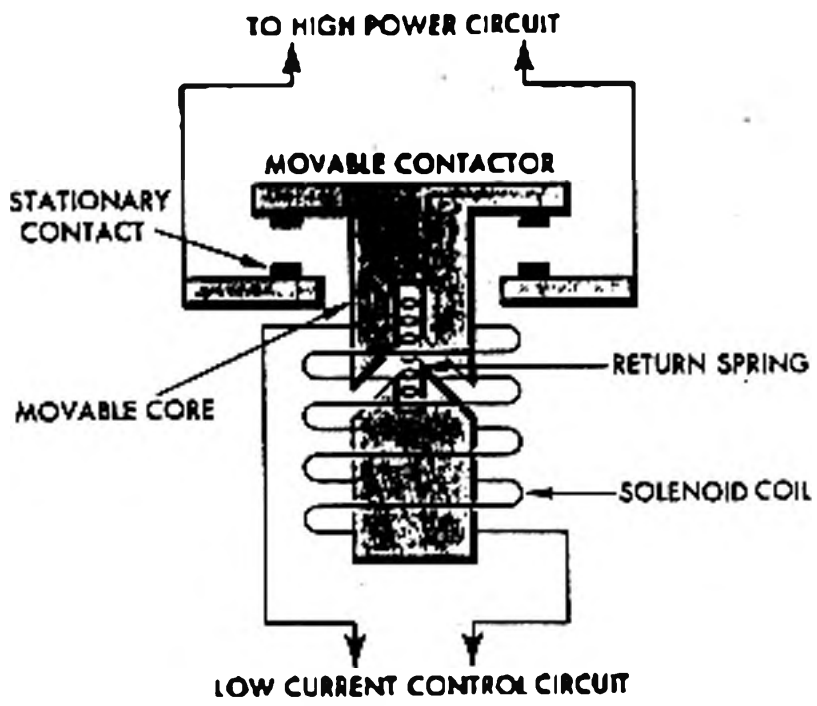


Figure 5-99. Relay Switch with Part of Iron Core Movable.

closed, the field around the coil causes the movable parts of the core to be drawn into the coil, closing the contacts and completing the heavy-current circuit. When the control switch is opened, the field around the coil collapses, and the return spring returns the movable core to its original position separating the contacts.

The quicker a relay carrying a large current is opened, the less it will arc and the less the switch contacts will be burned. Relay switches used to control the circuits of large motors have strong return springs which open the switches quickly.

Relay switches have either an insulating spacer on each coil terminal or an insulating spacer on one coil terminal and a metal spacer on the other. If a metal spacer is used, it grounds the terminal to the coil case. No ground wire to that terminal is required.

To permit the circuits controlled by heavy-duty relay switches to carry heavy currents and still protect them against short circuits, a special type of fuse called a current limiter is used. A current limiter permits the large overloads required for starting motors but "blows" before the circuit is damaged if the overload is continued.

In figure 5-100, two air-pressure input lines of an air-actuated relay are connected to the relay, one to each side of a diaphragm. This type of relay is used in a pneumatic-electric control system.

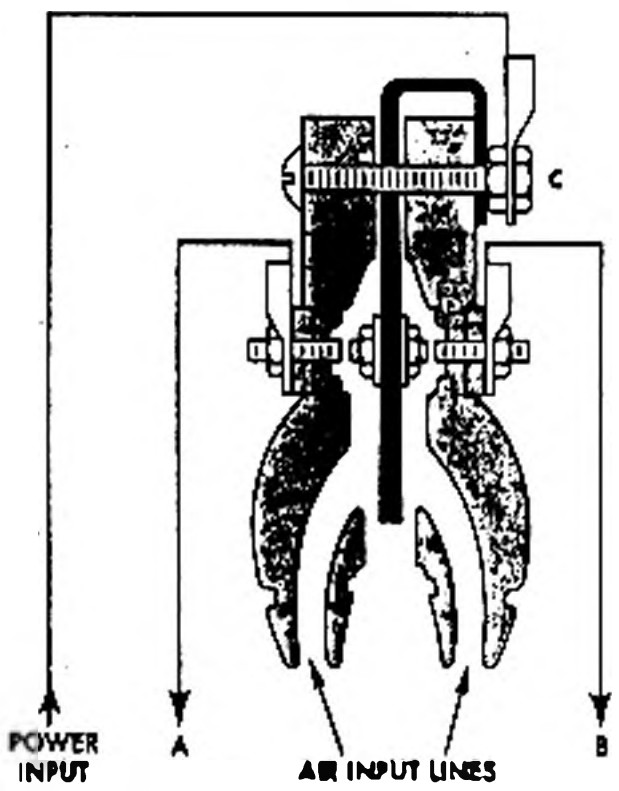


Figure 5-100. Air-Actuated Relay

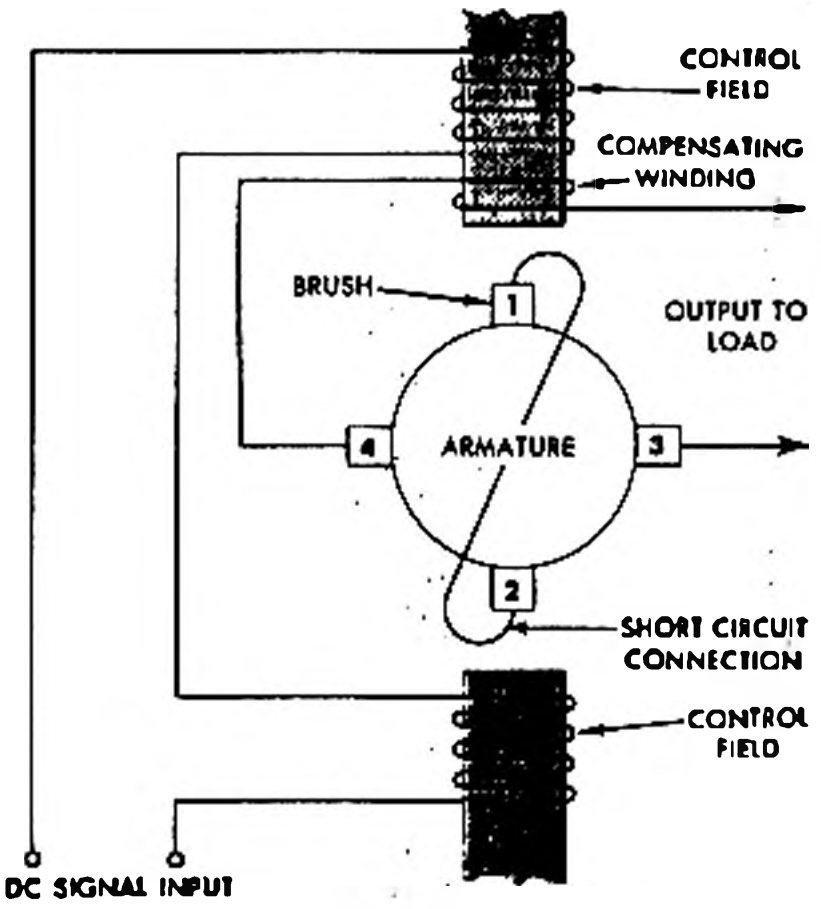


Figure 5-101. Basic Schematic of Amplidyne

When an error is present in the system, an air pressure is applied to one side or the other of the diaphragm. The diaphragm is forced over to complete an electrical circuit connection from terminal C to one of the other two terminals, A or B. In this manner, power is transferred across the relay into a suitable actuating circuit. Similar relays might be found in a hydraulic-electric system in which an electrical circuit is actuated by hydraulic pressure rather than air pressure.

AMPLIDYNES AS SPECIAL TYPE OF DC GENERATORS

An amplidyne consists of a DC generator rotated by an external motor. The DC generator can be considered an amplifier since a small amount of power applied to the field coil controls many times as much power in the output of the generator. An amplidyne schematic is shown in figure 5-101. The DC input control voltage is used to excite the field structure which is shown as a pair of poles above and below the armature. Rotation of the armature in the control field induces a voltage in the coils as they pass through this magnetic field. Since these coils are in a vertical plane when crossing the control field, the

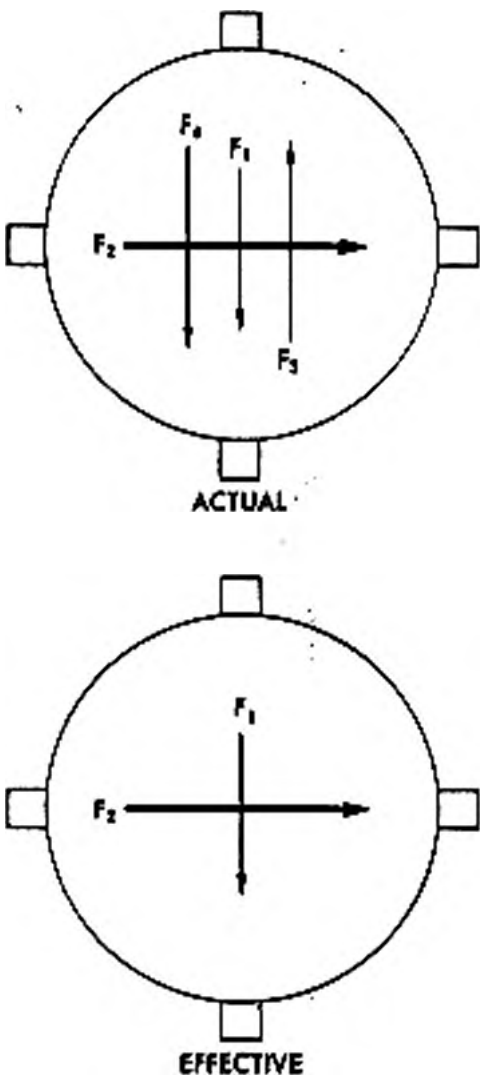


Figure 5-102. Magnetic Fields in Amplidyne

voltage appears at the upper and lower brushes, numbered 1 and 2. If a load were connected to these two brushes, the device would be an ordinary DC generator. Note that in the drawing, brushes 1 and 2 are tied together, or short-circuited. The resulting high current flowing in the armature produces a second magnetic field which is much stronger than the control field and at right angles to it. The lines of force of this second magnetic field lie in a horizontal plane. Armature coils cutting the second magnetic field have a voltage induced in them which appears across the second set of brushes, numbered 3 and 4, located on the horizontal axis. This induced voltage is the output of the amplidyne.

A problem of major concern relative to the amplidyne is the cancellation of the control field. When load current flows, it flows through armature coils which produce a magnetic field. The magnetic field tends to cancel the control field. This cancellation is an undesirable feature because it reduces the amplification and causes poor regulation of the amplidyne system.

This cancellation tendency can be corrected by adding auxiliary windings to the field structure and passing the output current through them. The magnetizing force produced by passing the output current through the auxiliary field can be made to balance that produced by the flow of output current through the armature coils. The various magnetic fields are represented by arrows in figure 5-102. As shown, magnetic fields in an amplidyne occupy certain positions and have certain directions, with respect to each other.

F_1 represents the control field produced by the input voltage. F_3 is the magnetic field produced by high currents passing through the shorted circuits. F_3 is the magnetic field set up by the coils through which the output current flows and which opposes the control field. F_4 is the magnetic field produced by the output current flowing through the auxiliary windings that have been added to the control field structure. This magnetic field, F_4, is equal and opposite to F_3 and balances, or cancels, the effect of F_3.

The operating principle of an amplidyne has been explained. Application of the amplidyne is represented by figure 5-103.

The amplidyne drive motor is connected to a DC supply which drives it at a constant speed and always in the same direction. The armature of the amplidyne, which is connected to the shaft of the amplidyne drive motor, is also driven at constant speed and always in one direction. The armature leads of the load-driving motor are connected to the amplidyne generator output brushes. The operation of the load-driving motor depends on the DC voltage generated by the amplidyne generator armature. In the illustration, one of the amplidyne control field windings is positioned on the field poles in such a way that if equal currents flow through both windings, the magnetic fields created will cancel. No voltage will be induced in the rotating armature since it is not cutting any lines of flux. When a greater current flows through the coil shown on the right (white lines) than through the coil on the left, a magnetic field, having north and south poles as shown, is established. A voltage is generated in the rotating armature. Current then flows through the armature of the load-driving motor causing it to rotate.

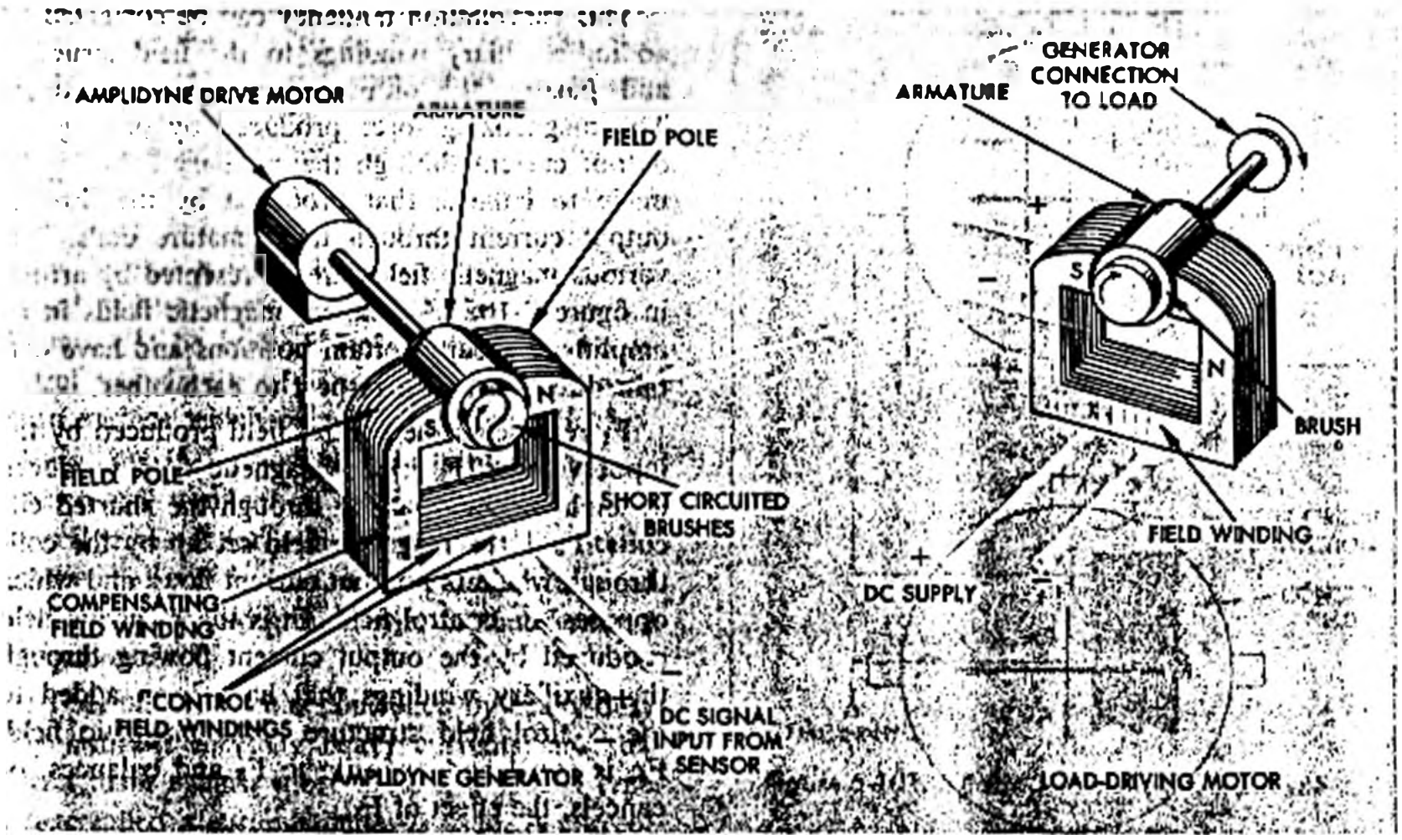


Figure 5-103. Amplidyne Generator Connected to Load-Driving Motor

After the "controller" block in the missile control block diagram comes the "actuator" block. Actuating units are covered in the next section.

SECTION F

Actuator Units

In any missile flight control system, the energy that originates with detection of an attitude error must be transformed into mechanical motion to actuate the appropriate control device. The control system component that accomplishes this energy transfer and transformation at the load end of the control system is the actuating unit, or actuator.

An actuator for a control system must be selected on the basis of the characteristics of the other components of the control system. It must be able to respond rapidly to the input error signals. The response involves the time element between reception of the signal and actuation of the control device. At the same time an actuator must produce an output of proper type and magnitude for a given input. The output of the actuator must be in terms of some function proportional to the error signal and must meet the power output requirement necessary to move the load.

Generally, actuating units employ one or more of the following methods of energy transfer; hydraulic, pneumatic, or electrical. Each method has certain advantages and each also has some disadvantages. The advantages and disadvantages of each method are explained in chapter 7, which deals with complete control systems.

HYDRAULIC ENERGY TRANSFER UNITS

The physical principle upon which the transfer of energy by hydraulic means is based is known as Pascal's law. This law states that whenever a pressure is applied to a confined liquid, the pressure is transmitted undiminished in all directions throughout the liquid regardless of the shape of the confining container.

Application of this principle has been widely and successfully used for many years. A great number of the tools and machines that you have occasion to use from day to day contain some

type of hydraulic energy transfer unit. Proof of this statement is found in the hydraulic doorstop, hydraulic jack, hydraulic car brakes, hydraulic car lift, and automatic transmission for the automobile. In the missile industry, applications of this same hydraulic principle are effectively used in the transfer of energy through automatic control systems.

Generally, hydraulic energy transfer units are simple in design and construction. The major advantage of a hydraulic system is that it eliminates the use of a complex arrangement of gears, levers, pulleys, etc., for the transfer of energy. Reaction time of hydraulic systems is relatively rapid because there is little slack or play to take up as compared to some mechanical systems. The forces generated at one point are transmitted rapidly and with small energy loss over considerable distances. Also, the liquid component (hydraulic fluid) is not subject to breakage, and the complete mechanism is subject to less wear as compared to a totally mechanical system.

Before studying any particular hydraulic unit as used in a missile control system, consider the construction and operation of a typical hydraulic lift.

Basically, the hydraulic lift consists of a suitable container fitted with two pistons and filled with a fluid which acts as the medium of energy transfer. This construction is shown in figure 5-104.

Note that an external force of 40 pounds is applied to the top side of the smaller piston. If the smaller piston has a surface area of 4 square inches, the 40 pounds of applied force produces a pressure of 10 pounds per square inch on the surface of the confined fluid. The larger piston, upon which the load is placed, has a surface area of 600 square inches. The system pressure of 10 pounds per square inch acts on each square inch of the large piston, producing a working force of 6000 pounds. You may conclude, then, that in addition to being used for the transfer of energy, hydraulic systems can also be used to produce a large output force with the expenditure of a much smaller input force.

Also note that the varying size of the container for the fluid and the irregular path of the delivery line do not (within limits) affect the transmission of the applied pressure. Another factor to remember is that the ratio of output piston area to input piston area equals the ratio of output to input force. For a given pressure, the force produced by the working piston is directly proportional to its area.

In certain applications of hydraulic systems, the distance moved by the output piston may be of primary importance instead of force amplification. For example, it may be desirable to have the output piston move some predetermined distance. This factor can be controlled by using pistons with properly related surface areas. Consider the hydraulic system in figure 5-105. Both pistons have surface areas of 2 square inches. If the force applied to the input piston causes it to move a distance of 3 inches, then 6 cubic inches of fluid

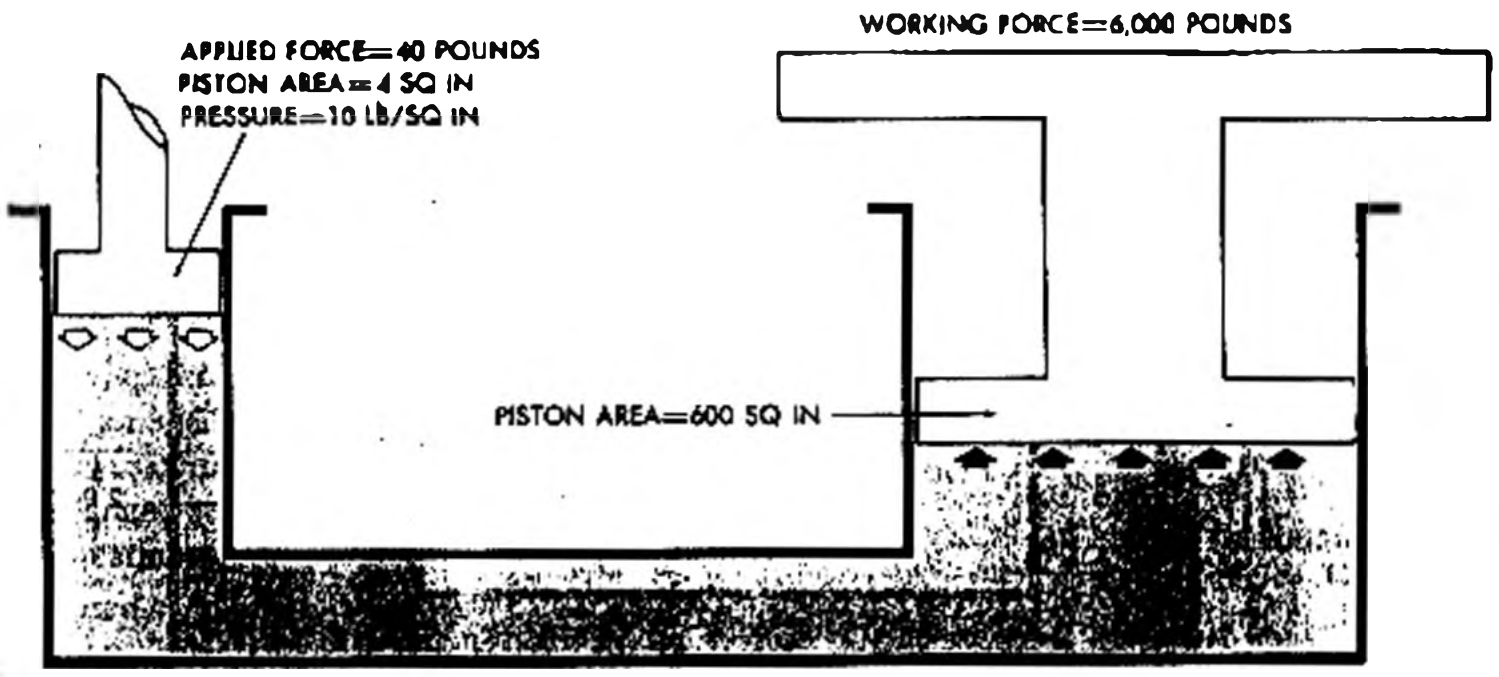


Figure 5-104. Basic Diagram of Hydraulic Lift

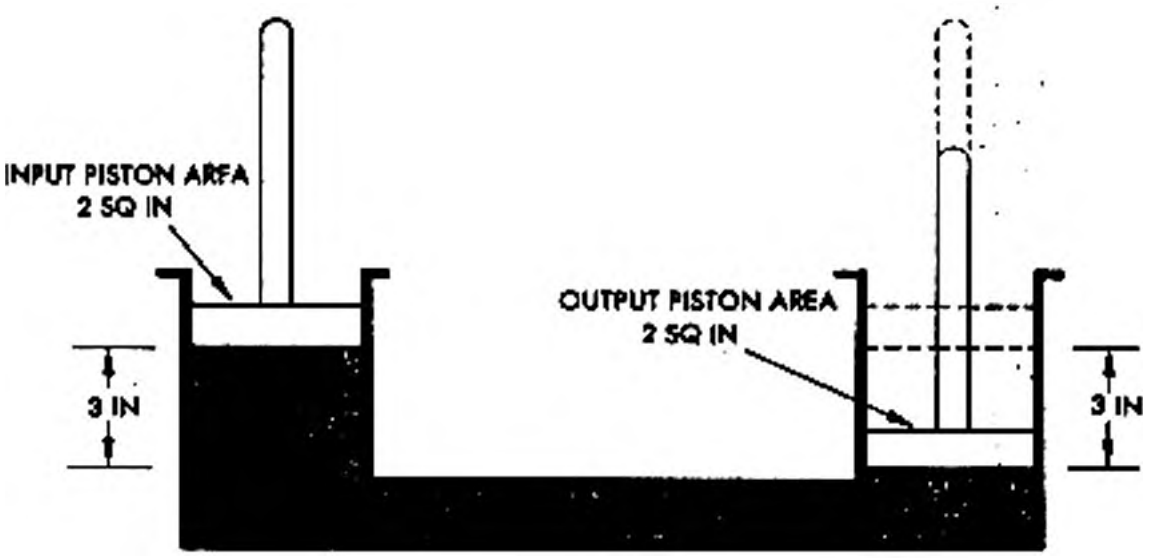


Figure 5-105. Hydraulic System with Equal Piston Displacement

have been displaced. In this instance, 6 cubic inches of fluid are in a cylindrical column with an end surface area of 2 square inches and a height of 3 inches. This column of fluid must go someplace, so it displaces the output piston. Since the surface area of the output is 2 square inches, it must be displaced a lateral distance of 3 inches to make room for the 6 cubic inches of fluid.

Now consider figure 5-106 in which the surface area of the input piston is 2 square inches and the surface area of the output piston is 3 square inches. If the force applied to the input piston causes the piston to be displaced a distance of 3 inches, 6 cubic inches of fluid will be forced into the output cylinder. The output piston, having a surface area of 3 square inches, will move only 2 inches to make room for the 6 cubic inches of fluid. Neglecting frictional losses within a hydraulic energy transfer system, the input force times the distance through which it moves equals the output force times the distance through which the force acts. This force-distance relationship is an important factor in the design of hydraulic actuating units.

Now consider the application of pressure transmission in fluids to the actuation of attitude control devices in missiles. Depending on the work to be performed, the parts of a hydraulic actuating system will vary in arrangement and physical dimensions. The illustration of a simple hydraulic system in figure 5-107 shows the relative location of the basic components. The direction of fluid flow through the system is indicated by the arrows.

Referring to the illustration, you can see that power supplied by the electric motor drives the hydraulic pump. The pump unit forces fluid under pressure through the delivery line to the directional control valve. The automatically positioned

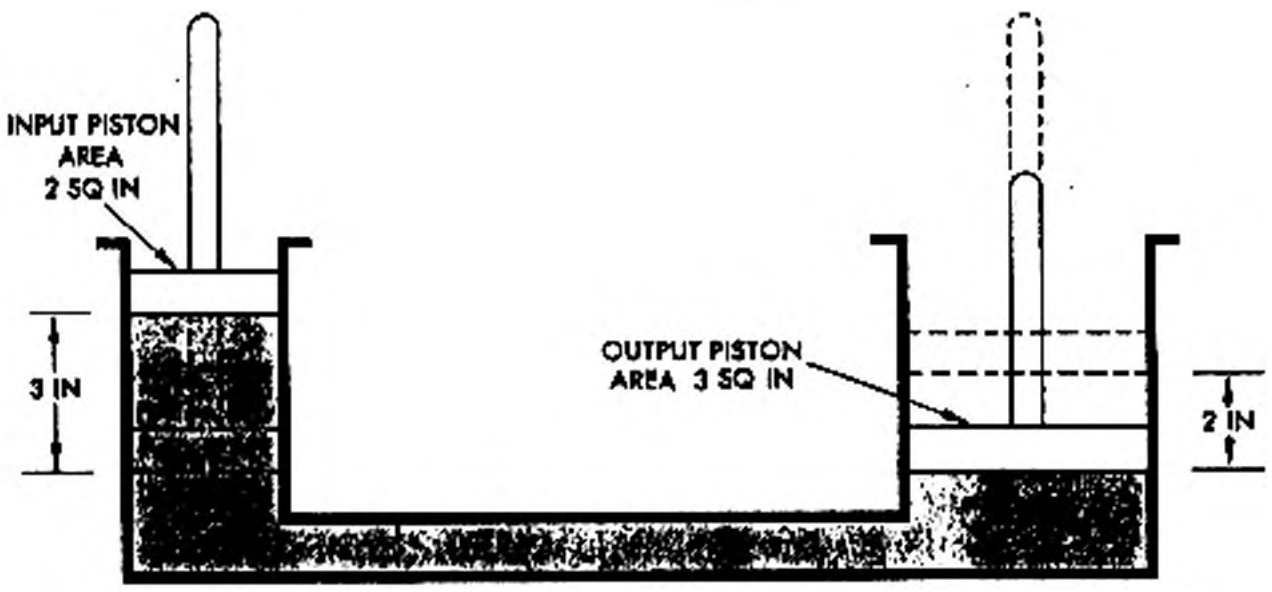


Figure 5-106. Hydraulic System with Proportional Piston Displacement

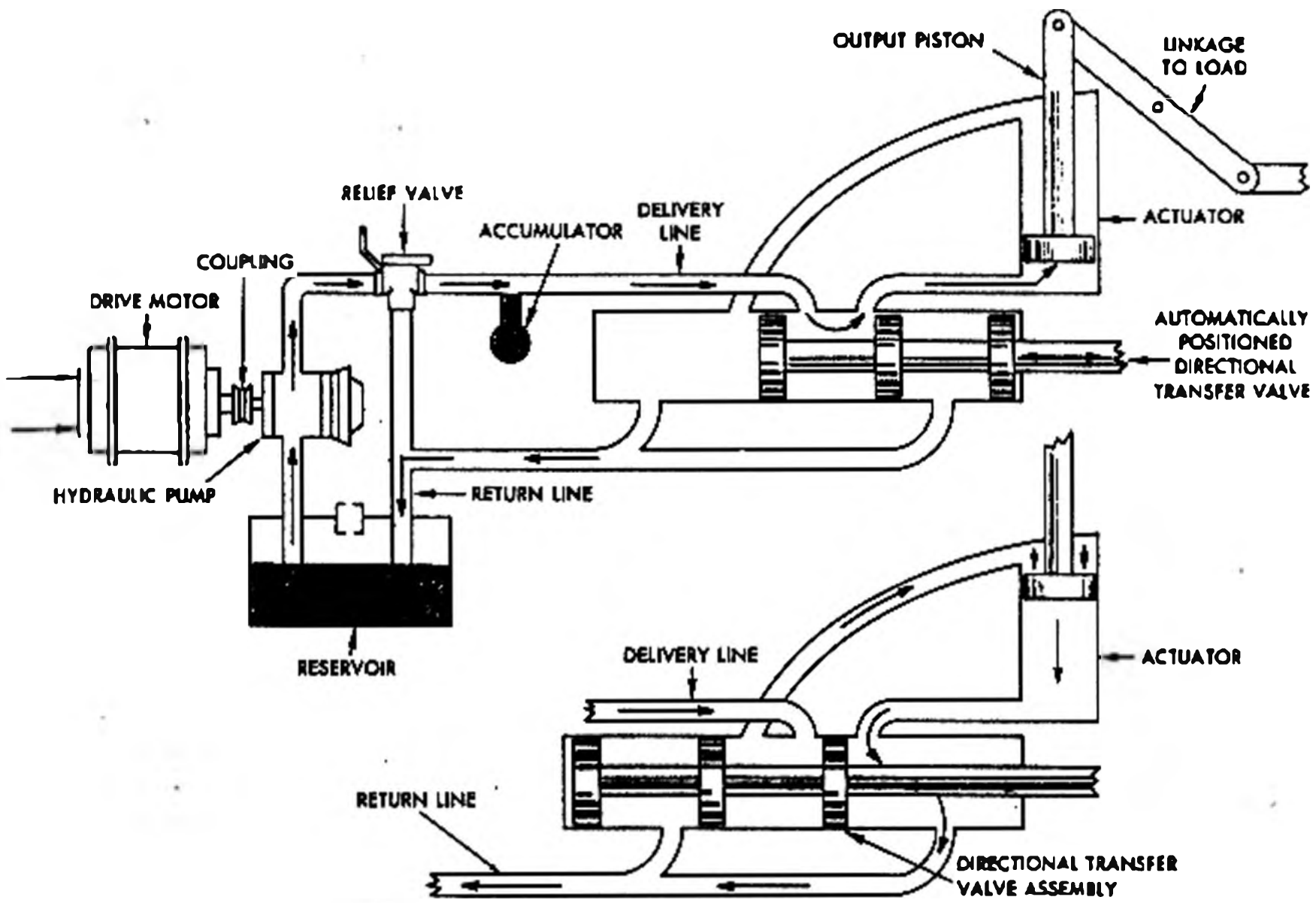


Figure 5-107. Components of Simple Hydraulic System

directional transfer valve determines the direction of fluid flow into the piston-type actuating unit. By means of the actuating unit arrangement, the motion of the piston and the force acting on the piston are transmitted to the load (control device) by the mechanical linkage. The following paragraphs cover the basic components of this hydraulic system.

Construction and Operation of Hydraulic Pumps

The primary energizing unit of the hydraulic system is the force pump. This pump can be driven by an electric motor or by some other energy source within the missile. Two types of pumps frequently used are the geared-type and the reciprocating or piston-type.

The geared-type pump consists of two tightly meshed gears which revolve in a housing as illustrated in figure 5-108. The clearance between the teeth of the gears and the housing is small. For operation, the intake port is connected to the fluid

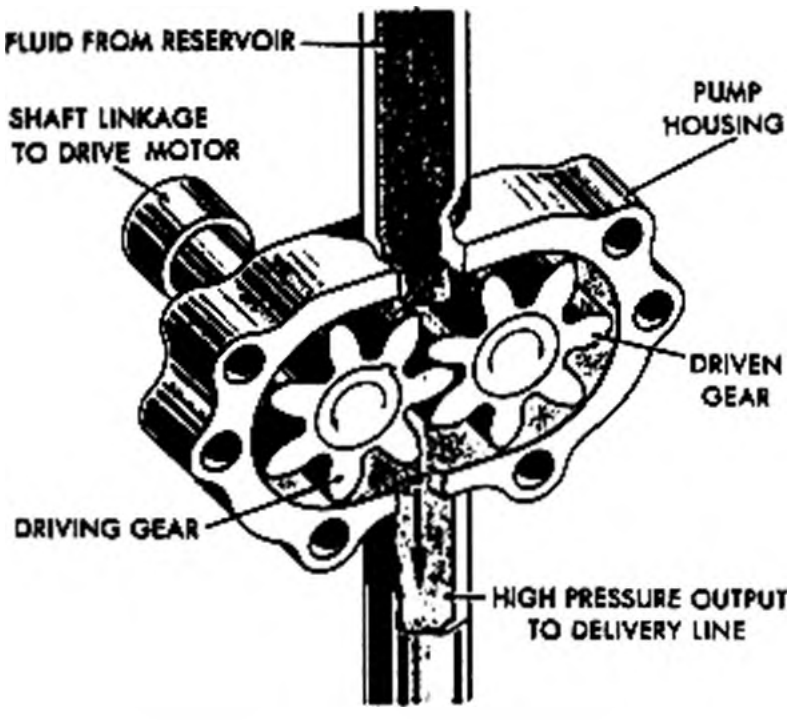


Figure 5-108. Gear-Type Hydraulic Pump

reservoir, and the output port is connected to the high-pressure delivery line of the hydraulic system. As the gear teeth pass the edge of the intake port, fluid is trapped between the teeth and the housing. This fluid is carried around the housing to the output port. As the gear teeth mesh in front of the output port, the fluid between the teeth is forced into the high-pressure delivery line. The hydraulic actuator and the load to which it is attached are displaced by the fluid being pumped through the system.

The second type of force pump that can be used in a hydraulic energy transfer system is the reciprocating or piston-type pump. Pumps of this type depend on the back-and-forth motion of moving parts, working in conjunction with suitable valves to force the hydraulic fluid into the high-pressure line. One advantage of the piston-type pump is its ability to develop higher pressures (relatively speaking) than other types of pumps. Figure 5-109 shows a simple double-acting pump which produces a continuous flow of fluid. A greater quantity of flow and less pulsation of the flow can be achieved by increasing the number of pistons in a piston-type pump.

Note in the diagram that the piston can move in both directions (double-acting). As the piston moves from right to left, a partial vacuum tends to develop in the right-hand section of the chamber. Immediately, fluid under atmospheric pressure in the reservoir forces valve 1 open and enters the chamber. At the same time, fluid on the discharge side of the pump attempting to reenter the chamber closes valve 2. The space left vacant by the piston as it moves from right to left is being charged with fluid from the reservoir. As the piston moves from right to left, it exerts a force on the fluid in the left-hand section of the chamber closing valve 3 and opening valve 4. Fluid passes through valve 4 to the pressure line. When the piston reaches the extreme left position, the chamber section to the right of the piston is completely filled with fluid. On the return stroke, from left to right, valve 4 closes, and valve 3 is forced open by fluid from the reservoir which enters the chamber to fill the space left vacant by the piston. At the same time, valve 1 closes, valve 2 opens, and fluid is forced into the delivery line.

To summarize this action, on the right-to-left piston stroke, fluid enters valve 1 and is discharged through valve 4, while valves 2 and 3 are closed. On the left-to-right piston stroke, fluid enters valve 3 and is discharged through valve 2, while valves 1 and 4 are closed.

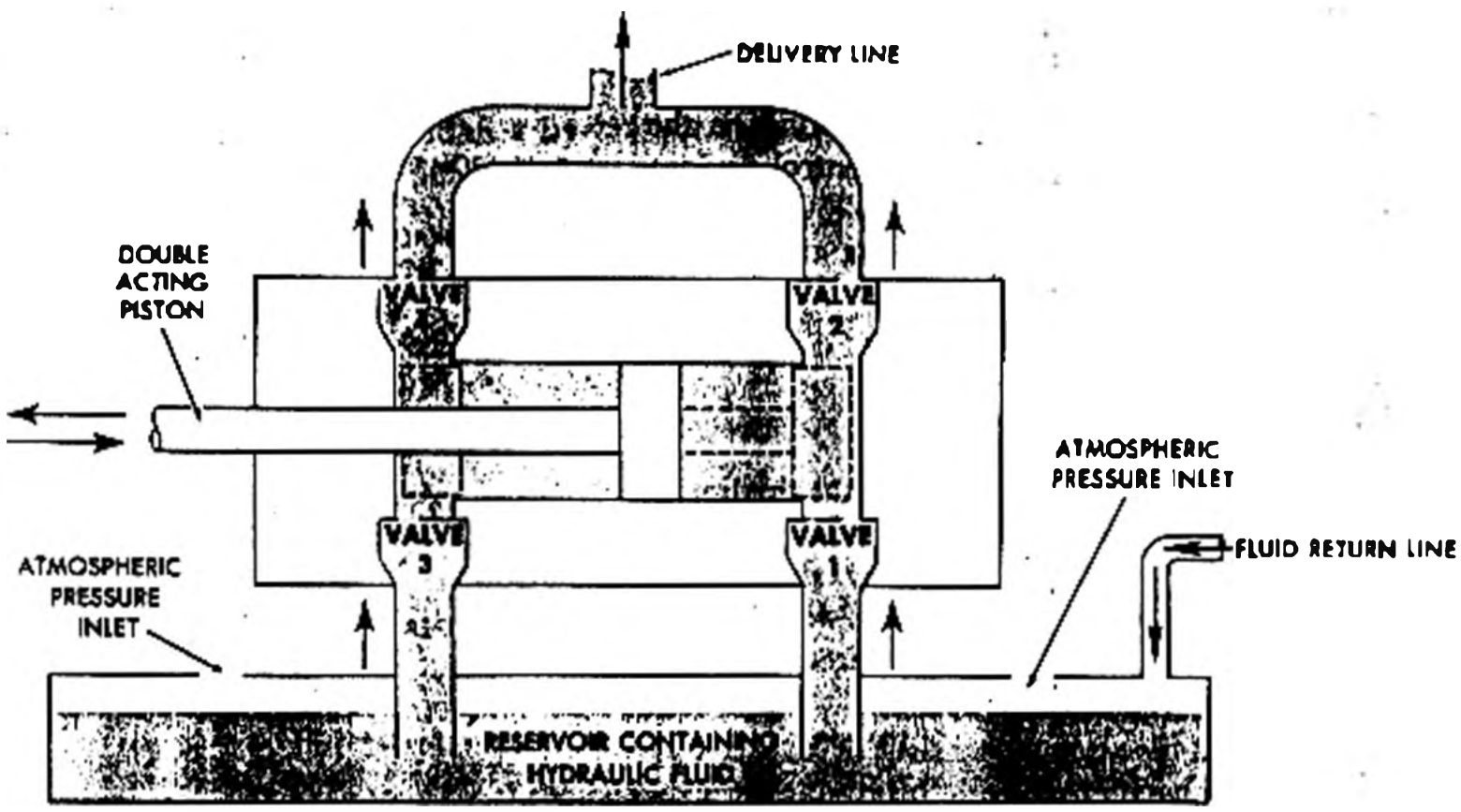


Figure 5-109. Double Acting Piston-Type Hydraulic Pump

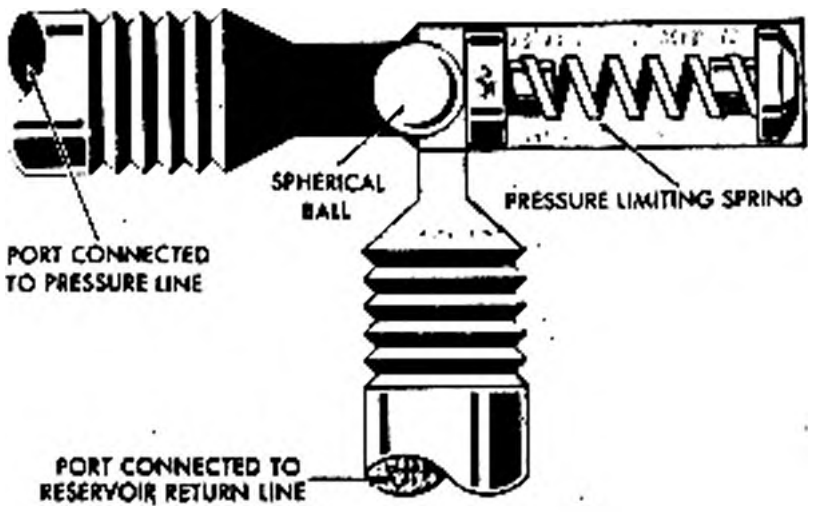


Figure 5-110. Typical Hydraulic Pressure Relief Valve.

Function of a Reservoir in a Hydraulic System

A reservoir is the storage place for the fluid used in a hydraulic system. Fluid flows from the reservoir to the pressurizing pump which forces the fluid throughout the hydraulic system. The reservoir also receives the returning fluid after it has performed the desired work on the hydraulic actuator piston.

Construction and Operation of a Typical Relief Valve

Pressure-relief valves are used to limit the pressure in a hydraulic system to some maximum value and prevent damage to parts of the system. Some hydraulic systems use hydraulic pressure-regulating switches instead of relief valves. Such switches control the power-pump operation in response to pressure changes within the system.

The typical pressure-relief valve shown in figure 5-110 consists of a metallic housing with two ports. One port is connected to the pressure line of the hydraulic system. The other port is connected to the reservoir return line. Notice that the spherical ball is held seated in the restricted section of the pressure line by a spring, which prevents fluid from passing to the reservoir return port under normal operating conditions. When the pressure at the pressure-line port becomes great enough to overcome the force exerted on the ball by the spring, the ball is moved off its seat. This action allows fluid to escape through the reservoir return port and reenter the reservoir. The pressure in the hydraulic system can never go above the value necessary to overcome the force exerted by the spring designed for that particular system.

Construction and Operation of a Hydraulic Accumulator

An accumulator acts as an auxiliary storage place for hydraulic fluid under pressure. In so doing, it tends to dampen out pulsations or pressure surges in the hydraulic system. Pulsating flow in a hydraulic system would cause vibration of components and unsteady operation of the control devices to which the actuators are linked.

Accumulators may be of either the floating piston type or the diaphragm type, both of which are shown in figures 5-111 and 5-112.

The floating piston-type accumulator consists of a metallic cylinder separated into a hydraulic fluid chamber and an air chamber by the floating piston. The diaphragm type consists of two hollow hemispherical pieces of metal separated into

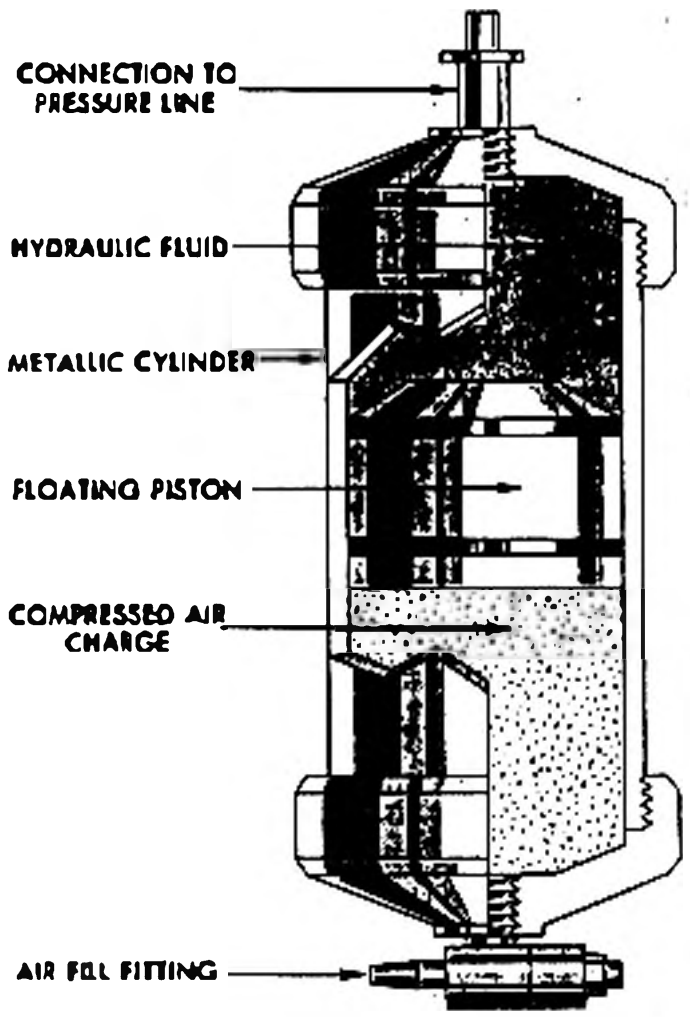


Figure 5-111. Floating Piston-Type Hydraulic Accumulator

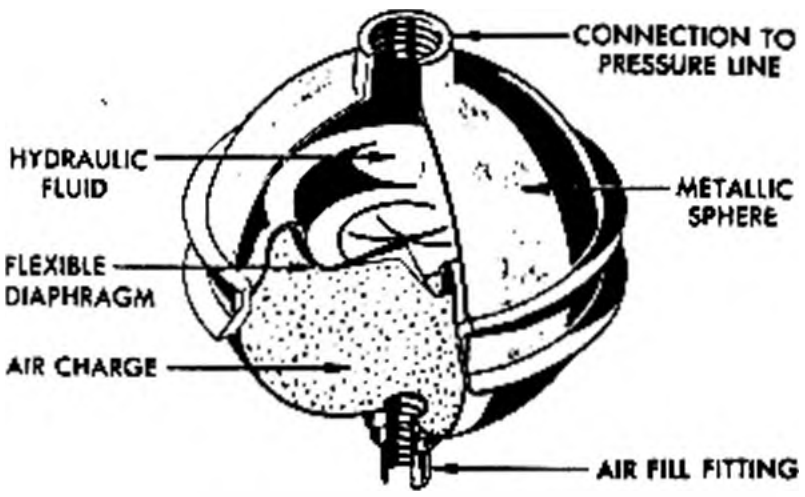


Figure 5-112. Diaphragm-Type Hydraulic Accumulator

a hydraulic fluid chamber and an air chamber by a flexible diaphragm. In both types, the air chamber is charged with compressed air to a pressure corresponding to the line pressure desired in the hydraulic system which exerts a force on the piston, or the diaphragm. If the line pressure builds up higher than the air pressure in the accumulator, fluid is forced into the hydraulic fluid compartment. This fluid pushes the piston down, or depresses the diaphragm, further compressing the air in the air chamber. During periods of peak load or power-pump lag, the compressed air forces fluid from the accumulator back into the hydraulic system. By building up pressure in an accumulator, variations in hydraulic-system line pressure are smoothed out.

Function and Operation of a Hydraulic Actuator

The purpose of a hydraulic actuator is to transform fluid pressure into the mechanical force necessary to move some type of control device. A basic hydraulic actuator consists essentially of a cylinder, with suitable fluid intake and exhaust ports, fitted with a piston and connecting rod. The actuator shown in figure 5-107 of a simple hydraulic system is a double-acting piston type. In the actuator, the pressurized hydraulic fluid can be applied to either side of the actuator piston, thus producing motion in either of two directions. On the upstroke, fluid under pressure enters the cylinder below the piston, forcing the piston up and forcing the fluid above the piston back to the reservoir. The downstroke of the actuator piston results when fluid enters the cylinder above the piston. In this instance, the fluid below the piston returns to the reservoir.

A second method of energy transfer involves the use of pneumatic units.

PNEUMATIC ENERGY TRANSFER SYSTEMS

Energy transfer systems that use air as the medium of energy transfer are referred to as pneumatic systems. Basically, the operation of a pneumatic system is similar to the hydraulic system just discussed. The most prominent difference between the two systems is that the medium of transfer in the pneumatic system is a gas while the medium of transfer in the hydraulic system is a liquid.

In a pneumatic system, air from a pressurized source passes through suitable delivery tubes, valves, and pressure regulators to do work upon some mechanical unit. The pressurized source generally consists of high-pressure air stored in metal tanks. The pressure energy, originally possessed by the air, is transferred from one point to another where it is transformed into mechanical work by a piston, or diaphragm, which is connected to the missile control device.

Unlike a hydraulic system, a pneumatic system does not re-use its transfer medium after it has performed work on the load. For that reason, the air must be stored at a pressure much higher than that necessary for actuating the load to maintain adequate system pressure as the stored air supply diminishes.

The pneumatic system has had very limited usage and there is only limited coverage in this text.

ELECTRICAL ENERGY TRANSFER SYSTEMS

Electrical energy transfer systems are numerous. The assembly of electrical components into systems for transferring energy varies widely, depending on the job the system must perform in a missile. The objective of such a system relative to actuating control devices is to transfer electrical energy from one point to another where it is transformed into mechanical motion of a control device. Generally, motors are used as the actuators in electrical energy transfer systems.

The type of electric motor used as an actuator depends primarily on the size of the load and the speed with which the load must be moved. In general, DC motors develop higher stall torque than AC motors and are more often used for driving heavy loads of the type encountered in high-speed missile control. An AC motor is inherently a constant-speed device and thus not suitable for the requirements of a servomotor where variation in speed of rotation is necessary. This factor also makes the DC motor more applicable than the AC motor as an electric actuator.

The armature of the DC motor is mechanically linked to the load and some means of controlling the speed and direction of rotation must be used. Two ways of accomplishing this control are by controlling the voltage applied to the field coils and controlling the armature voltage. Controlling the field voltage would permit control of armature speed only, upward from a certain minimum speed. To permit the speed of rotation of the armature to be zero at times, the armature method of control must be used. If no current flows through the armature windings, no rotation is produced. When current does flow through the armature windings, the speed and direction of rotation will be determined by the magnitude and phase of the error voltage. Usually, the servomotor armature voltage is controlled by feeding the error voltage into the field of a DC generator, the output of which is fed to the armature of the DC motor.

Figure 5-113 shows the schematic arrangement of the components of a DC servomotor. The DC generator armature is driven at a constant speed of rotation. When the DC generator field is energized by the amplified error signal current, a voltage is induced in the armature coils which are cutting the magnetic field. The output voltage induced in the armature coils is proportional to the speed of rotation of the generator armature and the magnetic field strength. Since the armature is driven at constant speed, the output voltage of the generator varies with the magnetic field. The magnetic field is proportional to the magnitude and polarity of the signal voltage used to excite the field. The variable output of the generator is fed to the armature of the DC servomotor. The field of the DC servomotor is excited by a constant DC source. The direction and speed of rotation of the armature is proportional to the polarity and magnitude of the armature current. The armature of the DC motor is mechanically linked to, and drives, the control device.

Use of a DC motor as an actuation unit controlled by a generator, as shown in figure 5-113, produces an undesirable reaction which must be eliminated if servo control is to be efficient. The problem arises because of the tendency of the magnetic structure of the generator to remain magnetized. It tends to remain magnetized after the magnetizing current has been diminished to zero by correction of the control misalignment that produced that current. This residual magnetic field tends to induce a small output voltage in the rotating armature coils even during the absence of field excitation current. This reaction is extremely undesirable for servo applications, because the DC servomotor will continue to develop torque (rotate) in the absence of an error

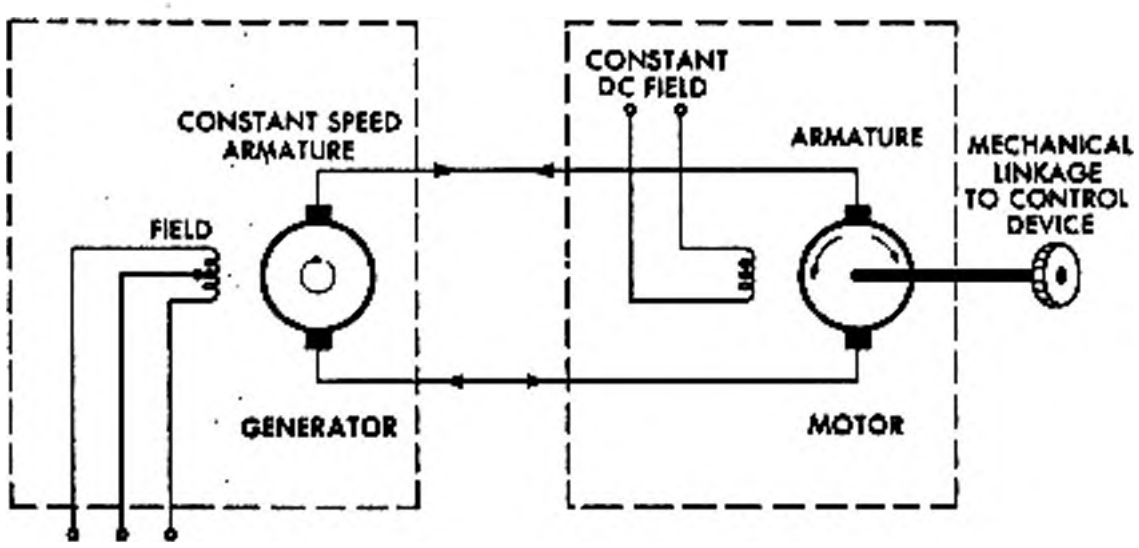


Figure 5-113. Basic Schematic of DC Motor used as Actuator

signal. A reaction of the type just described causes the load to move farther than desired.

This problem can be solved by using additional windings on the generator field structure. If these auxiliary windings are excited by low-level AC, the alternating magnetic field which results causes the average magnetization of the generator field to fall to zero when there is no excitation current flowing in the main field winding. The AC flowing in the auxiliary winding also causes small fluctuations in the main magnetic field. This reaction produces a small AC component in the generator output to the servomotor.

As a solution to this secondary problem, instead of the additional windings for inserting an alternating magnetic field, small permanent magnets are mounted on the generator armature in such a way as to revolve with the armature. This arrangement of permanent magnets provides an alternating magnetizing force for the generator field and permits the DC output of the generator to fall to zero when the main field is not excited. Since the permanent magnets have a fixed position relative to the armature coils, they cannot induce voltage in these coils. The generator output is free from AC components and the servomotor torque is more accurately controlled by the error voltage.

COMBINATION-TYPE ENERGY TRANSFER SYSTEM

In some instances, actuators for control devices employ two different methods of energy transfer to achieve the desired results. For example, an electro hydraulic actuator consists essentially of a hydraulically operated piston whose direction of motion is determined by an electrically positioned selector. Actually, this device could be considered a controller-actuator combination. The electrically positioned valve acts as the controller, and the hydraulically driven piston is the actuator. The compact arrangement of the components of this combination-type actuator justifies an explanation of it in this section.

Figure 5-114 illustrates the arrangement of the components of a typical electro hydraulic actuator unit.

Notice that the solenoid valve controls the input and output of hydraulic fluid to and from the actuating cylinder. The position of this valve is determined by the polarity of the error signal input. Also, the length of time that the valve remains in a certain position depends on the time interval during which the error signal of a certain polarity exists.

As illustrated in figure 5-114, the valve and piston occupy neutral positions. Assume that an error signal of a given polarity exists that will cause the valve to move toward the right as indicated by the dotted lines. This new valve position permits high-pressure fluid to pass through the valve into the actuating piston. The actuating piston is forced toward the left, and at the same time, fluid on the left side of the piston is forced

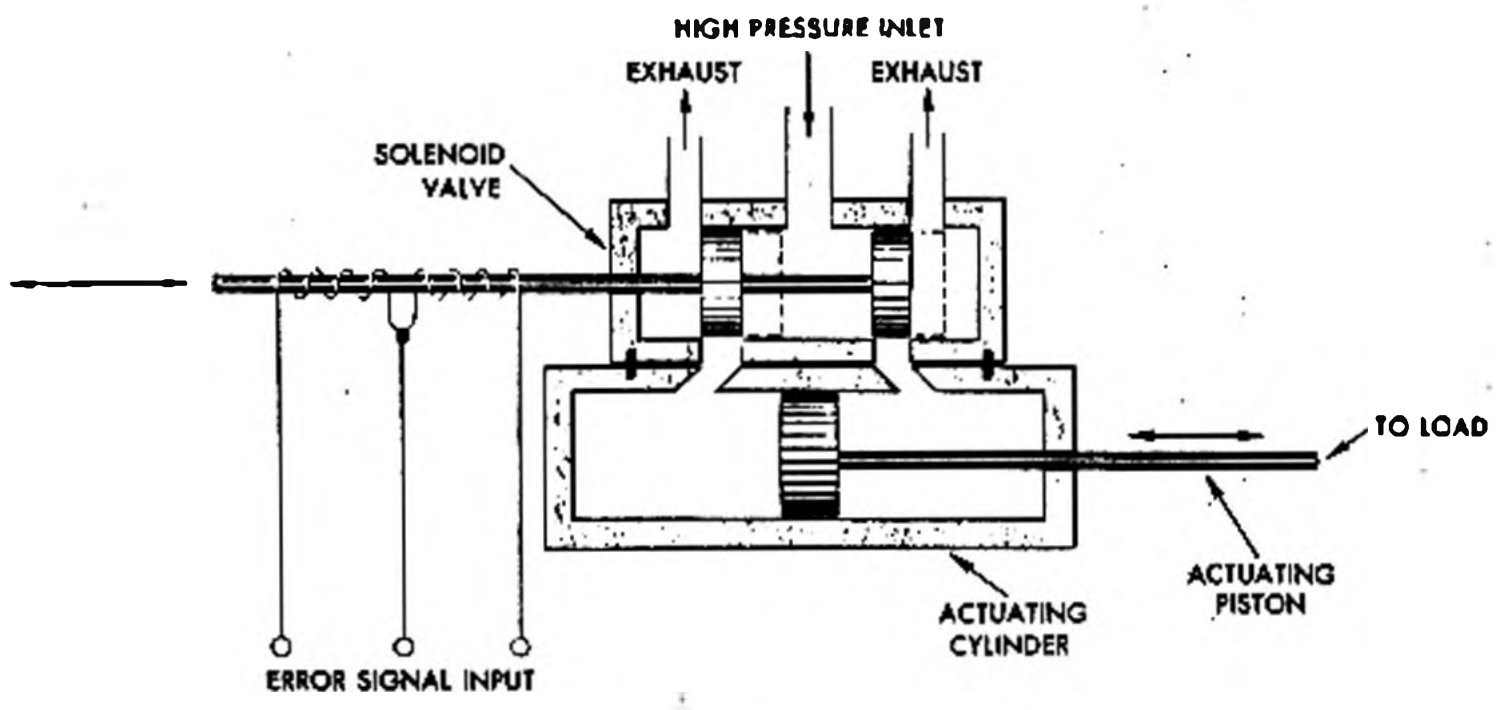


Figure 5-114. Basic Schematic of Electrohydraulic Actuator

out through the valve by way of the left exhaust port. During this action, the exhaust port located on the right is closed. When the attitude error has been corrected, the valve and actuating piston will again be in neutral positions.

When an error signal of opposite polarity exists, the solenoid valve moves toward the left. High-pressure fluid now enters the actuating cylinder on the left side of the piston and forces it toward the right. Fluid forced from the right section of the actuating cylinder passes through the valve and escapes by way of the right exhaust port.

Remember, a hydraulic system is a closed loop. If the rest of the system were shown in the illustration, the exhaust ports would be connected to the hydraulic reservoir. The fluid displaced by movement of the actuator piston would be returned to this reservoir.

USES OF MECHANICAL LINKAGES FOR ENERGY TRANSFER

To transfer energy effectively from the actuator to the control device in an attitude control system, some type of mechanical linkage must be used. With proper design, a mechanical linkage (in addition to transferring the energy) also permits a more advantageous application of a given force to the control device. This advantage is accomplished by changing the point of application as well as by producing an amplification of the torque moment.

To explain these statements further, consider a situation in which a hydraulically operated piston is being used to actuate a control surface designed to rotate about an axis as illustrated in figure 5-115A.

With the piston rod connected to the diametric center of the control-surface shaft as shown, no rotation of the shaft results, regardless of the force exerted by the actuator.

But, suppose the output force of the actuator is applied to a lever arm fixed to the control-surface shaft and the point of application is some certain distance (d) from the axis of rotation. This arrangement is illustrated in figure 5-115B. Now the control surface rotates because the force exerted by the piston is applied at a distance from the axis of rotation and produces a torque moment for a given force. In the diagram of an actuator

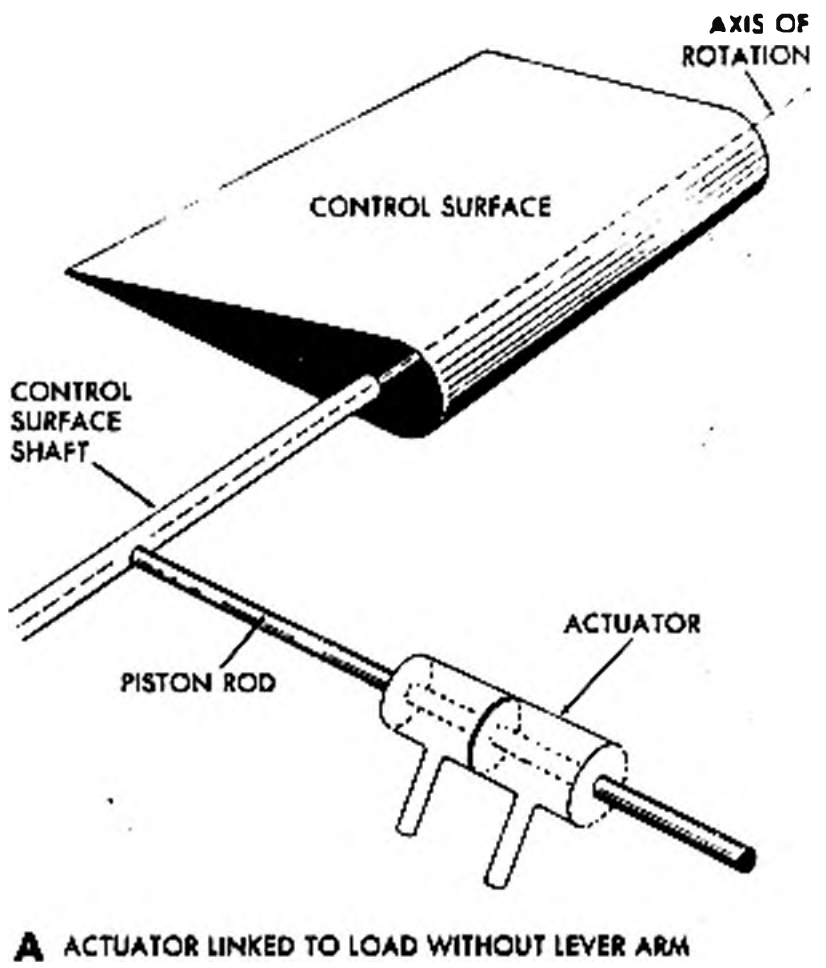


Figure 5-115. Mechanical Linkage

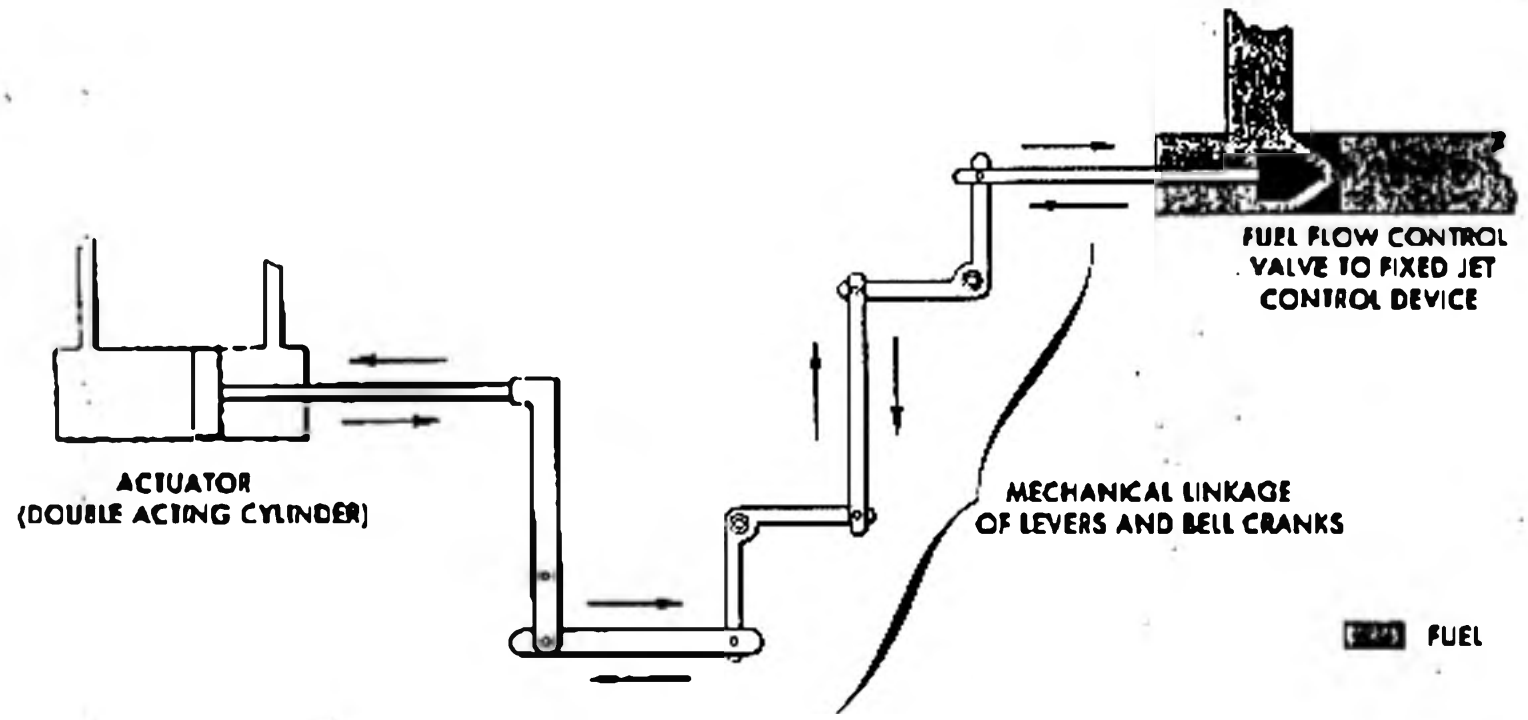


Figure 5-116. Levers and Bell Cranks used as Mechanical Linkage to Actuator and Load

without a lever arm, the force of the piston is applied at zero distance from the axis of rotation; the torque moment is zero.

Mechanical linking devices generally consist of an arrangement of gears, levers and/or cables. The linkage must be arranged so that the phase and magnitude of the actuator output will cause movement of the control device, of proper magnitude and direction, to correct the existing attitude error. Reversals in actuator output produce reversals in the direction of motion of the mechanical linkage components. Such reversal motion causes a reversal in the direction of motion of the load.

To sum up this section on mechanical linkages for energy transfer, study figures 5-116, 5-117, and 5-118. They show various arrangement for mechanically linking the actuator to the load. In each case, imagine that the actuator is moving in some given direction, and follow the resulting motion through the mechanical linkage to the load. Then reverse the direction of the motion of the actuator and again mentally follow the action of the mechanical linkage and direction of displacement of the load.

REVIEW OF ENERGY TRANSFER METHODS

The method of energy transfer used in an actuating unit depends on the other components of the control system. The merits of each method are considered in chapter 7.

The transfer of energy by hydraulic units is based on Pascal's law which states that whenever a pressure is applied to a confined liquid, that pressure is transmitted undiminished in all directions throughout the liquid regardless of the shape

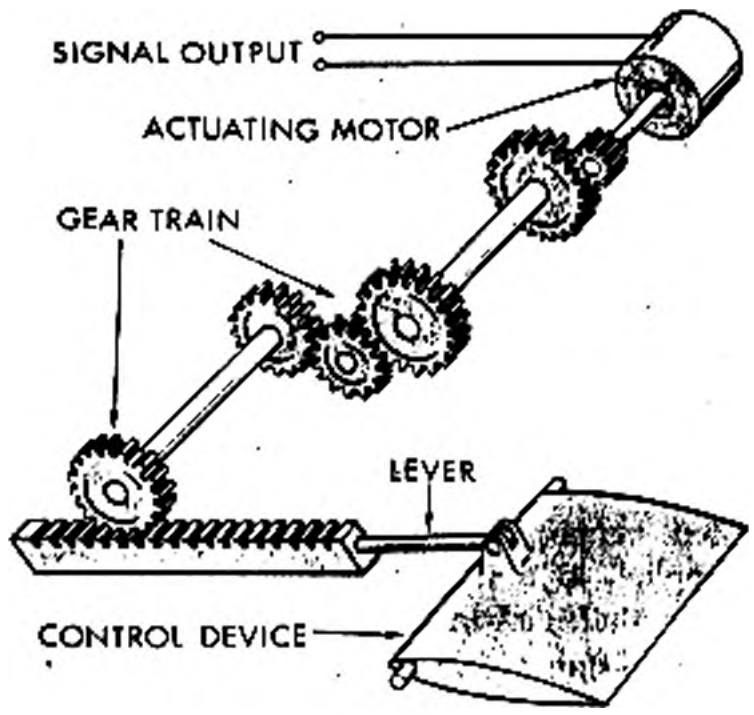


Figure 5-117. Gear Train-Type of Mechanical Linkage

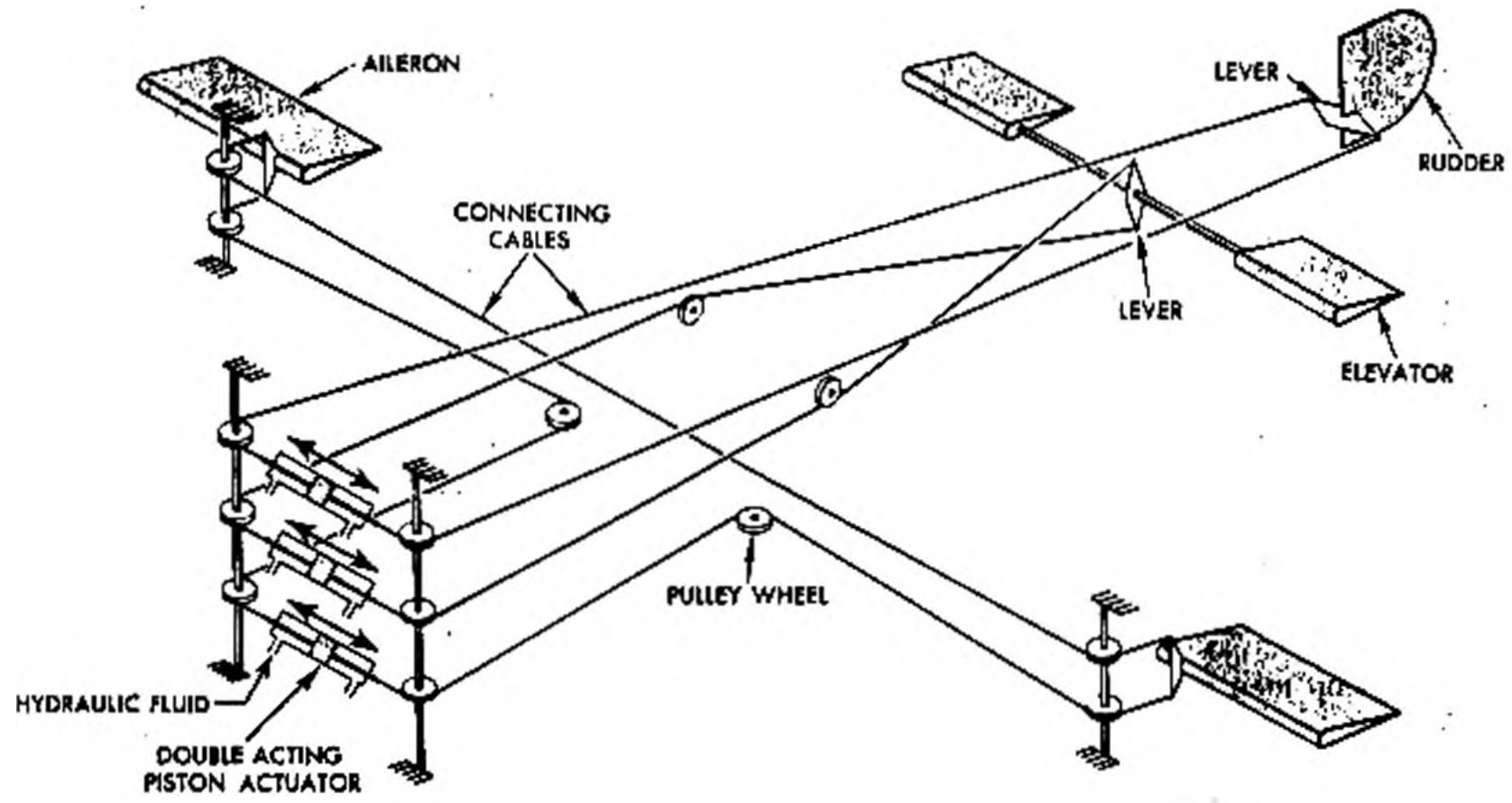


Figure 5-118. Cable linkage from actuators to control surfaces

of the confining container. Basically, a pneumatic system operates in a manner similar to a hydraulic system. A principal difference of the two systems is the medium used for transfer of energy. Liquid is used in hydraulic systems, while air is the medium in pneumatic systems.

The assembly of electrical components in missile electrical energy transfer systems varies widely according to the job the system must perform. These systems generally employ DC motors as actuators. Sometimes, two methods of energy transfer are employed in an actuator unit. An electrohydraulic actuator is such a combination unit. It consists of a hydraulically operated piston whose direction of motion is determined by an electrically positioned selector.

Energy transfer by mechanical linkages in missiles may be done by an arrangement of gears, levers, and/or cables.

SECTION G

Followup Units

A followup unit of a missile control system provides continuous information on control-surface position in relation to the airframe by means of a followup signal. This information is usually called followup since it indicates how the output of the control system is following an error signal.

FOLLOWUP SIGNALS INDICATE OUTPUT OF CONTROLLING EQUIPMENT

A followup signal is an indication of the output of the controlling equipment. The signal is approximately proportional to the surface deflection from streamline. It is an indication of the output of the controlling equipment. Missile deviation is represented when the signal combines with the error signal.

Without a followup signal, there would be nothing but varying air pressure to prevent the control surface from swinging its limit and "hitting the stops" any time an error signal from the sensor existed. The followup enables a surface to deflect an amount which depends on error signal strength. It also provides a signal to return the surface to streamline.

The followup signal combines with the error signal in a manner that opposes the error signal. The error signal is large enough to produce the necessary surface deflection. When the followup becomes as large as the error signal, the surface does not deflect any farther since the resultant of

the two signals is zero. (The resultant is the signal that operates the control surface.) If an error signal is strong, the surface deflection is large before the signals exactly counteract. Since the missile then starts returning to the desired heading, the error signal becomes less. The followup signal is then the larger, so the resulting voltage is reversed. This reversal moves the control surface in the opposite direction, until it is again streamlined. Such action provides smooth and rapid control which cannot be matched by an "on-off" system which affords intermittent control and overcontrolling.

FOLLOWUP LOOPS SUPPLEMENT MAJOR FEEDBACK SIGNALS

A followup loop supplements the major feedback signal of a control system. The followup loop has a generator which creates a feedback signal that is in addition to the major dynamic feedback path. The major path represents information on the reaction of the missile attitude fed back to the sensor. The angular movement of the aircraft completes this major loop. The minor feedback path (followup loop) returns information on the reaction of a control surface rather than the reaction of a missile. This return completes a minor loop which includes less equipment and is of less importance, in principle, than the major loop. Both loops are included in the basic missile control block diagram shown in figure 5-3. The followup loop is shown in figure 5-119.

The additional feedback loop increases the speed at which the missile responds to an error. The fundamental automatic control principle of using the reaction of the controlled item as a basis for further and continuous correction has already been mentioned. In a missile system the controlled item is missile attitude. If this reaction were the only guide to further correction, it would occur too late to provide fine control. The reaction would be late because of aerodynamic lag. When the control surface is deflected, an additional period of time elapses before the missile returns to the desired attitude. The deflection of the control surface determines the force tending to move the airframe, not the airframe position directly.

Without a followup loop, a missile would tend to oscillate about the desired heading. Consider a missile without followup turning to the right as the result of a gust of wind. The sensor detects the error and deflects the rudder to the extreme left to get the missile back on course. A left rudder position is maintained as long as this error is detected. After a short time, the missile returns to the desired heading. The sensor detects this fact by means of the major feedback loop, but the missile is now well on its way toward a left deviation which requires a right rudder. The operation repeats itself, creating cyclic movement dependent on the missile aerodynamics and the control-system characteristics.

The followup loop sends back earlier information since it does not include the lag of missile reaction. The signal is not without delay, however, since the control equipment is not instantaneous.

This system lag is shown in figure 5-120 which illustrates the combined error and followup signals. Consider any instant during the period that the error signal is increasing. At this instant the followup signal is less than the error signal. This

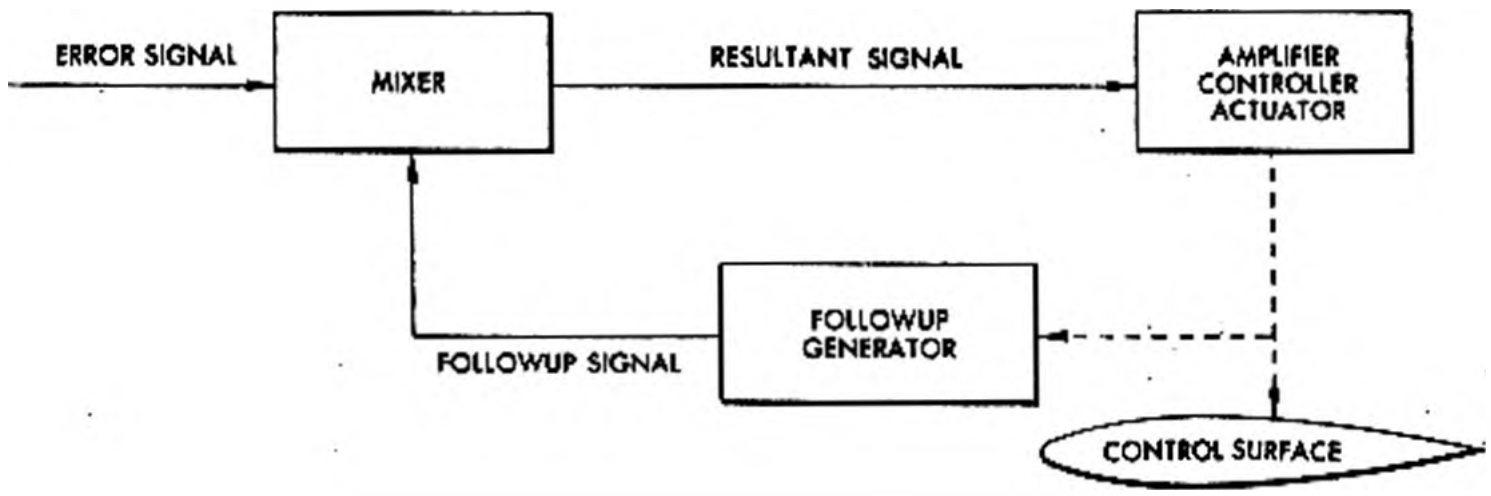


Figure 5-119. Followup Loop of Missile Control System

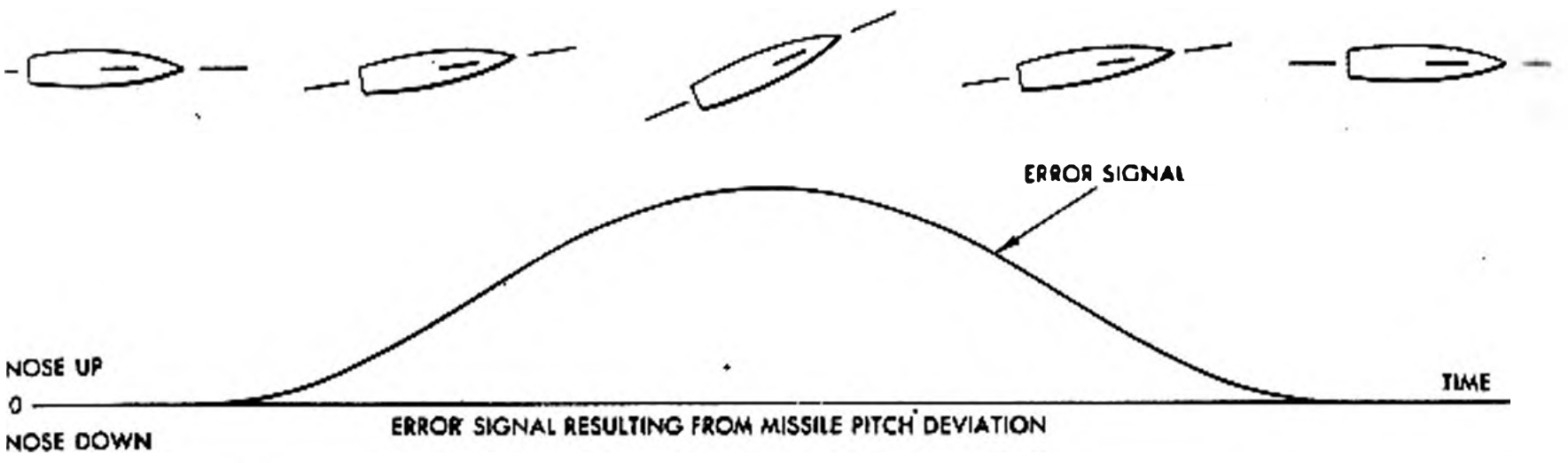


Figure 5-120. Relationships of Followup to Error Signal

produces a resultant signal. A certain lag time is required before the followup signal reaches the error signal amplitude which existed at the instant considered. This time is shown as "lag" in figure 5-120.

The figure shows that the followup signal is added to the error signal with such polarity as to oppose the error signal. This type of feedback is therefore degenerative. It also can be called inverse or negative feedback. It is comparable to the inverse feedback found in many electronic amplifiers.

The signals are combined to produce a resultant which is the output from the mixer circuit. Mixers are described in the section on computer units. The resultant is the corrector signal which operates the control surface.

FOLLOWUP GENERATORS DETECT SYSTEM OUTPUT

The devices which detect system output are called followup generators or control-surface pickoffs. They are position-detecting devices such as selsyn generators or resistance pickoffs.

Followup generators are mechanically connected to the actuator, control surface, or the linkages between them. A followup generator can be connected to jet vanes or movable steering jets if they are used to provide missile correction. Every

generator is nulled at the neutral position of the control surface.

Selsyn generators and resistance pickoffs are similar to the pickoffs used with gyros to detect missile attitude. These devices are explained in the section on sensor units.

Mechanical Followup

Followup information can be transmitted mechanically rather than electrically. In this type of system the control surface position is indicated by a force which is carried back to the controller by means of springs and levers.

This type of followup is not used extensively because of inherent lag in the system caused by play in the mechanical linkage. The electrical feedback circuit is preferred because of the instant response time of electrical signals.

Feedback Loops Complete Closed Cycle

The feedback loops complete the closed cycle of the missile flight control system.

Sensor units detect deviation from the desired flight path of the missile. Computer units change the sensor signals to represent the type of information needed. A control system would be unable to make desired changes in flight conditions without reference units. The fundamental purpose of amplifier units is to provide an enlarged reproduction of the input signal. Control units are necessary to control the amount and direction of movement of actuators. Actuator units transform and transfer energy at the load end of the control system. Followup units provide continuous information on control surface position in relation to the missile airframe.

The next chapter presents components used to make up a missile guidance system.

CHAPTER 6

Guidance Systems Components

We cannot proceed very far with missile development before realizing that we must give attention to the characteristics of the guidance system.

Guidance is the process of intelligent maneuvering that enables the missile to reach its specified destination. The general term "guidance" includes the entire scheme, both attitude control and path control; that is, guidance requires the sensing devices, computers, and servosystems necessary to convert the guidance commands into missile response. Guidance may be separated into phases of the flight path: initial (or launch), mid-course, and terminal.

The term "guidance system" is usually more restrictive. It is applied to a system that measures and evaluates flight information, correlates this information with target data, converts the results into parameters necessary to achieve the desired flight path, and furnishes this data in the form of commands to the missile control system.

Any guidance system must employ certain components. There must be an adequate number of stages in the system to accomplish the guidance function. The various stages, shown in figure 6-1, are taken up in the following sections.

The first block of the diagram, the sensor unit block, refers to the devices used to detect various forms of energy, such as electromagnetic, light, heat, or sound waves. The devices then translate the energy into a form usable within a missile. Sensors not only include the devices used to detect the energy radiations, but they also include the means by which the radiations are relayed to the other circuit elements. In addition, the sensor unit includes accelerometers used to sense changes in motion.

From the sensor unit, the information signal is passed to the computer unit, the next block in the diagram. Here the signal is "interpreted" so it can be applied to circuits that will actuate the control surfaces. Here are decoders, integrators, differentiators, signal mixing devices, and counting devices, to name only a few. This unit can be considered as the "brains" of the missile guidance system. The computer unit makes calculations and comparisons of external and internal information at microsecond intervals to keep the missile on course.

Also feeding into the computer unit is the reference unit. To obtain a set of guidance conditions,

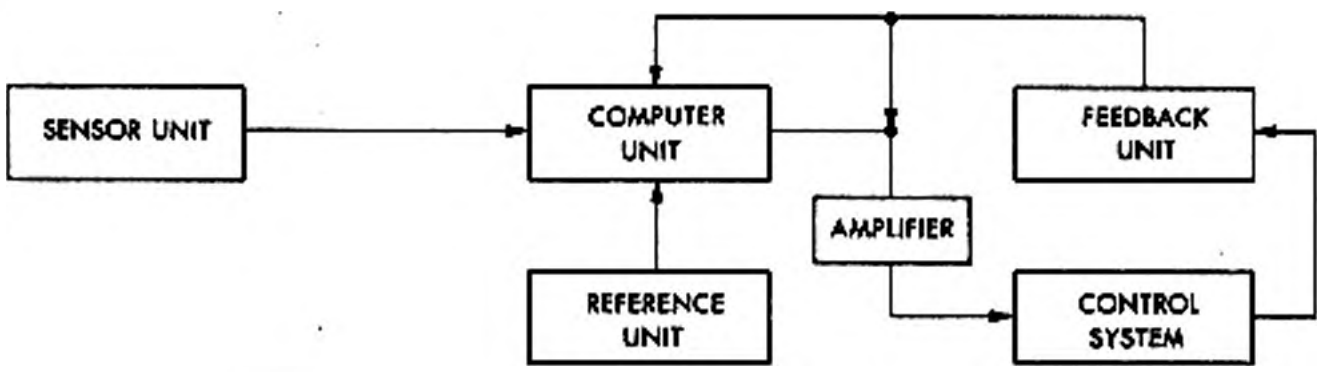


Figure 6-1. Stages of Missile Guidance System

two sources of information often must be compared. This information may come from an outside source or it may be some form of recorded information that was put into the missile prior to launch. The function of the reference unit is to establish the conditions and refer the signals obtained from them to the other circuitry for computation, amplification, and power conversion to the control system.

The guidance signals, after proper recognition, must be amplified sufficiently to become usable in the remaining units of the missile. The amplifying block contains circuit elements that give "energy" to the system. These amplifying devices may react to voltage, current, frequency, or combinations of these, and then increase their operational effect.

The "muscles" of the guidance system are contained in the control system. Relays, solenoids, servomotors, and other actuating units go to work, putting into operation the result of numerous calculations, comparisons, and amplifications of the initial guidance request made to the system.

The control system is sometimes considered as a system in itself since its operation is separate from the guidance system. This is true in the sense that it may control the thrust unit and cause movement of jet vanes, swiveling engines, aerodynamic surfaces, or the like, to maintain attitude stability and to correct deflections caused by wind gusts or other disturbances. One of the functions of the control system is to respond to commands from the guidance system. So it is an integral component of the guidance system.

The function of the feedback unit depends upon the type of guidance system. It can be thought of as the part of the system that helps to distinguish signals that are right from signals that are wrong, and to smooth out the actions of the system. It constantly observes the guidance signals and attempts to increase or decrease their effects to obtain desired results.

We should understand that the components of the block diagram not only relate to the functions that take place within the missile itself but also to various outside factors as well. With this in mind, you will later be able to see how the reference unit may refer to a ground base unit that is setting up the reference. The computer section may or may not be contained within the missile, or the section could be a combination of both outside equipment and missile-contained equipment that is tied into the operation of a missile.

SECTION A
Sensor Units

This section presents the first block of the guidance system, the *sensor unit*, which detects the various forms of energy or phenomena.

Guidance can be performed by any one, or combination, of several different methods. Methods, as used here, refer to the principal overall guidance system. Some of the systems are presented in relation to the sensing element used in conjunction with them. The general methods of guidance are listed below:

- Preset.
- Command.
- Beam rider (direction along a beam).
- Hyperbolic (direction by navigational network).
- Celestial navigation.
- Terrestrial reference (including earth phenomena).
- Homing.
- Intertial.

METHODS USED IN MISSILE GUIDANCE

To understand and appreciate the problems that may arise in choosing a specific guidance system, we should first have an overall picture of the guidance methods listed above.

In the *preset method* of guidance, the control equipment is contained wholly within the missile. All adjustments to course, altitude, speed, and dump-point must be made prior to launch time. Having a missile fly a predetermined path in such a manner has both advantages and disadvantages. A major advantage of the preset method is that countermeasures against such a missile are quite limited. A major disadvantage is that after missile launch, no correction can be made for factors not previously anticipated.

In the *command method* of guidance, the intelligence for missile guidance comes from outside the missile. The missile contains a receiver that receives directions from a ground station or mother

aircraft and executes these commands through the control system and the control surfaces. Usually, in the command system, several channels are operated from the control point by modulating the transmitted signal at various frequencies. Changes in altitude, direction, speed, or any other factor may be made as desired by the controlling operator.

Sometimes a television system is used to give additional accuracy to this method, especially when the mother aircraft is endangered by accompanying the missile over hostile terrain. The controller then has the same viewpoint as if he were physically in the missile, guiding it to the target. The use of a television repeat-back system is limited to target areas that are not overcast. A disadvantage to the command system is that it is subject to enemy jamming.

When using the *beam-rider method*, the missile contains equipment which enables it to follow an electronic beam. In this method, the missile is launched into the beam. As the beam is directed toward the target, the missile tends to keep itself centered on the beam until it hits the target. This beam should not be confused with the radar beam that follows (tracks) the target. The two beams are illustrated in figure 6-2.

The *hyperbolic method* can be used as a long- or short-range navigational network. Such a system consists of master and slave stations putting out low-frequency pulses at constant intervals. The slave station is triggered by the master station and transmits its pulse a few microseconds after the master pulse. By using automatic computers contained in the missile, the missile establishes position by an imaginary line of position set up by the master and slave stations.

The *celestial navigation method* is highly complex. It consists of a mechanism that takes celestial fixes and, through electronic means, keeps the missile on course. This method is used primarily for long-range guidance. Its general operation is somewhat as follows:

Star-tracking telescopes take a fix on some predetermined star, and the information obtained is fed to an electronic computing device that determines the position of the missile relative to the earth. Necessary controls are activated by the error voltages developed if the missile does not coincide with predetermined values set into the missile. This system would only be used on long-range missiles since it is necessary for the missile to fly at high altitudes above the weather where the stars are constantly visible. This method is illustrated in figure 6-3.

In the method involving *terrestrial reference*, missile position is determined through the use of charts, and some characteristic property or phenomena of the earth. A system of map matching is commonly used. A comparison of maps that

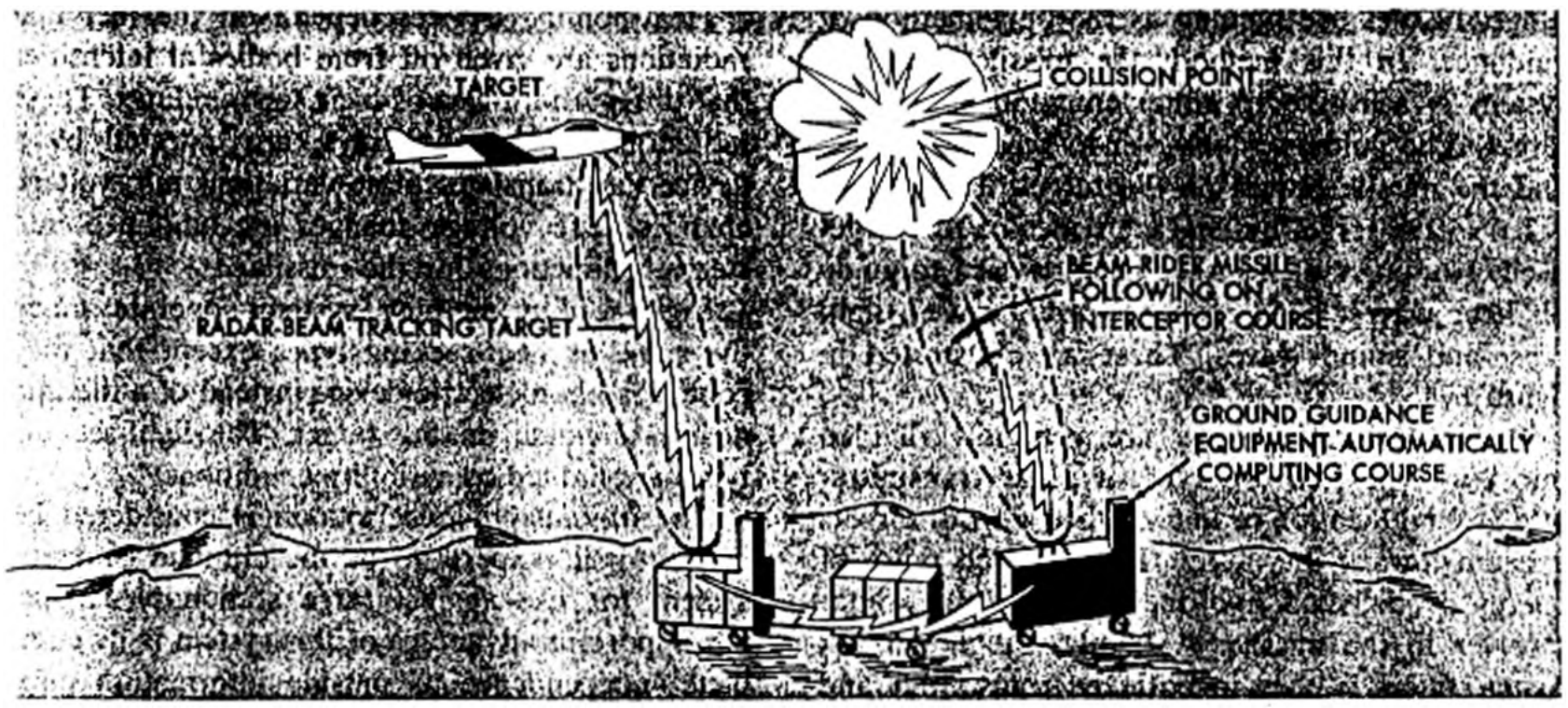


Figure 6-2. Beam-Rider Method of Missile Guidance

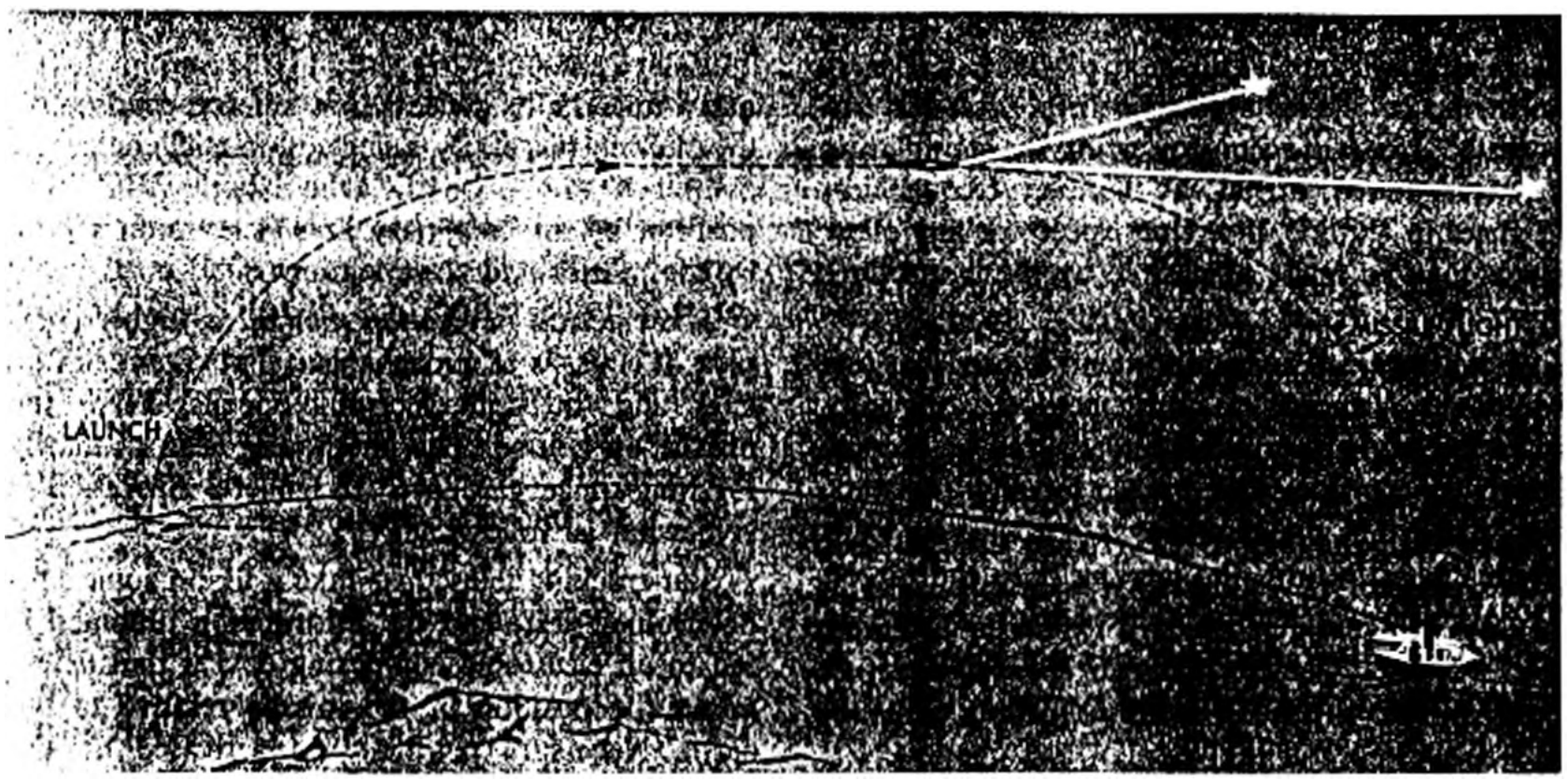


Figure 6-3. Celestial Navigation for Long-Range Missiles

were previously obtained by reconnaissance missiles to the target area is made by electronics equipment contained in the missile.

Systems are also being devised which guide missiles by using the earth's gravitational, magnetic, and electrical fields. Much cosmic research work is being done to determine the effect that these cosmic rays have on the earth's magnetic and electrical fields.

A *homing system* may be used during the terminal phase of flight on most missiles. Many short-range missiles use homing as the only method of guidance. In this method, the missile guidance system is sensitive to some emissions from the target. The guidance locks on to these emissions and the missile homes on the target. The missile is led to the target by characteristics of the target itself. Homing systems have been devised to guide on heat waves, radar echoes, radio waves, light rays, and sound waves. These are illustrated in figure 6-4.

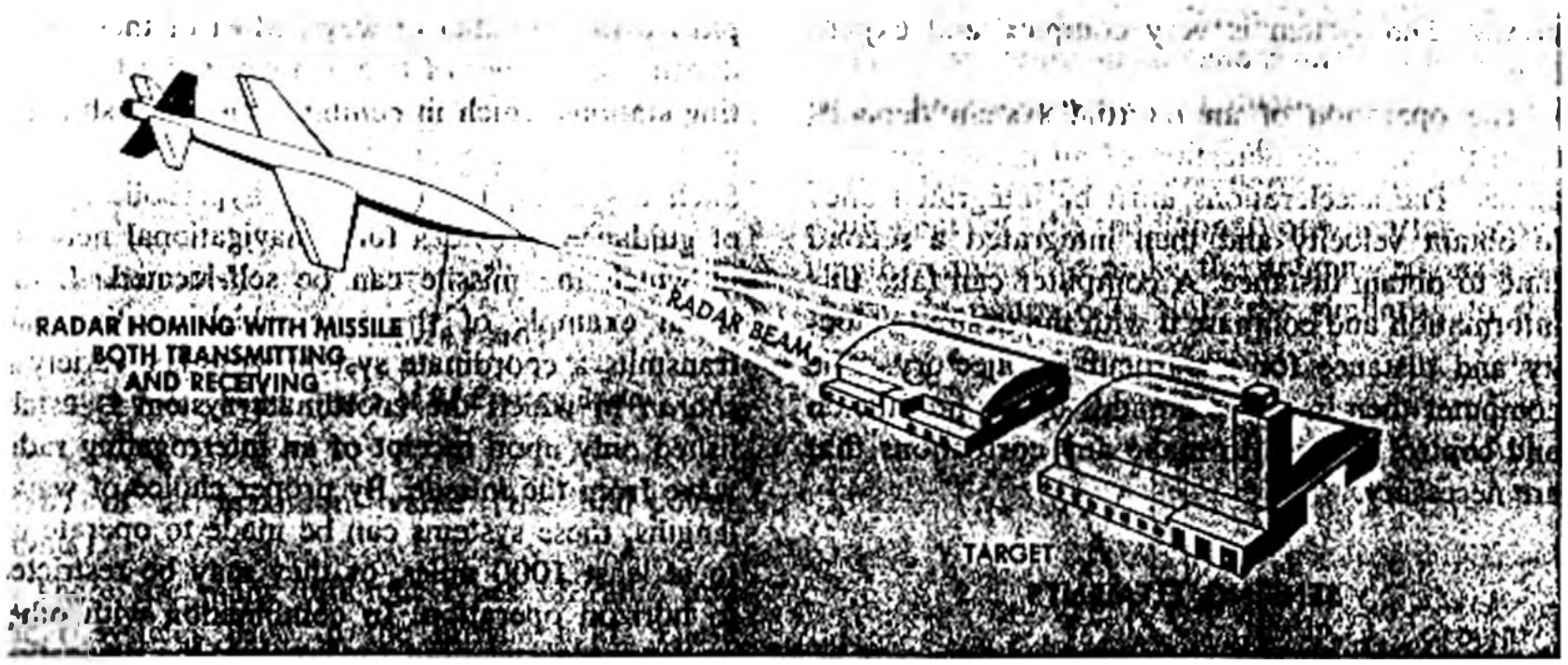
Radar homing systems are generally divided into two types. In one type, a radar transmitter at a ground location or in a mother aircraft directs a beam toward the target. The reflected pulse of energy is detected and measured by the radar receiver in the missile. A highly directional antenna system is used in both the transmitter and the receiver installation. The main disadvantage is that the transmitting equipment needs to be within line-of-sight distance from the target.

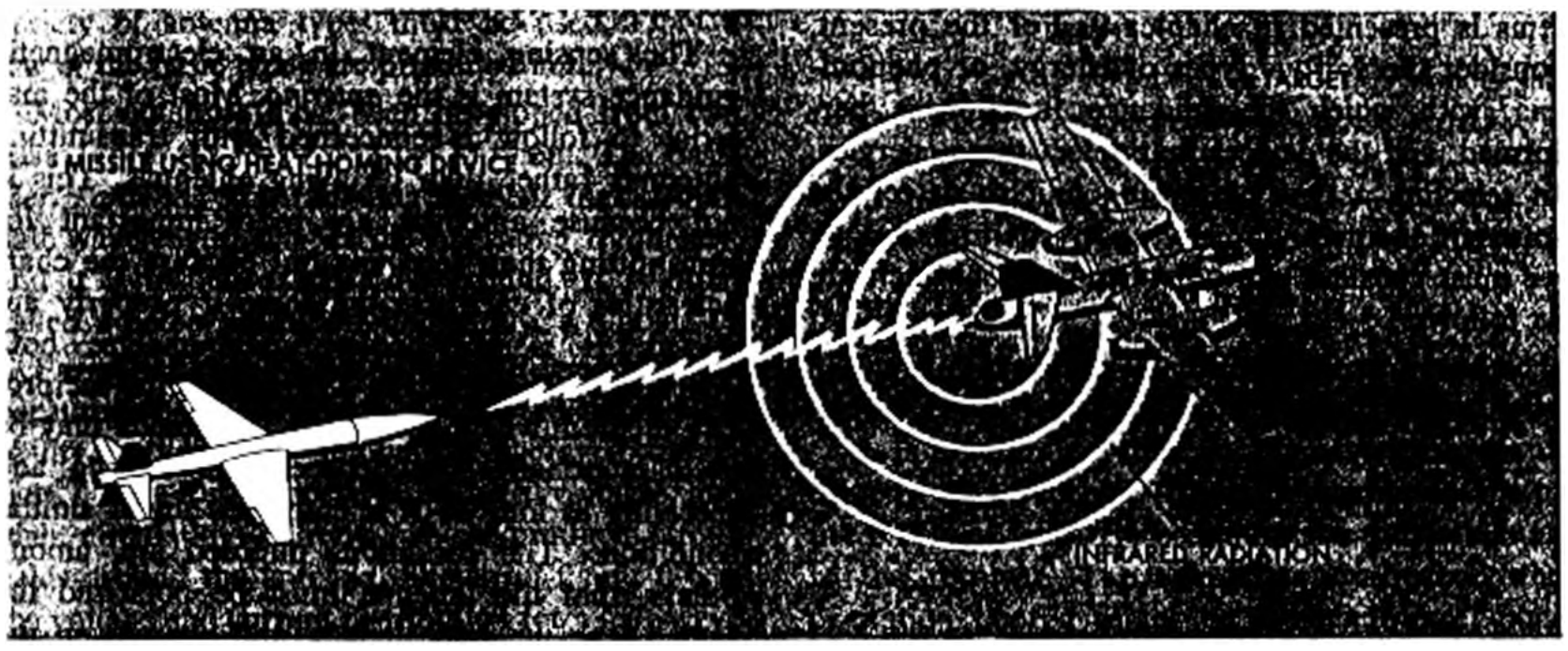
In the other type of radar homing, the missile itself transmits and receives the radar pulses as illustrated in figure 6-4. Here again, the target is "illuminated" by the radar transmitter, giving the missile a target that stands out well from the target background. This system of homing is well adapted to both short- and long-range missiles. Radar homing systems are subject to enemy jamming.

Heat-homing devices depend on the fact that radiations are given off from bodies at intensities that differ from those of their surroundings. These radiations, or infrared rays, are detected by a device that translates these variations into voltage changes. The voltage changes are amplified into steering directions for the missile.

Light-seeking homing systems operate much the same as heat-seeking types except that they contain a photoelectric device instead of a thermal device to home on the target. The light-seeking types are limited by visibility conditions.

The *inertial guidance system* is considered by many missile experts to be the ultimate guidance system for use against large stationary targets. An important advantage of this system is its complete independence of outside information once the missile is launched. Because of this, the system





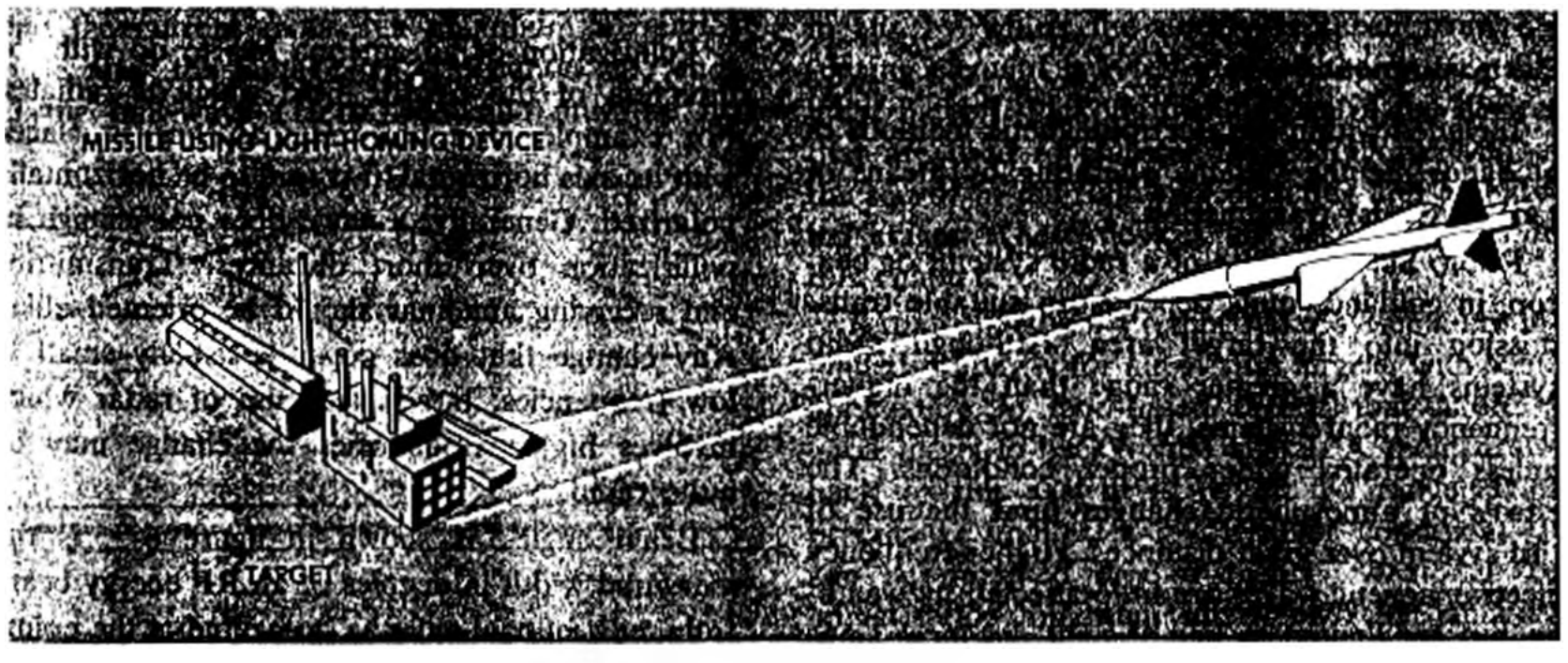


Figure 6-4. Homing Systems

can neither be detected nor jammed by external means. The system is very complex and expensive.

The operation of an intertial system depends upon the accurate detection of all missile accelerations. The accelerations must be integrated once to obtain velocity and then integrated a second time to obtain distance. A computer can take this information and compare it with the desired velocity and distance for a particular trajectory. The computer then issues commands to the propulsion and control systems to make any corrections that are necessary.

SENSING ELEMENTS

A more detailed operation of the guidance systems is presented in chapter 9. For the present, consider the various types of energy, or phenomena, with which the guidance function is performed, and the devices that detect these energies. In selecting a type of sensing unit to detect the forms of energy and guide the missile to its target, we must consider many items, some of which are as follows:

- Maximum range.
- Information required.
- Accuracy.
- Operating conditions.
- Viewing angle of sensor.
- Sensor size and weight.
- Type of target.
- Relative speed of target.

Electromagnetic Sensors

Included within the electromagnetic methods of guidance are the systems operating by means of radio or radar transmission.

RADIO SENSORS. Use of a radio system, or beacon, in guidance units depends on suitable transmission into the target area. Although radio systems have a greater range than the higher-frequency radar systems, they are not dependable unless operated under proper conditions. The effects of atmospheric changes limit the use of this system over great distances. Higher-frequency electromagnetic waves are little affected by weather, but are limited to line-of-sight distances since they are not refracted by the ionosphere.

Guidance by the use of radio can be accomplished in a number of ways. Most of these ways depend on the use of two or more radio transmitting stations which in combination establish some coordinate system on the surface of the earth. Such a system, known as the hyperbolic system of guidance, provides for a navigational network in which the missile can be self-located. *Loran* is an example of the type which continuously transmits a coordinate system. Another variety is *shoran* in which the coordinate system is established only upon receipt of an interrogating radio wave from the missile. By proper choice of wavelengths, these systems can be made to operate up to at least 1000 miles, or they may be restricted to horizon operation. In combination with other guidance devices used for homing, the navigation network is very versatile.

The sensing element in an electromagnetic guidance system is the *antenna*. Most of the material to follow is concerned with transmitting antennas rather than receiving antennas. This is not because antennas are more important for transmitting than receiving but because the characteristics of each are the same if properly analyzed. It is not necessary to analyze both operations. This likeness comes from the reciprocity theorem which states that an antenna may be used interchangeably between transmitting and receiving as long as it is employed under similar conditions. The conditions imposed are, among others, that the antenna must be polarized the same, have the same length, and operate on the same frequency.

Polarization of an antenna has to do with the direction of the electric field which it radiates. For example, if a half-wave antenna is placed with its axis horizontal, it is said to be horizontally polarized. Generally, a wave does not change its polarization over short distances. Transmitting and receiving antennas should be oriented alike. Any change that does occur is usually small at low frequencies, but in the case of radar which involves higher frequencies the change may be quite rapid.

Beam angle, as shown in figure 6-5, is the measured width in degrees of an RF energy beam from a directional antenna, measured at the points away from the antenna where the electric field is

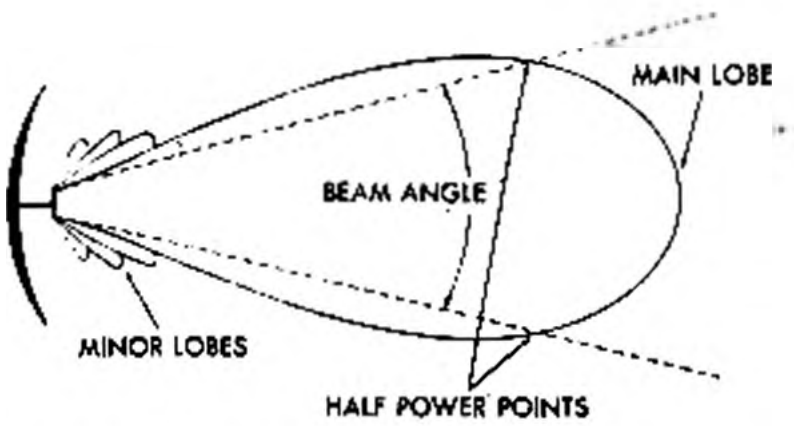


Figure 6-5. Beam Angle of RF Energy Beam

0.707 of its maximum value (the half-power points).

There are many different types of radio and radar systems used in the missile field. These operate over a wide range of frequencies. A large variety of antenna types must be used. Only a few of the basic types are presented here.

Most antennas have been developed from two basic types, the Hertz and the Marconi. The basic *Hertz antenna* is a radiating element that is one-half wavelength long at the operating frequency, and it is insulated from the ground. It is often called a half-wave dipole or, simply, a dipole.

Large radiation occurs in the half-wave element because of its resonant characteristics and its ability to store large amounts of energy in induction fields. Resonance causes high voltages and high circulating currents which produce strong fields around the antenna. Shown in figure 6-6 is the voltage and current distribution for a half-wave antenna. An examination of the waveforms shows that the current reaches a maximum a quarter cycle after the voltage. In application, the ends of the half-wave (Hertz) antenna must be insulated for the high voltages and the center must have low resistance to eliminate high I^2R losses.

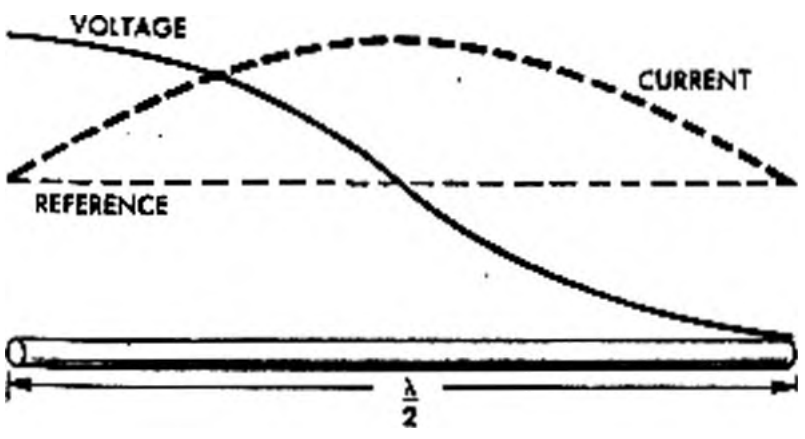


Figure 6-6. Voltage and Current Distribution on Half-Wave Antenna

The basic *Marconi antenna* is one-quarter wavelength long and is either grounded at one end or connected to a network of wires called a counterpoise. The ground or counterpoise provides the equivalent of the additional one-quarter wavelength required to make the antenna resonate as shown in figure 6-7. Note the similarity to the half-wave element when the image antenna is included.

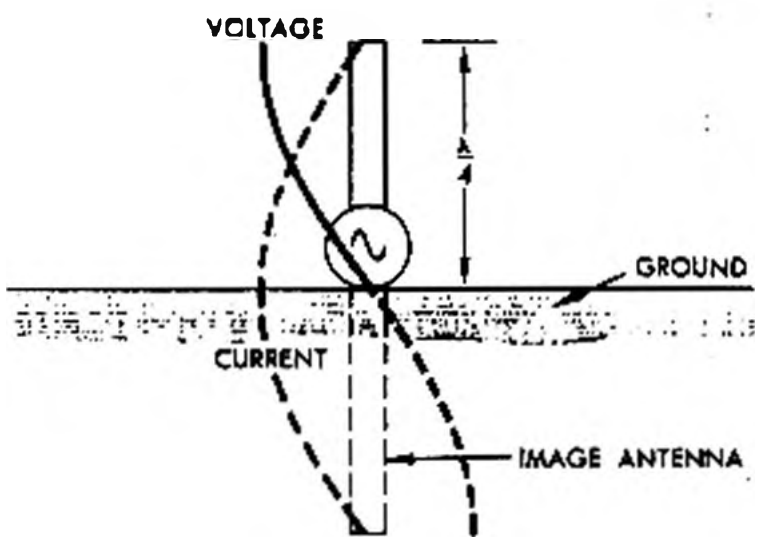


Figure 6-7. Standing Wave Amplitude of Current and Voltage on Marconi Antenna

It is seldom possible to connect an antenna directly to a receiver. Instead, it is necessary to use RF lines to carry the energy from the antenna to the receiver. The RF lines which carry the excitation energy might be resonant lines, nonresonant lines, or a combination of both. The means by which the energy is taken from the sensing element is referred to as the sensing element pickoff. Pickoffs include any of the devices that are used to transfer the energy received at the sensor to the following detecting and amplifying stages. These devices are taken up later in this chapter.

RADAR SENSORS. Present radar systems are intended primarily for air or ship targets, with some application to land-based targets that afford a good target-to-background signal. A main drawback of this system is its limited range. It is practically unaffected by weather conditions, it can be operated day or night, and it can be adjusted for minimum interference by jamming transmitters.

At higher frequencies, the primary functions of radar antennas are to concentrate most of the transmitted energy in one direction and to give a usable directional pattern in the same direction. The directional features of radar antennas are their most important characteristics. A directional antenna system makes it possible to illuminate a specific target area and receive the reflected pulses from a selected direction. Directivity of both the transmitting and receiving antennas improves the overall efficiency of the system and minimizes enemy countermeasures.

The basic method of obtaining directivity in an antenna system is to space two or more simple half-wave elements so that the fields from the elements add in some directions and cancel in others. A set of antenna elements is called an *antenna array*. Among the common types of arrays are the broadside, end-fire, collinear, and the parasitic or Yagi array.

Another simple method of obtaining directivity of an antenna system is the use of *parabolic reflectors*. In this type of antenna system, the driven element is placed at the focal point of a parabolic reflector.

The antenna is then working into a constant-phase reflecting surface. The pattern of field intensity is altered by changing the focal length of the paraboloid, distorting the reflector, or varying the field intensity. Four types of parabolic reflectors are shown in figure 6-8.

Another means of concentrating power is by *wave reinforcement*. This concentrating of power is accomplished by one or more rows of dipole radiators, spaced some distance apart, radiating at a fixed phase with respect to each other. Dipoles may be arranged along the length of a supporting rod, as shown in figure 6-9, and excited to yield a narrow endfire, or the dipole arrangement may give a nondirectional pattern in the horizontal plane and have a narrow pattern in the vertical plane.

When slots of the proper size and shape are cut into a waveguide, they act as radiators in the same way as a dipole antenna. The slot arrangements are generally found in higher-frequency radar systems. Collinear slot arrays are used in place of dipole radiators to obtain a smoother aerodynamic surface, as illustrated in figure 6-10.

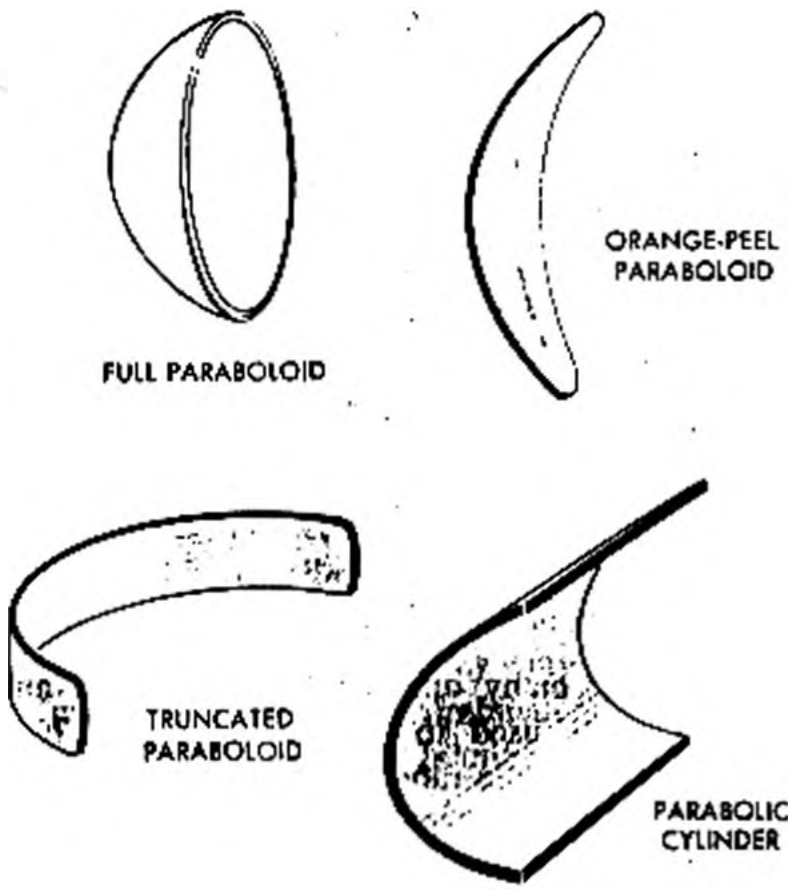


Figure 6-8. Parabolic Reflectors

Energy is also beamed by the use of horn-type radiators (see figure 6-11), all points of the wavefront being in phase at the mouth of the horn. The horn is generally used in place of the dipole and reflector arrangement. In general, horns are best adapted for use with waveguide-type feeds, while dipoles are used with transmission line feeds.

As higher speeds were attained in missiles, new antenna units were developed to eliminate drag. For this reason, slotted, waveguide termination, and recessed antennas are used.

Mention already has been made of the slot type. Energy can also be radiated from the waveguide itself if it is properly terminated. The beam angle and radiation pattern are governed mainly by the angle at which the waveguide is terminated. These waveguide termination types of antennas, pictured in figure 6-12, are generally called zero termination.

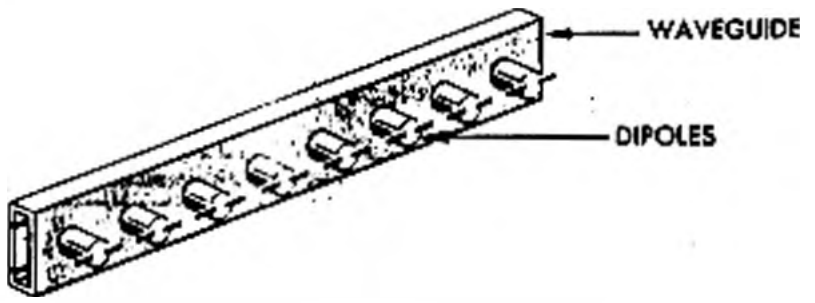


Figure 6-9. Collinear Array of Dipole Radiators

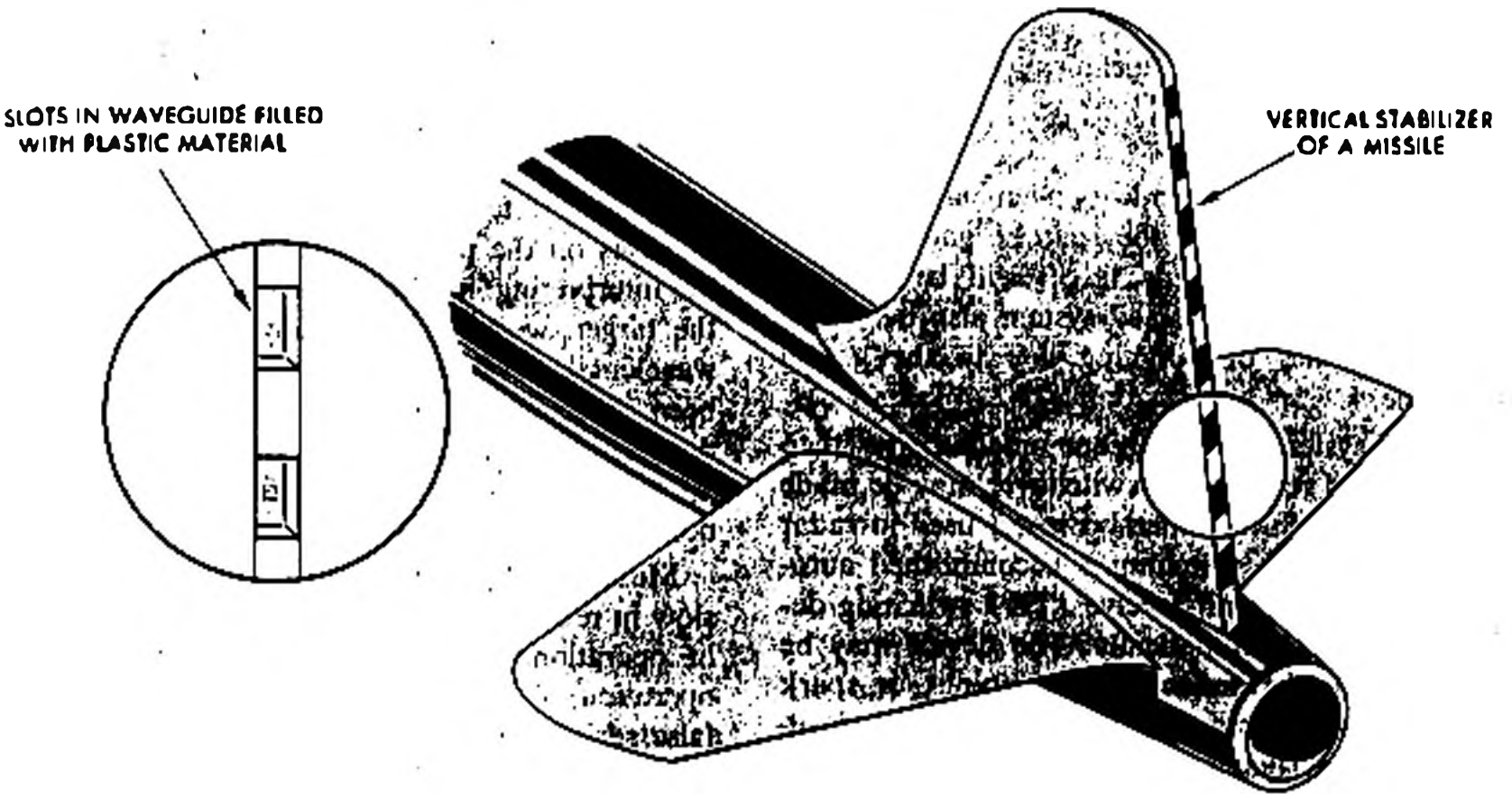


Figure 6-10. Collinear Slot Arrays Permit Smoother Aerodynamic Surfaces

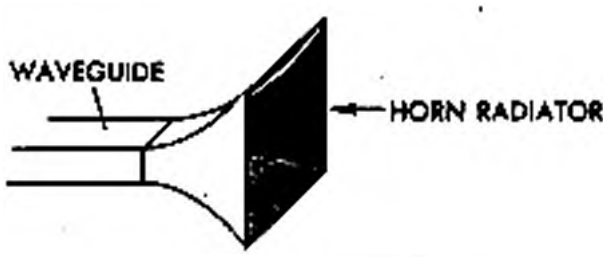


Figure 6-11. Horn-Type Radiators

If a radiating stub is made to lie within a surface rather than protrude, it can radiate a pattern much the same as a protruding type. Such a recessed antenna is illustrated in figure 6-13. Since they offer little air resistance, recessed antennas are invaluable for missiles designed to travel at high Mach numbers.

The antennas just discussed are only samples of the numerous types that could be used as the electromagnetic sensing elements of missiles.

For the purpose of guidance, there are three general types of radar systems that are used. These are as follows:

- Pulsed.
- Continuous wave (CW).
- Frequency modulated (FM).

Important factors to be considered when choosing equipment are the relationships between wavelengths of transmission, size of antenna, maximum range, and angular discrimination of each. Generally, FM radars operate on longer wavelengths than pulsed and CW radars. FM radars require large antenna arrays to obtain the narrow beam needed to give good angular discrimination.

Angular discrimination and maximum range increase as frequency increases; antenna size decreases with wavelength if angular discrimination, maximum range, and other factors remain con-

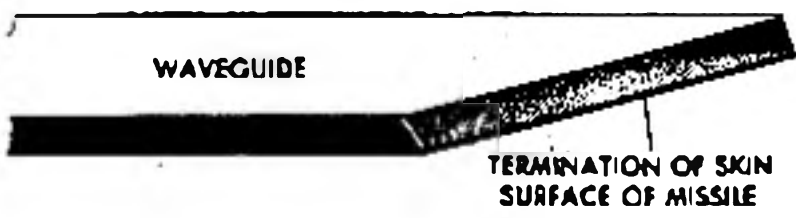


Figure 6-12. Zero Termination Antenna

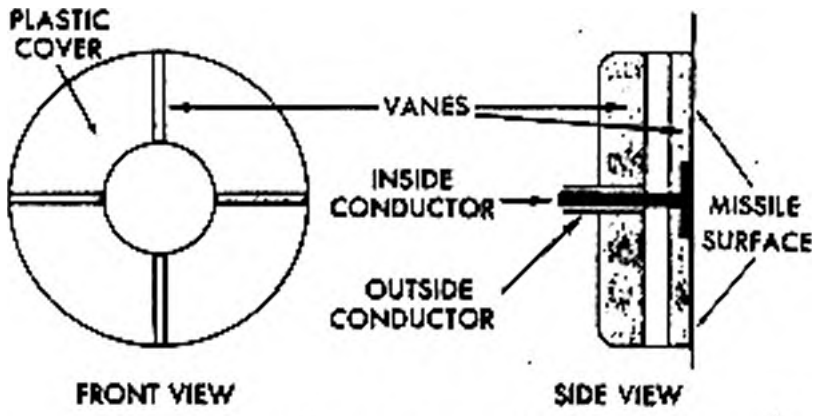


Figure 6-13. Recessed Antenna

stant. The higher-frequency systems are usually not desirable because technical difficulties and echoes from clouds, cold fronts, and rainstorms increase with frequency.

Pulsed systems provide convenient means of target selection in that they can be gated so that the echo from only one target will be used to guide the missile. A pulsed system also decreases the possibilities of jamming due to the gating principle.

In RF systems in which one antenna is used to transmit and receive, the switchover may be made mechanically. Such a method is not used in radar systems in which switching is accomplished automatically. A transmit-receive (TR) switching device functions automatically. The device may be in the form of resonant lines, T-joints, or spark gaps. These devices permit transmission and reception of a signal by a single antenna system without subjecting the receiver to damage by the high power being transmitted. It is also necessary to "cut off" transmission at the proper time in order to receive the reflected signal.

Illustrated in figure 6-14 is a typical spark-gap TR box. Upon transmission, a spark is created. The spark causes an open circuit to be reflected at the waveguide leading to the receiver. None of the high-power energy is allowed to enter the receiver. After transmission, the spark gap is extinguished. The physical dimensions of the waveguide then reflect a short circuit in that portion of the waveguide leading to the transmitter. This reflection causes the incoming signal to be diverted into the receiver section.

Infrared Sensors

To successfully use a guidance system that operates on the principle of receiving intelligence by interpreting infrared radiations (heat) from the target, we must first study the intensity and wavelengths of these radiations. Once this has been done, a simple detecting device may be designed that requires no transmitter, relying only on the radiations emanating from the target for detection.

Most of the infrared seekers are comparatively slow in reaction time and quite limited in distance of operation. Fog and rain limit their effective operation, and decoy fires or flares may cause false seeking. Considering the altitudes that some missiles attain, the heatseeking device may not be sensitive enough to discriminate between target and background when the predump point is reached.

Infrared radiation in the electromagnetic spectrum is the band of frequencies immediately below visible light. Any object that is at a temperature different from its surroundings radiates heat in various quantities and at characteristic frequencies. Heat energy has wavelengths from approximately 300 to 0.7 micron (1 micron = 10^{-4}cm) or frequencies between about 1 million megaHertz and 400 million megaHertz.

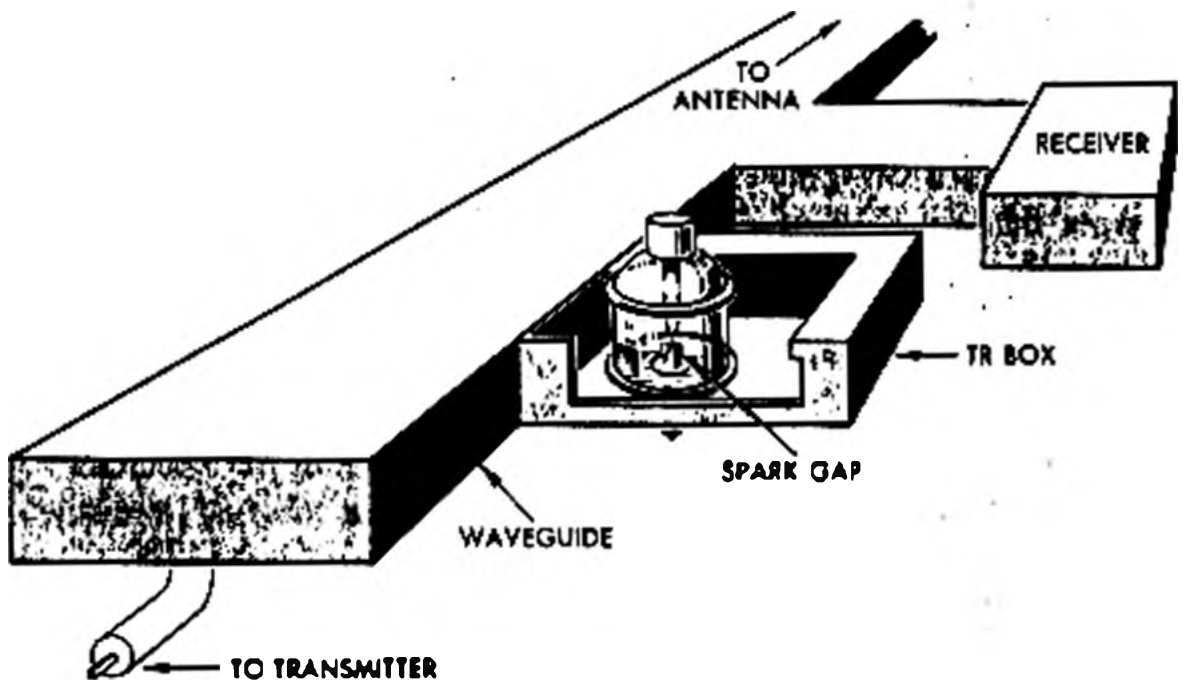


Figure 6-14. Spark-Gap-Type of Transmit-Receive Switching Device

Radiations of infrared energy can be detected with a variety of sensing elements, but the most common are the thermocouple and the bolometer.

A *thermocouple* is a joined pair of conductors of dissimilar metals as shown in figure 6-15. Any

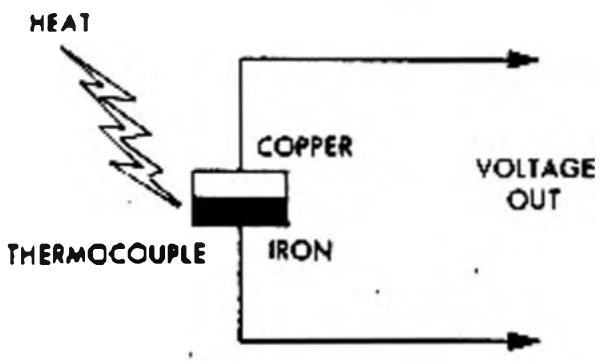


Figure 6-15. Thermocouple

change of temperature at the junction causes a measurable voltage difference to exist between the two.

Although the voltage difference at the junction of a thermocouple is quite small, the sensitivity can be increased to a point where it becomes a useful detector. This increase is accomplished by using a number of such devices in series with an appropriate circuit. When several thermocouples are joined in series, they are referred to as a thermopile. A thermopile detector may be mounted at the focal point of a parabolic reflector to increase the sensitivity and to provide a means of determining the direction of the radiation.

A *bolometer* is another type of sensing element. It is a heat-sensing element which depends on the change of electrical resistance of a material for its action. The first bolometer was made in 1880. It was made of two thin strips of platinum, which formed the two arms of a Wheatstone bridge. The strips were blackened on one side. As heat was applied to the strips, the absorption of energy by one of the strips caused an increase in its electrical resistance. This change of resistance caused a change of current flow when an external voltage was applied to the circuit.

The bolometer, shown in figure 6-16, has proven to be a very efficient sensing unit for infrared homing systems. (This instrument has been used to measure the temperatures of stars.) It usually consists of very thin strips of black-coated (oxidized) nickel mounted in an evacuated cell or chamber into which has been injected a small

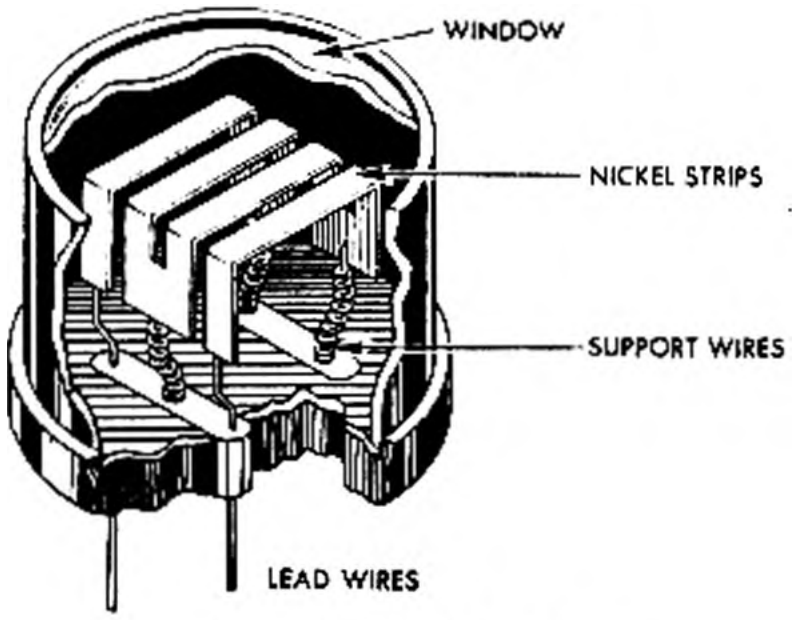


Figure 6-16. Bolometer

amount of hydrogen gas under low pressure. The hydrogen gas tends to increase the heat-absorbing property of the bolometer and causes it to respond more rapidly to slight variations in the amount of infrared radiation focused upon it. Bolometers possess the advantages of small size, rugged construction, and relatively large current-carrying capacity, in addition to a high degree of sensitivity.

Heat radiations cover a wide band of the electromagnetic spectrum. The sensing device must be especially responsive to the radiations that it is likely to receive throughout a wide range of atmospheric conditions such as variations in humidity and dust concentration. Research has proved that infrared radiation of from 8 to 14 microns in wavelength (8 to 14 millionths of a meter) comprise a large percentage of the total heat radiation of a body. This radiation is obstructed by moisture particles to a lesser degree than radiations of other wavelengths. The bolometer is designed to be most sensitive to radiation within the 8- to 14-micron band.

Light-Sensitive Sensors

Many types of light-sensitive tubes, such as those used in star trackers and as sensors in light-seeking homing systems, are available. These cells or tubes all operate on the same basic principles. Figure 6-17A is an example of a *phototube*. When light rays strike the photocathode, they cause electrons to be emitted from its surface. These electrons are attracted to a postively charged anode, in the center of the photocell, producing a

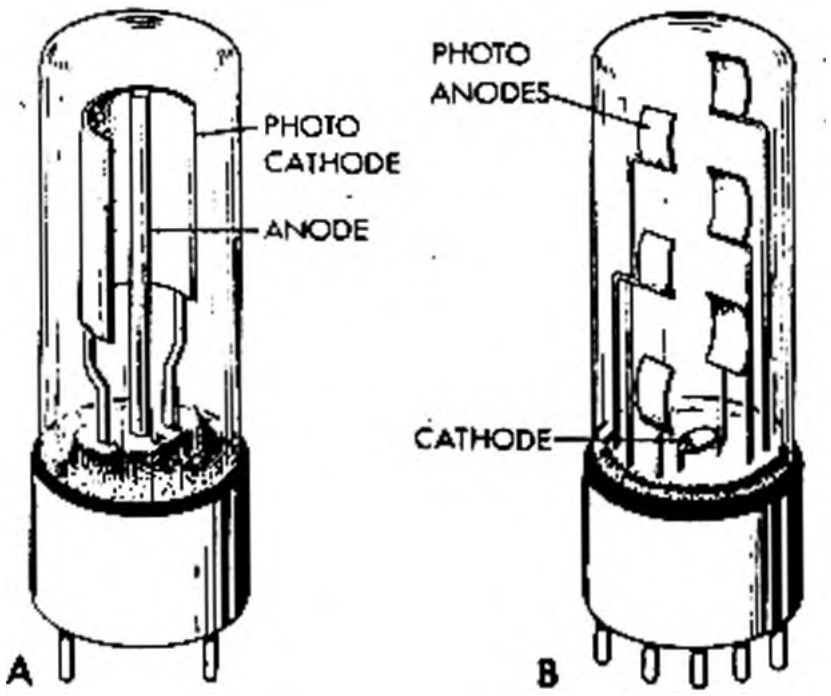


Figure 6-17. Phototube and Photomultiplier Tube

current flow in the cathode-to-anode circuit. The amount of current produced varies in proportion to the amount of light striking the photocathode. The composition of the coating on the photocathode determines its sensitivity to a particular color or frequency of light. (Photocells with different cathode coatings are used in the field of spectrum analysis. A single cell supplemented by a set of color filters can also be used for this purpose.) If the characteristics of the light emitted by a target are known, a sensing element may be selected which is most sensitive to the light with the known characteristics. By using filters to exclude light with unwanted characteristics, the selectivity of the system can be increased to a higher degree.

The *photomultiplier tube*, shown in figure 6-17B, is similar to the type used in star trackers. The photo anodes are arranged so the electrons emitted by the cathode are attracted to a positively charged anode from which an increased number of electrons are released by secondary emission. These electrons are attracted to the next anode, which is at a higher positive potential that the preceding anode. A greater number of electrons are released. This process continues throughout the photomultiplier tube until the final anode is reached, which may be at a very high positive potential. In this manner, a relatively large current flow can be produced from an initially small amount of photoemission. The use of photomultiplier tubes is restricted to systems in which high-anode voltages are available. They are highly efficient as light-sensing devices due to their sensitivity and amplifying characteristics.

Camera tubes, such as the iconoscope, are actually photosensitive tubes designed to provide more complex information from an illuminated object than merely the intensity or wavelength of the light reflected from it. Homing guidance systems employing television or facsimile (map matching) principles may use camera tubes as sensing elements. Their use is generally restricted to relatively large missiles.

One system of missile guidance is the celestial navigation system. In this system, *star-tracking telescopes* (one shown in figure 6-18) are used to sense the positions of celestial bodies. By determining the positions of these bodies, spatial coordinates can be established. The coordinates are used as a guidance reference network. A set of telescopes which have a constant fix on at least two reference points is used, forming an automatic navigation system.

The celestial system of guidance is considered a light-sensing device, the light source in this case being a reference rather than a target. Star-tracking telescopes are usually used as a corrective reference for an inertial guidance system in long range missiles. The term used for this system is "stellar-supervised."

Filters used in conjunction with the sensing units previously described are either light or

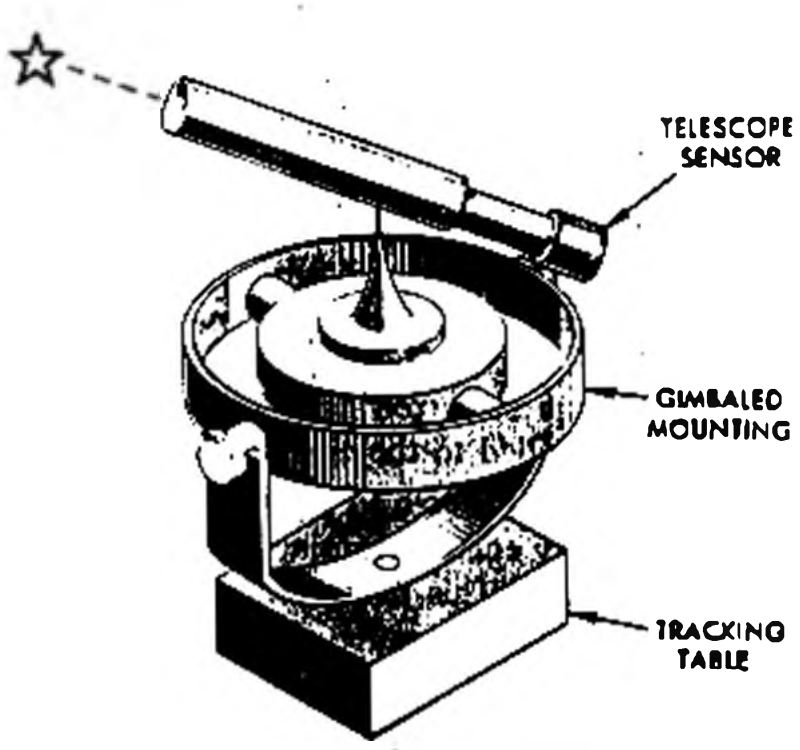


Figure 6-18. Star-Tracking Telescope

electrical filters, depending upon the type of sensing element with which they are used.

Very few materials absorb all wavelengths or colors of light uniformly. When the desired wavelengths or colors of light are determined, a filter can be placed between the source of light and the sensor to absorb the undesirable wavelengths. (The principles of light filtering are used in photography to produce desired contrasts between light and shadow or between various colors.) Filters are generally made of colored glass, but they may be of other translucent materials possessing the desired qualities. For use with an infrared sensor, a filter may be selected which would absorb the violet, blue, and yellow light. This would permit the infrared light to pass through to the sensing element. Usually, no single filter screen will produce the desired result; so two or more filters may be necessary to remove all or most of the undesired wavelengths. Filters should be selected that will not reduce the intensity of the desired form of light below the sensitivity range of the sensing unit.

Acoustic Sensors

The Navy has used the principles of sound detection for many years to determine the presence and position of submarines or other ships. This equipment, known as *sonar*, uses tracking circuits similar to radar with the exception that the input frequency is ultrasonic (sound waves) rather than high RF frequency. Missiles that fly at velocities greater than the speed of sound cannot use such a system as no reliable acoustic signals would be available.

The most common acoustic sensors are *microphones* that will pick up and amplify sound waves. When used as a sound-seeking homing system, the microphone serves as the sensing device. Although microphones vary greatly in design and configuration, they all employ the same basic principles of operation.

One type operates by varying electric current. The variation is made by changing the resistance between loosely joined electrical conductors, such as grains of carbon. Another type operates by varying the reluctance of a magnetic circuit. The microphone used in a sound tracker should be very sensitive. This calls for a large diaphragm or surface to intercept the sound waves. The diaphragm should possess a high degree of resiliency or freedom of movement so that sound waves of low intensity will produce relatively large movements of the diaphragm. The diaphragm itself may be an integral part of the electrical circuit or it may be mechanically linked to the electrical circuits involved.

The *capacitor microphone* is shown in figure 6-19A. Here the diaphragm acts as one electrode of a capacitor. When sound waves strike the outer

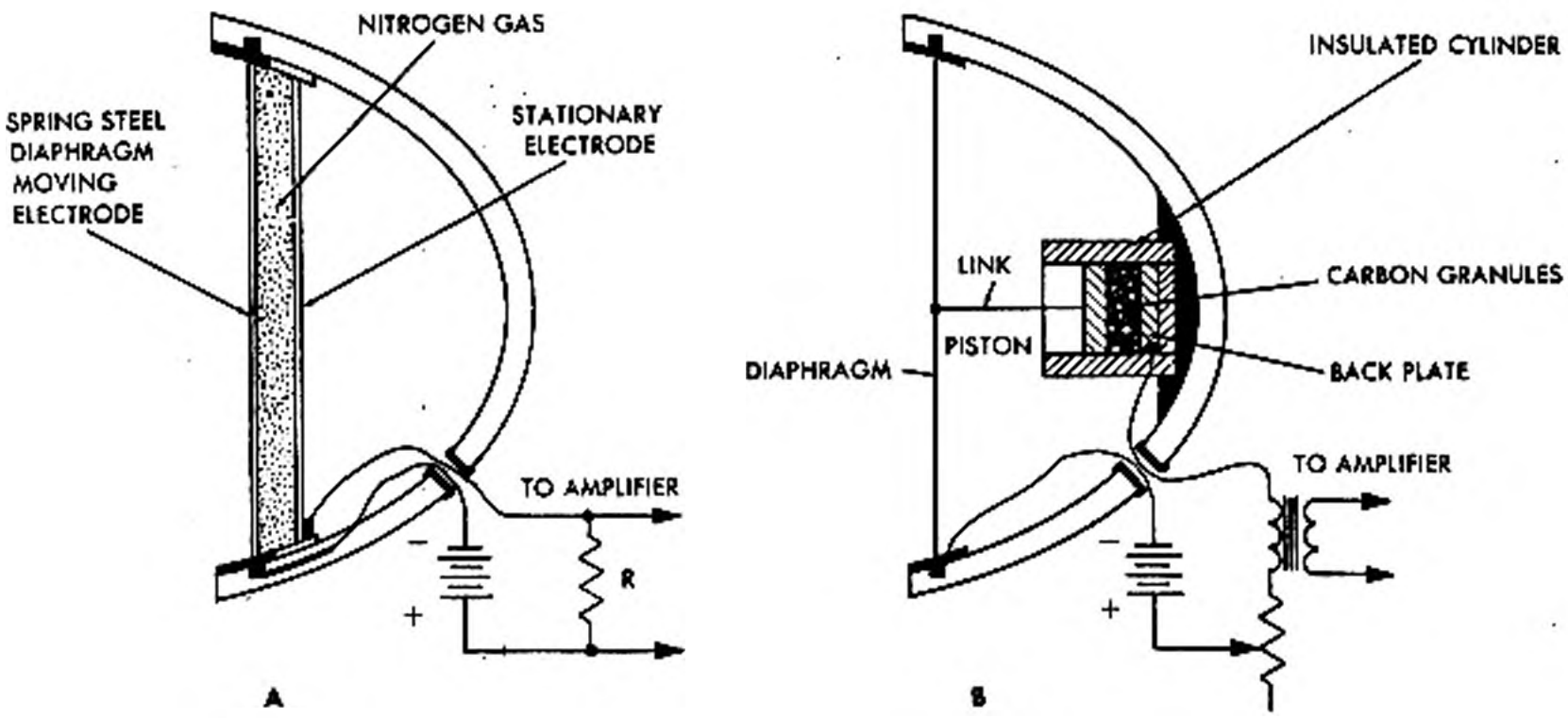


Figure 6-19. Capacitor and Carbon Microphones

surface of the movable electrode, they tend to push it inward toward the stationary electrode, thus reducing the space between the two. This increases capacitance, causing current to flow and developing a voltage drop across the resistor R. Voltage variations across R will result from the variations in capacitance produced by movements of the diaphragm.

The *carbon-grain microphone* is shown in figure 6-19B. It is one of the earliest types developed and is similar to those used in telephone circuits. It consists of a thin diaphragm of spring steel, mechanically linked to a piston that moves within an insulated cylinder containing loosely packed granules of carbon. The piston serves as one electrode, and the backplate of the cylinder serves as the other. Vibration of the diaphragm moves the piston so that it exerts greater or less pressure upon the carbon granules, producing variations in their resistance.

When the electrodes are connected in a circuit, the variations in the resistance of the carbon granules cause variations in the current flowing through the limiting resistor and the primary of the transformer. The current induces variations of voltages across the secondary of the transformer. These secondary voltages vary proportionally with the movements of the diaphragm and represent similar variations in the amplitude and frequency of the sounds received. The design of the microphone cavity or housing determines the directional properties of the microphone. This device should be shock-mounted to prevent pick-up of undesirable vibrations.

A single microphone may be mechanically pointed like a rotating antenna to pick up sounds from the direction desired or to locate the direction of the source of a sound. Several stationary microphones, each pointed in a different direction, may be used with a switching system or commutator so that the output of each microphone may be sampled successively and the direction of the sound source determined from the output of greatest amplitude.

To increase the selectivity of a sound-sensitive homing system, a conventional *bandpass filter* can be used. Such a filter is designed to eliminate or cut off signals of frequencies above and below the frequency band of the desired signal. For example, suppose a sound-seeking homing system is used to guide a missile to a target which emits a characteristic sound. The sensing element might be receiving sounds of many frequencies of even greater intensity than the sounds emitted by the target. The characteristic sound probably lies within a narrow band of frequencies. The system must respond only to the particular frequency within that band.

An electrical filter, to possess the desired bandpass characteristics, must include at least two circuits. One branch should offer very high impedance to signals of higher frequencies than the highest frequency present in the desired signal. The other branch should offer very high impedance to signals of frequencies lower than the lowest frequency present in the desired signal. Circuitry and formulas for filter design may be found in most electrical engineering handbooks.

Magnetic Sensors

Magnetic systems have been used with some success. Variations in the earth's magnetic field, effects of magnetic disturbances, and difficulties in dip and magnetic components add to the problems of developing a dependable and accurate magnetic system. The use of such a system lends itself well to long-range missiles and is relatively free from jamming. Magnetic devices are used sometimes as auxiliary equipment to various other systems.

Magnetic systems employ flux-gate sensing units that sense variations of the earth's magnetic field due to deposits of iron in certain localities. Variations in the earth's magnetic field are sensed as flux density changes in the flux gate and are transferred into an AC signal that positions the missile to a predetermined course. The use of the magnetic system is limited to regions away from the poles. The limitation exists because the polar regions have frequent magnetic storms that would disrupt the system. The magnetic field about the polar regions is not well enough defined for accurate guidance.

A common sensor used in the magnetic guidance system is the *magnetometer*. Generally, three such devices are positioned mutually perpendicular to each other to sense the missile's position along a fixed magnetic line.

A principal advantage of such a system is that it is capable of sensing objects under water or ground.

Inertia Sensors

Gyros and accelerometers are the basic components of any inertial guidance system. These devices have been highly developed for many other applications and their characteristics are well used as sensors in missiles. Gyros are usually associated with the control system. In a guidance system, they are used primarily to establish a stable platform for the accelerometers.

To detect the degree and direction of trajectory change, accelerometers are used. An accelerometer is an inertia device. A simple illustration of the principle involved in accelerometer operation is the human body in an automobile. If the automobile is subjected to sudden acceleration, the body is forced back in the seat. If the auto is stopped suddenly, the body is thrown forward. When the auto goes into turns, inertia force on the body is away from the turn.

The basic principle of operation of an accelerometer consists of the measurement of the inertia reaction force of a mass to an acceleration. Many different methods can be used to measure this force. The choice of method mainly depends on the frequencies present, whether or not acceleration is changing, and the type of output desired.

There are two principal types of accelerometers. In the first type, the inertial force of the mass causes a displacement of the mass in an elastic mounting system. The displacement is then measured by any one of several methods. The second type of accelerometer operates on a fundamentally different principle. The force that counteracts the inertial reaction force of the mass is not supplied by an elastic mount but is supplied by an electric current, a stream of air, or any other system which can produce a controllable force. In this system, a small deflection of the mass is detected, and a force is instantly applied to the mass to prevent any further motion. Acceleration is indicated by the magnitude of the force applied to the mass by the electric current or whatever other means is used to produce the balancing force.

Whenever a mass which is supported by an elastic mounting system is displaced due to some force, the mass tends to oscillate about its rest point. This oscillation is known as simple harmonic motion. The number of oscillations during a period of time are limited by damping action of the system. Damping limits the amplitude of each oscillation. The number of oscillations is limited. The damping action may be increased by the use of a device especially for this purpose, called a damper.

In figure 6-20, if the spring-suspended mass were displaced and then released, it would tend to oscillate in simple harmonic motion. If there were no retarding action by the spring, it would oscillate as illustrated in the "undamped" condition. A spring does offer resistance to the oscillation; consequently, the oscillations will diminish as represented in the "damped" condition.

To have an efficient accelerometer, the amount of damping applied to the system must be just great enough to give a nonrestricted displacement and yet prevent any oscillations from existing.

Consider the accelerometer in figure 6-21. If the missile suddenly accelerates in the direction

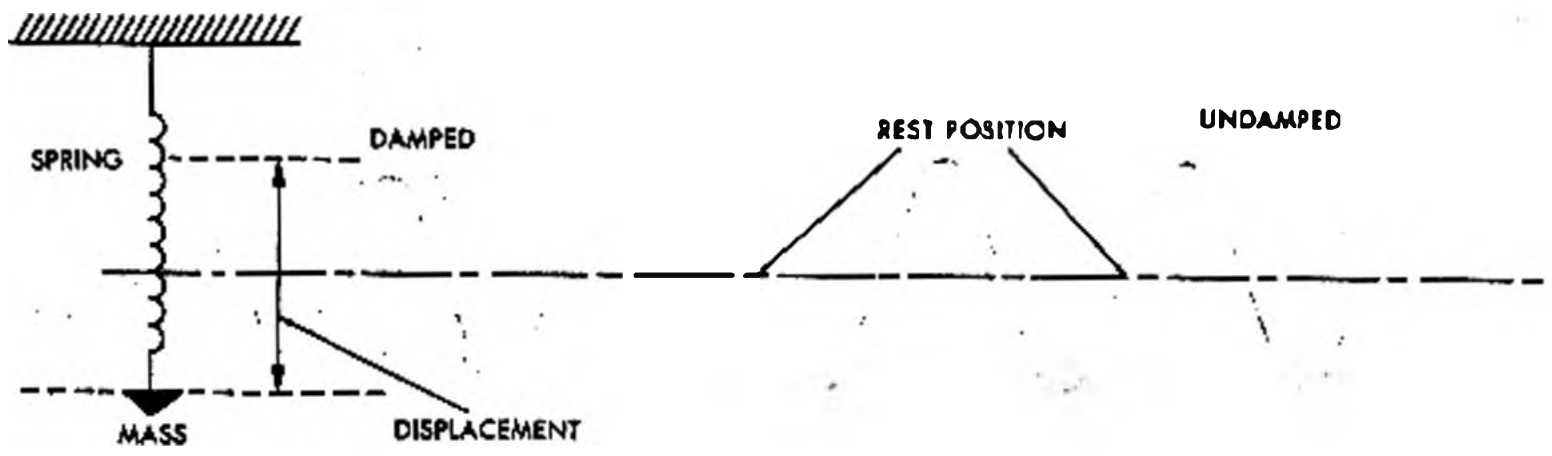


Figure 6-20. Effect of Damping on Spring-Suspended Mass

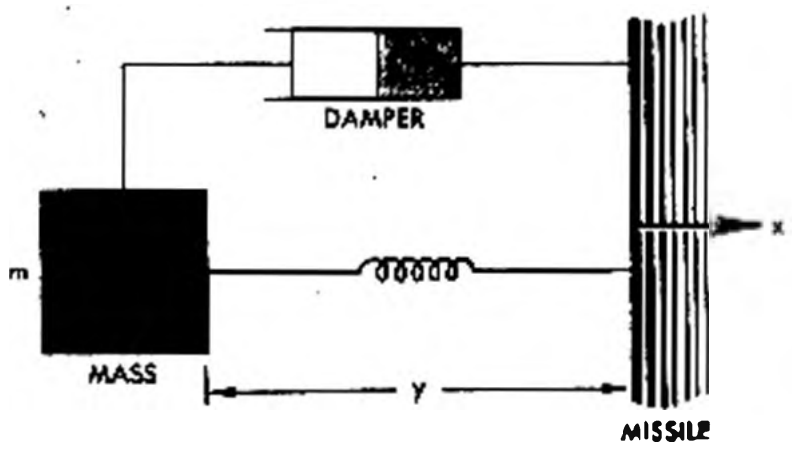


Figure 6-21. Accelerometer Function in Inertia-Sensor Unit

of X, the mass m is displaced (relative to the missile) some distance in terms of y. The relative acceleration of the mass during this time indicates a force present because F = ma; where F is force, m is mass, and a is acceleration.

There is a certain displacement of the spring per unit of force applied. The spring at the end of this period of time tends to reverse the direction of force on the mass. This would lead to simple harmonic motion if allowed to continue. The damper produces an action that limits this oscillation. The damper must attain a certain velocity at the end of this time so that its retarding action will counteract the force set up by the spring as a result of the displacement of the mass. The damper will again present such a counteracting force when the mass has reached its maximum displacement in the opposite direction. Displacement in the opposite direction is due to the force exerted by the spring.

There are three results that may come about in an accelerometer;

1. The system may be overdamped (it will not oscillate). This condition occurs when the damping action is greater than the tendency of the spring to sustain oscillation.

2. The system may be critically damped (limiting condition of oscillation). This condition occurs when the above-mentioned forces are equal.

3. The system may be underdamped (will oscillate with decreasing amplitude). This condition occurs when the damping action is less than the tendency of the spring to sustain oscillation.

Under certain conditions, the displacement of the mass relative to the missile can be used to measure one of three things:

1. Displacement proportional to acceleration.

2. Displacement due to the velocity, as indicated by the damping action.

3. Displacement proportional to distance moved by the missile.

To measure the acceleration for missile use, the accelerometer should operate in the underdamped region, bordering the critically damped condition. This gives a system that is sensitive, and at the same time the system tends to prevent any transient oscillations. The accuracy of the system depends on the method used to measure the displacement of the mass relative to the missile and on the linearity of the elastic mounting. The sensitivity of the system may be increased by reducing the frictional forces in the mounting and by improvement of the indicators used.

As previously presented in the chapter on physical principles, the average acceleration required to displace an object a distance s in t seconds, assuming zero initial velocity, was given as $s = \frac{1}{2} at^2$; where s is the distance in feet, a is the acceleration in feet per second, and t is the time in seconds.

The accelerometer sensitivity required for accuracy in a given system may be found by the use of this equation. For example, if a missile has a velocity of 1500 mph, the required time to travel 400 miles is roughly 960 seconds, and the average acceleration error that will cause an error of 3000 feet in this time can be found by substituting the time and distance into the formula. This substitution would give the acceleration error as 0.0065 ft/sec/sec. The measured acceleration, in error by 0.0065 ft/sec/sec, all other errors being zero, would cause a system error of 3000 feet in 400 miles.

From this example, the acceleration error appears to be quite low. If the maximum acceleration experienced by this missile was 1 g (32.2 ft/sec/sec), the error is only about 1 part in 5000, and if the maximum acceleration was 10 g, the error is only 1 part in 50,000. However, the error in distance is determined not only by the error in acceleration but also by the square of the time. If the range is increased (and, therefore, the time of flight), even this small error in acceleration becomes too much for the extreme accuracy required by present-day missiles. For ballistic missiles with intercontinental range, even greater

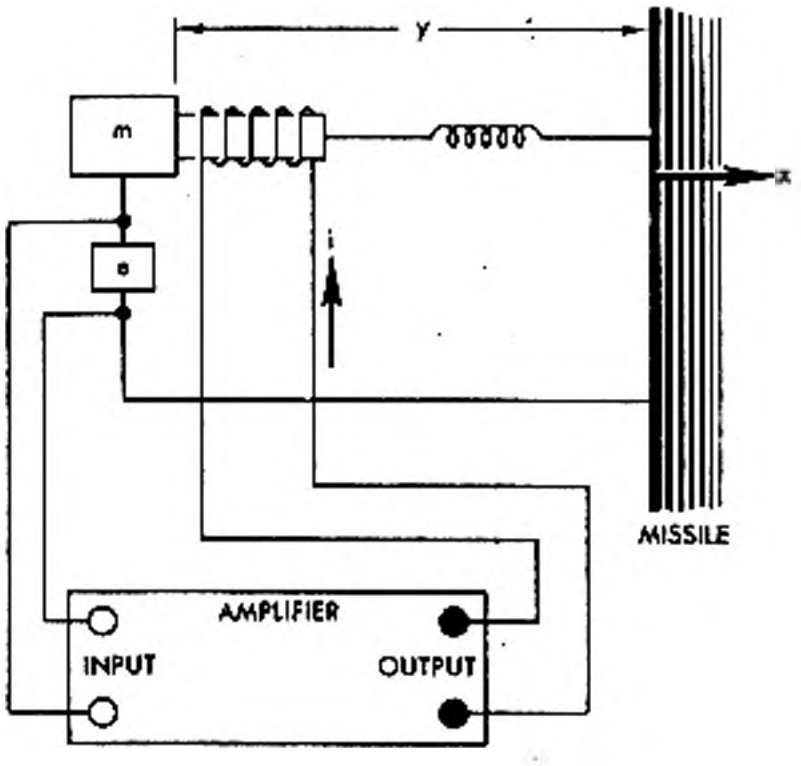


Figure 6-22. Accelerometer with Inertia Reaction Force of Mass Supplied by Electric Current

accuracy and sensitivity are required of the accelerometers.

Shown in figure 6-22 is the second type of accelerometer in which the force that counteracts the inertial reaction force of a mass is not supplied by an elastic mount but by an electrical current.

Let e be the voltage that is developed due to the displacement of the mass relative to the missile and i the output of the amplifier (the restoring current fed to the coil).

If e depends on the displacement of the accelerometer mass (m), then the output of the amplifier (i) depends on the displacement (y). If the missile accelerates, there is a certain voltage applied to the amplifier input. After necessary amplification, there is a proportional current applied to the coil so as to counteract a certain amount of the displacement. As the mass (after initial acceleration) tends to move in the opposite direction, another restoring current is generated.

The acceleration could be measured by the displacement, by the input to the amplifier, or by the restoring force of the current fed to the coil.

The displacement allowed to the mass is small, which means that errors arising from the nonlinearity of the electric field are mostly eliminated. The sensitivity of such a system can be increased with little effect on the range of operation.

There are many ways of obtaining a signal. Obtaining a signal is a function of the displacement of the mass in the first type of accelerometer.

Another method has been developed by the Massachusetts Institute of Technology. This method uses a change of inductance between two coils as their separation is varied. The coils form part of an inductance bridge. As the inductance changes, the bridge is unbalanced and an AC voltage output results.

Another accelerometer uses wire strain gages as the suspension elements for the mass. The strain gages form the arms of a bridge. A change of acceleration causes a change of the electrical resistance of the circuit, giving an AC output that is an indication of the acceleration.

An interesting example of the second type of accelerometer is the vibrating accelerometer, which was developed at Fort Bliss. In this accelerometer, the mass and coil vibrate, causing periodic currents to flow through a coil. An acceleration causes the average momentum of the system to change. The average current output of the system is proportional to the average acceleration.

Still another principal type of accelerometer is the manometer. In this type, accelerations are controlled by the electrolyte flowing toward one or the other ends of the manometer. This action provides current control between pairs of electrodes. The venturi, shown in figure 6-23, damps the response of the manometer by controlling the electrolyte movement. This damping prevents os-

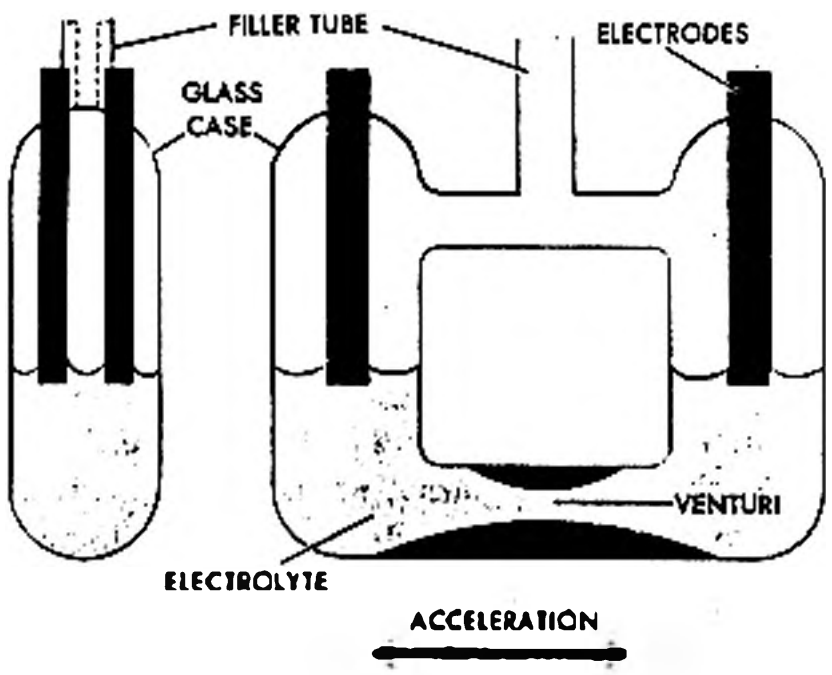


Figure 6-23. Manometer Accelerometer

cillation of the electrolyte, which prevents oscillation of currents fed to the control system.

Another accelerometer in use is of the vibrating-reed type. In this type, a change of frequency indicates the acceleration along the axis of the reed. The reeds are used facing in opposite directions so that the difference in frequency between the reeds is proportional to acceleration in either direction. Tuning forks have been used instead of simple reeds to eliminate dead spots when the reeds are vibrating at the same frequency. The two tines of a tuning fork always vibrate 180° out of phase, and the energy of vibration would be theoretically canceled out at its base.

The gyroscopic principle has been used to build a very accurate integrating accelerometer. An intergrating accelerometer sums up accelerations and has an output representing velocity.

In the gyro integrating accelerometer, shown in figure 6-24, an unbalanced weight is mounted on the gyro output axis (accelerometer input axis). When the accelerometer is subjected to an acceleration along its output axis, a torque is set up that causes the gyro to precess. The precession is detected by the pickoff and a signal is produced. This signal is amplified and applied to a servomotor which drives the gyro case at an angular rate just fast enough to produce a second torque that balances out the torque produced by the acceleration.

Since the angular rate of gyro precession about the accelerometer output axis is proportional to the linear acceleration along the same axis, the total angle through which the gyro rotates about this axis is the integral of the rate of turning over the time interval. The output of the synchro transmitter is proportional to the linear velocity.

There are many other ways of measuring the acceleration of a missile. The methods just discussed are just samples of the many ways used to accomplish this function.

Remember that when a guidance system is chosen for a missile, various advantages and disadvantages of each system must be considered before making a selection. A combination of two or more of the systems may be used to attain guidance throughout the flight. No matter what systems are chosen, the sensor element is the eye to an otherwise blind flight.

SIGNAL PICKOFFS FOR SENSING ELEMENTS

To transfer an incoming signal from the sensor unit to later stages of amplifying and computing, a sensing element pickoff is used. The sensing-

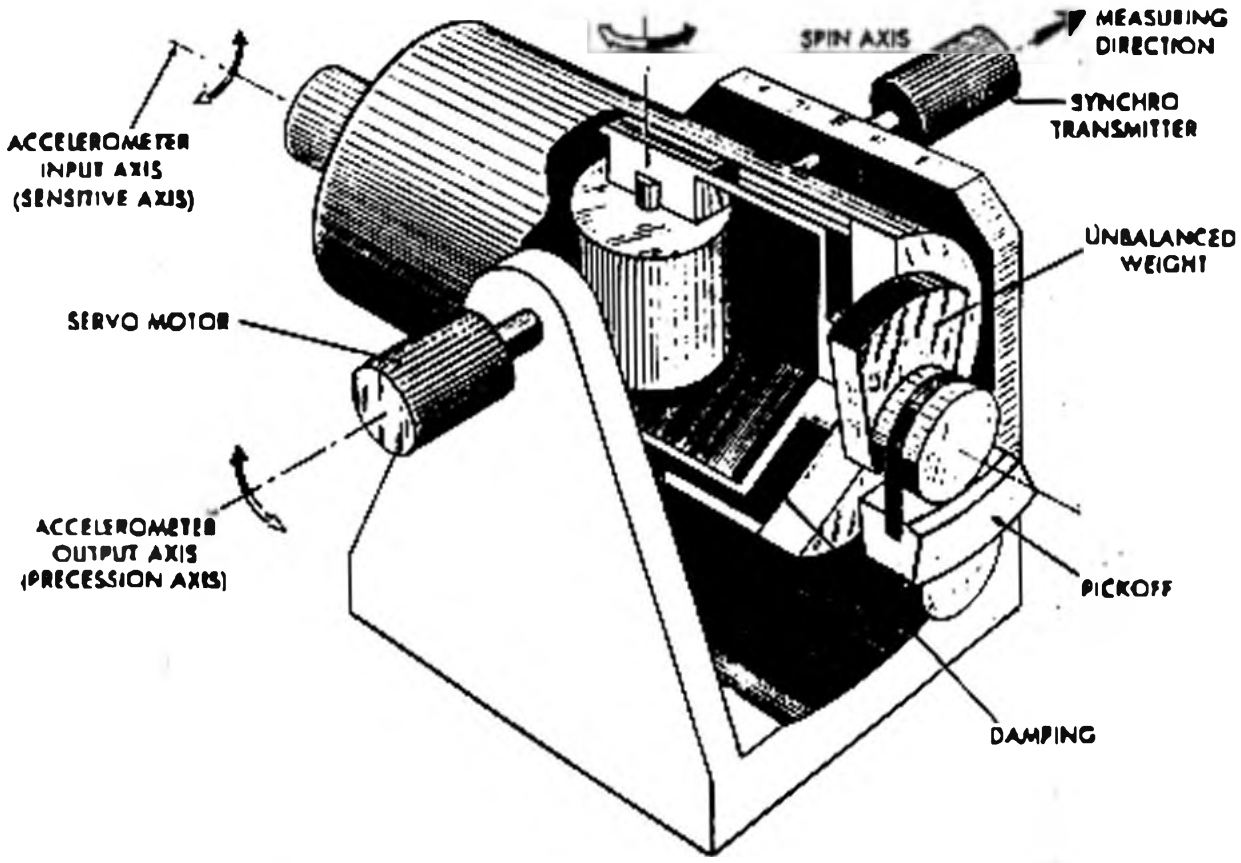


Figure 6-24. Gyro Integrating Accelerometer

element pickoff is any unit or combination of units that takes a received signal and transfers it in a usable form to the following stages.

Transformer Pickoff

To transfer regular RF, the usual sensing-element pickoff is the transformer. The primary of the transformer is connected to the sensing element, and the load is connected to the secondary. An input transformer refers to a transformer that couples a low-impedance source of energy to the grid of a tube. The input transformer generally has a high step-up ratio so the magnitude of the incoming signal will be such that it may be amplified in later stages.

In the RF system of guidance, an antenna is considered as the sensor, and the unit that relays the signals to later stages is the sensing-element pickoff.

In considering figure 6-25, note the line which indicates the detector as a portion of the sensing-element pickoff. It is often difficult in an electronics system to define clearly the limits of a particular unit. Visualizing the detector as the pickoff unit rather than the input transformer may be easier. Or perhaps both the input transformer and the detector should be considered as the sensing-element pickoff. The important point to remember is that the sensing-element pickoff takes a received signal from the sensor and transfers it in a usable form to the following circuitry of the system.

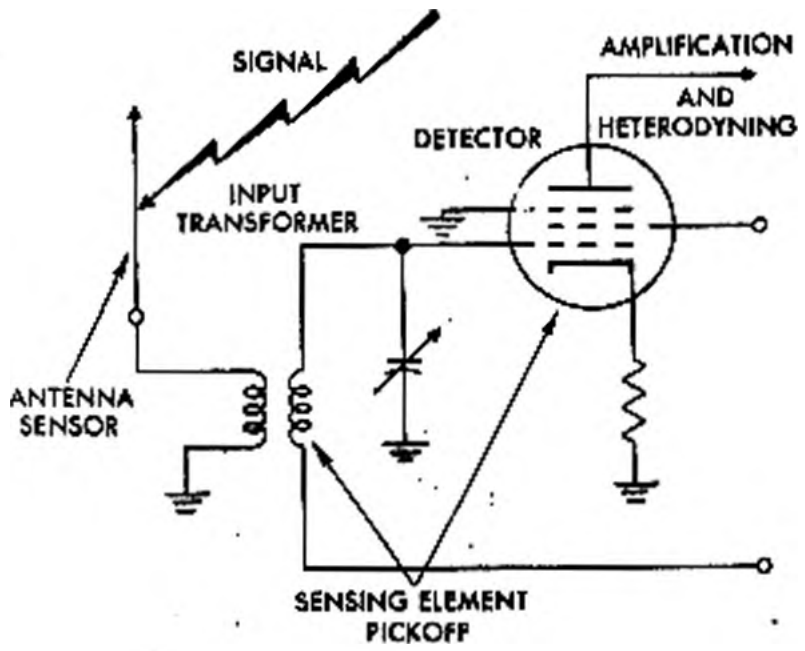


Figure 6-25. Sensing Element Pickoff for RF

Pickoffs for Radar Systems

In a radar system, the energy that is received by a waveguide after leaving the antenna is transported along the waveguide to a point where it must be transferred once more to a usable form by the receiver. This energy may be fed directly into a resonant cavity where oscillations are initiated to excite amplifying stages. The energy also may be detected by crystals or tuned stages.

If the energy is to be collected from the waveguide, a pickup probe is often used. Figure 6-26 shows a cutaway sketch of a waveguide section and pickup probe. A probe will receive a maximum of energy if it is inserted into the waveguide at a certain point and position.

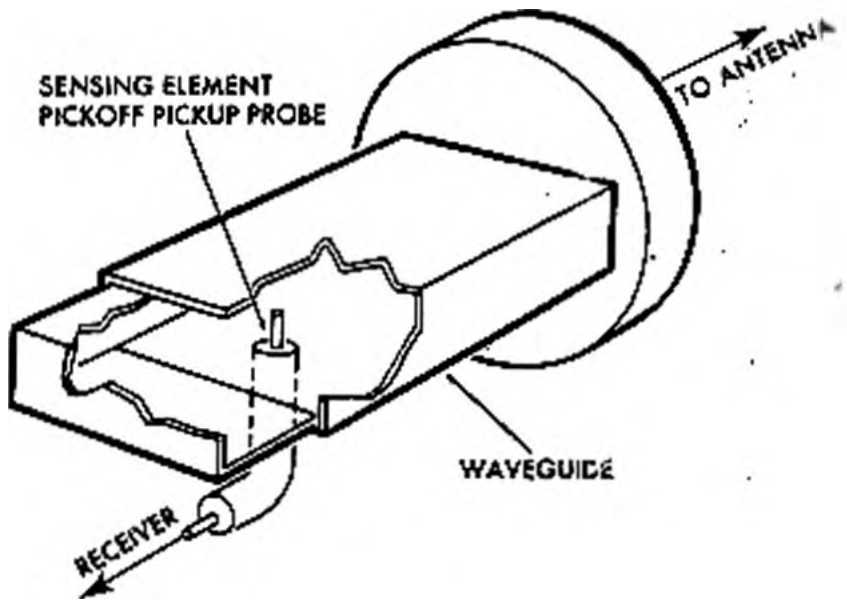


Figure 6-26. Waveguide Section and Pickup Probe

The energy returned from the target is available for the receiver. The receiver accomplishes the heterodyning, detecting, amplifying, and computing necessary to cause the target selector to operate at the proper time to provide a signal to the controls unit.

The amount of energy available to the receiver is proportional to the effectiveness of the antenna sensing unit.

Clear definite limits are difficult to determine when referring to the sensing-element pickoff. In the illustrated case, the waveguide, the pickup probe, and possibly a tuned circuit may be combined as a unit and referred to as the sensing-element pickoff.

Pickoffs for Infrared Sensing Elements

The infrared sensing elements have a number of different types of pickoffs as is true in the case of RF and radar systems. Common use is made of a bridge circuit as part of the sensing-element pickoff. The bridge circuit is useful in the heat-detecting systems because sensors for these systems generally operate on the principle of resistance changes.

Figure 6-27 shows a typical bridge circuit used with a bolometer sensor. The changing temperatures applied to the bolometer are accompanied by a fluctuating resistance so that the bridge circuit is thrown off balance. The transformer primary in two arms of the bridge acts as an impedance match between bolometer and the high impedance input to the amplifier. The transformer also converts the signal to an approximate sinusoidal waveform before passing it on to further amplification stages.

As in any of the sensing-element pickoffs, the infrared pickoff should be sensitive enough to relay quickly any received change of signal and should be subject to a minimum of operational difficulties.

Some types of sensing-element pickoffs were mentioned previously in the chapter on control system components. Some of these could be applicable to guidance systems as well. One such device, for example, is the thermistor.

A thermistor is a variable resistance element. The resistance of a thermistor varies inversely as the temperature; that is, as the temperature increases, the resistance decreases. The resistance of the thermistor is varied by thermal changes in the

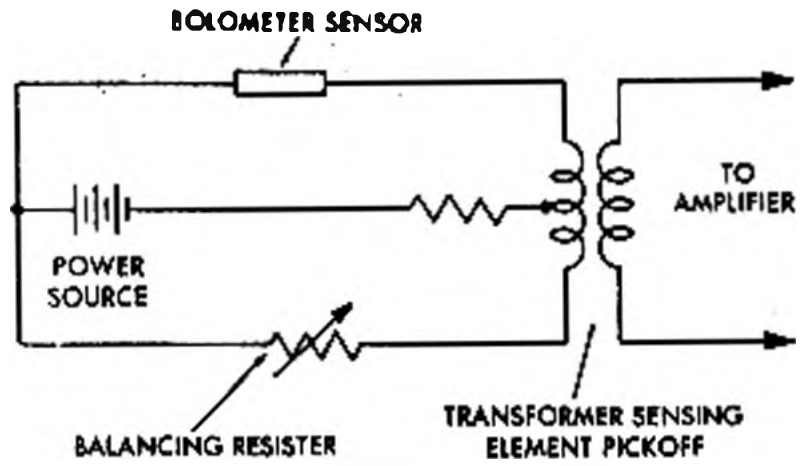


Figure 6-27. Typical Bridge Circuit used with Bolometer.

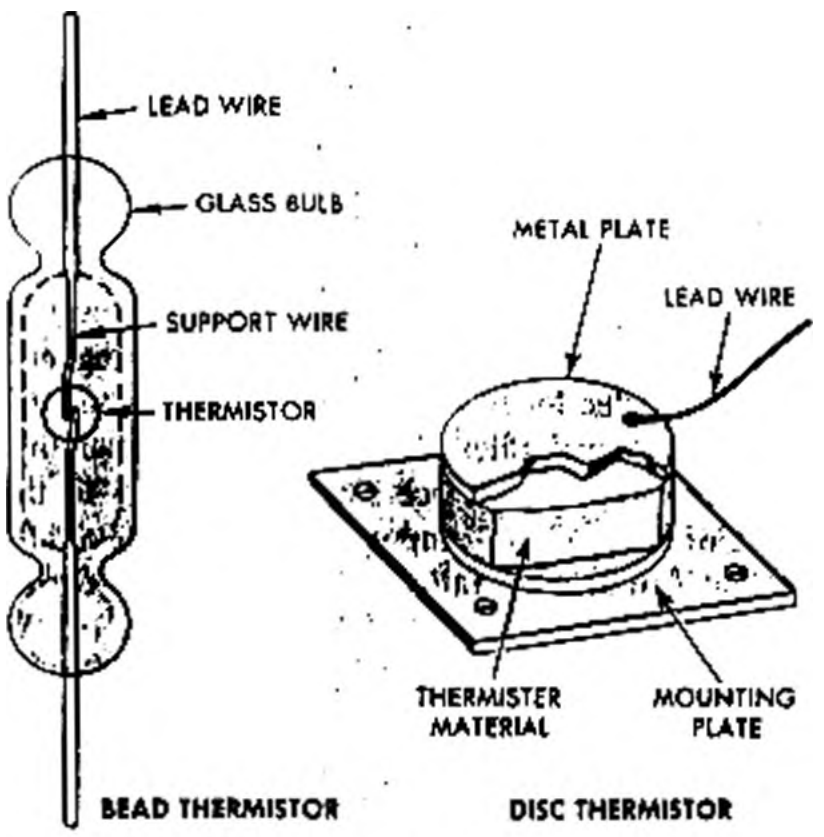


Figure 6-28. Thermistors

environment, by AC or DC flowing through it, or by absorbed RF power.

Two types of thermistors are illustrated in figure 6-28. The bead thermistor used for power measurement has a small mass and is affected by the changes mentioned in the previous paragraph. The disc thermistor, used in compensating networks, has a large mass, and its resistance is relatively unaffected by the current flowing through it. Its resistance depends primarily on the ambient temperature.

In the case of these thermistors, the bead thermistor could possibly be used in RF systems as a sensing-element pickoff unit as well as in a control system. The disc thermistor would find its primary use in control systems.

This common use of units in both the guidance and control systems brings to light another situation that is common when dealing with missile systems; that is, the line of demarcation between the control function and the guidance function is often narrowly and arbitrarily drawn. The two functions are always dependent on each other and closely integrated, and the problems that arise in one may not be analyzed properly without due consideration to the other. Guidance and control must function as a unit to have a satisfactory missile flight.

Commutation Principles

In conjunction with many of the sensing-element pickoffs, there is a device which connects the information channels in the circuit to the receiver at periodic intervals for relatively slow cycle information. In any guidance system numerous information signals as to range and altitude are being received, and the guidance system could not distinguish between the various received signals without first segregating them before passing them on to the remaining amplifying stages. Since the order in which the channels are sampled is known, a commutator can be used to separate the information so that it may be interpreted by the receiver.

The commutator in this case is the stationary member while the brushes are the rotating part. Information is fed into the current-carrying segments of the commutator as shown in figure 6-29 and then picked off by the rotating brushes. This gives a means of systematically measuring and transmitting the measured information so that various factors may be indirectly observed by a ground-control operator.

A synchronizing pulse usually indicates to the receiver the starting and stopping point of a cycle. Commutators are usually driven by a geared-down dynamotor for a smooth operating system. Gearing the commutator means voltage changes felt by the dynamotor are not felt as strongly by the commutator. For example, if a dynamotor is operating at 9000 rpm and the commutator is geared for 300 rpm operation, the voltage changes appearing at the dynamotor are proportionally reduced at the commutator.

The pickoffs discussed in this chapter are just representative. There are other types of pickoffs, and some of the ones mentioned here with respect to a particular guidance system may be used on other guidance systems. The major consideration in choosing a sensing-element pickoff is whether or not it is electrically applicable to the system.

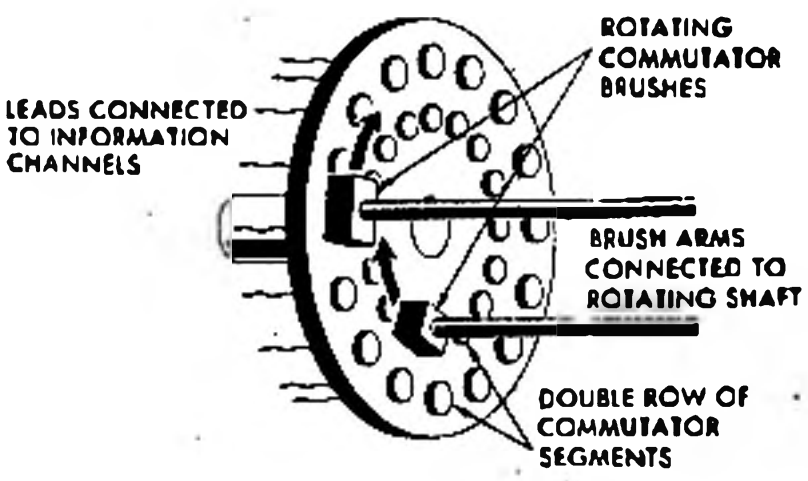


Figure 6-29. Operation of Commutator

So far, various means of sensing an incoming signal and relaying that signal to other circuitry have been presented. These units that have been presented generally are not used solely for receiving an energy signal. They also are used to transmit a signal. The basic units used for radiating energy would have similar devices used to give off information or guidance signals.

To give off or receive a signal by one of the particular guidance systems, there must be a systematic way of determining the relative position of the missile with respect to the target and/or the base control station. The way in which the sensor "looks" for its position is referred to as scanning.

SCANNING METHODS

Effective search for a target or a missile that is to be guided can be made only when the whole area of interest is scanned without gaps. Therefore, the search must be rapidly carried out in some methodical manner. A radar system may have a line-of-sight path and a maximum range far exceeding the range of a missile and yet be useless if it fails to pick up, track, and guide the missile early in its trajectory.

The scanning procedures discussed here are not limited to radar systems alone. Scanning procedures may also be employed in infrared, light, or any other system that focuses a beam of energy.

A spiral or sawtooth scan, illustrated in figures 6-30 and 6-31, can be used to search the volume of space. A beam which does not fill either dimension of the solid angle covers the volume of space.

A common type of scanning used in beam-rider systems is that of conical scanning, shown in figure 6-32. Conical scanning occurs when a beam is slightly off center and rotates in a manner which generates an elongated football-like volume. The signals are strongest within the generated volume. The longitudinal axis of the beam passes through the radar antenna and the target. When a missile travels along the axis of the "football," all signals from the rotating beam are neutralized in the de-

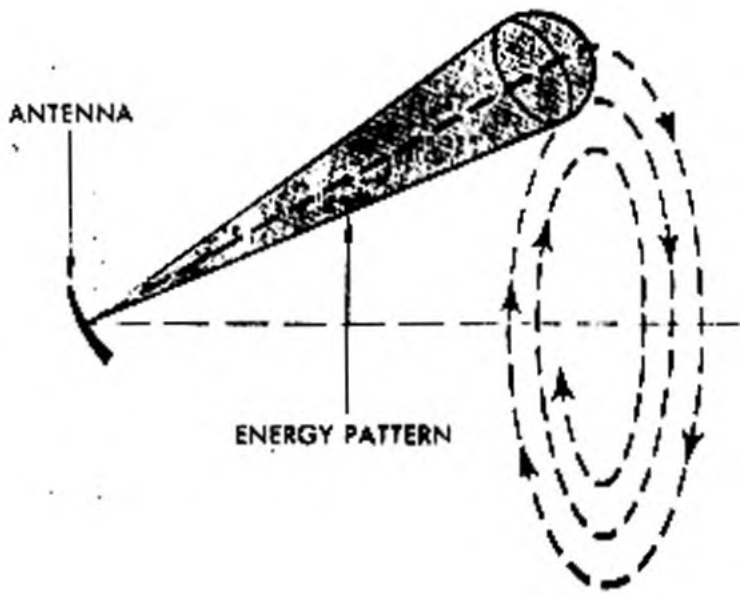


Figure 6-30. Spiral Scanning

tector carried within the missile. Should the missile get off the axis of the beam, the signals on one side would become stronger in proportion to the other side. This condition activates a mechanism to steer the missile back on the axis.

To obtain 360° of azimuth search, the helical-type scan shown in figure 6-33 is used. In the helical scan, the antenna tilts one-half beamwidth in elevation for each complete revolution in azimuth. Normally, the scan is not carried above 30° in the vertical direction because a large portion of the search time would be spent in this region. The region represents a small portion of the volume of space in which objects may be found.

To insure solid coverage, the separation between turns of the scanning cycle should be approxi-

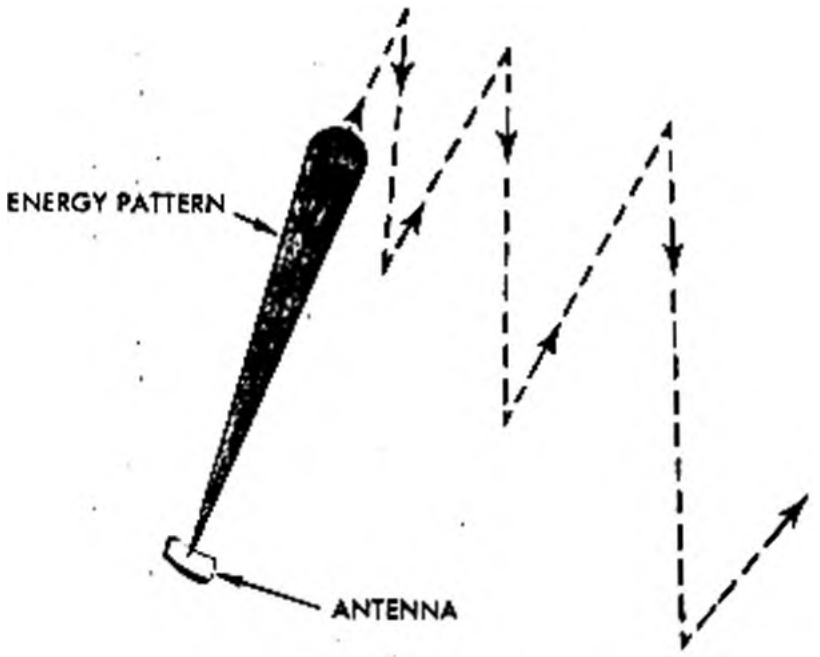


Figure 6-31. Sawtooth Scanning

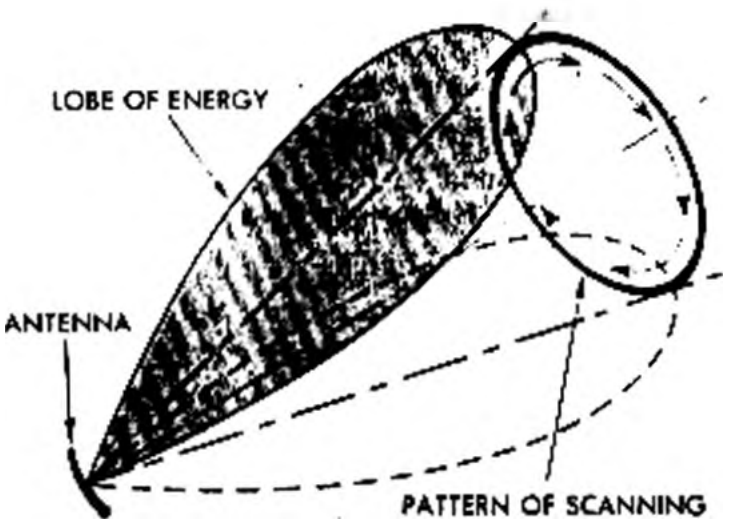


Figure 6-32. Conical Scanning

mately half of the beamwidth. This means that each succeeding scanning cycle will overlap the previous one. In the case of the spiral scan, the radial motion must have a fixed relationship to the circular rate. The horizontal motion of the sawtooth must be related properly to the vertical scan speed. The main factors to consider in a scanning process are as follows:

1. The minimum scan period required for a given solid angle of search at the maximum range.

2. The maximum angular velocity at which the beam will successfully search at the maximum range.

3. The distance that a missile flying away at a certain speed along the line-of-sight of the radar will travel between two successive scans during search at maximum range.

Another means of positioning a missile with respect to a transmitter is by lobe switching. In this system, two overlapping beams are used to increase the angular deviation of the beam. (Note figure 6-34.) The lobing is accomplished by using two different antenna systems or by the use of one antenna system which is made to oscillate by mechanical or electrical means.

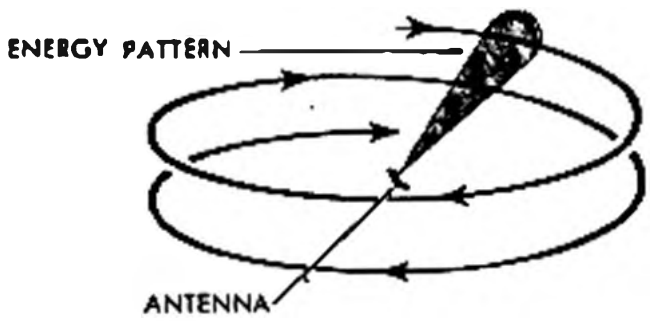


Figure 6-33. Helical Scanning

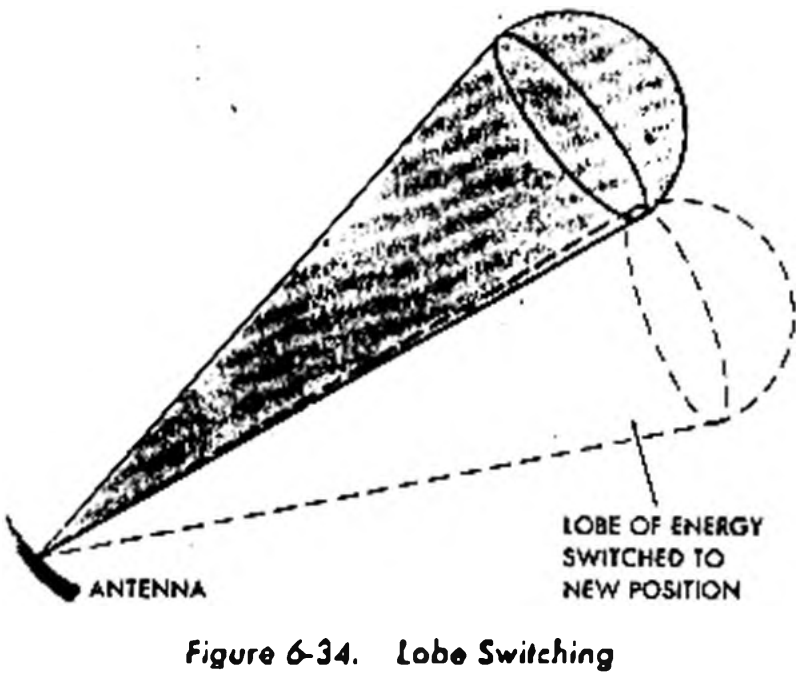


Figure 6-34. Lobe Switching

Various devices are used to obtain the desired type of scan for the guidance system. In the radar system, the lobing is accomplished by shaping the reflector and then causing it to rotate about the fixed antenna. A similar pattern can be attained by keeping the reflector stationary and causing the radiator to rotate. Offset antenna patterns can be attained also by the lobe-switching method.

For the infrared and light-seeking devices, vibrating mirrors are used to scan the area. As a mirror rotates or vibrates, it causes the focal point to shift about the sensing element, thus giving an indication of the point where the target signals are the greatest. Such a mirror is shown in figure 6-35.

For star-tracking systems and some of the light-seeking systems, split or slotted discs are used to interrupt the light at a definite rate. The use of a split disc fixes a definite area to be scanned by the system and permits time for the circuit to recover from radiations that were scanned on the previous rotation. Pictured in figure 6-36 are various types of slotted discs used in missile systems.



Figure 6-35. Vibrating Mirror used for Scanning

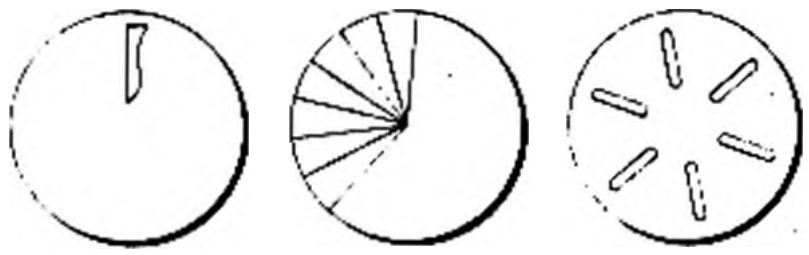


Figure 6-36. Slotted Discs in Star-Tracking and Light-Seeking Systems

POLARIZATION METHODS

Polarization of antennas is important in missile guidance systems. As stated earlier, polarization has to do with the direction of the electric field. Electromagnetic waves are made up of two fields: the electrostatic and magnetic fields. The electrostatic and magnetic lines of force travel at right angles to each other in a plane that is moving in a direction perpendicular to the lines of force. The arrows in figure 6-37 indicate an instantaneous direction of the fields of a vertically polarized wave. The direction of travel of the fields is into the page. If either of the fields were to be reversed, the direction of travel would be reversed. When both fields reverse, which occurs every half-cycle, the direction of travel remains the same.

The electromagnetic wave is vertically polarized because the electrostatic lines of force are perpendicular to the earth. The wave would be horizontally polarized if these lines of force were parallel to the earth. These two general types of polarization are referred to as linear polarization.

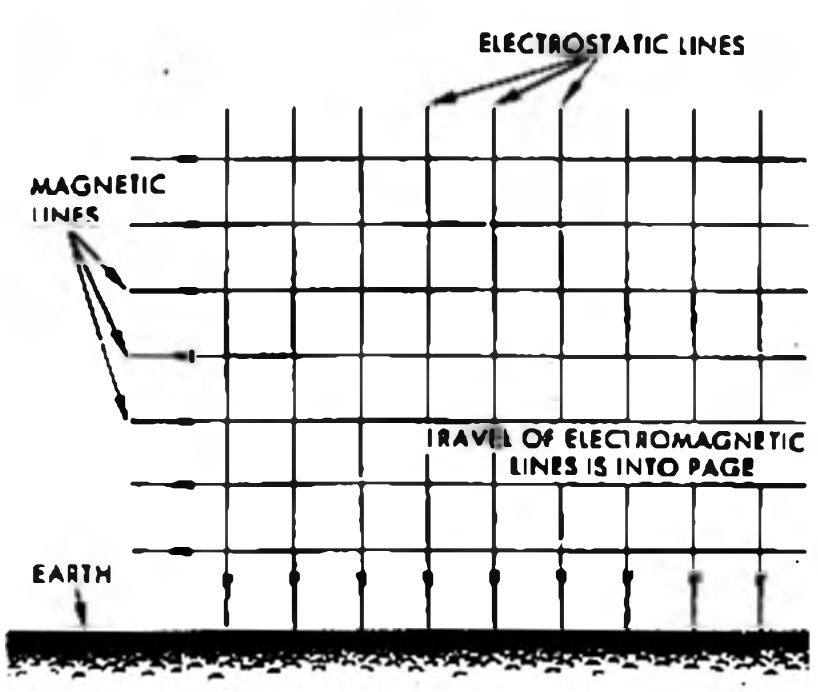


Figure 6-37. Instantaneous Direction of Electrostatic and Magnetic Fields of Vertically Polarized Wave

The electrostatic field of any wave tends to change polarization. This change is small at the lower frequencies but tends to be more pronounced as the frequency is increased. Since it does tend to change its polarization, it takes on a rotational characteristic of either sense. Due to phasing of portions of the waves when the field encounters a target, the sense of rotation may be in both directions. This is particularly the case when it encounters an asymmetrical target. The phasing in this case would tend to give an elongated field of electrostatic energy. The wave would be elliptically polarized.

The field becomes circularly polarized when the phase difference of the electrostatic fields is 90°. The field of electrostatic energy is circular in nature and may contain either directional sense.

When the polarization of a field is changing, it is difficult to detect a maximum of usable signal. If a device is used in which reception of either directional sense can be detected, the signal definition may be greatly improved. The general case in radar systems is that the symmetrical areas of scan will return a signal circularly polarized, while an asymmetrical area will return an elliptically polarized signal. The picture seen on a radarscope is greatly improved if both senses of field rotation are detected.

Much research is being done in the field of antenna polarization. Development of predictable wavelength sections (rectangular and circular) and phasing devices are under continuous study.

SECTION B
Computer Units

Computer units in missile guidance systems are found in various forms. A computer unit may be simple, such as a mixing circuit in an airborne vehicle, or it may be complex, as in long range inertial guided missiles. Computers are also used in ground installations where the complete flight of the missile is determined. The many computers used in the moon flights were seen by television viewers.

An important function of the computer unit in many guidance systems is *coding and decoding information* related to the missile flight. Enemy countermeasures or the guidance of more than one missile at a time may make this function necessary. Discriminating circuits select pulses of particular width, amplitude, frequency, phase, and/or time difference and reject all others. Noise interference is reduced to a minimum by these decoding circuits so as to develop reliable flight signals.

Another function of computer units is *mixing signals* from sensor and reference units to produce error voltages. These signals are sometimes mixed in a certain ratio or according to programmed instruction. Error voltages are fed to the amplifier unit which increases their magnitude. Then the error voltages are introduced into the control system to correct any deviation from the computed flight path. The amplified error voltages also are passed through the followup unit to be reprocessed in the computer.

Components in some guidance computer units compare two or more voltages to produce error signals. These components may be in the form of voltage or phase comparator circuits. The autosyn units discussed in the preceding chapter are used in guidance computer units to resolve signal voltages into components for processing. In some guidance systems, autosyn resolver circuits convert earth-reference coordinates to space-reference coordinates.

Airborne computers may be classified in terms of phases of the missile flight. They may be in the form of distinct units or combinations of the following computer units:

- Prelaunch computer.
- Launch computer.
- Azimuth computer.
- Elevation computer.
- Program computer.
- Dive-angle computer.
- Ballistic computer.

Another method of classifying computer units is to consider the operating principle. Both airborne and ground-based units use two classes: analog and digital. Their major differences are presented briefly in this introduction. The two operating principles involved in their individual component operation are discussed later in the chapter.

An analog computer manipulates physical quantities that represent the mathematical variables of the particular problem under study. In the mechanical type of analog computer, for example, the machine variables may be rotating shafts driven

by gear trains. The angular displacement of the shafts is measured to produce the solution to the equation or mathematical operation. Computers for solving navigation and bombing problems have used these principles for years. Early radar gun directors also made use of this type of operation.

Differential analyzers in both mechanical and electronic form have been applied to solve problems in the missile field. The first large-scale differential analyzer was operated at M.I.T. in 1930. This original machine used mechanical principles. Current types of electronic differential analyzers are used in almost every phase of missile dynamics to obtain information without constructing actual prototypes of the vehicle. Although the differential analyzer is often called an analog computer or simulator, the computing elements are not usually direct analogs of the physical quantities in the problem, but are analogs of the mathematical equations describing the problem.

A digital computer performs the mathematical operations with numbers expressed in the form of digits. The machine essentially is composed of counters which register and add in discrete steps. Desk calculating machines are familiar forms of the mechanical type of digital computer. The first automatic sequence-controlled calculator used mechanical counters controlled by magnetic clutches and relays. Commercial construction of this type of automatic calculator began in 1939.

Electronic digital computers had their start in the electronic numerical integrator and calculator. This machine was the first to use electronic circuits as the actual computing elements. It was designed primarily for step-by-step numerical integration of the equations of external ballistics. Although this computer used a scale of ten, the airborne digital computers used in missiles operate on the binary system of numbers. The binary system is presented later in this chapter.

CODING AND DECODING UNITS

Familiar examples of coding and decoding systems include teletype, transoceanic telephony, television, identification of friend or foe (IFF), radio, and radar beacons. All of these systems use a synchronizing key or code at the receiving end to properly match the coded transmission. This matching is necessary to extract the intelligence.

Types of coded transmission are described in chapter 4. Each type of pulse modulation—amplitude, width, rate, position, frequency, and count—requires a decoding circuit that reproduces the original signal and keeps interference and noise at a minimum. The decoding circuits described here are examples of those now being used in the missile guidance computer units.

The three systems described in the following paragraphs are used for decoding proportional information for lateral control of the missile and ON-OFF control for dump signals. Information is transmitted by discrete or continuous changes in pulse repetition rate, frequency modulation of the pulse repetition rate, and pulse time modulation.

In the pulse-repetition rate system (PRR) shown in figure 6-38, the coder determines the repetition rate of the transmission of radar pulses. For proportional control, the repetition rate varies from 1800 to 2000 pulses per second according to the DC error-voltage input. The on-off channel is operated by a preselected discrete pulse-repetition frequency (PRF) of 1500 cycles per second.

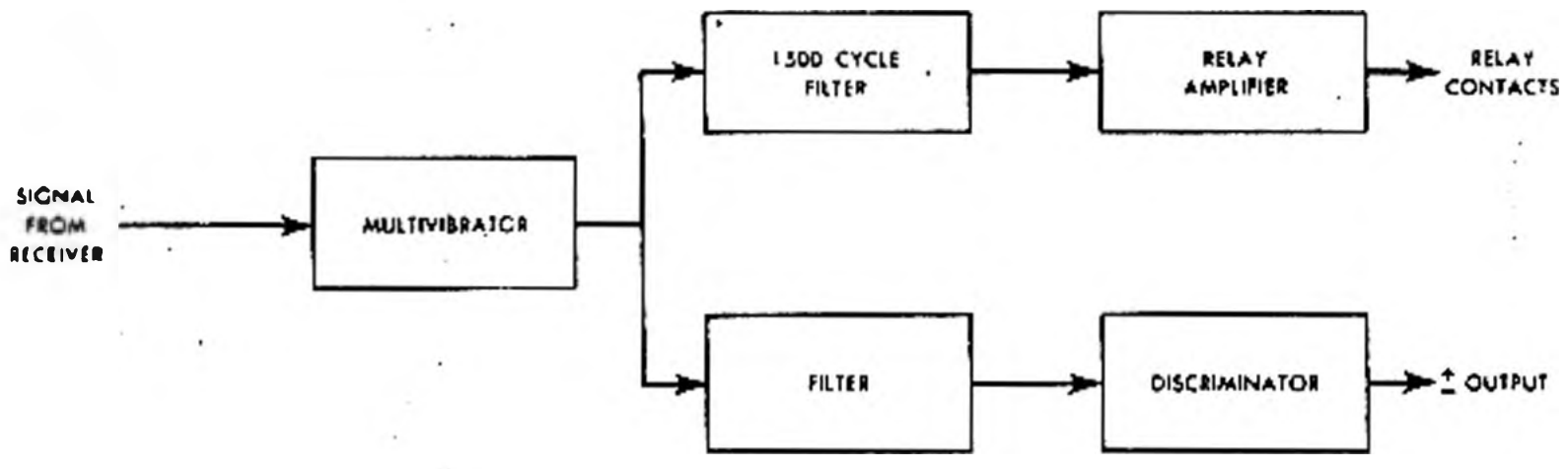


Figure 6-38. Decoder for PRR Systems

In a system using frequency modulation of the pulse-repetition rate, referred to as pulse-frequency modulation (PFM), a subcarrier frequency modulates the radar repetition rate. Conventional frequency-modulation circuits are used for detecting the command information. Frequency instability is a problem in this system.

In the pulse-time modulation (PTM) system, illustrated in figure 6-39, three pulses are transmitted by radar. The first two are coded for beacon response. The third pulse is shifted in time according to intelligence being transmitted. At the receiver, the first two pulses are decoded and used as a reference and for beacon triggering. A radar is required that will supply three pulses, with the third pulse variable in time to provide the information for proportional control. This system is used in one type of a command guidance set.

The PFM and PTM systems have about the same reliability and simplicity and are more critical in operation than the PRR system. The PRR system is reliable and simple. It requires no adjustment in the airborne equipment and only two adjustments in the ground equipment.

Pulse-Width Discriminator (PWD)

A discriminator circuit which passes pulses of greater width than a certain predetermined width and rejects all narrower pulses is shown in the block diagram of figure 6-40. Preceding circuitry develops input pulses of constant amplitude but varying widths. There are output pulses from the decoder only if the input pulses exceed a certain preset value which is adjusted by a potentiometer setting. Noise and interfering pulses of narrower widths are rejected and produce no output.

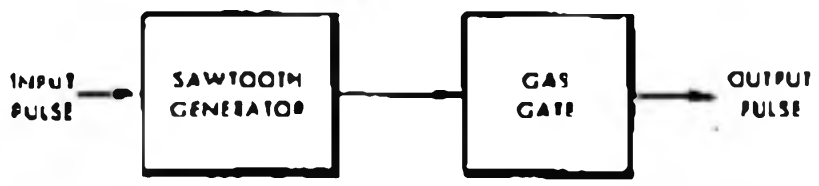


Figure 6-40. Pulse-Width Decoder

The stages shown in the schematic in figure 6-41 consist of a sawtooth generator (V_1) and a gas gate (V_2). Tube V_1 is normally conducting with a resulting low plate voltage. Negative-going input pulses of sufficient amplitude to cut off V_1 would normally allow the plate voltage to rise to B + and to develop an amplified and inverted version of the input signals in the plate circuit. The integrating condenser (C_2) in the plate circuit of V_1 develops a sawtooth output. The amplitude of the sawtooth output increases with the increase in pulse width.

Tube V_2 is biased below cutoff by R_8 negative voltage. The setting of this potentiometer determines the amplitude of the positive-going sawtooth signal which allows V_2 to conduct. This gas tube ionizes when the grid signal sufficiently decreases the negative voltage on the grid of V_2. The tube V_2 will conduct through the output cathode resistor (R_7) until the plate voltage decreases to the point at which the tube deionizes. The output pulse from the cathode of V_2 is delayed with respect to the input pulse as shown by the waveshapes. This delay is caused by the time required for the sawtooth signal to decrease the negative voltage on the grid of V_2 enough to allow conduction in the tube. Potentiometer tap R_4 can be adjusted for a smaller

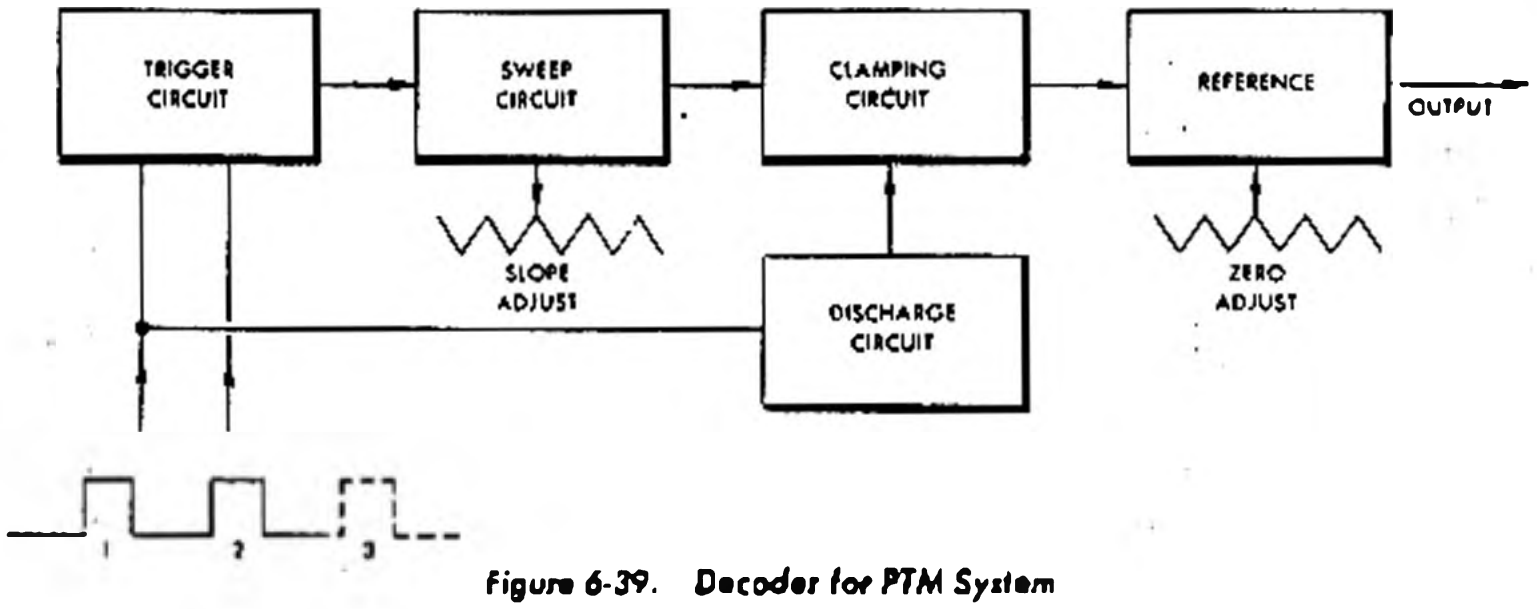


Figure 6-39. Decoder for PTM System

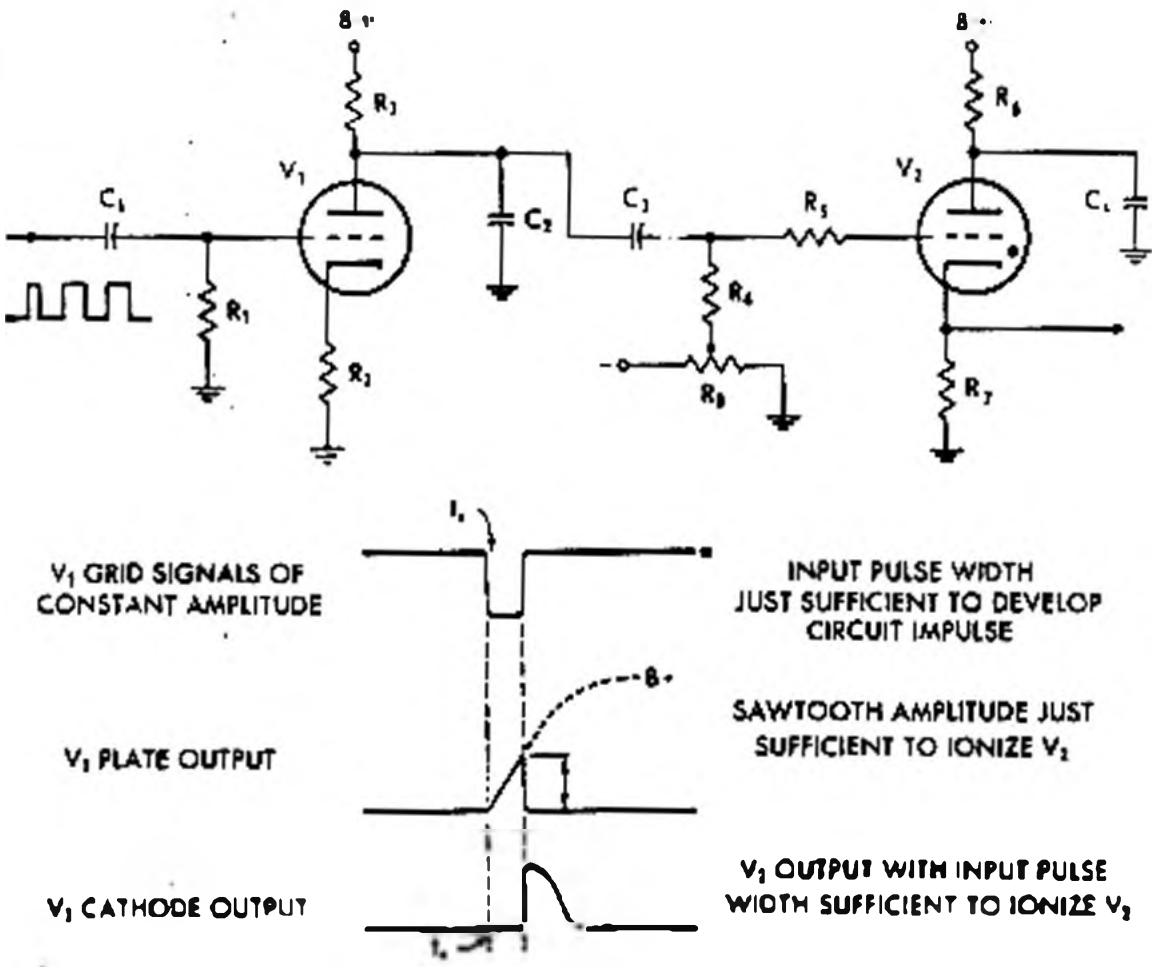


Figure 6-41. Sawtooth Generator and Gas Gate

or greater negative voltage to require a different pulse width to trigger V_2.

PWD USING COINCIDENCE TUBE. The schematic in figure 6-42 shows a pulse-width discriminator circuit using a coincidence tube. To make the circuit operative, the pulse that is being selected must be changed into a sawtooth waveshape. These waveshapes are illustrated as waveshapes A and B. This change would be performed by a sawtooth generator stage before the input of this circuit.

Note that all of the tubes are biased below cutoff by means of a negative-bias supply voltage. The bias on tubes V_1 and V_2 can be varied by moving the wiper arms on potentiometers R_1 and R_4, respectively.

Assume that this circuit is to be adjusted to pass pulses of 1.0 ± 0.1 microsecond duration and reject all others. This means that 0.9 microsecond is the minimum and 1.1 microsecond is the maximum pluse width to be passed. Tube V_1 is used to determine the minimum pulse width. The bias on V_1 is adjusted so that it will begin conducting when the signal voltage (E_1) on its control grid is equivalent to the value at 0.9 microsecond after the start of the sawtooth wave. (See waveshape B.)

A surge of current through the RF choke L_1 would be shunted by R_{8}, causing a damped oscillation or ringing effect at the natural frequency of the choke plus the capacity of the circuit. The small resistor across the choke reduces the oscillation to only one complete cycle. (See waveshape C.) The signal on the control grid of V_2 is one sine wave, first negative and then positive, starting in coincidence with the time at which the negative-bias voltage is overcome by the positive sawtooth wave. If the pulse is too narrow to cause the sawtooth to reach an amplitude great enough to overcome this bias. V_1 will not be triggered and will thus discriminate against that particular pulse.

Tube V_2 is being used as the maximum pulse-width discriminator. Again, as in the case of V_1, the negative bias has to be set by moving the wiper arm of R_4, but in this case the setting will be a different voltage. Resistor R_4 must be set so that V_2 conducts at the sawtooth voltage (E_2) equivalent to 1.1 microsecond. (See waveshape

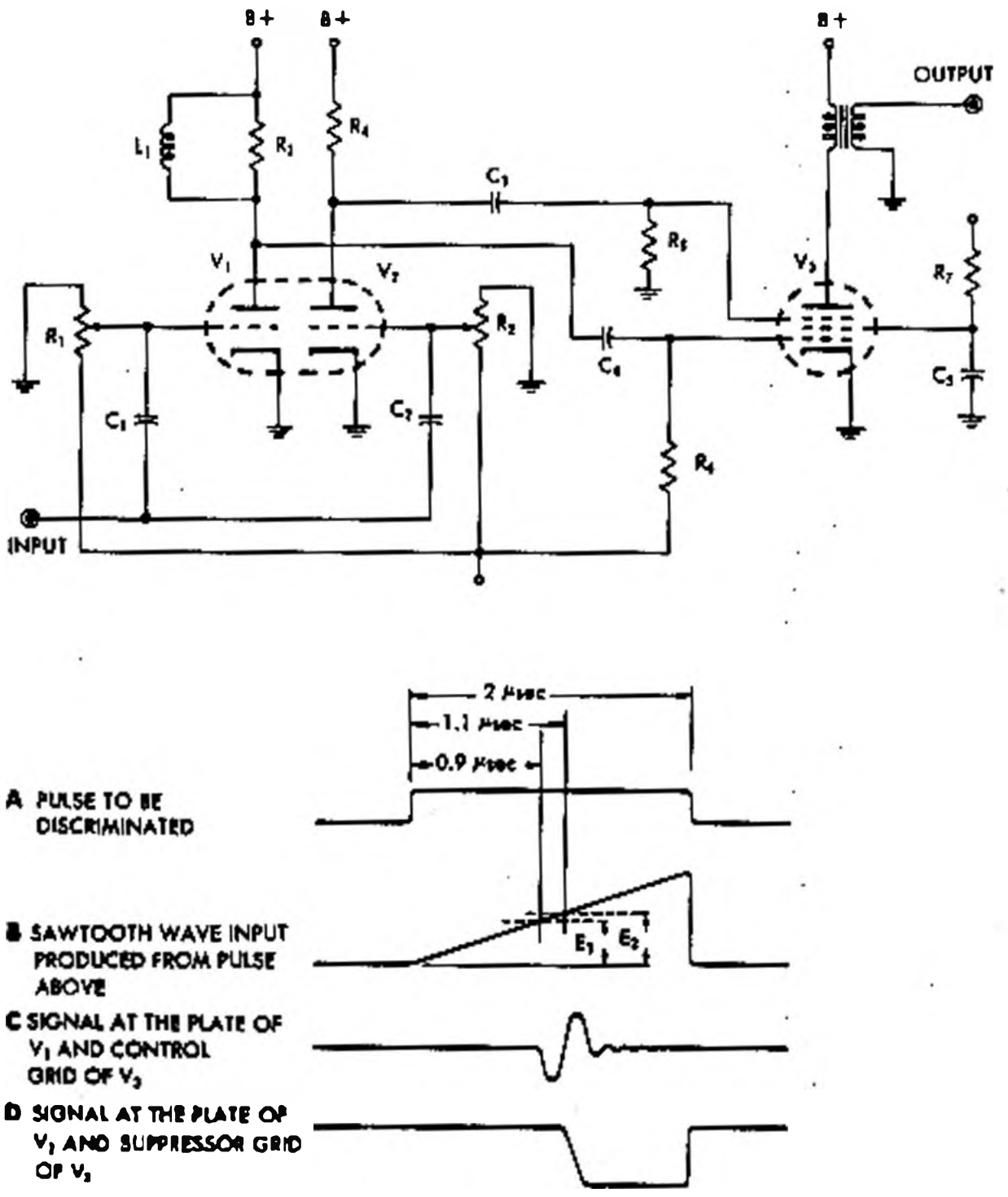


Figure 6-42. Pulse-Width Discriminator Circuit using Coincidence Tube

B.) The plate load (R_4) is very large. This causes V_2 to be easily overdriven (saturated), and the signal on the plate approximates a square wave. (See waveshape D.) This signal is fed to the suppressor grid of V_3, and it is large enough in amplitude to keep the tube from conducting, even though the control grid might be above cutoff voltage. If there is a pulse that is too wide, both V_1 and V_2 will conduct, and both signals will show in coincidence on the grids of V_3. The time relationship between the two signals, if they are present, is shown in waveshapes C and D. Tube V_3 is also biased below cutoff by means of the negative-bias supply. Since the sine wave first goes negative, the signal does not put the tube into conduction until the positive half-cycle is reached. If the pulse is too wide, the negative gate appears on the suppressor grid at the same time that the positive half-cycle of the sine wave appears on the control grid. The negative gate keeps current from flowing in the plate circuit, while the positive half-cycle takes the control grid above cutoff. This means that the screen grid will draw current and there will be no output signal from V_3.

When a pulse of the desired pulse width comes in, V_1 conducts and V_2 will not conduct. The only signal to V_3 is the sine wave on the control

grid, which produces one surge of current in the plate circuit during the positive half-cycle of the sine wave.

If a change in the pulse shape is desired, a blocking oscillator or a one-shot multivibrator could be used to give a square pulse.

Because of the critical operating points of the grids of V_1 and V_2, a stabilized negative bias supply is required to give accurate pulse-width discrimination.

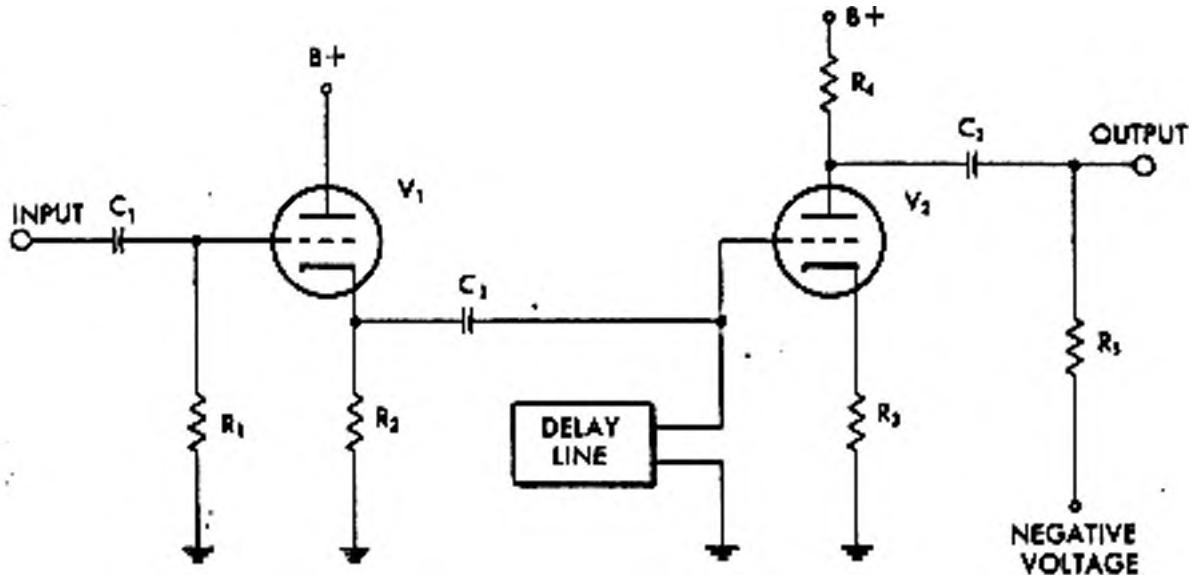
PWD USING DELAY LINE. Another pulsewidth discriminator employs the use of a delay line, as shown in figure 6-43. As before, assume that the circuit is set to pass 1.0 microsecond pulses and no others.

The network C_1 and R_1 differentiates the input pulse, forming a positive peak on the leading edge and a negative peak on the trailing edge. Note waveshapes A and B of the illustration. These two pips (the positive and negative peaks) are separated by a time lapse equal to the duration of the pluse. Tube V_1 acts as a cathode follower; the pips appear across R_2 and are coupled to a delay line and the control grid of V_2 by condenser C_3.

The delay line is grounded on the other end so that a negative reflection of the positive pip appears 1.0 microsecond after the pip appears at the grid of V_2. (The positive pip travels for 0.5 microsecond to the shorted end of the delay line where it is inverted, and in 0.5 microsecond it travels back to the input end of the line.) If the reflection coincides with the negative pip from the trailing edge of the input pulse (see waveshape C), a large-amplitude negative pip results from their summation (waveshape D).

Tube V_2 inverts this signal, and the signal is then coupled to the next stage by means of C_4 and R_5. Resistor R_3 is connected to a negative voltage which biases the following stage below cutoff.

The positive signals at the grid of V_2 would result in negative pips at the plate of V_2. These could not make the following stage conduct. The



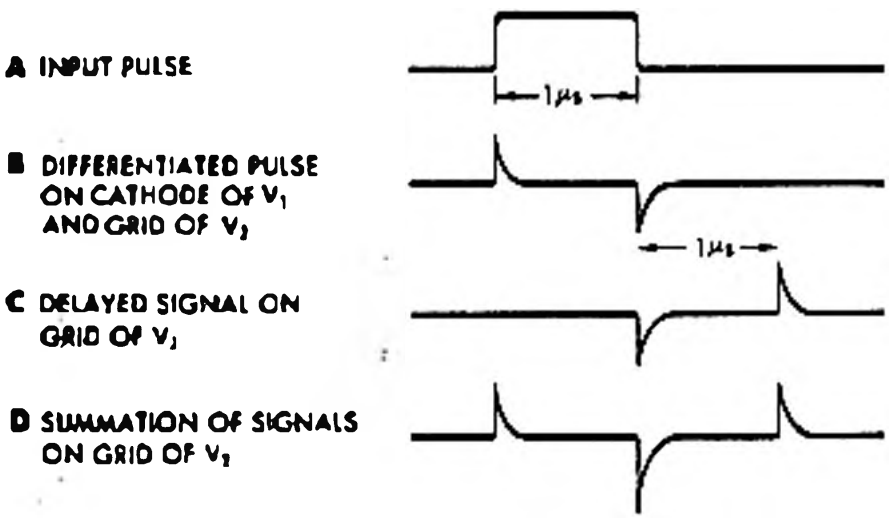


Figure 6-43. Pulse-Width Discriminator using Delay Line

negative pips at the grid of V_2 are added to each other, resulting in positive pips at the plate of V_2. The pips are then of sufficient amplitude to make the following stage conduct.

This following stage could be part of a pulse-forming circuit which is heavily biased to eliminate reaction to all but the large signal developed above.

Pulse-Spacing Discriminators

Two types of pulse-spacing discriminators are presented here. Basically, the first type consists of a delay line and a coincidence circuit, and the second consists of a multivibrator, a blocking oscillator, and coincidence tubes. The applications of these two circuits are presented after the circuitry has been examined.

The pulse-spacing discriminator shown in figure 6-44 makes use of a delay line and a coincidence circuit. The cathode follower stage, tube V_1, is necessary to maintain a constant impedance to the delay line. The leading edges of pulses to be passed are spaced 3.5 microseconds apart, and the pulses themselves are 0.5 microsecond wide. Note waveshape A in the illustration.

The signal is fed to the control grid of V_2, the first tube in the coincidence circuit, and it

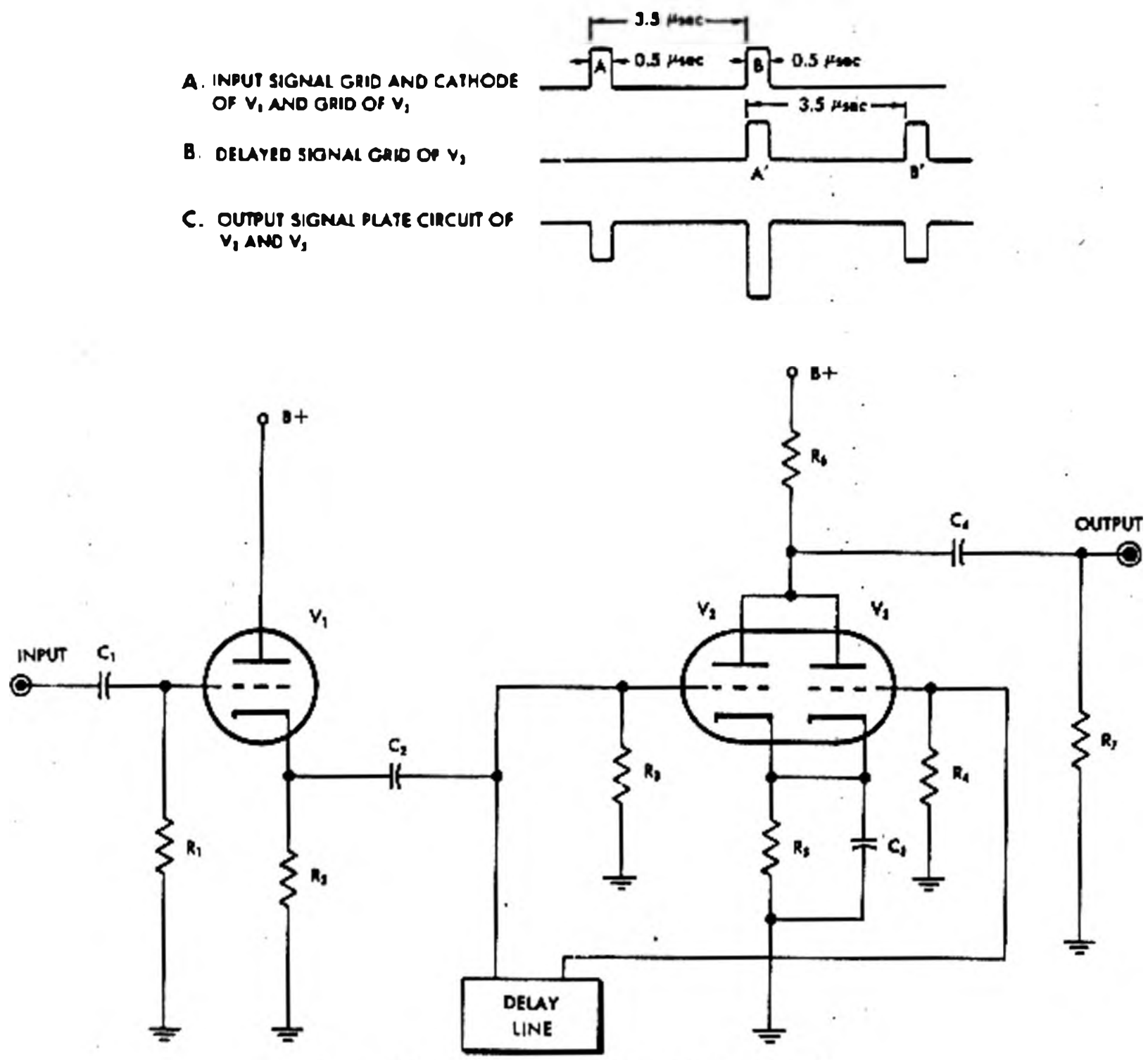
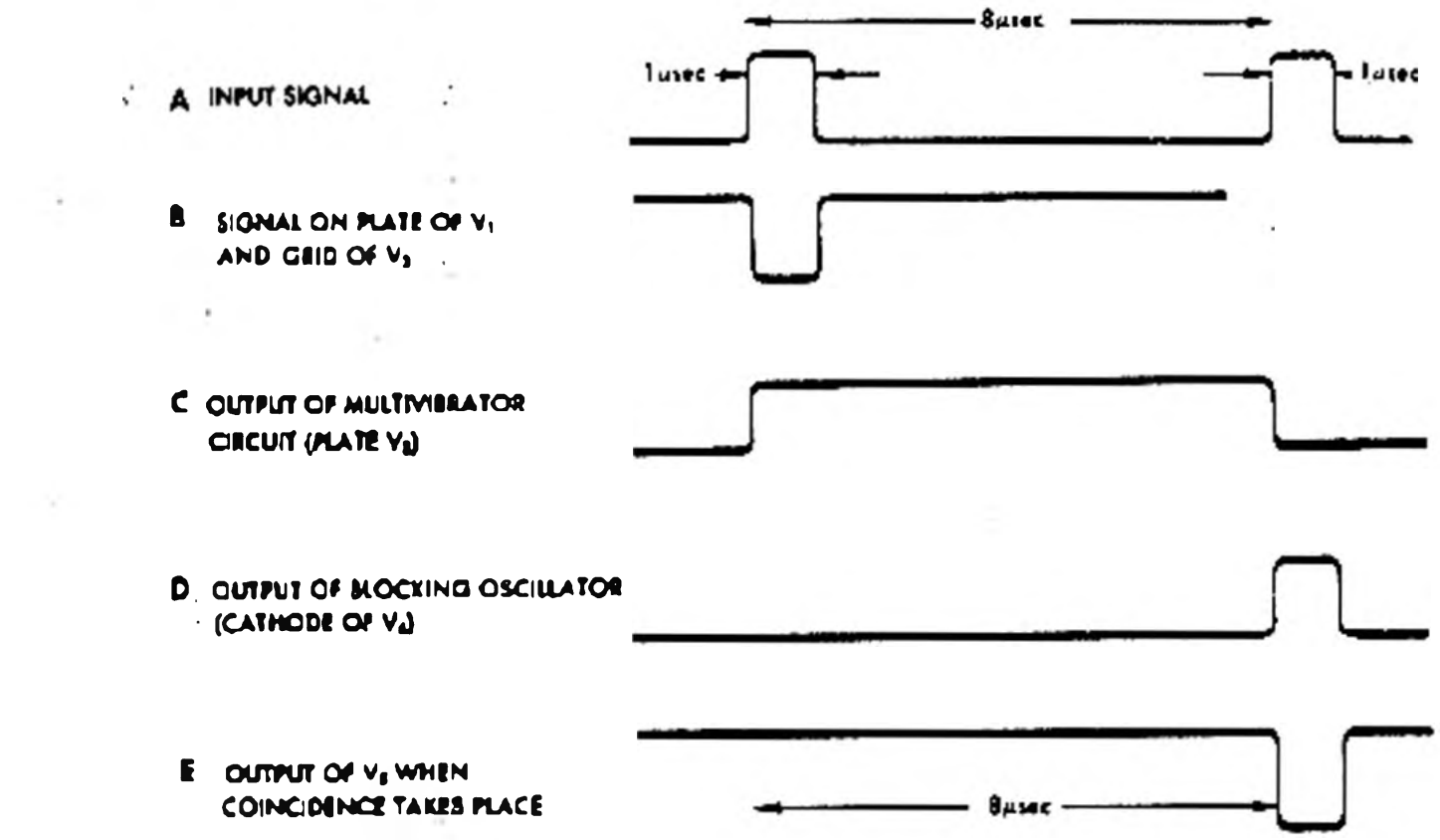


Figure 6-44. Pulse-Spacing Discriminator-Type 1



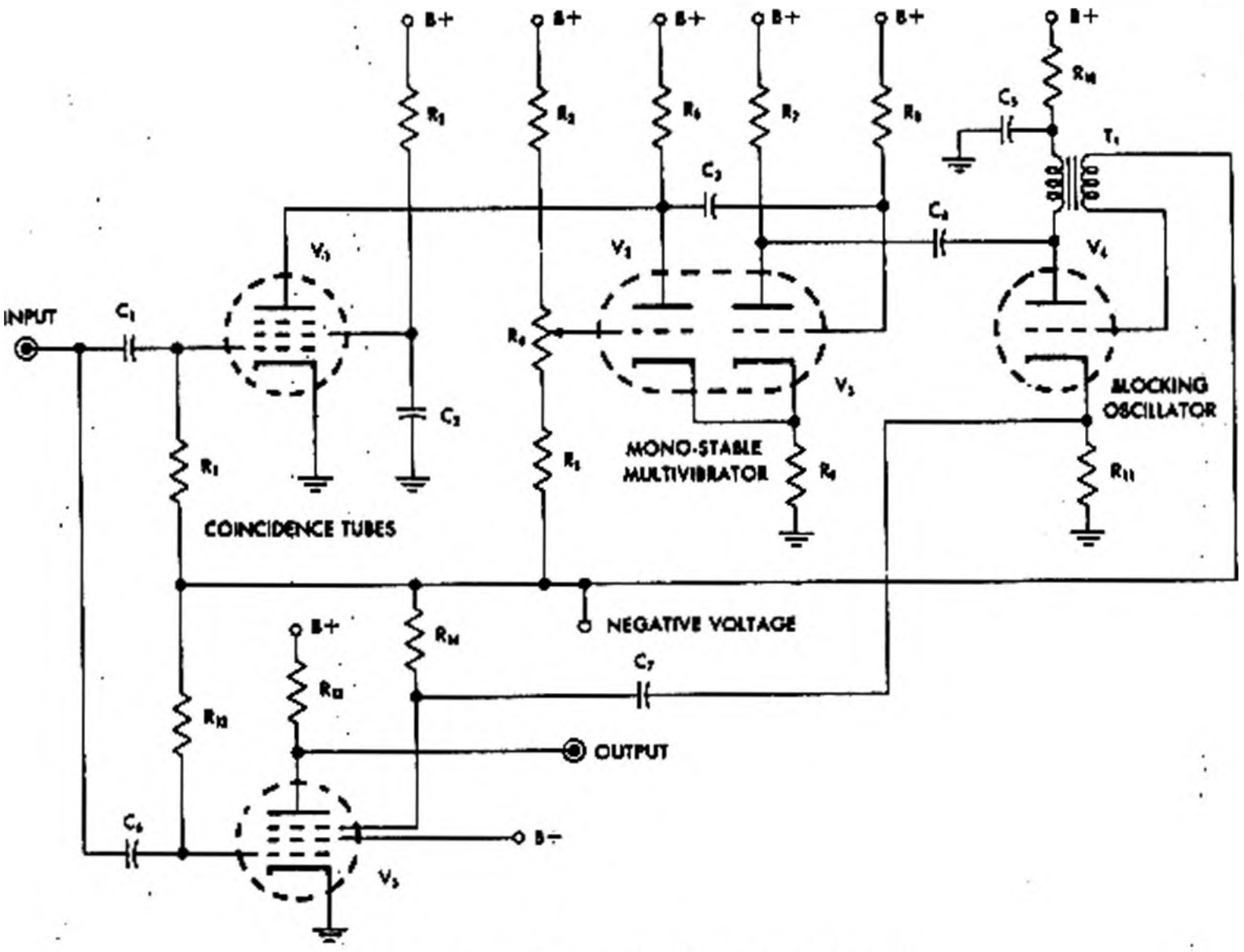


Figure 6-45. Pulse-Spacing Discriminator-Type 2

also is fed to the delay line. The delay line is tapped in such a manner that the signal is delayed by an amount exactly equal to the spacing of the two pulses (waveshape B) and then fed to the control grid of V_3, the second tube in the coincidence circuit.

Both of these signals change the current flowing through the plate load resistor R_6. The first pulse (A) on the grid of V_2 causes the plate voltage to drop, giving a negative pulse at the plate. The second pulse (B) on the grid of V_2 also causes a negative pulse at the plate but, at exactly the same time, the first pulse (A') of the delay signal appears on the grid of V_3, causing a further drop in plate voltage (waveshape C). When the second pulse (B') of the delayed signal appears on the grid of V_3, it causes a negative pulse at the plate. This pulse is the same amplitude as the first pulse A on the grid of V_2.

The accuracy of this circuit is good, being approximately ± 0.05 microsecond.

This signal on the plate of the coincidence circuit can be used to trigger a number of different circuits to indicate that pulses of that spacing have been received at the input. The signal could be used as a positive trigger by inverting it through an amplifier. Any circuit triggered by the signal would have to be biased so that only a signal composed of two pulses appearing simultaneously on the grids of V_2 and V_3 would have sufficient amplitude to trigger that circuit. Therefore a pulse applied to only one grid (V_2 or V_3) will not yield an output to trigger the next circuit. The pulses which are applied to the discriminator must be uniform in amplitude. This uniformity of amplitude can be accomplished with a limiter stage.

The second type of a pulse-spacing discriminator is shown in figure 6-45. Tubes V_1 and V_4 are coincidence tubes, and tubes V_2 and V_3 compose the multivibrator circuit which is designed to pass pulses which are spaced 8 microseconds apart. The accuracy of this circuit is not as good as the accuracy of the other pulse-spacing discriminator.

The output of the multivibrator is taken off the plate of V_3 and is a positive gate (waveshape C). The width or duration of the gate can be varied by potentiometer R_4.

The gate is differentiated by coupling capacitor C_4 and the low-impedance winding of the blocking oscillator transformer. The differentiation of the trailing edge results in a negative pip which triggers the blocking oscillator. The signal, a square pulse (waveshape D) is taken off the cathode of V_4 and coupled to the suppressor grid of V_5. At the same instant, the second pulse of the input signal is appearing on the control grid of V_5. Tube V_5 conducts and a negative pip appears on the plate (waveshape E). This negative pip is the output signal, and it indicates that two pulses of the proper spacing have been received. Both the control and suppressor grids of V_5 are biased below cutoff. There will be no current in the plate circuit of V_5 unless there is coincidence between a signal on the control grid and a signal on the suppressor grid.

Pulse-Recurrence Frequency Discriminator

The circuit of a pulse-recurrence frequency discriminator functions as a decoder by accepting only a limited band of pulse-recurrence frequencies (PRF) as synchronizing signals. This band of PRF can be preset by a potentiometer adjustment. Pulses at recurrence frequencies outside of this limited band will not produce a synchronized output from the discriminator and can be effectively gated out of succeeding circuitry by delay-line and coincidence-tube combinations.

Tubes V_1 and V_2 of figure 6-46 and their circuit components comprise a free-running multivibrator. Positive-going input pulses within the synchronizing range produce negative-going output pulses of the same recurrence frequency. The discriminator output is used to trigger the following circuit, such as a blocking oscillator.

The setting of poteniometer R_3 in the plate circuits of V_2 determines the band of PRF desired to synchronize the discriminator. As the tap on R_3 is raised toward B+, decreasing the resistance through which C_3 discharges, the band of PRF that will synchronize the circuit is raised in frequency. Choice of component sizes determines the range of the frequencies. At the higher-frequency bandwidth setting, the range of synchronizing frequencies is also increased.

The switch (S_1) in the grid circuit of V_2, when closed, decreases the bandwidth from two to three times. Closing the switch makes the discriminator more selective to variations in the pulse-recurrence frequencies.

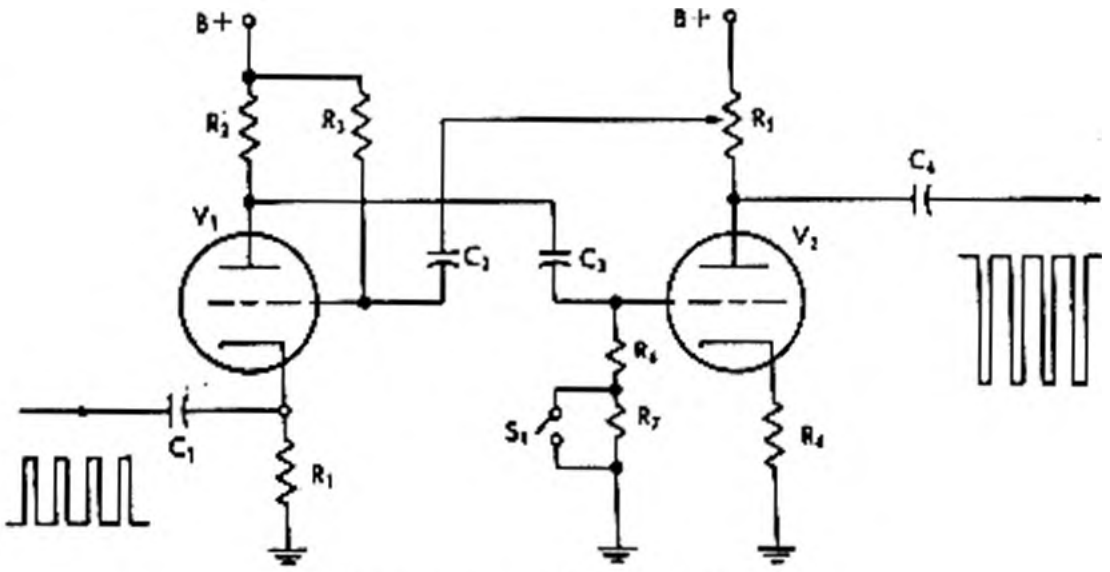


Figure 6-46. PRF Discriminator

Phase Comparator

A phase comparator detects deviations from a pre-selected value of time difference between two sets of input pulses. There is no output from the circuit when the input pulse trains are of the desired time difference. The output is negative if the time difference between the pulses is greater than the selected value. The output is positive if the time difference is smaller than the selected value. The magnitude of the output voltage is proportional to the magnitude of deviation. The polarity of the output represents the direction of deviation.

A sawtooth voltage, input No 1 in figure 6-47, is developed by a pulse train of constant pulse-recurrence frequency. The middle of each sawtooth, which is of zero voltage value, is considered as zero time reference (t_0). Input No 2 consists of pulses 180° out of phase but of the same PRF as input No 1. The pulses at input No 2 are developed through transformers or paraphase amplifiers according to the deviation and are displaced a certain amount from the zero time reference. The combined effect of input No. 1 and input No 2A results in zero output from the discriminator, representing an on-course condition.

The phase comparator circuit in figure 6-48 is compared to a bridge circuit in the schematic of figure 6-49. Without the sawtooth voltage input at terminal No 1 of figure 6-48, the bridge circuit is similar to a balanced bridge with no output.

In this case, the input No 2 pulses allow the four diodes to conduct equally and develop no voltage difference between points A and B (figure 6-49). With no voltage difference between these junction points, no current flows in R_1 and R_5 resistors, resulting in zero output voltage. This result is duplicated when the sawtooth voltage input at No 1 and the pulse inputs at No 2 are of zero time difference as shown in the case of input No 2A of figure 6-47.

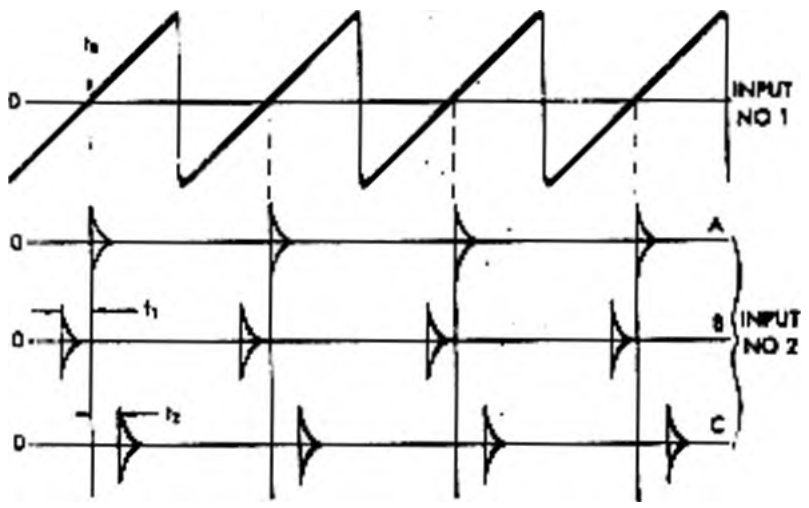


Figure 6-47. Pulse Time Difference Relationships

Consider the time when the input No 2B pulses are applied to the input terminals while the sawtooth input No 1 is negative. Tube V_3 conducts more than V_1, starting current flow around the circuit R_3, R_2, R_4, V_2, the output resistor R_5, and then back to V_3 cathode through R_1. This current flow develops a negative voltage at the output terminal, junction of V_2-V_4, representing an off-course condition.

If the No 2 input pulses are applied to the discriminator while No 1 input is positive, as in the No 2C case, current flow can be traced through

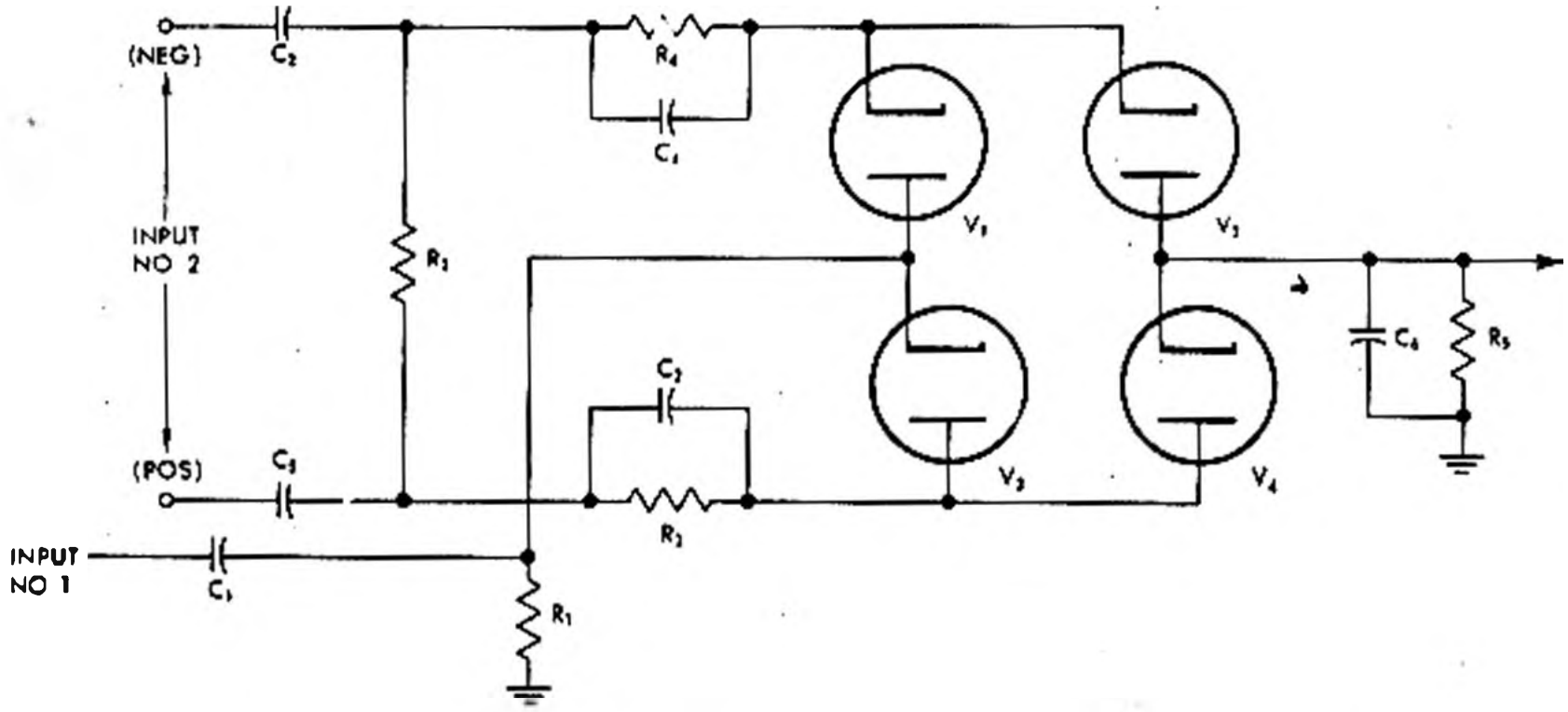


Figure 6-48. Phase Comparator

V_1, R_1, R_5, V_4, R_5, R_2, and R_4. The output voltage is now positive, corresponding to an off-course condition in the opposite direction.

Autosyns as reference units in control systems are described in the preceding chapter. The autosyn units in guidance computers are used to shift the phase of one or more voltages. The autosyn resolvers, as they are called, often function as coordinate converters.

In some systems of reference, polar coordinates are converted to plane Cartesian coordinates, or plane Cartesian coordinates are converted to polar form.

Autosyn resolvers are essentially transformer

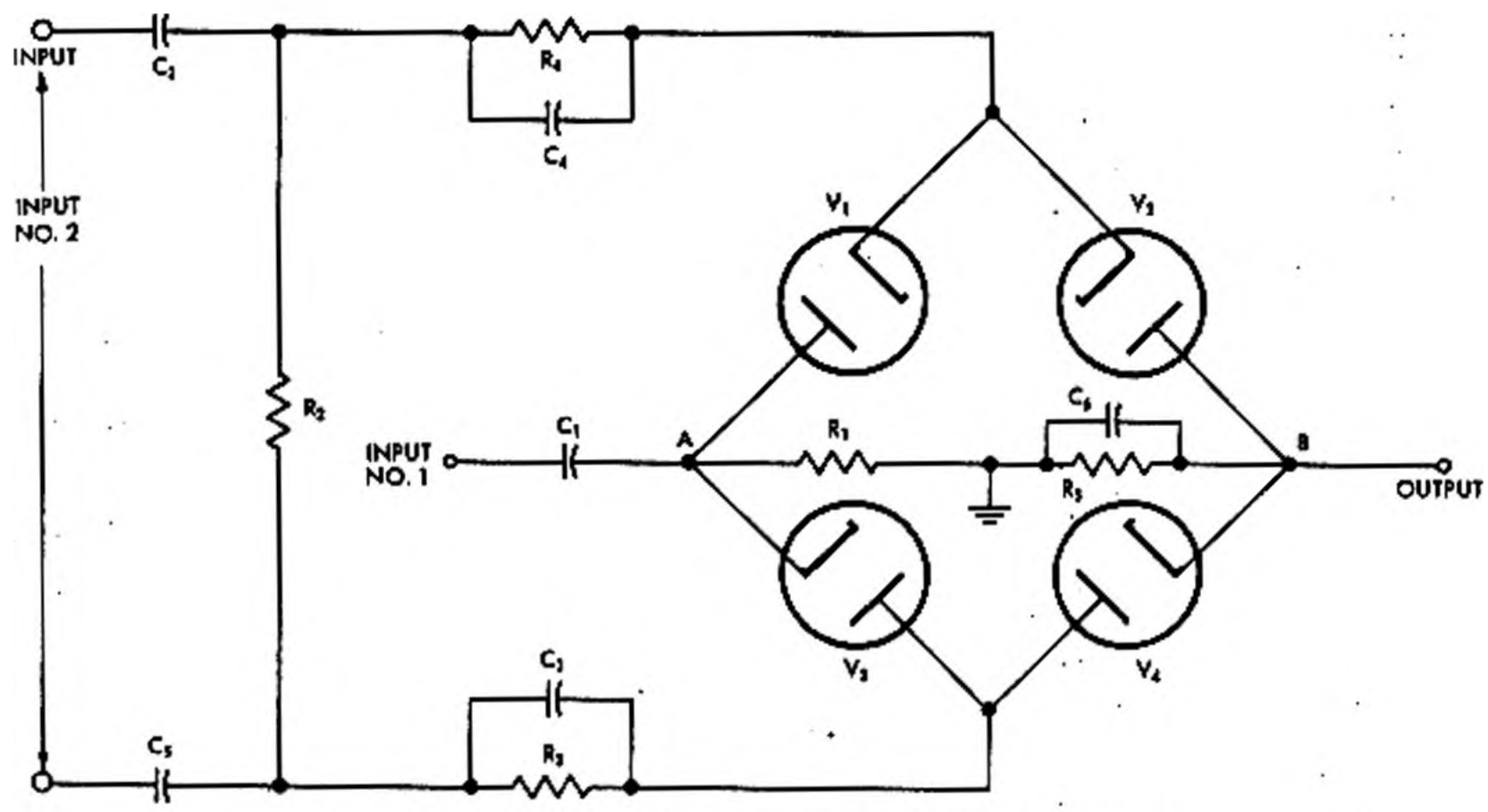


Figure 6-49. Bridge Circuit Representation of Phase Comparator

arrangements with two primary or stator windings and two secondary or rotor windings. In figure 6-50 representing a resolver schematic with the rotor in the zero position, voltage applied to the P_1-P_3 winding induces maximum voltage of the same phase in the S_1-S_3 secondary winding. Voltage applied to the P_2-P_4 primary induces maximum voltage of the same phase in the S_2-S_4 secondary. With a small positive rotation from zero position, the voltage induced in S_1-S_3 is in phase with the P_2-P_4 excitation, and the voltage induced in S_2-S_4 is 180° out of phase with the P_1-P_3 excitation.

The output of each resolver secondary depends on the primary voltages and the rotor position. In the computing units, positive voltages representing positive quantities are in phase with some reference voltage. Negative voltages representing negative quantities are 180° out of phase with this reference voltage.

Figure 6-51 shows the resultant field produced by the vector addition of the fields of the primary windings. These fields induce voltages into the secondary windings as shown in figure 6-52.

The action described in the preceding paragraphs is used in converting from one set of coordinates to another. A brief review of fundamental polar and plane Cartesian coordinate properties follows.

In figuring these coordinates, a reference point in a plane and a reference direction are chosen. The position of any other point in the plane can

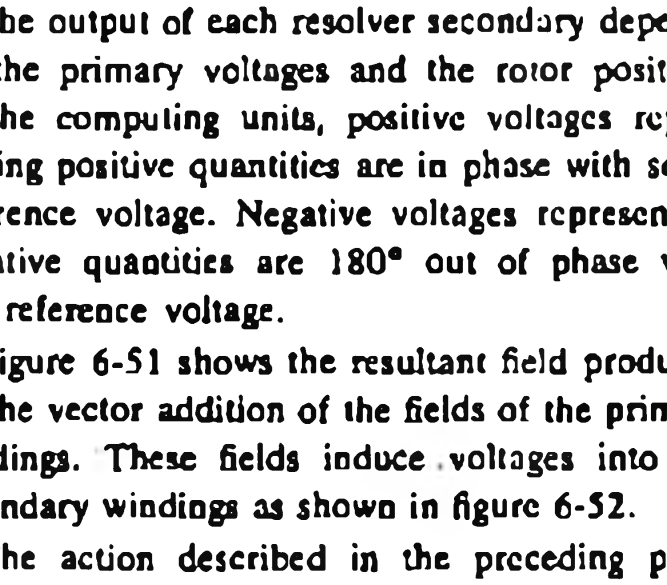


Figure 6-50. Resolver Schematic with Rotor in Zero Position

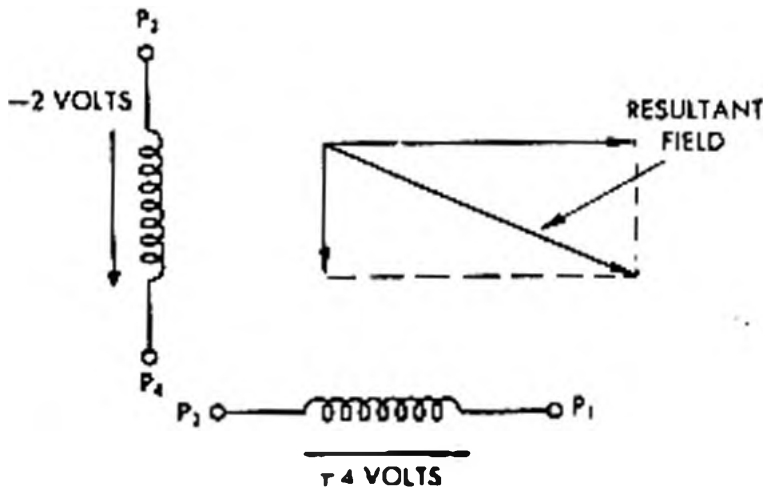


Figure 6-51. Resultant Field of Primary Windings

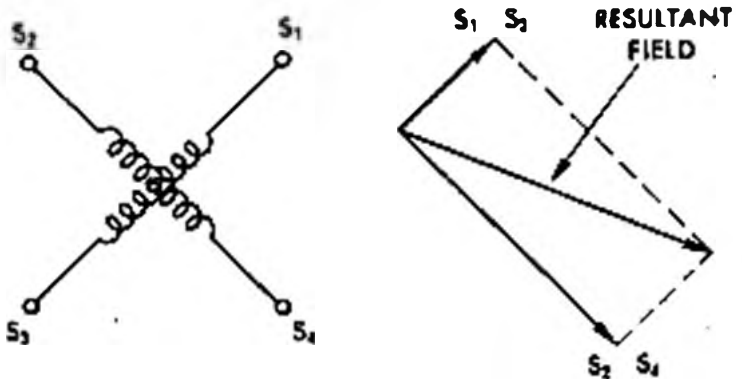


Figure 6-52. Secondary Voltages

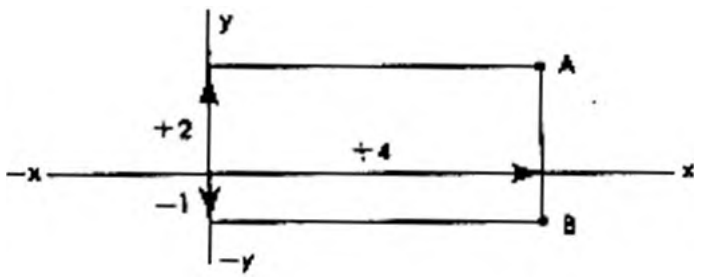


Figure 6-53. Plane Cartesian Coordinates

be located by two numbers called coordinates. In Cartesian coordinates, both of the "x" and "y" coordinate numbers represent distances. In polar coordinates, one coordinate corresponds to a distance and the other to a direction.

In figure 6-53, point A is represented by +4, +2 coordinates, and point B coordinates are +4, −1 from the origin 0.

In figure 6-54, the distances of the points A and B from the origin is the range, which is always positive. As shown, when the direction angle is measured from a reference direction in a clockwise rotation, the angle is called a bearing. When measured from north clockwise to a point, the angle is sometimes called the azimuth.

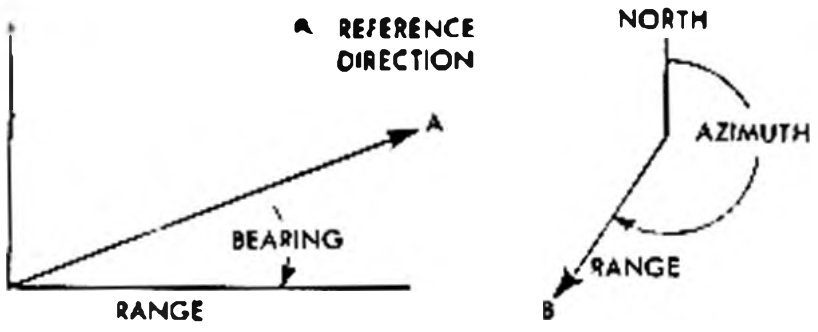


Figure 6-54. Polar Coordinates

Polar coordinates can be developed from a set of plane Cartesian coordinates by the method illustrated in figure 6-55. Voltages representing the east coordinate and the south (minus north) coordinate are vectorially added to form the polar coordinate, labeled range. In the resolver, the primary voltages may be considered the plane Cartesian coordinates. The resultant magnetic field develops voltages in the secondary windings that are the Cartesian vector components for a different set of Cartesian axes.

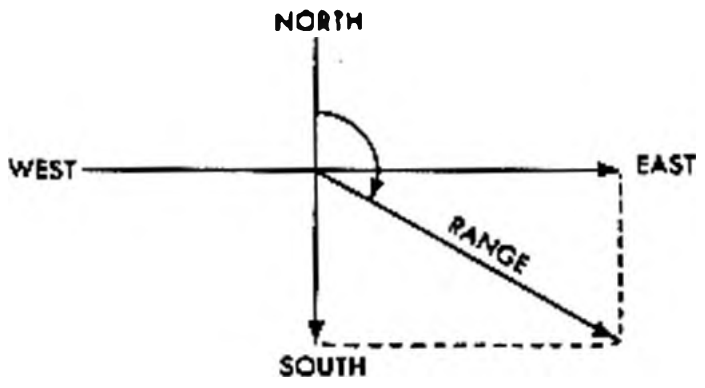


Figure 6-55. Coordinate Conversion

Figure 6-56 illustrates the angles involved in rotating coordinate axes. First, the plane Cartesian coordinates are converted to polar coordinates as shown by the solid line, labeled range. The reference direction of this polar coordinate system is rotated through any desired angle. The new polar coordinates are then converted to plane Cartesian coordinates.

Autosyns are used in computer units as resolvers to produce voltages proportional to the sine or cosine of a measured angle. A voltage is applied to the stator winding which is to be resolved into components. This voltage amplitude represents a particular line-of-sight distance.

Voltage is induced in the rotor winding which is proportional to the angular displacement be-

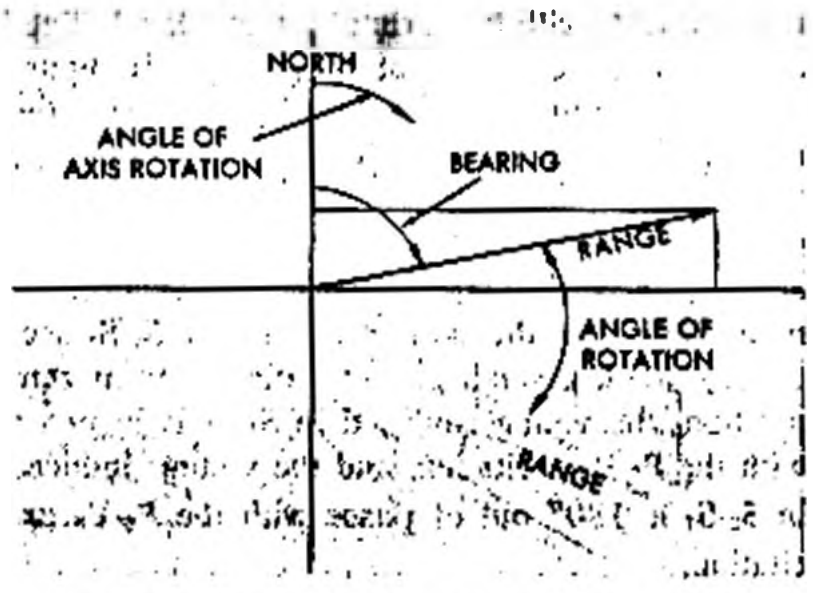


Figure 6-56. Rotating Coordinate Axes

tween stator and rotor. Assume an angle of value A as indicated in figure 6-57. Line a represents the value S sin A, and line b represents C cos A from the trigonometric relationship.

An autosyn with its resistance-capacitance network is often referred to as a goniometer. Three types of goniometers are used. An uncompensated type without a counter is used in equipment to set up delays, to synchronize pulses, and to set up guidance pulses. Accurate time measurement of the pulses is not necessary, but the delays are variable and can be maintained stable.

A compensated goniometer with counters calibrated in microseconds is used in ground equipment to set up target information in a missile. There is also a motor-driven goniometer, rotated at a constant rate to simulate changing time differences.

All these goniometers have the same circuitry, which is shown in figure 6-58. Coils L_1 and L_2 are wound physically at right angles to each other. Coil L_3 can be rotated through 360°.

ANALOG COMPUTERS

Examples of mechanical devices that use analog computing principles include the slide rule, the automobile speedometer, and the automobile differential gearing. In these examples, physical quantities are made to obey mathematical laws.

On the slide rule, logarithms of numbers are represented by lengths on the scales. Multiplication, division, raising to a power, or taking a root

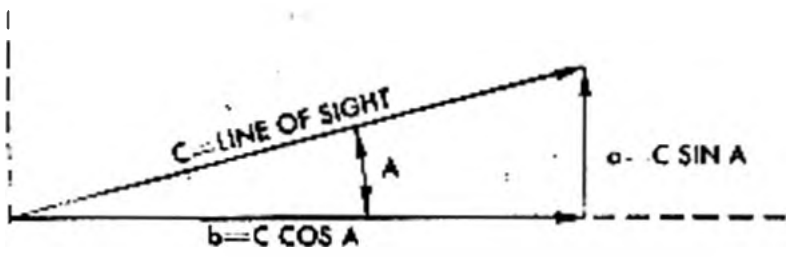


Figure 6-57. Sine and Cosine Functions Resolved Graphically

of a number is relatively accurate and rapid by operating on these quantities.

The automobile speedometer indicator is moved into position by the generation of electrical energy that is proportional to the speed of the drive shaft. This operation of indicating the rate of change of the automobile position is one of differential calculus.

The automobile differential is a mechanical subtracting device. The amount of subtraction is determined by the gear ratio of the differential gears. Extra speed gained by one wheel is subtracted from the other. This principle is used to add two quantities in mechanical differential analyzers.

The two quantities to be added are supplied to the ends of an adder. The sum is obtained by measuring the revolutions through which the body of the adder turns. Quantities to be subtracted can be represented by rotating the shaft in the reverse direction. Mechanical differential analyzers have been used for years to solve missile trajectories and problems in stability design.

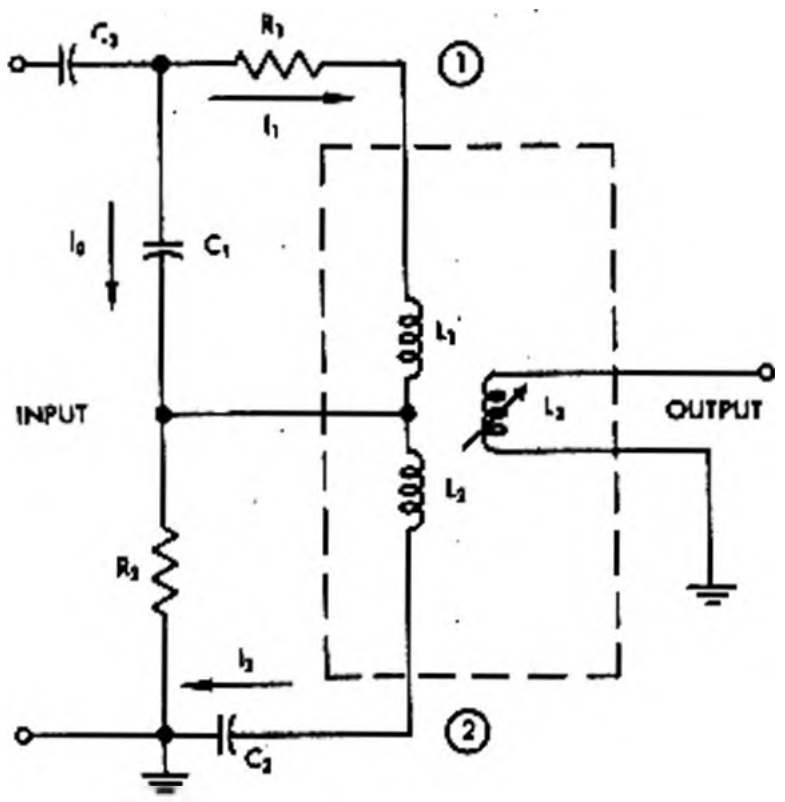


Figure 6-58. Autosyn with Associated Circuitry

Electronic differential analyzers are used in almost every phase of missile dynamic study to obtain information without constructing actual prototypes. A typical missile goes through several stages of development, such as airframe design, design of the automatic control system, selection and design of the guidance system, analytical checks of performance, and actual checks of performance or flight-test analysis. The analog-type machines are able to study the flight trajectory without breaking it into separate phases because they can handle continuous variables and, in particular, they can handle integration as a continuous process. One component common to the electronic analog computers is the DC feedback amplifier. This unit and its associated circuitry provides the approach to the principles of analog computing.

First, consider how one of the simulators or electronic analog computers is used in the missile field. This particular 4000-electron-tube machine simulates the characteristics of both missile and target. Through its use, the need for test firings of new missile designs is greatly reduced. Each of the machine's computing sections handles some particular phase of a complete missile system problem. The major sections are composed of the simulator, the target simulator computer, the aerodynamic computer, the guidance computer, and the recording and display devices. More recent simulators and computers have become solid state in design, eliminating electron tubes.

The guidance computer receives information from the simulator and the target simulator. Its output is fed to the aerodynamic computer.

The aerodynamic computer also receives information from the simulator in the form of missile altitude, velocity, and attitude. The guidance computer introduces missile fin-deflection information. The aerodynamic computer section determines which aerodynamic forces and torques should be delivered back to the missile simulator section.

The target simulator generates the target trajectory in the form of target positions in earth coordinates. Target maneuvers can be simulated by changing target speed, climb, and turn values.

Plotting units are normally used for missile and target trajectory plotting in both horizontal and vertical planes. Any 18 desired variables may be

recorded by photoelectric recorders. There are two display devices: a model of the missile assumes the altitude and fin deflection of the simulated missile, and a trajectory model moves objects representing the missile and the target in three dimensions.

The following brief description of the inputs and outputs to the missile simulator computer illustrates the complexity of the circuits involved. Inputs in the form of voltages represent the aerodynamic forces along the roll, yaw, and pitch axes of the missile; the aerodynamic torques about these three axes; the initial position of the missile with respect to the earth's axes; the initial linear velocities along the three missile axes; and the initial attitude of the missile axes with relation to the earth's axes.

Outputs from the same missile simulator section are indications of the axes of the missile, initial linear velocities along the earth's axes, missile linear velocities along the missile axes, angular velocities about the missile axes, and the missile attitude in terms of the direction cosines of the missile with respect to the earth's axes.

Basic Operation of Feedback Amplifiers

One factor common to all electronic analog computers is the DC feedback amplifier. This basic component with its input and feedback networks is described as it is used to perform the basic computer operations. One method indicating the basic circuit is shown in figure 6-59.

One of the most difficult design problems in DC amplifiers is the reduction of drift. Special compensating circuits are used in combination with well-regulated power supplies to minimize this source of error in the output voltage. In some computer installations the vacuum tubes are aged to reduce the drift resulting from changes in cathode emission. Balancing the output voltage to zero for a zero value of input voltage is a regular procedure in some analog circuits.

Characteristics of most DC feedback amplifiers include high input impedance, low output impedance, and very high gain tubes.

Multiplication of Voltages

One of the simplest methods of multiplying a voltage by a constant coefficient is to use a voltage divider or a potentiometer. In figure 6-60, the proportion of input voltage e_i present at the output terminal e_o is determined by the ratio of R_2/R_1.

A DC amplifier also can be used with fixed resistors to multiply a variable voltage by a constant coefficient. In figure 6-61, the output voltage e_o is proportional to the ratio of the resistor values but with the sign changed to take into account the phase inversion.

Multiplication of two variables is commonly accomplished by using a servomechanism. Some type of servomechanism is used to position the potentiometer arm so that the amount of mechanical rotation is proportional to one of the input signals (e_2 in the schematic of figure 6-62). Input signal e_1 is connected across the potentiometer. The output voltage of e_o is proportional to the product

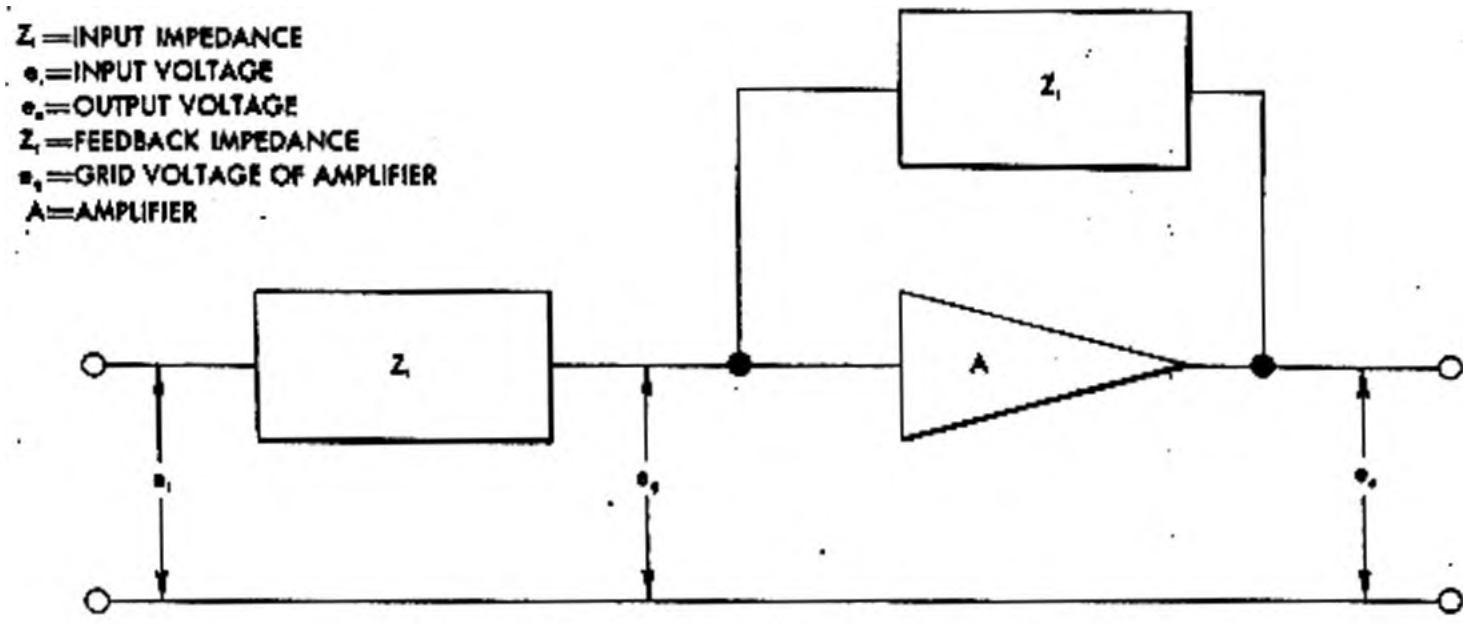


Figure 6-59. Basic Feedback Amplifier Circuit

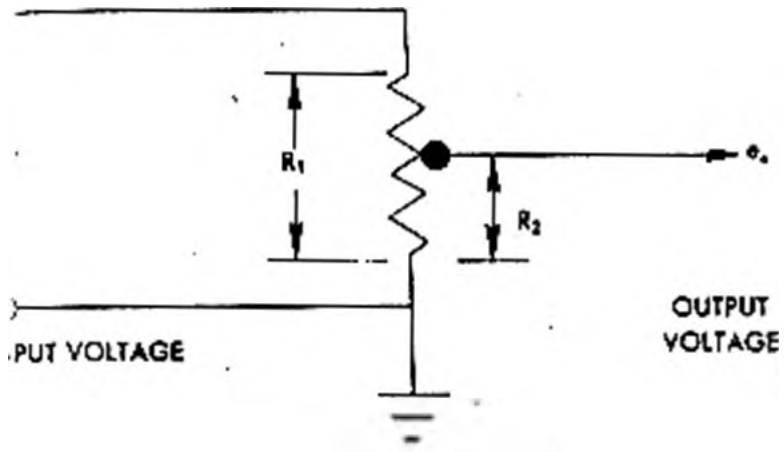


Figure 6-60. Multiplication of Voltage by Constant Coefficient

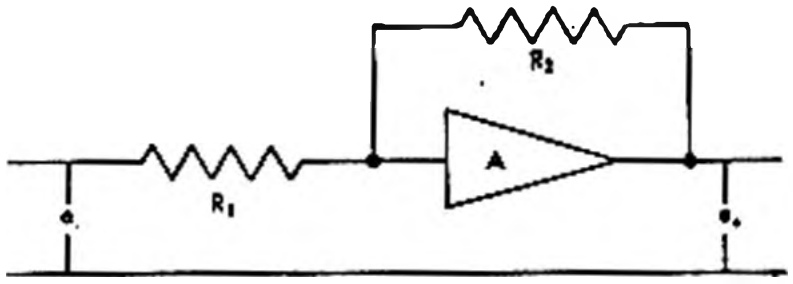


Figure 6-61. Multiplication of Variable Voltage by Constant Coefficient

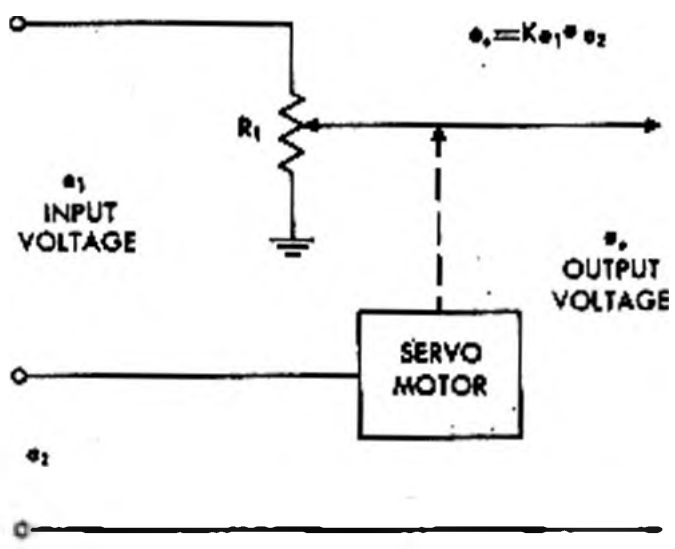


Figure 6-62. Servo Multiplier

of the two input voltages. The dotted line in the schematic represents a mechanical connection.

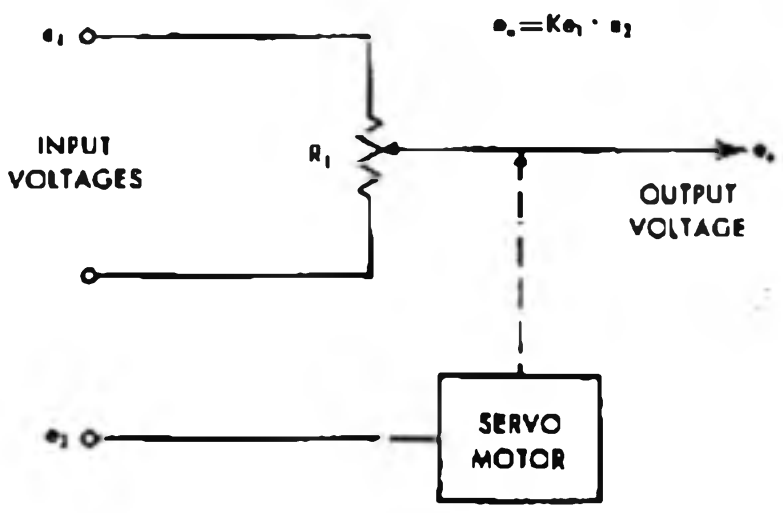


Figure 6-63. Servo Multiplier allows Multiplication of Two Variables for All Combinations of Signs

The connection shown in figure 6-63 allows multiplication of two variables for all combinations of signs. Values of e_2 represented by the position of the potentiometer arm can be positive or negative in sign. The computation speed is limited by the rate at which the motor driving the potentiometer arm can follow the changes in the voltages e_2.

Linear circuits alone cannot be used to multiply because the product of two quantities is not a linear type of expression. Multipliers must use devices in which gain depends on one or more of the applied voltages.

Electronic multipliers follow a much more rapid change of input signal than a servomechanism type. The multiplier unit is one of the basic forms of non-linear computing components. Linear amplifiers may be used to feed the variables into the multiplier.

The circuit shown in figure 6-64 uses two sign-changing amplifiers. The circuit is connected so that both negative and positive values of input voltage e_1 are introduced to make the sign of the product correct.

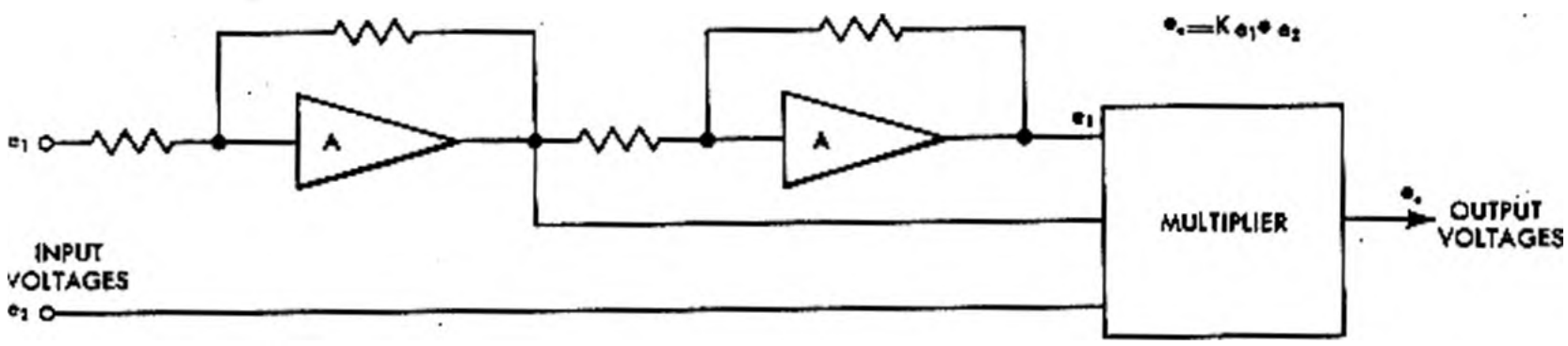


Figure 6-64. Multiplication of Two Variables using Two Sign Changing Amplifiers

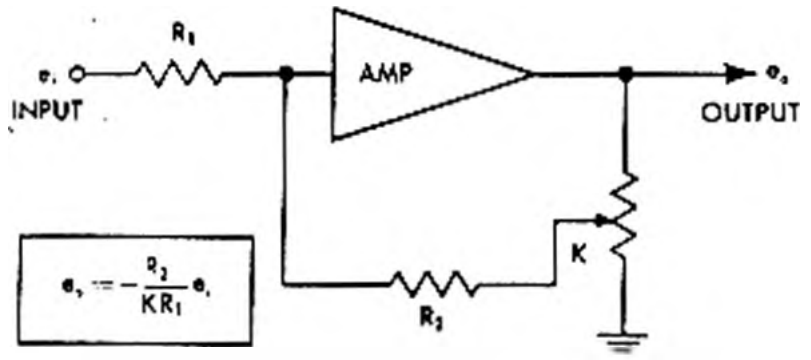


Figure 6-65. Division of Constant using High-Gain Negative Feedback Amplifier

Division of Voltages

One method of dividing a variable voltage by a constant is to multiply the variable voltage by the reciprocal of the constant. When using this method, multiplying devices serve as dividers.

Another method of dividing a variable by a constant uses a high-gain negative feedback amplifier.

In figure 6-65, the constant K is determined by the setting of the feedback potentiometer arm in the amplifier output circuit. The loading effect of R_2 must be considered when the potentiometer is set. The constant must not approach zero. If it did reach zero, the result of the operation would not be accurate because the amplifier output would be limited.

A high-gain amplifier is also used in a circuit which determines the quotient of two variables. An unknown quantity (Z) is assumed so that the product of Z times X, one of the variables, is made equal to Y, the other variable. The unknown Z and the variable X are fed into the multiplier circuit as shown in figure 6-66.

The output of the multiplier unit and the negative Y value are summed in the amplifier. The output of the amplifier when fed back becomes the quotient of Y divided by X.

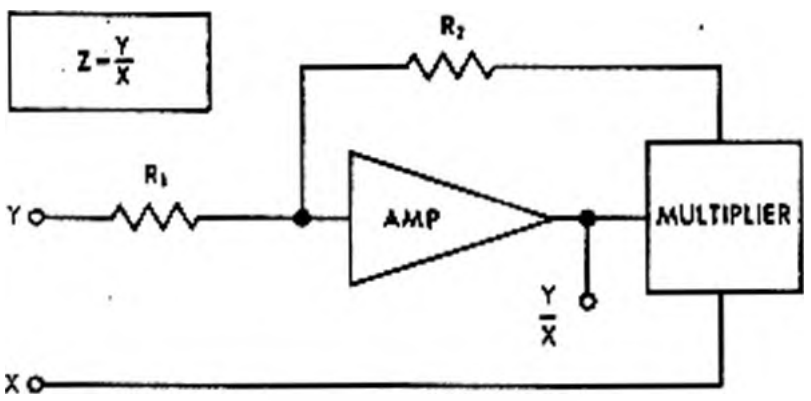


Figure 6-66. Division of One Variable by Another using High-Gain Amplifier

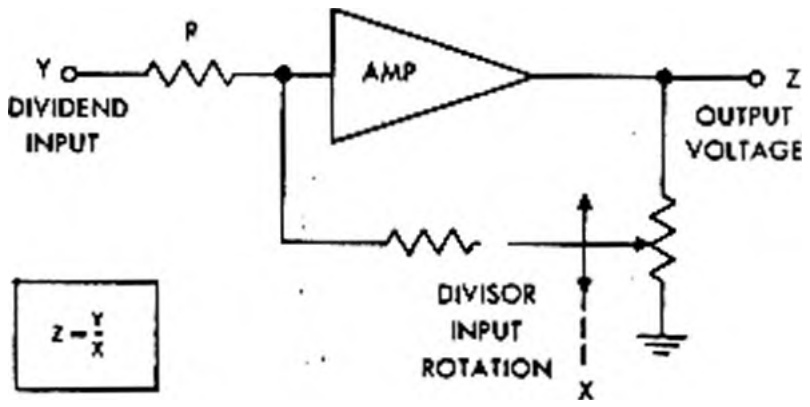


Figure 6-67. Division of One Variable by Another using High-Gain Negative Feedback Amplifier with Potentiometer

Another method of dividing one variable by another uses the high-gain negative feedback amplifier with a feedback potentiometer. The dividend, or voltage to be divided, is supplied to the amplifier input terminal. The divisor, or voltage dividing, is used to position the arm of the feedback potentiometer mechanically.

In figure 6-67, Y is divided by X. The amplifier input is –Y and the amount of feedback is proportional to X. Since the gain of a negative feedback amplifier is inversely proportional to the feedback, the output is inversely proportional to X, giving the result of dividing Y by X.

Summation of Voltages

Summing networks are used to add two or more voltages. The output voltage is proportional to the sum of the input voltages when the summing resistors are all equal. Note figure 6-68.

The voltage division between the output resistor R_o and each summing resistor R_1, R_2, or R_3 produces an attenuation of each input voltage in the

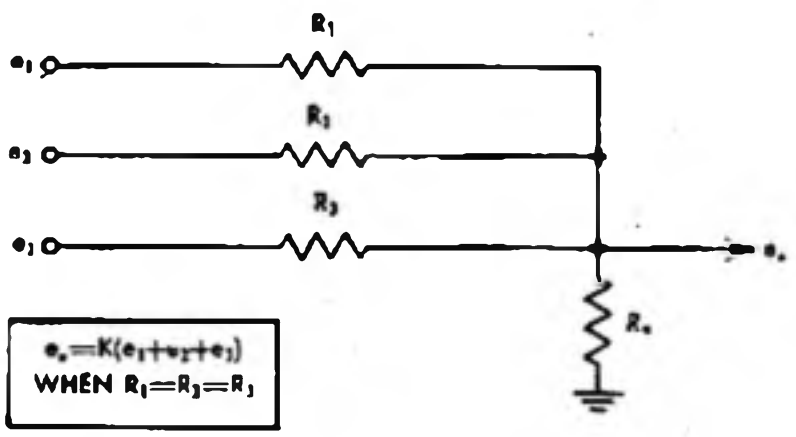


Figure 6-68. Summing Resistor Network

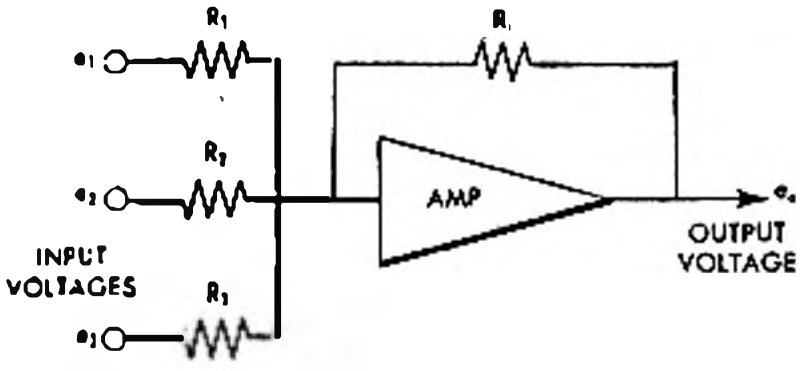


Figure 6-69. Summing Amplifier

circuit. There also may be a loading effect on the input circuits because of the use of the parallel network.

The value of the constant of proportionality K is determined by the ratio of the summing resistors to the output resistor. Scale factors are set up in the computing units so that quantities can be combined in operations in the correct relative values. In general, there is a different scale factor associated with each variable in a problem.

When the summing resistors in the summing network are unequal so as to represent different scale factors, the output voltage e_o is proportional to the sum of e_1/R_1, e_2/R_2, and e_3/R_3.

A summing amplifier also can be used to add several voltages.

The output (e_o) of the circuit shown in figure 6-69 is inverted. The circuit can also be used to multiply by a constant.

If the summing resistors have the values of $R_1 =$ 1 megohm, $R_2 = 0.50$ megohm, and $R_3 = 0.25$ megohm, the input voltage terminals correspond to coefficients 1, 2, and 4, respectively. The output voltage is then proportional to the sum of $1 \times e_1$, $2 \times e_2$, and $4 \times e_3$.

A summing amplifier with the coefficients described above is shown in block diagram form in figure 6-70.

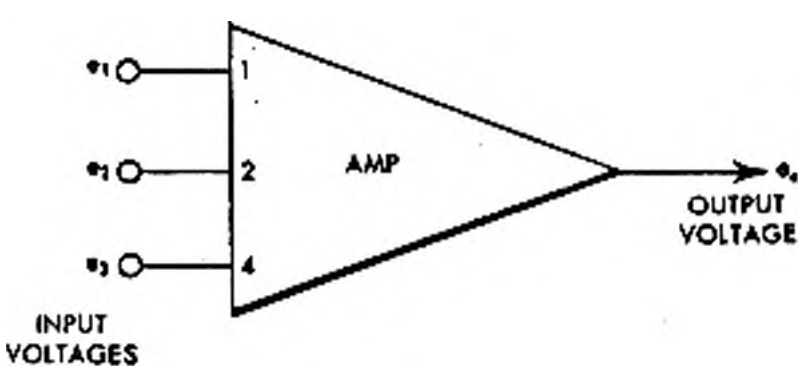


Figure 6-70. Summing Amplifier with Coefficients 1, 2, and 4

A summing amplifier may be used to multiply several input voltages by constant coefficients. This is made possible by paralleling input circuits as shown in figure 6-71. In this particular example, the output voltage is proportional to the sum of $1 \times e_1$, $3 \times e_2$, and $11 \times e_3$.

If the input voltages are positive, the output is negative. By means of potentiometers in the input circuits, this type of summing amplifier produces sums or differences of the input voltages with any combination of preset coefficient values.

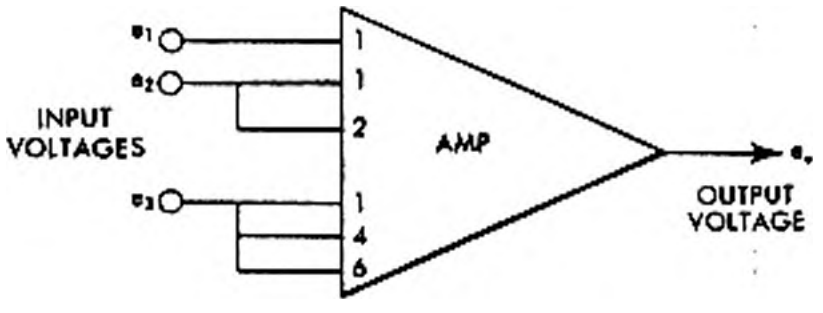


Figure 6-71. Parallel-Input Summing Amplifier

DIGITAL COMPUTERS

In missile guidance, a large amount of complex data needs to be evaluated quickly. Not only are computers essential to the guidance of the missile, but they are also quite necessary in the research and development of missiles.

Some telemetered information must be instantaneously evaluated for the range safety officer, while still other data must be evaluated so the findings of missile flights may be incorporated in future launchings.

A digital computer is assigned a certain numbering system. It then performs a mathematical operation on these numbers (expressed in the form of digits) to yield an answer, also expressed in digits.

Most of the development on modern calculating or computing devices, as they are known today, was started in the late 1800s. These first machines were quite limited as to the operations they could perform and the speed at which calculations were made. The dial telephone was one of the first type of machines that could handle multiple operations successfully. The dial system is capable of handling several million telephone calls by successive relay

operations according to assigned digits of the phone numbers.

Another modern development of calculating machines is that developed by the International Business Machines Corporation (IBM). A predecessor to the IBM machine was developed by Herman Hollerith in 1890. He made extensive experiments with punched cards that would permit electrical contact to be made wherever a hole appeared in the cards. The device that developed from these experiments counted or added, singly or in combination, as desired. This punch-card technique is still a major component in many modern computers.

The computer must be capable of storing information. This storing of information is commonly referred to as the *memory unit* of the computer. Here, numbers and instructions are retained until the unit is called upon to use the information to complete the problem. As described later, the memory unit may retain the information for long or short periods of time depending upon the type of device that is employed for this purpose. In addition to storing information, the memory unit needs some means of transferring the information to other parts of the system. These links or channels by which the information is passed from one unit to another may be called the *data transmission system.*

The next unit in the system is the *logical unit,* which receives the transmitted material and performs the arithmetical operations on the problem. The logical unit adds, subtracts, multiplies, and divides. It takes square root, attaches the proper sign, and gives the proper comparison of quantities.

A *control system* is the next item that must be included in the computer. The control system tells the machine when to perform a certain operation. It also governs the type of operation that is to be performed. It governs whether the operation is to be addition, multiplication, or any other arithmetical calculation.

There must be some provision made for the reception of data and the supply of answers in the machine. These units, representative of their functions, are called *input and output units,* respectively.

Although a digital computer is versatile in problem handling, there are some "traits" it does not possess. It lacks intuition, which means that every problem must be presented to the system before it will operate properly. It does not possess the ability to make guesses or draw conclusions. It only gives answers according to the accuracy of the data presented. The computer is not capable of rearranging the sequence of operations or of developing any of its own instructions. It cannot make any changes in the programming of the material. And it cannot form any situations that are not pertinent to the problem as given in the problem's original form.

Not having intuition, the mechanical brain can only deal with facts and deliver answers based on these facts. No extraneous matters intrude upon its straightline "thinking."

Figure 6-72 shows the functional diagram of a digital computer. The components that carry out the various functions in the computer may not be physically distinct. The same components or types of components may be used for more than one function. For example, a component that performs arithmetical operations may also be used as a memory unit.

Binary Numbers

Most of the digital computers use a binary system of coding. If a question is asked of a machine, the smallest unit used would be one to represent the truth of the statement. The answer may be in

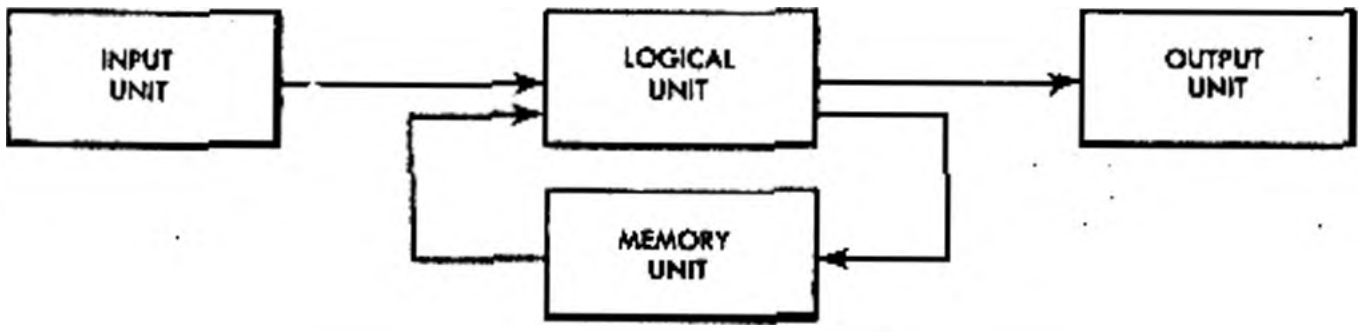


Figure 6-72. Functional Diagram of Digital Computer

the form of a yes or no (a checkmark or no checkmark) or, as in the case of electronic instruments, an on or off (an impulse or no impulse). In short, there are just two conditions that need to be considered to determine whether a statement is true or false.

In the binary digital computers, use is made of this fact by employing what is known as a binary representation of numbers. The two-fold difference of representing information makes it possible to use devices that represent only the go or no-go conditions. For example, a relay in a closed position could represent a go condition, while the same relay in an open position would represent the no-go condition. The punch-card system used in computers lends itself well to this system. In this case, the presence of a hole would represent the go (yes) position and the absence of a hole would indicate the no-go (no) condition. These two conditions are indicated in the binary numbering system by the two digits "0" and "1," respectively.

Everyone is familiar with the everyday usage of the scale-of-ten (decimal) representation of numbers. Assume the number 8501 and determine how it is made up by powers of tens. This number may be represented by what is called the *powers of ten,* that is with 10, 100, 1000, 10,000—or multiplied by itself as many times as desired. The number 8501 indicates:

$(8\times1000) + (5\times100) + (0\times10)) + (1\times1)$, or

$(8\times10^3) + (5\times10^2) + (0\times10^1) + (1\times10^0)$.

In the decimal representation of a number, the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 are used. In the binary representation of a number, only two digits, 0 and 1, are used in various combinations.

In the binary system, a sequence of 0s and 1s indicates the presence of the number 2 raised to some power. For example, figure 6-73 shows that the digits 1101 indicate $(1\times2^3) + (1\times2^2) + (0\times2^1) + (1\times2^0)$. This would equal 13 in the decimal system.

Shown in table 6-1 are some decimal numbers and their equivalents in the binary system.

Examine the table for a moment. Consider the first four decimal numbers 0, 1, 2, and 3. To use the decimal system of numbering in a computing device, you would need to have four registers to record the information presented. As shown in

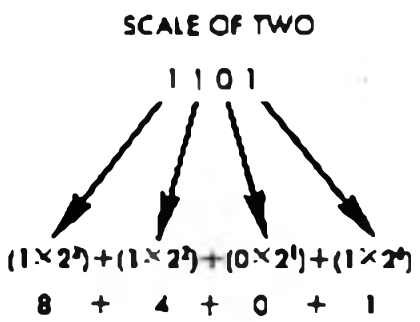


Figure 6-73. Decimal and Binary Representations of Numbers

table 6-1, the binary system needs only two digits, the 0 and 1, to represent the first four numbers. In this case, four conditions of information can be represented by the use of two digits.

You can see where an ON-OFF electronic system fits into the binary pattern very well. Say that you wish to indicate the decimal number 3 in a computing function. This can be accomplished in the binary digital computer by representing it as 11. The binary digits (11) would be represented in a relay system by both of the relays being in the closed (go) condition.

Once again referring to the table of decimal and binary numbers, three digits of the binary system may represent eight pieces of information; with four digits, 16 pieces of information may be represented. By the time you have seven binary digits, 128 pieces of information may be represented.

Table 6-1. Decimal and Binary Numbers

DECIMAL	BINARY	DECIMAL	BINARY
0	0	13	1101
1	1	14	1110
2	10	15	1111
3	11	16	10000
4	100	17	10001
5	101	18	10010
6	110	19	10011
7	111	20	10100
8	1000	21	10101
9	1001	22	10110
10	1010	23	10111
11	1011	24	11000
12	1100	25	11001

Boolean Algebra

Before a computer can operate properly and can satisfactorily employ some numbering system, such as the binary system, it must have the ability to:

- Learn and remember data presented to it.
- Make a choice based on previous results.
- Make long chains of operations.
- Determine if the answer is correct.
- Determine when one problem is finished and when to start on another.

For the computer to be capable of accomplishing the foregoing items, the proper information must be given to it. From this point on, it will determine the answer by what is known as logical truth.

Logical truth differs from ordinary truth in that it deals not only with facts but also in suppositions based on facts. For example, the statement "sugar dissolves in water" is an ordinary truth. Certain things are understood. The amount of sugar is much smaller than the amount of water. If a whole bag of sugar is mixed with a teaspoonful of water, all of the sugar would not dissolve. The computer cannot operate by understanding such limiting conditions unless these conditions are given to it so that it may form a logical truth pattern to work from.

A logical pattern may be of the following nature:

1. All people are animals.
2. All animals are mortals.
3. Therefore, all people are mortals.

Such statements as the above are called logical syllogisms. They are truth statements that result in a truth statement. It should be understood that all syllogisms may not be logical. For example, examine the following statements where statement 3 is not true even through statements 1 and 2 are:

1. All animals have legs.
2. All people have legs.
3. Therefore, all animals are people.

Many logical truth patterns are used by people in everyday life without their realizing that they are using them. Most of the simpler logical patterns are distinguished by words such as and, or, if, else, not, then. In mathematical terms these logical patterns may be indicated by plus, minus, times, divided by, or combinations of these terms.

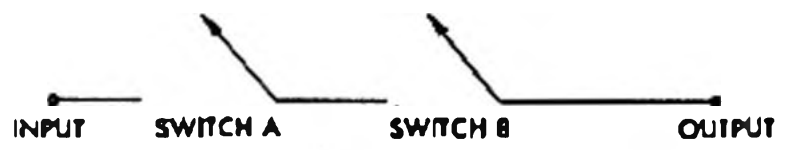
If a set of statements is made, a table can be set up that will determine what is known as the truth values of these statements. In setting up this table, some of the logical connectives, such as and, or, if, etc., must be used. When the table is completed it would list the truth or falsity of each of the conditions of the statements that were considered.

We can set up mathematical relations of the truth values of statements and perform mathematical operations on these statements. This study of mathematical logic is called the algebra of logic. George Boole, an English mathematician, introduced this algebra of logic in the book, *The Laws of Thought,* in 1854, thus the term Boolean algebra.

This algebra of logic can be used in the study of electrical circuits and switching. By setting up appropriate circuitry, the truth values of such statements as would imply connectives (or, and, not, etc.) can be determined mathematically. Each of these connectives can be represented by circuitry, and the condition of the circuitry designated by a number. For example, if the truth of a statement is shown as the digit 1 and the falsity of the statement as a 0, certain facts concerning the switching circuit of figure 6-74 can be determined.

In this circuit if a switch is open, it would be shown in the table as 0 (current can not pass); if it is closed, it will be shown in the table as 1 (current can pass). In the output, the 0 will indicate no-flow of current and the 1 will indicate flow of current.

By referring to the table in figure 6-74, there is just one condition of the switches that will allow current to flow to the output. This condition is when switches A and B are closed. Here is simu-



SWITCH A	SWITCH B	OUTPUT
0	1	0
0	0	0
1	0	0
1	1	1

Figure 6-74. Series Switching Circuit showing Truth Value of Statements that use "AND"

lated the truth value of statements that use the connective "and." The digital notation is 1-1-1.

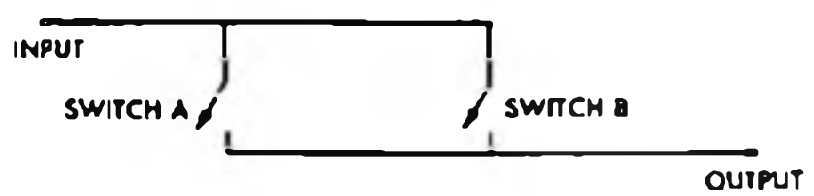
To determine the truth value of statements that use the connective "or," a circuitry similar to the parallel switching circuit of figure 6-75 can be set up.

In this case, as contrasted to the conditions of the series switching circuit, the current flows in all conditions of the switches except one. The truth value of the statement "current will flow in the output if switch A or B is closed" has been determined.

The truth or falsity of these statements can be indicated in a digital manner by the use of a computing device that can scan the truth values of statements containing a variety of logical connectives.

The results of a digital computation in a guided missile may be used in any phase of the flight. For example, data (statements) concerning the airspeed and altitude may be entered into a computing system to regulate the amount of fuel flow to the engine. Or perhaps the point is reached where the engine is to be shut off, and the missile is to dive into the target. The computer is called upon to determine the proper time for this to happen.

Now that the requirements and principles involved in having an electronic device that makes accurate computations in fractions of a second have been covered, the actual devices and some of the basic circuitry that make up a computing system may be examined. These circuits are presented from the overall standpoint of their possible uses. It is next to impossible to list all of the applications of any one device or to say for sure



SWITCH A	SWITCH B	OUTPUT
0	1	1
0	0	0
1	0	1
1	1	1

Figure 6-75. Parallel Switching Circuit showing Truth Value of Statements that use "OR"

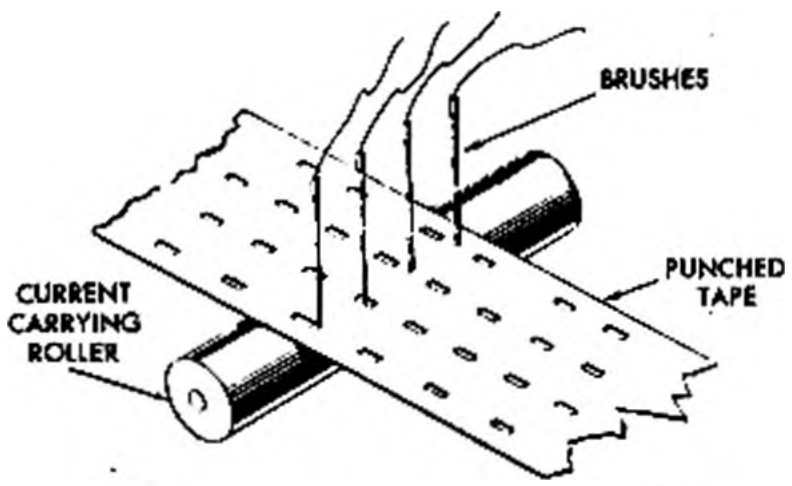


Figure 6-76. Punched-Tape System for Digital Computers

that one particular device would satisfy the requirements of a problem better than some other device.

Input Unit

All the known data (statements) of a problem are injected into the computing system via the input unit. This input is usually supplied to the system by means of punched paper tape, photographic tape, magnetic tape, or wire on which the data and instructions have been recorded.

TAPES. In using a punched paper tape, such as illustrated in figure 6-76, the original information is transferred to the tape according to a pattern that represents certain digits used in the computer. In the case of the binary digital computer, these digits would be 0 and 1 (indicating the various powers of 2). The punched tape passes through a reading unit that permits electrical contact to be made whenever a hole is present under one of the current-carrying brushes.

The punch card or paper tape sorts, lists, selects, and copies information. It makes comparisons and selections according to instructions, and it adds, subtracts, multiplies, and divides. To use a punched tape, the original information is converted into patterns on the tape by means of some typewriter-like device. The use of photographic tapes in the motion picture industry is well known. In this case, the sound patterns are recorded on film, then photoelectric tubes are used to take this information from the film.

Another means of supplying the computer with information is by use of magnetic tape or wire,

upon which alternating currents have been recorded. In this case, the variations of the recorded signal bring about the operation of various components. The magnetic tape and wire reading (or pickup) units are quite similar to those of the commercial varieties. Note figure 6-77.

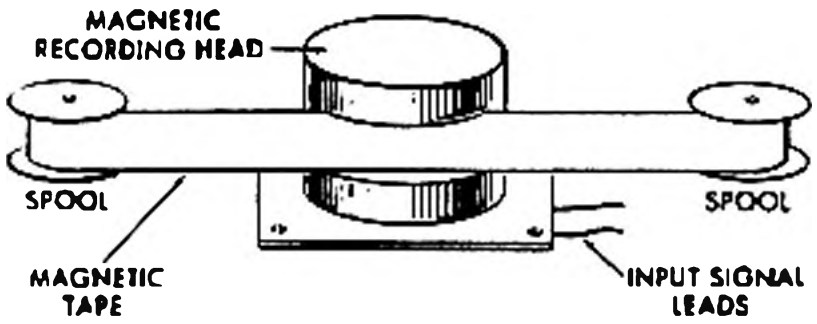


Figure 6-77. Magnetic Tape Recording

In addition to the recorded types of information that may be employed at the input, an operator can originate the instructions to the computer by using push buttons, switches, or some other means of supplying data and instructions to the computer.

In missile equipment, the computer may receive data originating from error signals. These errors could be fed directly to the computer so that rapid corrections could be made.

DECODING. As stated earlier, the process of decoding information is an important function that must be carried out by the computer. Information received at the computer must be put into a form that is suitable for the operation of other circuits. This preparation of information may be accomplished by any of several types of demodulation that are covered earlier in the text. A type of decoding that lends itself well to a digital-type computer is that of the pulse-decoding relay circuit pictured in figure 6-78.

The figure illustrates a method by which a command signal composed of narrow and wide pulses may be decoded in a relay circuit. When narrow pulses with long spacing are received, relay A closes. The second relay B does not close since the duration of the pulse is of insufficient time for its shunt capacitor to charge through the series resistor R.

When signals consisting of wide pulses with short spacing are received, both relays close. The operation of these types of relay action are covered more thoroughly in the chapter dealing with relays.

Here, as in other digital devices, a means of forming a logical pattern is available. In this case, the connectives "if" and "and not" are introduced. This can be explained by considering the time when the pulses are applied. If a narrow pulse comes in, relay A closes "and not" relay B.

In the case of the wide pulse input, a connective of the form "both" is used. If a wide pulse is applied "both" relays close.

In addition to the guidance and reference signals, signals may be fed back from other points of the guidance system. The computer input must be capable of handling a large quantity of information.

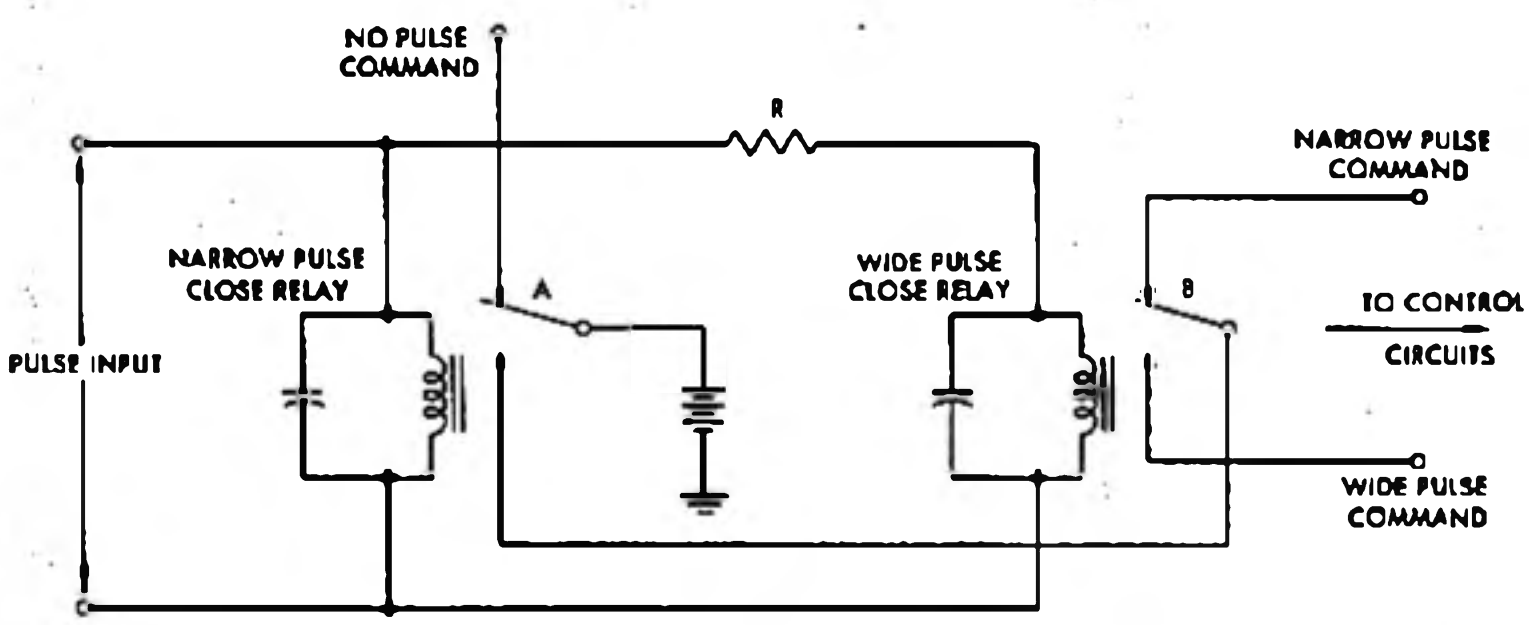


Figure 6-78. Pulse-Decoding Relay Circuit

Logical Unit

As presented earlier, the simplest type of computer is one in which the basic components have but two positions. This simple computer is easier to adapt to a binary numbering system rather than a scale-of-tens system. The switch is the basic example of a device that easily notes the condition of either "go" or "no-go." Other units are designed to simulate the action of a switch. These components that exhibit two stable positions are generally referred to as gates or gating circuits. The gate, as the name implies, permits or prevents the flow of current in a circuit.

GATES. Most gating circuits have an output and two or more inputs and are designed so that the output is energized only when certain conditions are met. The following stages are activated only when the proper conditions exist at the gating circuit. Triodes, tetrodes, and pentodes are often used for this purpose. Since these are amplifier tubes, the circuit may be called a coincidence amplifier—an amplifier that has an output only when two pulses are applied simultaneously to the tube.

The circuit in figure 6-79 shows a tetrode used in a gating circuit. Notice that the tube has fixed cathode and screen grid bias and inputs at both the control and screen grids. With the correct bias voltages applied, a positive pulse on the control grid will not cause the tube to conduct because the screen grid is still negative. If the screen grid swings positive at the same time the positive pulse appears at the control grid, the tube will conduct and the inverted and amplified pulse will appear at the output. The right conditions must exist at both input A and input B before an output will appear at the plate of the tetrode.

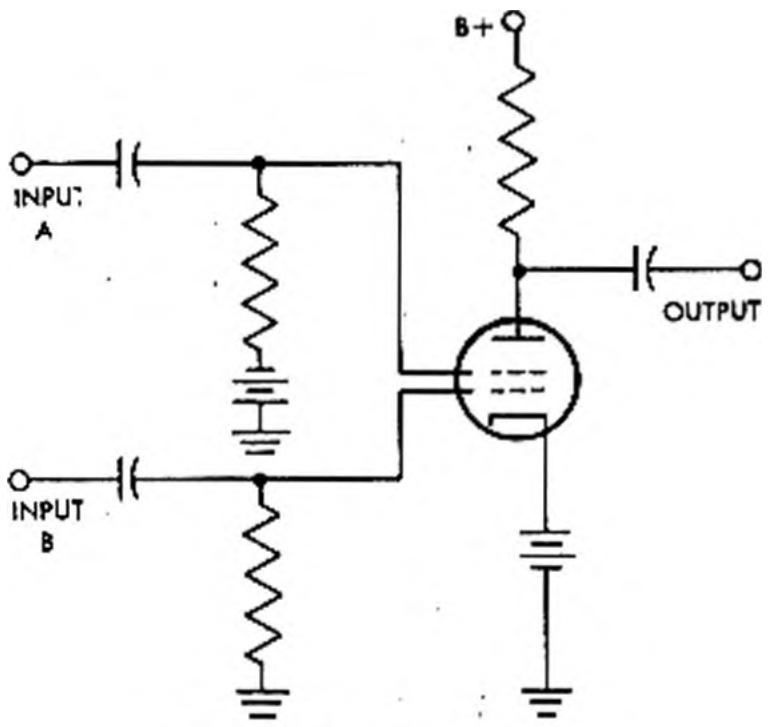


Figure 6-79. Tetrode Gate

Since diode tubes permit current to flow in only one direction, they may be used satisfactorily in gating circuits. Current flows through a diode only when the appropriate voltages are present. This flow can cause other tubes or circuits to be influenced accordingly. Shown in figure 6-80 is a diode-coupled gating circuit using two diodes and a triode.

In the normal condition with no inputs, both diodes conduct and the grid of the triode is slightly positive with respect to ground. The fixed bias at the cathode of the triode prevents the triode from conducting until the control grid goes above the cutoff potential. If V_1 or V_2 is cut off by a positive pulse at input A or input B, the grid of V_3 will not rise about cutoff voltage. If both V_1 and V_2 are cut off by positive pulses arriving simultaneously, the grid voltage will allow V_3 to conduct, and an output will appear at its plate.

These are but two of the many possible electronic circuits that can be used to develop the gating pulses required in computers. Tapes may also be considered as gating devices because they allow following circuits to operate only during certain periods. Solid state circuitry using transistors is replacing vacuum tubes in newly designed circuits.

COUNTING. Another important function that takes place in the logical unit is that of counting. In any computation, it is necessary to have a means by which the number of operations may be determined. This is the function of the counter.

This function is performed in connection with the punched tape of the magnetic recording. In these cases the repetition of pulses would be recorded along with the other operational functions.

Once again, an electromechanical switching principle may be used in the functions of counting. In this case, a multiposition switch is actuated by some electromechanically operated clutch system to give any desired number of pulses in a given period of time.

The counting function may or may not be a continuous operation in a computing system.

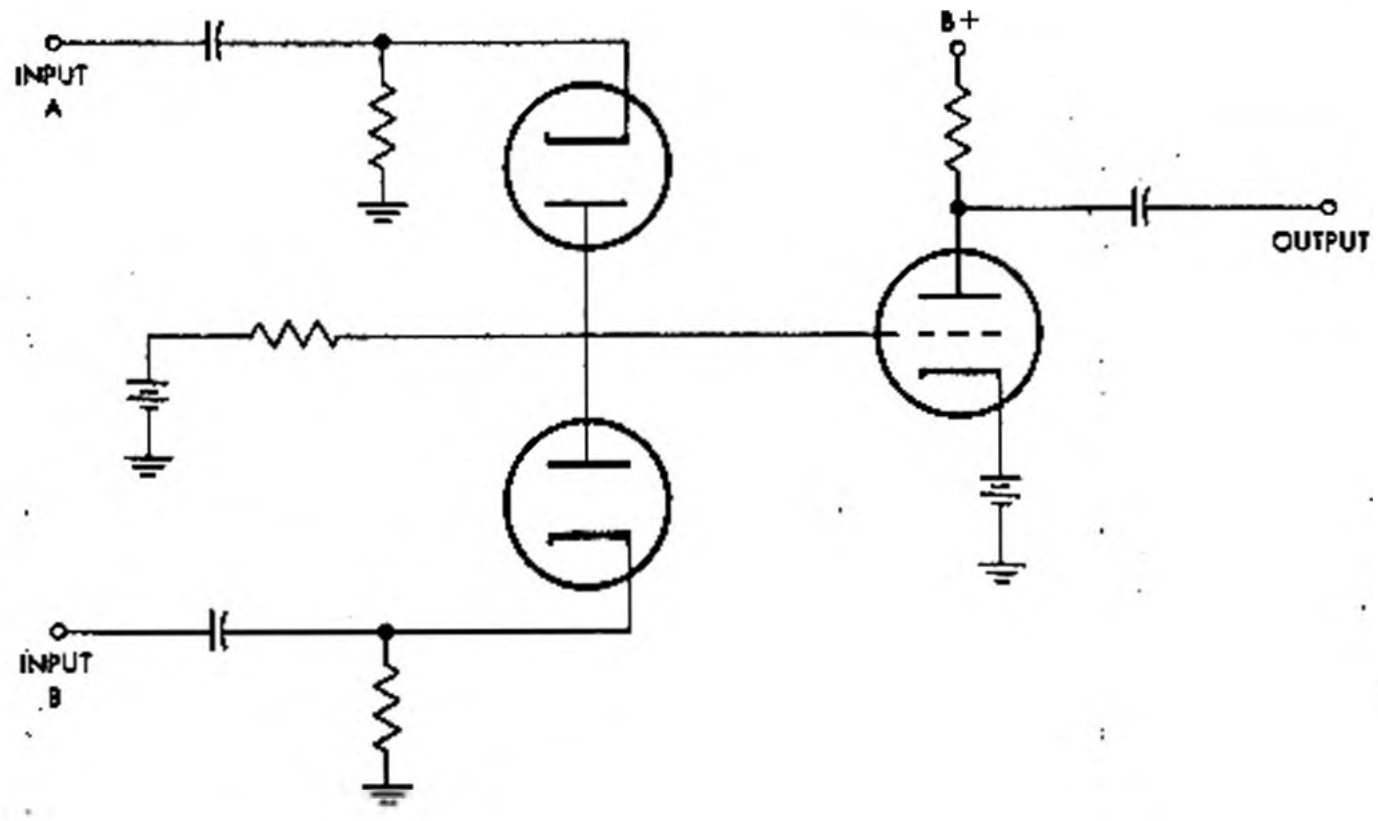


Figure 6-80. Diode-Coupled gate

When the counting takes place depends on the state in which other components of the system are set—that is, it may be necessary to count a certain number of other operations that have taken place before some new operation is to begin.

One of the basic means of having a continuous or interrupted system of counting is by employing the multivibrator or flip-flop principle. One of the basic multivibrators is the Eccles-Jordan type. Originally designed and used as a vacuum tube circuit using two triode tubes, the tubes have been replaced by transistors and the circuit has become solid state. This bistable multivibrator is shown in figure 6-81. This circuit requires two input triggers to complete one cycle. The circuit is in a stable state when either transistor is conducting while the other is cut off. The states of the transistor are switched with the application of a properly applied trigger pulse.

When DC power is first applied, one transistor conducts while the other is cut off. The conducting transistor quickly saturates. Each transistor remains in this stable state, held there by the condition of the other, until a trigger pulse is applied. A negative trigger applied to the base of the conducting transistor will not change the state of operation. The same negative pulse applied to the base of the cut off transistor will change the state of the circuit. The transistor that was conducting will now be cut off, and the transistor that was cut off will now be conducting. This circuit could also be triggered by applying a positive pulse to the base of the conducting transistor.

The method used to trigger the circuit is determined by the polarity and magnitude of the trigger pulse available and the desired repetition rate of the output.

For circuit operation of the multivibrator shown, assume transistor Q1 is conducting and transistor Q2 is cut off. A negative trigger pulse applied to the base of each transistor at time T0 causes transistor Q2 to conduct. The increasing collector current causes the collector voltage (V_{c1}) of transistor Q2 to decrease (become less negative). This change in voltage is coupled to the base of transistor Q1 and reduces its forward bias. Conduction in transistor Q1 begins to decrease. As the collector current decreases, collector voltage (V_{c1}) changes from zero to a negative value (approaching the value of battery voltage (V_{cc}). This change in voltage is coupled to the base of transistor Q2, making the base more negative and increasing the conduction of the transistor. The regenerative feedback continues until transistor Q2 is saturated and transistor Q1 is cut off.

Transistor Q2 continues conducting and transistor Q1 remains cut off. They remain in this

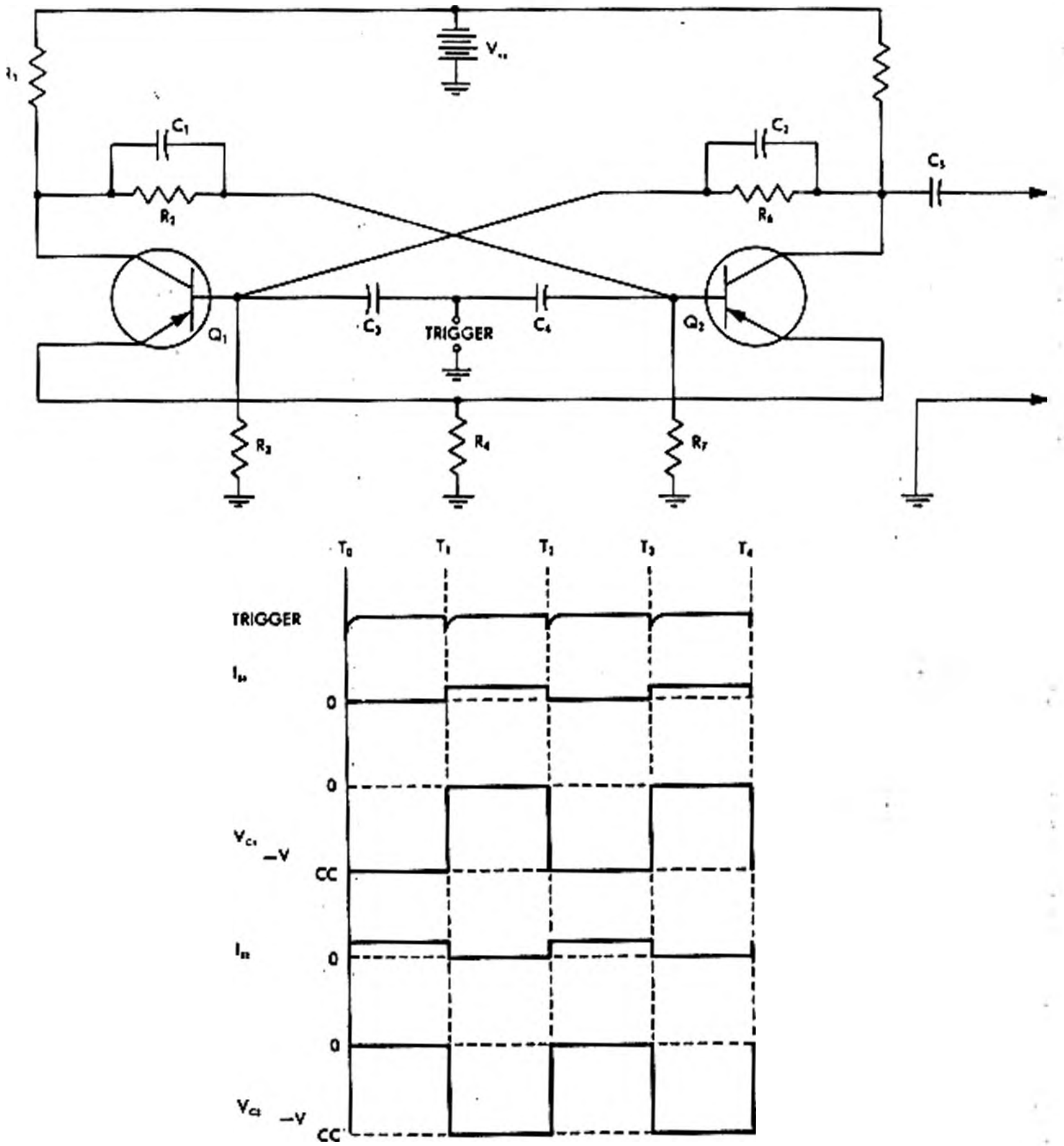


Figure 6-81. Transistor Eccles-Jordan Multivibrator and Waveforms

state until the next negative trigger pulse is applied to the base (time T1). At time T1 transistor Q1 conducts and Q2 is cut off. This completes one full cycle of the output square wave. The time constant of capacitor C1 and resistor R2 and that of capacitor C2 and resistor R6 essentially determine the time from conduction to cutoff of transistors Q1 and Q2, respectively. In addition, the capacitors rapidly couple the changing voltages to the bases to insure rapid switching action of the transistors.

The output signal of the bistable multivibrator is coupled through a capacitor (C5) to the load. Although the output is taken across transistor Q2, it could have also been taken across transistor Q1. Frequency division of this transistor multi-

vibrator is the same as that in the electron tube version, that is, a ratio of 2 to 1.

Resistors R1 and R5 provide the DC collector load for their respective transistors. The necessary forward bias for transistors Q1 and Q2 is provided by a voltage divider network consisting of resistors R3 and R6 for Q1, and R7 and R2 for Q2. Resistor R4 has a twofold purpose: to provide emitter temperature stabilization and to insure a bistable condition within the circuit. Capacitors C3 and C4 serve to couple the input trigger pulse to the base of each transistor.

The multivibrator circuit lends itself well to a binary system using the digits 0 and 1. Nonconduction of one of the transistors could represent the 0, while the conduction state could represent the 1.

There are many adaptations of the flip-flop principle. As presented in the next part, this principle can be used equally as well in the memory circuits of the computer.

Memory Unit

A memory unit stores information until it is called for in the computation. A memory unit is actually the "backbone" of the computer. The term "delay circuit". is often used in connection with the memory unit of the computer. This delay is in actuality a form of storage—a short-time storage system. The usage of the term "delay" is usually with regard to a particular operation that is to be performed after some other operation. "Storage" refers specifically to the retention of some information that will be needed in later calculations.

Here again, in memory circuits, a diversity of components are used in a computer. Some of the components common to the input and logical units are also used in the memory circuit. This is true of the punched tape, magnetic recording, and the switch. There are some circuits that are quite limited as to their uses, and some of these that find their specific use in the memory unit are presented in the following paragraphs.

One of the most adaptable means of high-speed reference storage is the sonic delay line illustrated in figure 6-82. Sonic delay lines, as the word implies, are lines that function by transmitting sound pulses through a medium of liquid, solid, or gas. The one in the illustration is a liquid-type delay line, using mercury as the transfer medium.

In this delay line, two quartz crystals are used, one as a transmitting element, the other as a receiving element. Quartz crystal is an efficient material in that it exhibits excellent piezoelectric effects, and the acoustical impedance of quartz and mercury are comparable. Matching of acoustical impedances indicates that there will be an optimum transfer of sound energy relayed between the mercury and quartz. Suppose, for example, that a signal is fed into the delay line and that it has the coding of 101011 (in accordance with the binary system). The quartz crystal receives this code as "pulse, no pulse, pulse, no pulse, pulse, pulse." The quartz crystal, in turn, sets up sound waves within the mercury that have the same rhythm, or repetition. The sound waves in striking the quartz crystal at the opposite end of the delay line reproduce the electrical impulses. These pulses of energy are then amplified and returned to the input of the delay line where this cycle of energy transfer is repeated.

This cycle of transmitting and receiving the impulses is continued until a gating circuit is operated that will either transfer the energy to other components or clear it from the system.

Another means of storing information is by electrostatic storage—cathode-ray tube storage.

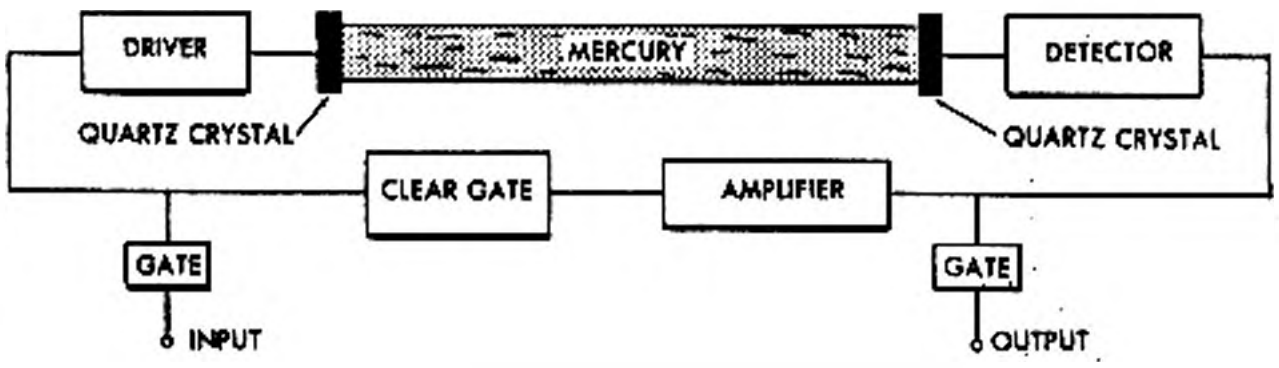


Figure 82. Mercury-Tank Sonic Delay Line

Shown in figure 6-83 is the basic electrostatic storage tube. In this case, a beam of electrons is directed upon a surface composed of conducting and non-conducting segments. As the electrons strike this surface, an electrical charge is created.

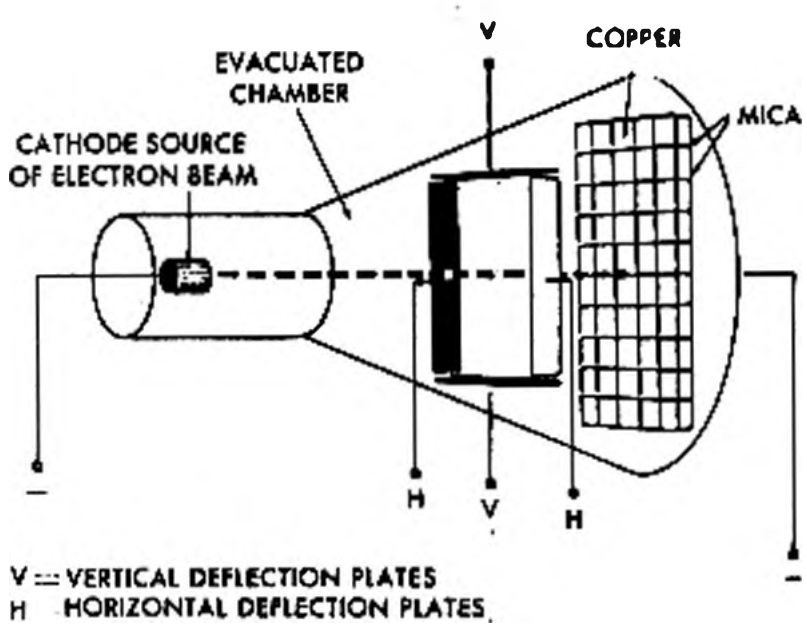


Figure 6-83. Electrostatic Storage

The presence or absence of the electrical charge gives an indication of 0 and 1 conditions. Thousands of units of information can be handled readily by the electrostatic tube. The information can be stored or erased readily, and in conjunction with other circuits, this information can be retained for long periods of time.

Phosphorescent materials are also used for digital storage. In this case, a cylinder (drum) is coated with a phosphorescent material and energized with electromagnetic radiation. The irradiating spots are then read off by means of a photoelectric cell. Because of difficulties with light shielding, with heat interference from the system, and with optical interference, this system of storage is limited in application.

Output Unit

After all computations have been completed, there must be some means of presenting the answers to the operator. Any of the devices that are used for putting information into the computer may also be used to take information from the machine. This includes recorded tapes of all types, and relays whose positions may be indicated by light bulbs.

Geared indicators, similar to those on the common desk-type calculator may also register the results.

In an airborne missile system, the results of the computer may be used directly to influence the control system or, perhaps, the sensor unit in a directional system of guidance.

Mechanical Systems

So far the operational techniques and a few of the common types of components used in the digital computers have been presented. Table 6-2 lists the various functions of a digital computer along with the common types of devices that may be used in each.

Note that during the discussion no mention was made of any mechanical systems used within the

Table 6-2. Functions of a Digital Computer

INPUT UNIT	LOGICAL UNIT	MEMORY UNIT	OUTPUT UNIT
Tapes	Switches	Tapes	Tapes
Punched paper (or cards)	Gates	Punched	Punched
Photographic	Vacuum tubes	Magnetic	Magnetic
Magnetic tape or wire	Tapes	Photographic	Photographic
Buttons	Punched	Sonic delay line	Switches (lights)
Switches	Magnetic	Electrostatic	Relays (lights)
Relays	Photographic	Phosphorescent	Geared indicators
Mechanical	Flip-Flop (multivibrator)	Mechanical	Mechanical
Error Signals from missile	Relays		Missile correction of errors
	Mechanical		

digital computing machines. These were not omitted because of unimportance, but because the mechanical types are generally peculiar in their construction to fit a particular job. For example, it may be desirable to use some type of a gearing mechanism to perform a calculation. This gearing may involve several types and arrangements to fit the purpose.

Many mathematical functions such as addition, subtraction, multiplication, or division may be handled within the computer. Consider the addition processes performed by a computer. The first step in making ready the information for the binary digital computer is to put it into the binary numbering system. This binary form of the information is then fed into circuits that perform the logic operation of "and." The memory units come into operation whenever there is any carrying to be done. For example, if the two decimal numbers 9 and 5 are to be added, they are converted to their binary equivalents 1001 and 101, respectively. Then the last two digits of the binary number enter an "and" circuit. This 1 plus 1 gives the number 2 or, in binary notation, 10. Therefore a 1 has to be stored for a short time. Further "and" circuits then operate to give a final binary answer of 1110. This process is illustrated in figure 6-84 for further clarity.

The process of subtraction would be similar to that of addition except that circuits employing "not" and "or" logical patterns would have to be used. These circuits may use flip-flop circuits or relay systems in addition to adders. Another method of subtraction in computers is to take complements of numbers and then perform algebraic addition operations.

Any multiplication process can be, and usually is, performed by a series of additions. Likewise, the division of a group of numbers may be accomplished by the repeated process of subtraction.

Any mathematical process can be duplicated by the use of digital computers. Problems involving square roots, trigonometric functions, or differential equations can be handled with great speed and accuracy.

COMPUTING DEVICES IN THE MISSILE FIELD

Many computing circuits and devices are used in the missile field. In some cases, they are contained in separate and distinct units and are identified as computers. Other units in missile systems often have computing circuits and devices as integral components. These components are often not readily apparent and identifiable. In future study of missile systems, keep in mind that many missile electronic, electrical, and mechanical systems and subsystems contain elements for performing some computing functions.

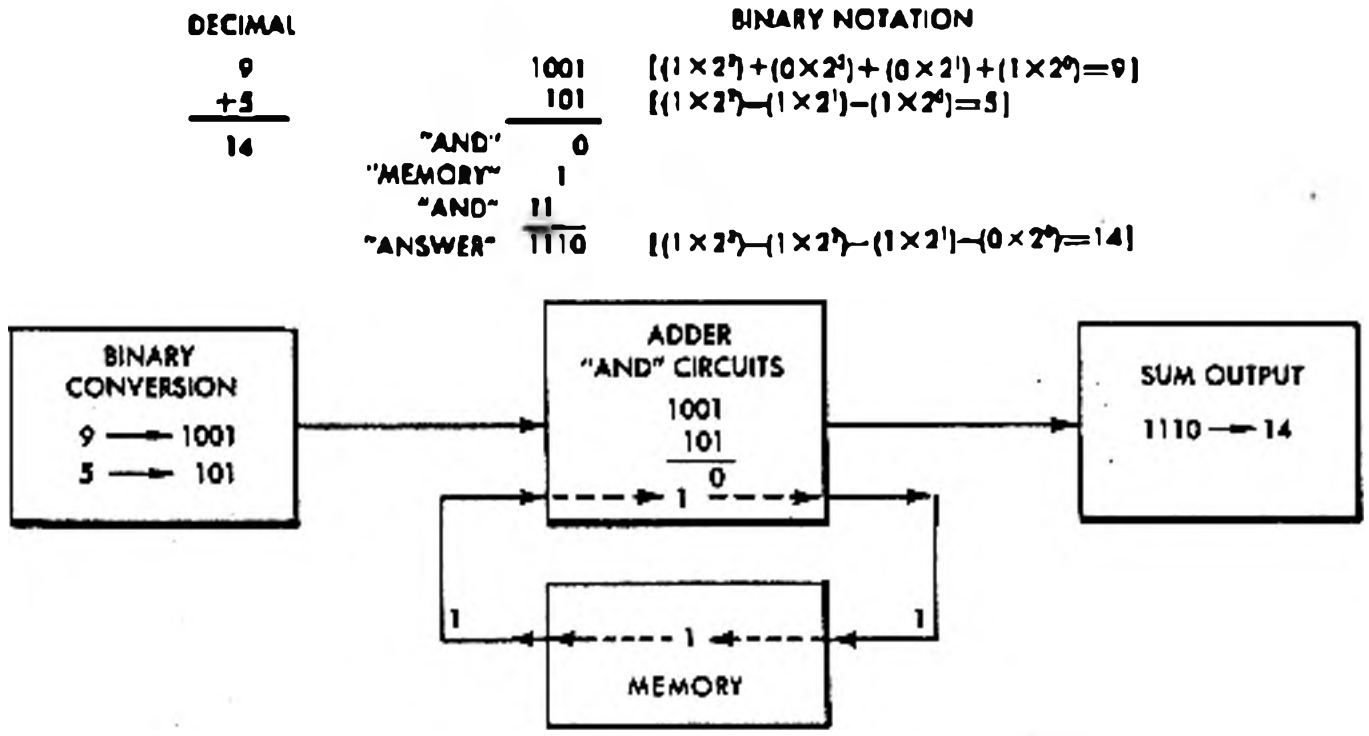


Figure 6-84. Binary-Digital Addition

SECTION C
Reference Units

The reference unit employed in a missile guidance system provides standards for generation and synchronization of electrical impulses and for timing of electronic circuits to insure proper functioning of the guidance system.

Guidance references are classified broadly into two groups:

1. External references (associated with the base station, ground radar, and transmitters).
2. Internal references (located within the missile).

EXTERNAL GUIDANCE REFERENCES

External guidance references are considered to be those that are outside the missile but essential to the operation of the system in controlling the course of the missile. Included in this category are radio and radar beacons, celestial bodies, and topographic or gravitational characteristics of the earth.

Radio and radar beacons find primary usage as reference units in short range missiles. In systems employing radio or radar reception, the position of the missile is accurately established by one or more beacon transmitters.

Reference, in some cases, is established by homing on an enemy transmitter. It also is possible to plant a transmitter in enemy territory. A small missile is used to plant the transmitter. After the transmitter is planted, a large missile carrying a heavy payload can be accurately guided to the target.

Hyperbolic grids (navigational networks) are another type of radar external references. These references are generally used in establishing the initial part of the flight only.

Celestial bodies also afford an accurate external reference. Using precise star tracking systems, along with predetermined information, a missile is programmed throughout its course. Stars make an excellent reference for long-range missiles.

The topographic and gravitational characteristics of the earth are additional reference standards used either for short- or long-range systems. Topographic features of the earth are used for reference in all map-matching systems. Television systems may also be considered as using the topography of the earth.

Gravitational influences of the earth serve to establish the reference for all of the inertia types of guidance systems. Plumb bobs, gyros, accelerometers, and other inertial devices establish the position of the missile by making reference to the gravitational influence of the earth.

The various units that may be used as external reference are not taken up in detail here. Some of these units are described in a later discussion on guidance systems. The rest of this presentation of reference units relates to the internal reference units—those found within the missile itself.

INTERNAL GUIDANCE REFERENCES

Although some of the internal reference units considered here, such as stabilized platforms, make use of outside references, they are covered as internal reference units because they are contained within the missile.

Timing Controls

Consider first the various methods of timing. One of these methods or various combinations of these methods might be used in a guidance system.

OSCILLATORS. Oscillators used as timing controls and frequency standards must possess a high degree of stability under all operating conditions. This means that they must be relatively insensitive to variations in power-supply voltages, temperature, and pressure.

They must provide output voltages of constant frequency and amplitude to serve as standards or time bases for such guidance-system units as trajectory playbacks, scanner motors, fixed fields of gimbal and platform torquers, gyro rotors, gyro torquers, and microsyn signal generators. All of these applications demand a high degree of accuracy at relatively low operating frequencies.

Crystal-controlled oscillators possess most of the desired features except that of low-frequency operation. Because of piezoelectric characteristics, they are restricted generally to frequencies of 100 kHz or higher and are susceptible to variations in temperature.

Relaxation-type RC oscillators employing gas-

cous tubes operate at very low frequencies but tend to be somewhat erratic. They require a constant-voltage power supply. Multivibrator-type oscillators also are subject to these limitations.

To meet the requirements of a missile guidance system reference, many modifications of the foregoing oscillators have been tried. Two such modifications are crystal-controlled oscillators with frequency dividers, and multivibrators with regulated and compensated power supplies.

One satisfactory time-base generator developed to date is a tuning-fork oscillator. At a frequency of 960 Hertz, this type of oscillator can be stabilized to a small frequency error over a wide range of temperature.

The tuning-fork oscillator shown in figure 6-85 is controlled by a 960-Hertz tuning fork in the grid and plate circuits of the 6AU6 tube. As the tuning fork vibrates, the inductance of the coil in the grid circuit changes, producing a signal on the grid of the 6AU6 pentode. This signal is taken from the plate of the pentode and coupled to the grid of the triode-connected 6AU6 amplifier.

From the plate of the triode-connected amplifier, the signal is capacitively coupled back to the driving coil of the tuning fork which is wound to produce a voltage of the same phase in the grid winding. This provides the regenerative feedback necessary to sustain oscillation.

A portion of the signal on the plate of the triode-connected 6AU6 is coupled through a 0.1-μf condenser and a varistor, then back to the cathode of the pentode 6AU6. This provides degenerative feedback to control the oscillator. If the oscillations at the plate of the triode tend to increase because of a change in temperature or some other transient condition, more degenerative voltage is fed back to the cathode of the pentode stage, thus decreasing its gain.

The 200K potentiometer in the grid circuit of the oscillator controls the amplitude and phase of the grid voltage. This serves as a frequency adjustment, preventing a frequency change that may be caused by too much current in the driving coil of the tuning fork.

One sinusoidal output from the 960-Hertz tuning-fork oscillator is fed through a twin-tee or parallel-T filter designed to pass only a 960-Hertz signal. The output is then applied to a gyro signal generator or other circuitry which uses synchronizing or timing signals of this specific frequency.

Another output is passed through a buffer amplifier to frequency-dividing or frequency-multiplying circuits to produce other timing frequencies such as 60, 120, 240, 480, 1920, 3840 Hertz, etc.

Various other time-base generators suitable for applications requiring a lesser degree of accuracy are discussed later in the text.

FREQUENCY CHANGERS. Frequency changers are either mechanical or electronic devices used to

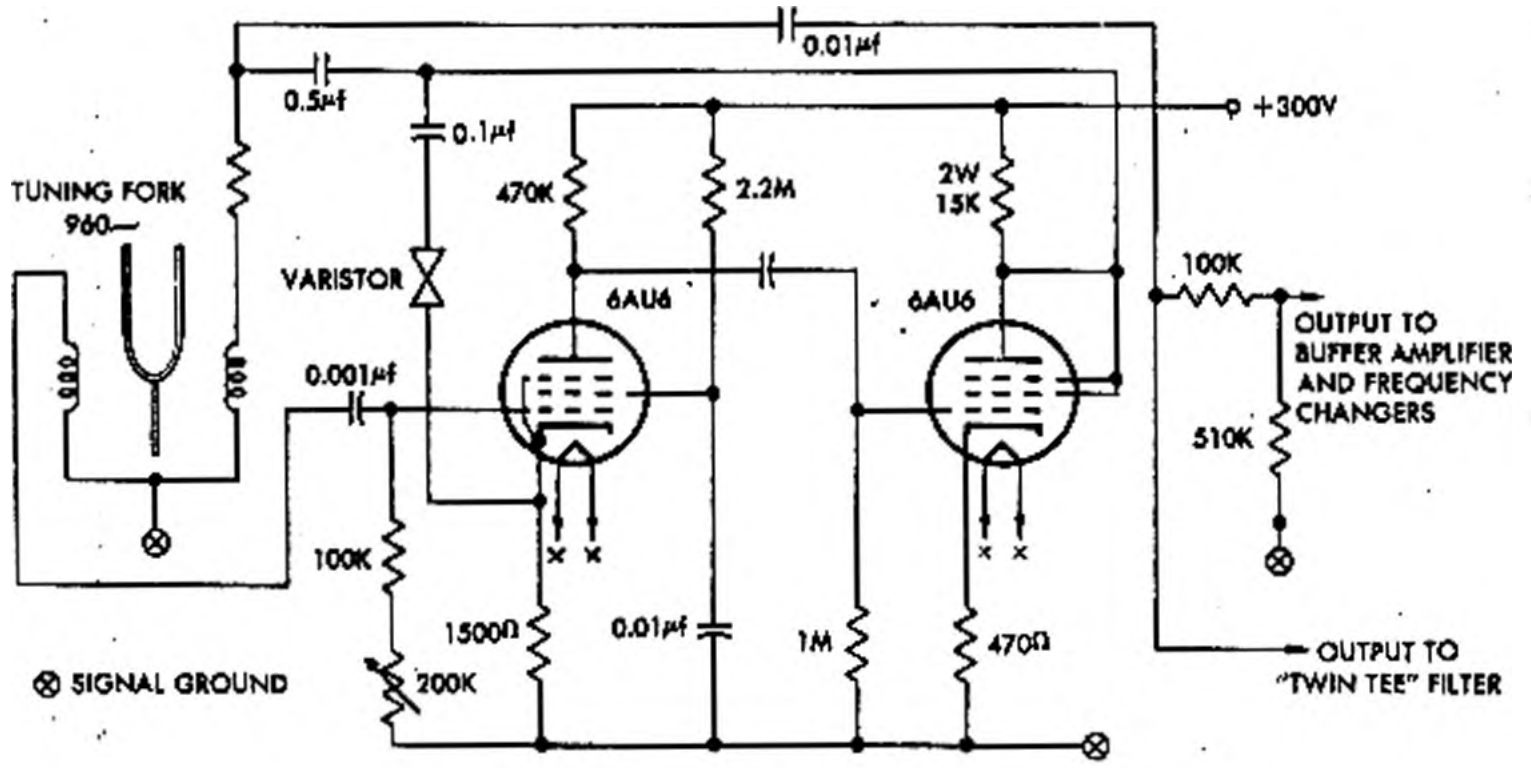


Figure 6-85. Tuning-Fork Oscillator

divide or multiply the output frequency of a standard signal generator. If the application requires a signal frequency within close tolerances, the frequency-changing device must possess the same degree of accuracy as the standard time-base generator. This requirement restricts the use of mechanical devices, such as choppers and vibrators, as timing controls or frequency standards in missile guidance reference units.

Where even submultiples are desired such as f/2, f/4, f/8, etc., of a standard reference frequency, a bistable multivibrator can be used as a frequency divider. This device produces a squarewave output with a frequency equal to one-half that of the input signal, and either half of the multivibrator may be conducting when no input signal is present. Two or more such devices can be cascaded to produce even submultiples of the standard frequency signal. The bistable multivibrator shown in figure 6-86 can be used for dividing frequency by two. The division is determined by the values of the resistors and capacitors in the circuit.

Assuming a 960-Hertz input to the multivibrator, the output frequency would be 480 Hertz. These signals could then be coupled to cathode followers through harmonic filters so that the 480-Hertz sine waves are obtained from cathode followers for reference purposes in synchronous and test goniometers, guidance pulse generators, or any other circuitry requiring a reference signal of 480 Hertz per second. This 480-Hertz signal can be applied to a similar circuit for producing a 240-Hertz reference signal and so on down to frequencies of 60 or 30 Hertz, below which this circuit tends to become too unstable for reference purposes.

Frequency multipliers also may be mechanical or electronic, but as in the case of frequency dividers, they must possess stability and accuracy in the same degree as the primary frequency reference when used as standards of reference. Few mechanical devices possess these desired qualities, and consequently, they are not suitable for application in missile guidance reference units. At radio frequencies, electronic multiplication is relatively simple. It is accomplished by the use of amplifier stages with tuned grid and tuned plate circuits. The grid circuit is energized by the primary reference signal, and the plate circuit is tuned to a harmonic of the fundamental frequency. This system is widely used. Where power requirements are small, harmonics as high as the 55th have been used. Any inaccuracy in the primary signal is multiplied proportionately in this type of frequency multiplier and may render the signal unsuitable as a reference.

One method of frequency multiplication which

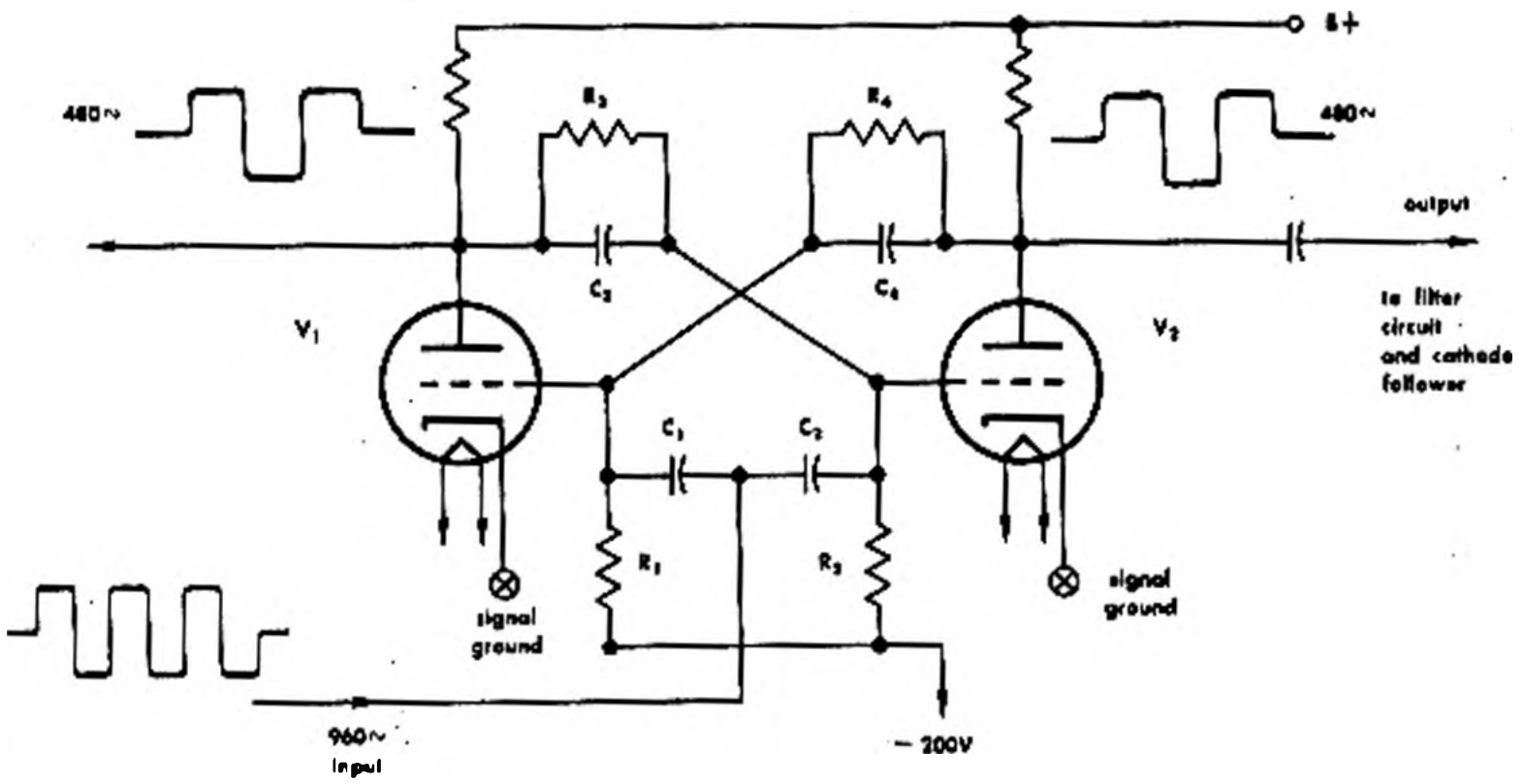
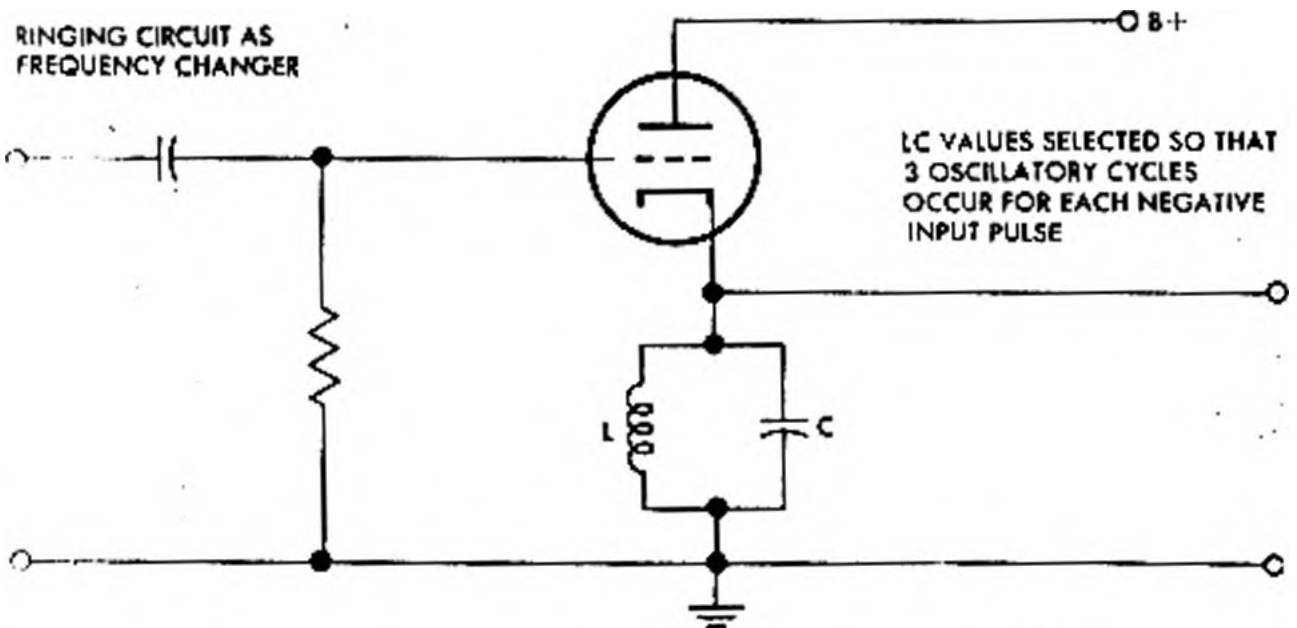


Figure 6-86. Divide-by-Two Multivibrator



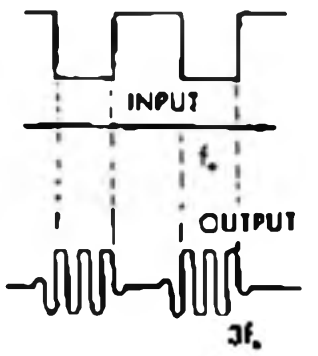


Figure 6-87. Ringing Circuit

is widely used for producing time-bases and range markers in radar equipment employs a circuit known as a ringing circuit, or ringing oscillator. The ringing circuit shown in figure 6-87 employs a suitable amplifier tube with a tunable LC circuit in series with its cathode. A transistor could also be used in place of the tube. The tube is normally conducting until the negative portion of the signal cuts it off. At this time, the field produced by the normal cathode current, flowing through the inductor (L), collapses. Its current flows into the capacitor (C), initiating an oscillatory action. The values of L and C determine the frequency of oscillation; therefore, the primary reference signal is used to start and stop the ringing action of the LC tank circuit.

The number of oscillations produced in the tank circuit is controlled so that the frequency of the oscillations is any desired multiple of the frequency of the triggering signal (primary reference). The controlled frequency must be within the tuning range of the LC tank circuit.

DELAY CIRCUITS. Missile guidance systems frequently employ circuits which require a finite time for their actuation. The timing and sequence of operation of each circuit must be closely controlled with respect to the master signal or data pulse. To obtain the desired sequence of operation, the pulse or waveform is delayed or stored by means of some mechanical or electronic device which has a definite period of transmission between its input and output. An electronic circuit which performs this function is called a delay circuit or storage network. Delay circuits also are employed in radar applications to measure time intervals and form pulses, to synchronize sweep circuits and time-base calibration in oscilloscopes and sweep generators, and to provide channel separation in time sequence in multiplex communication systems.

The type of delay circuit employed depends on the characteristics of the data to be delayed and the time delay required. Some circuits commonly used as delay circuits are artificial transmission lines, start-stop multivibrators, and the phantastron circuit. These circuits and their applications are presented in this section.

Since any conductor possesses inductance and capacitance, a transmission line is considered to be an LC circuit. The characteristics of a transmission line can be duplicated by proper combinations of L and C used as a delay circuit suitable for delay time within the range of 0.01 to 100 microseconds. Some artificial transmission lines consist of wave-filter sections mutually coupled. For long time-delay requirements, so many sections might be needed to provide the required delay and sufficiently high cutoff frequency that some other delay device might be more practical.

For delay time longer than 100 microseconds and up to several thousand microseconds, a pulse is delayed by transmitting it via supersonic waves through a liquid medium, such as mercury. In some applications, such as electronic computers, pulses are stored in mercury ultrasonic tanks for relatively long periods of time.

Transmission lines, such as coaxial cables and parallel conductor lines, are used as delay lines

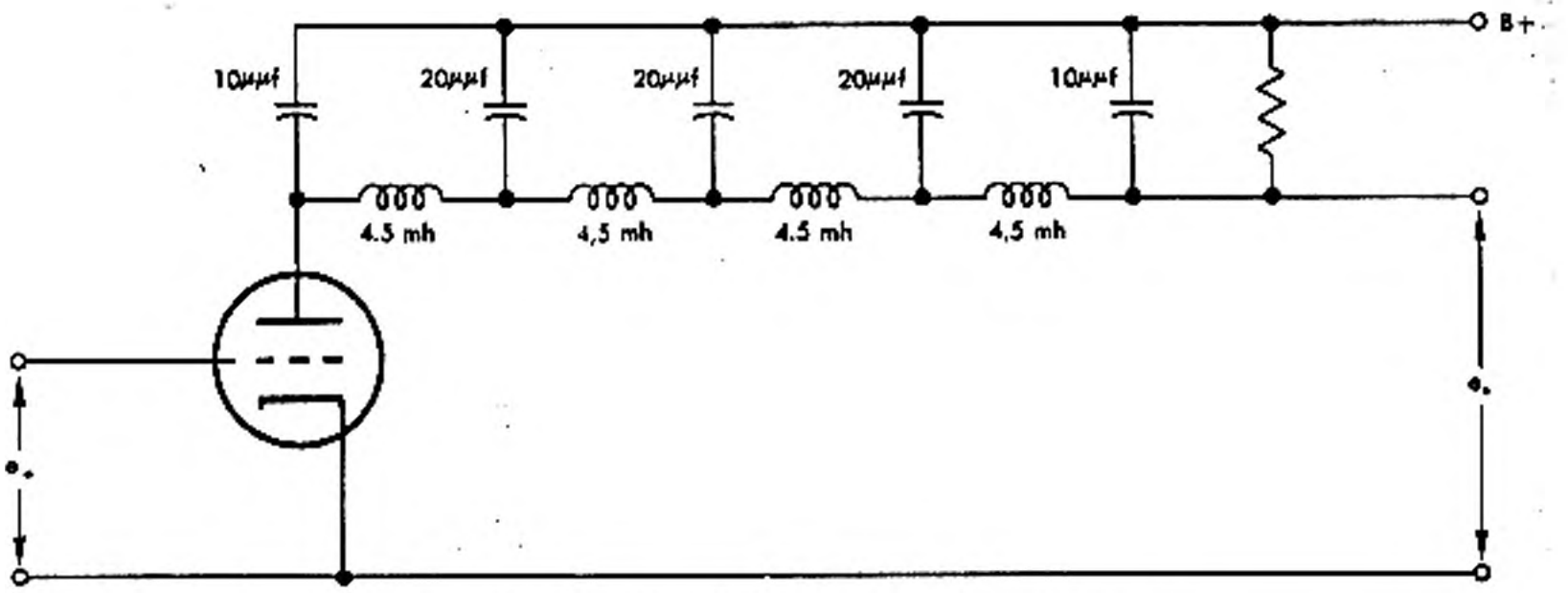


Figure 6-88. Circuit Employing Artificial Transmission Line

when delays of the order of a fraction of a microsecond are required.

The time delay (Td) of the line in figure 6-88 can be calculated by means of the equation

$$Td = N\sqrt{LC}$$

where Td is the time delay in seconds, N the number of sections, L the inductance per section in henries, and C the capacitance per section in farads. Solve for Td as follows:

$$Td = N\sqrt{LC}$$

$$Td = 4\sqrt{4.5\times10^{-3}\times20\times10^{-12}}$$

$$Td = 4\sqrt{9\times10^{-16}}$$

$$Td = 4\times3\times10^{-7}$$

$$Td = 12\times10^{-7} = 1.2\times10^{-6}\ \text{sec}$$

$$Td = 1.2 \text{ microseconds}$$

The characteristic impedance, Z_o, of the delay line in the above example is found from the formula

$$Z_o = \sqrt{\frac{L}{C}}$$

where: L = inductance per section in henries

C = capacitance per section in farads

$$Z_o = \sqrt{\frac{4.5\times10^{-3}}{20\times10^{-12}}}$$

$$Z_o = \sqrt{\frac{4.5}{2\times10^{-8}}} = \sqrt{2.25\times10^{8}}$$

$$= 1.5\times10^{4} = 15{,}000$$

$\therefore Z_o$ of the delay line = 15,000 ohms

If this delay line is terminated in a 15,000-ohm resistor, pulses applied by the tube to the input of the line travel to the output end, and are not reflected. The operation of the tube is the same as if a 15,000-ohm resistor were connected directly in its plate circuit.

When an artificial transmission line is terminated in its characteristic impedance, voltages applied at its input end are dissipated by the resistance termination without reflection. If the line is open-circuited at its terminal end, any voltage applied to the input is reflected back from the open end without change of phase, just as in the real line.

Figure 6-89 illustrates an application of an artificial transmission line used to introduce a delay so that V_2 is operated at a slightly later time than V_1. The delay is determined by the values

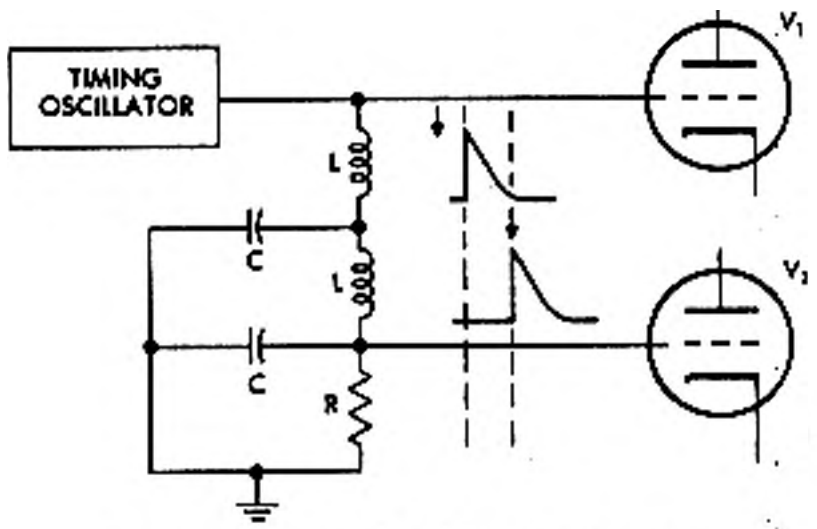


Figure 6-89. Artificial Transmission Line to Introduce Delay

of the lumped L and C constants using the previously mentioned formula:

$$Td = N\sqrt{LC}$$

where N is the number of sections of line used. Since R is the terminating resistance of the line as well as the grid input resistance of V_2, it must be equal to the characteristic impedance of the line to prevent reflections.

An artificial transmission line has characteristics similar to a low-pass filter. Such a filter tends to pass all frequencies below its cutoff value (F_c) and to reject all frequencies higher than the cutoff value. The greater the number of sections in the line, the more sharply defined is the cutoff frequency, which may be expressed as

$$F_c = \frac{1}{\pi\sqrt{LC}}$$

Synchronizing pulses usually either are rectangular in shape, or they consist of a sudden rise in voltage followed by a slower exponential delay. Such waveforms include a considerable number of high-frequency components. If the delay line is to pass such a wave without distortion, its cutoff frequency must be high enough to pass the necessary frequency spectrum.

In the guidance systems of missiles, radar pulses are frequently used to initiate the operation of various circuits within the missile.

When the same pulse is used to trigger more than one circuit, and where some specific time interval or sequence of operation is desired between activation of the circuits, a variable delay circuit must be provided. The start-stop (monostable) multivibrator is suitable for producing the desired delay in many of these applications.

The start-stop multivibrator is usually triggered by a pulse of short time duration, so a square or rectangular wave must first be differentiated and the undesirable portion clipped before it is suitable as a synchronizing pulse. In figure 6-90 the sync pulse is negative, and a transformer is used to reverse its polarity so that a positive pulse is applied at the grid of V_1.

In this circuit, V_1 is normally cut off by the cathode bias developed across R_k, as V_2 conducts heavily. When the positive pulse is applied at the grid of V_1, it will allow V_1 to start conducting, thus decreasing its plate voltage. Capacitor C_1, which was charged to the B+ voltage, will now start discharging through R_2, making the grid of V_2 negative with respect to its cathode. Current flow through V_2 will decrease, which decreases the current through R_k and the voltage drop across it. This allows V_1 to conduct even more, dropping its plate voltage still further and making the grid of V_2 more negative. This action is repeated until V_2 is cut off as shown by its plate voltage waveshape in the illustration. The rise from low plate voltage to B+ voltage indicates that this action, though accumulative, is practically instantaneous.

Tube V_2 will remain cut off, and V_1 will continue to conduct until C_1 discharges sufficiently toward the lowered value of plate voltage of V_1 to allow the grid of V_2 to rise from its lowest value to cutoff voltage. At this point, V_2 starts to conduct, increasing the voltage drop across R_k and reducing the current through V_1. The decreased plate current of V_1 causes the plate voltage of V_1 to increase. This increase, coupled to the grid of V_2, causes V_2 plate current to increase still further. This action is repeated until V_1 is driven to cutoff and V_2 is conducting heavily. Once again the action, though accumulative, is practically instantaneous.

The plate voltage from V_2 is applied to a transformer with a short L/R time constant, so the square wave is differentiated in the secondary. The output voltage has a negative pulse that has been delayed a certain period of time with respect to the input pulse. By the proper selection of circuit elements, the desired time delay can be achieved. The time delay in this circuit can be varied by changing the resistance of R_2, and precise final adjustments can be made by changing the resistance in the plate circuit of V_1.

PHANTASTRON CIRCUIT. The phantastron circuit is a medium precision delay circuit which is similar in operation and output to the "flip-flop" circuit. It possesses an advantage over multivibrator circuits in that it is quite stable under power-supply voltage variations.

This circuit is frequently employed as a delay circuit for timing the sequence of operations in missile guidance systems and in pulse-decoding systems.

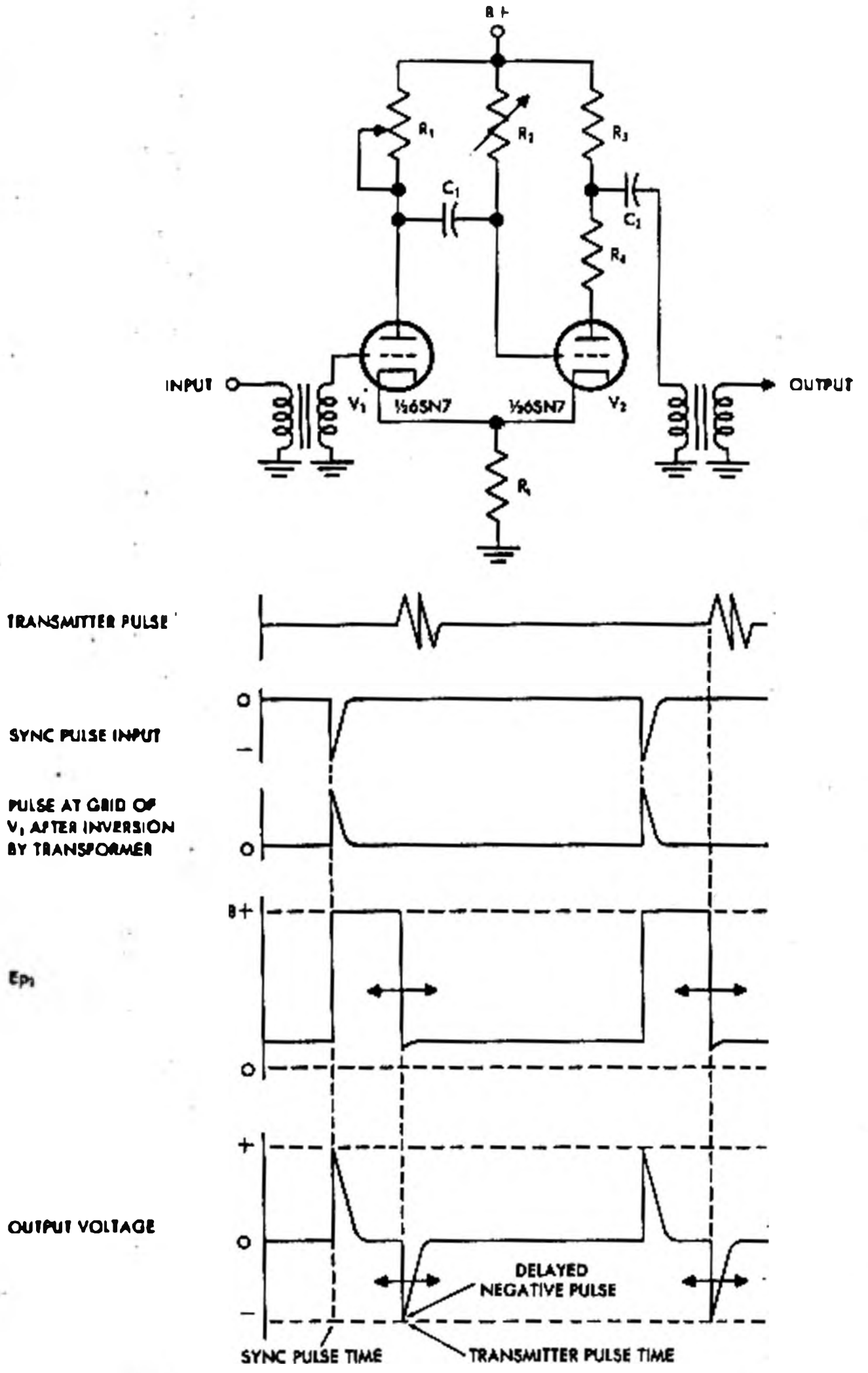


Figure 6-90. Start-Stop Multivibrator as Delay Circuit

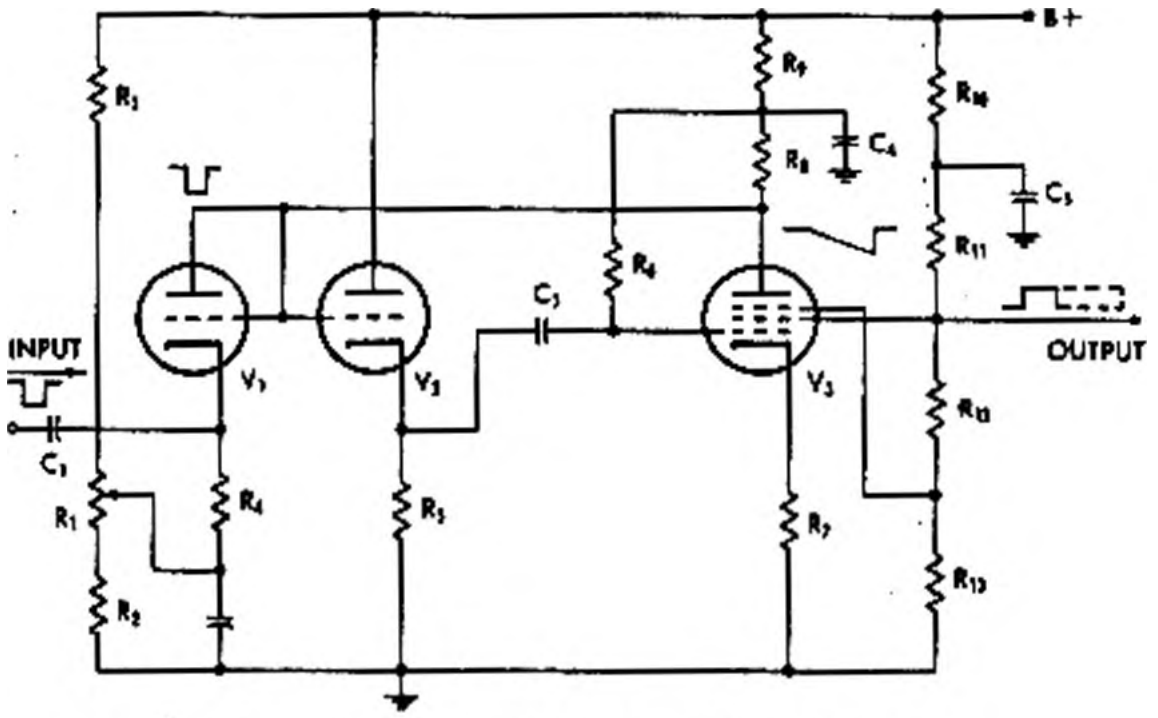


Figure 6-91. Phantastron Pulse Delay Circuit

Figure 6-91 illustrates a phantastron circuit employed to delay a negative input pulse over a range of approximately 50 to 350 microseconds.

Control-grid initiation of delay is used in this circuit instead of the conventional suppressor-grid method. Plate voltage is set by R_1. Tube V_1 is conducting and holds the control grid of the phantastron (V_2) at a constant potential until the signal-pulse appears.

The suppressor-grid voltage is determined by the bleeder chain R_{10}, R_{11}, R_{12}, and R_{13}.

The time delay is adjustable by the setting of R_1, which determines the plate voltage and the initial charge on C_3. The network C_3 and R_6 comprises a Miller integrator between the plate and grid. Tube V_1 serves as an input control diode and V_2 as a cathode follower input.

The operation of a similar circuit employing the Miller integrator is covered later in this chapter. The Miller integrator circuit is used as a pulse shaper rather than as a delay circuit in the amplifying units.

GONIOMETER. A goniometer unit consists essentially of autosyn phase shifters connected by common gearing with external hand control for driving the gear train.

The signal input to the unit is 100 kHz, 10 kHz, and 2 kHz sine wave voltages from the timing

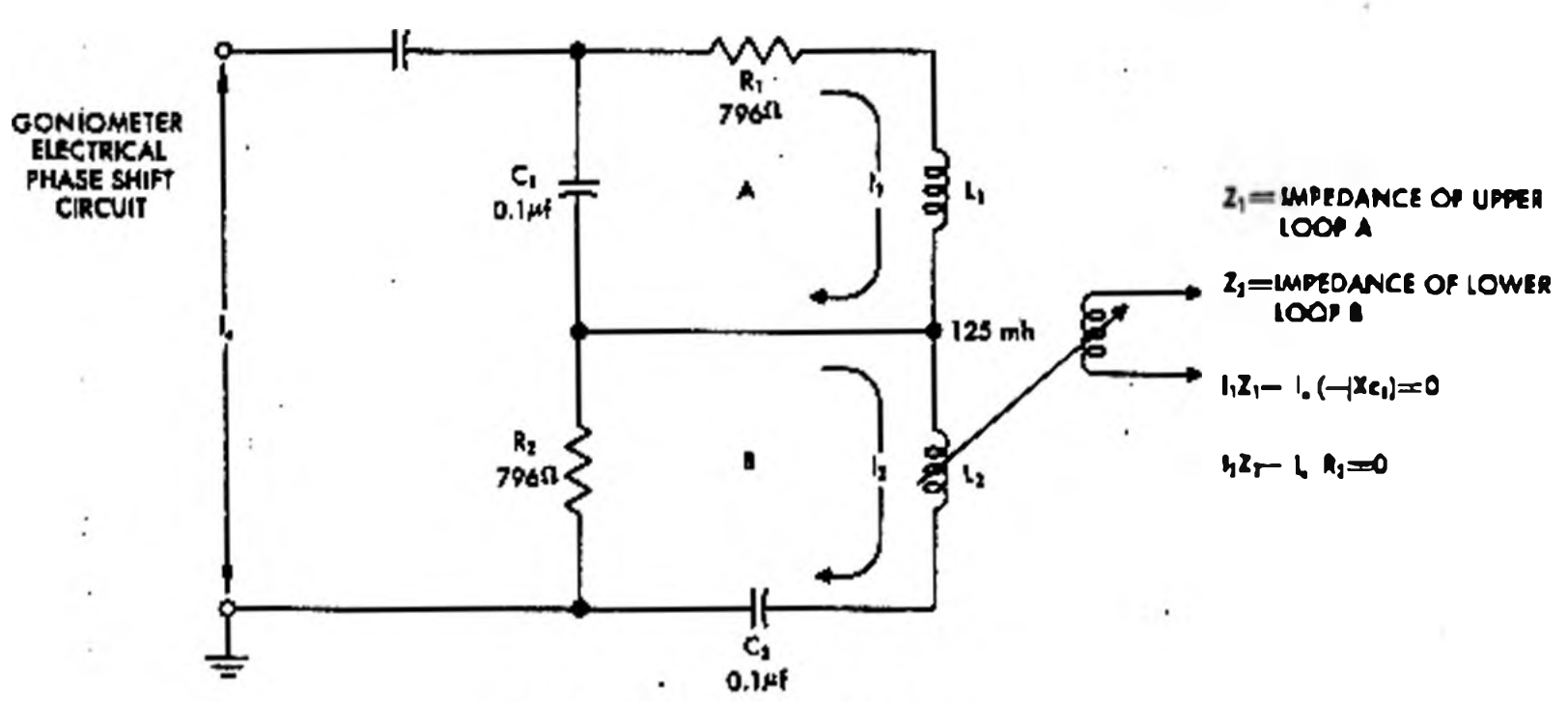


Figure 6-92. Goniometer Circuit

unit, while the output consists of the same sine wave voltages displaced in phase with relation to the input. This electrical phase shift is accomplished by means of an autosyn resolver for each frequency of the input voltage. The autosyns are connected by a gear train in such a manner that the electrical phase shift of the 100 kHz sine wave voltage is 10 times the phase shift of the 10 kHz sine wave voltage and 50 times the electrical phase shift of the 2 kHz sine wave voltage. The phase-shifted output voltages of the unit, when used in conjunction with the pulse generator, originate a video pulse which can be delayed in time by any amount.

A handwheel drive for the goniometer gear train is located on the front panel of the unit to provide a means for varying time delay. A second control is provided for step delays of 100 microseconds.

The 2 kHz sine wave signal is fed into an impedance coil (L_1-L_4 in figure 6-92), which makes the circuit look like pure resistance, and into an RC network across the two coils of the rotor which are wound mechanically 90° apart. The current through capacitor C_1, the coil L_1 of the autosyn, and the resistor R_1 will lead the current 90° through resistor R_2, capacitor C_2, and the other coil, L_2, of the autosyn.

To prove that the voltages are 90° out of phase: The capacitive reactance is 796 ohms; therefore, the currents in loop A of the figure are equal to those in B displaced 90° as follows:

$$\text{Dividing: } \frac{I_1Z_1}{I_2Z_2} = \frac{I_o(-JX_{C1})}{I_oR_2} = \frac{JX_{C1}}{R_2}$$

Since: $R_1 = R_2 = X_{C1} = X_{C2}$; $X_{L1} = X_{L2}$; $Z_1 = Z_2$.

$$\text{Then: } \frac{I_1Z_1}{I_2Z_2} = \frac{JX_{C1}}{R}; \; I_1 = JI_2; \; EL_1 = JE_{L2}.$$

The above formula is the same for any of the three autosyn networks. The output voltages of constant amplitude are furnished to the synchronous pulse generator and the monitor scope.

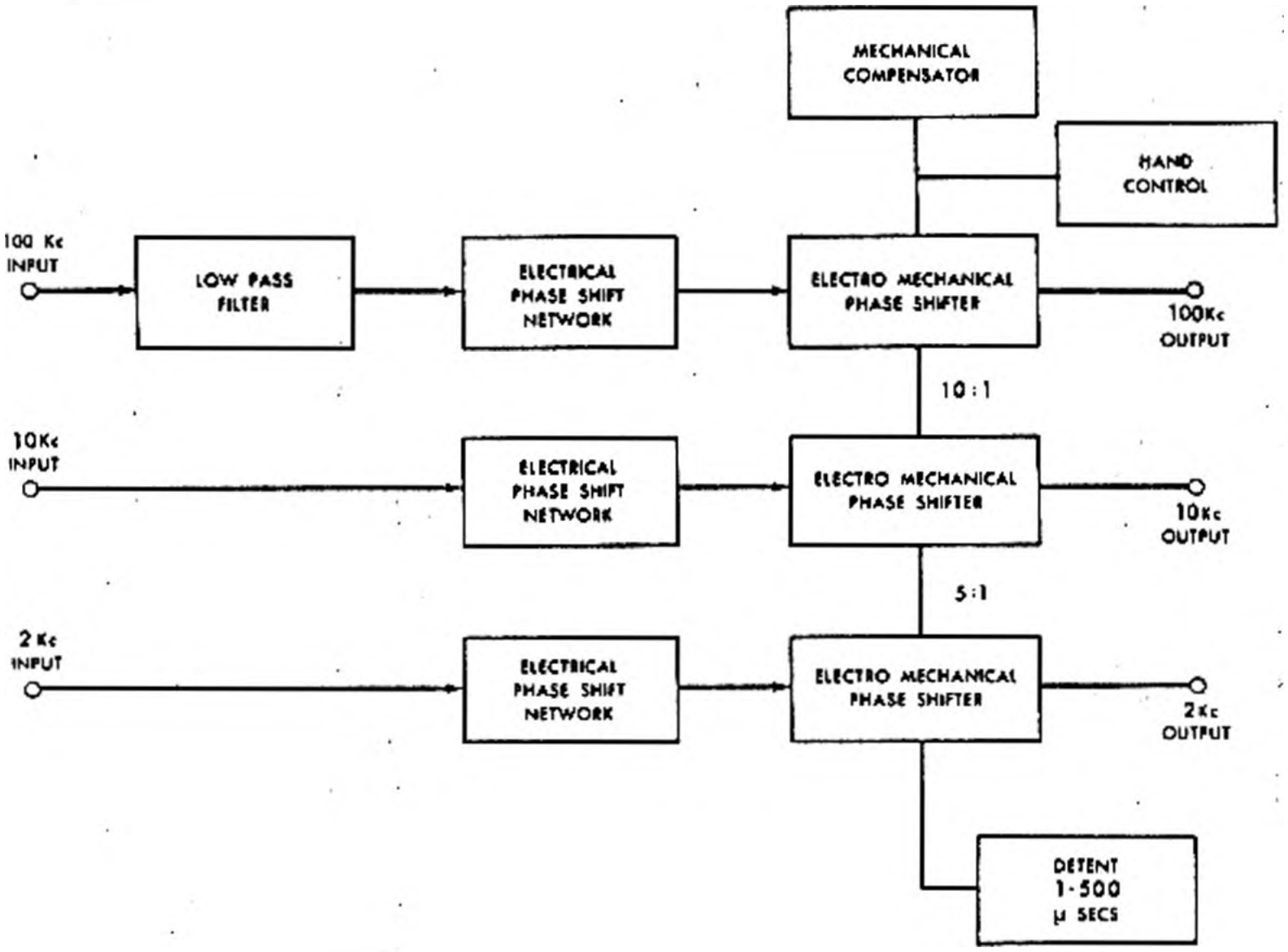


Figure 6-93. Goniometer Unit Diagram

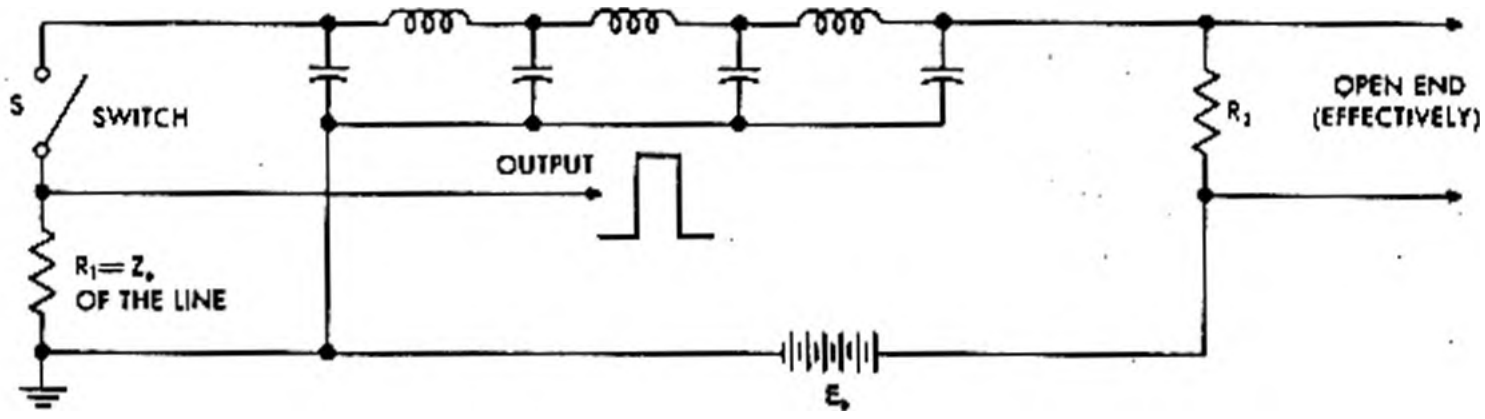


Figure 6-94. Pulse-Forming Line with Mechanical Switch

Figure 6-93 illustrates in block-diagram form the goniometer unit discussed in the preceding paragraphs.

PULSE-FORMING CIRCUITS. To transmit short, powerful bursts of RF energy, the transmitting tube of a microwave radar system is modulated (or keyed) by high-voltage rectangular pulses of short duration. A pulse of rectangular shape is desirable because it allows the transmitter tube to operate with maximum efficiency and frequency stability.

An artificial transmission line is used to create the modulating pulse either at a low-voltage power level (after which it is amplified by the modulator system) or directly at a high power level. Figure 6-94 illustrates a pulse-forming line with a mechanical switch.

The four condensers of the line are charged up to the potential of E_b through a high resistance (R_2). Resistor R_2 has a resistance so much greater than the characteristic impedance (Z_0) of the line that this end of the line may be considered effectively as an open circuit.

When switch S is closed, the line abruptly begins discharging at a constant rate through R_1, which is equal to the characteristic impedance of the line. If the network consisted of the four condensers only, the discharge would follow an exponential curve, and the current flow through R_1 would not be constant. By using a well-designed artificial line which contains both inductance and capacitance, the discharge rate can be held to a substantially constant value as long as the line is discharging.

At the instant the switch is closed, the discharge wave starts traveling down the line toward the open end, causing the wave to lose half of its charge. Upon arriving at the open end, the wave is reflected back toward R_1. When it reaches R_1, the line is completely discharged, and the current flow through R_1 abruptly ceases, terminating the rectangular voltage pulse appearing across that resistance. When the switch is open, the line again charges to the potential of E_b.

A more practical version of the foregoing circuit is shown in figure 6-95. Note that the me-

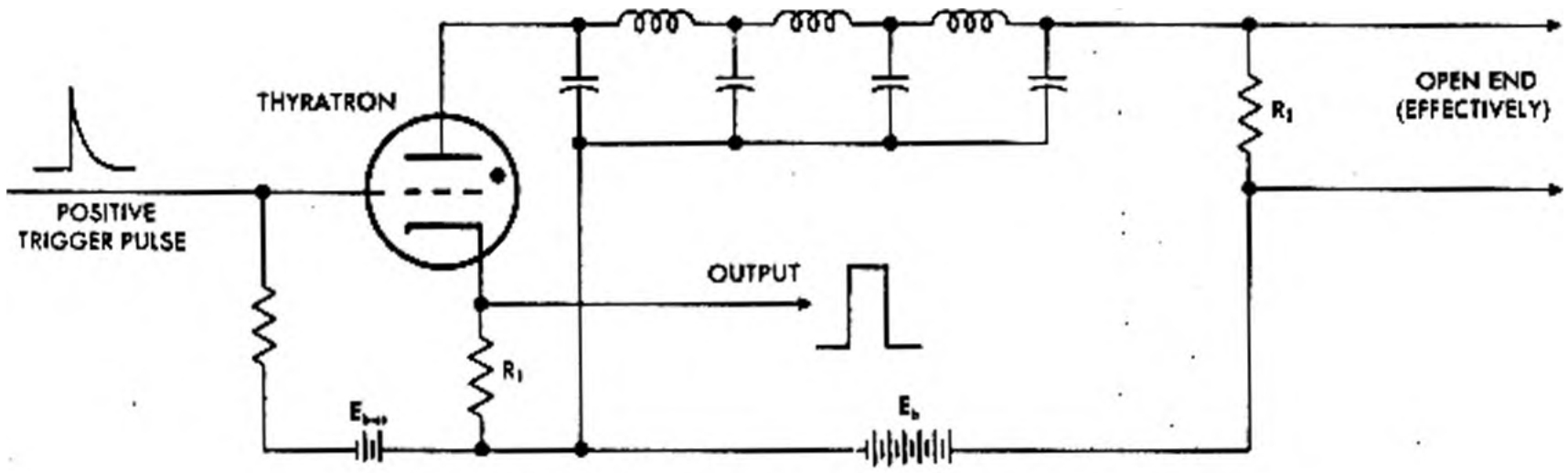


Figure 6-95. Pulse-Forming Line with Thyratron Switch

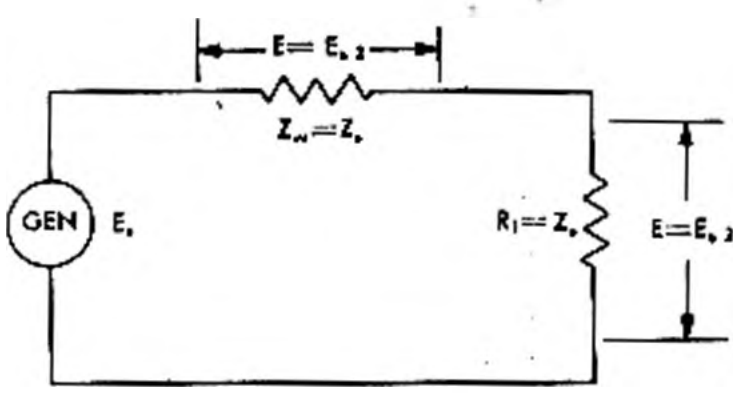


Figure 6-96. Equivalent Circuit of Pulse-Forming Line with Thyratron Switch

chanical switch (S) of the previous circuit has been replaced by a thyratron tube, which permits the microsecond timing required in a radar system. The gas tube is normally non-conducting because of the negative bias on its grid. This allows the line to charge up to the value of E_b through R_2. The line is switched to its terminating resistance R_1 when a positive trigger pulse drives the gas tube into conduction. The current for this conduction is furnished by the line at a constant rate, causing a rectangular wave of voltage to appear across R_1 as long as the line discharges.

As the line becomes completely discharged, the plate voltage of the thyratron falls below the ionization point. As a result, the tube becomes nonconductive because, in a short period of time, E_b is unable to supply enough current through the high resistance of R_2 to maintain the plate voltage. When the thyratron ceases to conduct, the line again charges up to the potential of E_b. A relatively long charging time can be allowed since it normally takes place during the interval between transmitted pulses.

The charged line can be considered as a generator with an internal EMF of E_b and an internal impedance of Z_0. At the time the thyratron fires, the line (or generator) is effectively thrown across resistance R_1, which is equal to Z_0 and in series with it. This equivalent circuit is shown in figure 6-96.

A voltage equal to one-half E_b appears across R_1, and at the same instant a voltage wave $-E_b/2$ starts down the line. This wave is reflected back from the open end in phase and arrives back at R_1, having completely discharged the line. At this time, the thyratron ceases to conduct, and the voltage across R_1 drops abruptly to zero. The potential has been maintained across resistor R_1 during the time required for the voltage wave to travel down the line and return. The thyratron merely starts the action and the time of its conduction is determined by the characteristics of the pulse-forming line.

It was stated earlier that for a delay line, the time delay is equal to $N\sqrt{LC}$. In the case of a pulse-forming line charged to a given potential, the voltage wave must move down the line and back, producing a time delay twice as great. The width of the voltage pulse developed across the terminating resistance is determined by

$$T = 2N\sqrt{LC}$$

while the pulse repetition rate depends on the fre-

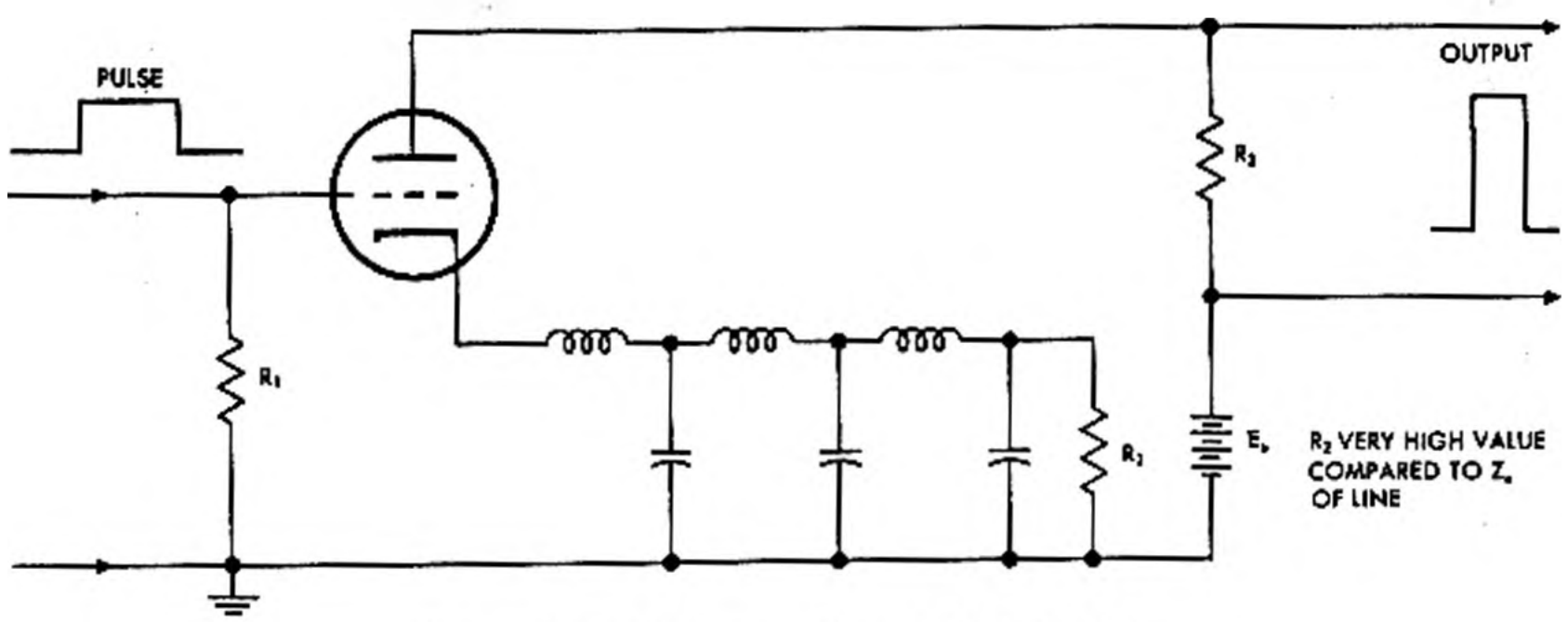


Figure 6-97. Pulse-Forming Line with Triode Amplifier

quency of the timing oscillator which triggers the thyratron switch tube.

A variation of the pulse-forming circuit is shown in figure 6-97. In the circuit, the pulse-forming line is in the cathode circuit of a triode amplifier tube.

Assume that the amplifier tube is normally nonconducting and that it is desired that the tube produce a rectangular pulse of short duration at the output. The cathode resistor of the amplifier is replaced by an artificial line, and the tube is brought into a conducting state by the sharp rise in voltage as the leading edge of a rectangular wave is applied to its grid. At this instant, plate current increases sharply and establishes a voltage wave between cathode and ground. This voltage starts down the line, and upon reaching the effectively open end (R_2), it is reflected back without change in polarity, arriving between cathode and ground after a time $2N\sqrt{LC}$. This sudden increase in voltage on the cathode is sufficient to cut off the tube. The leading edge of the output pulse (appearing across R_3) is formed as the tube is placed in conduction and coincides with the leading edge of the pulse on the grid.

The constant amplitude of the output pulse is maintained by the characteristics of the line. The trailing edge of the output pulse is formed as the line cuts off the tube at the time $2N\sqrt{LC}$ and bears no relationship to the grid voltage. After the tube is cut off, the line discharges slowly through R_3 until the cathode voltage becomes low enough to allow the tube to conduct, or until a pulse again drives the tube into conduction, initiating a new cycle of operation.

The foregoing circuit does not necessarily have to be an amplifier stage. It could be an RF oscillator, or a transmitting tube, with an artificial line in its cathode circuit. Such an application would determine the width of the RF pulse transmitted whenever the transmitter circuit is triggered by the timing oscillator.

While the standard artificial line is quite satisfactory to produce a time delay, it has some limitations as a pulse-forming device because each section of the line tends to produce a hump, or ripple, in the flat top of the rectangular-wave output. This defect can be minimized by using a large number of small units to obtain the required amount of inductance and capacitance. The resulting humps, or ripples, while more numerous, would be of much smaller amplitude. This would result in an artificial line more closely approximating a real transmission line. The artificial line composed of a large number of small LC units is less compact and is more complex in design.

When pulses are formed by artificial lines at low power levels and then amplified, some deviation in the pulse shape can be tolerated. Limited deviation is permissible because amplifiers can be operated so as to effect a small amount of plate limiting and thus clip the uneven flat top before it is applied to the transmitter tube.

When pulses are to be developed initially at high power levels for direct application to the transmitter tube, little can be done to improve the pulse shape when using a standard artificial line as the pulse-forming medium. To meet the rigid pulse-forming requirements of this latter application, a special type of line, known as the Guillemin line, has been designed. This line is shown in figure 6-98 as employed with a mechani-

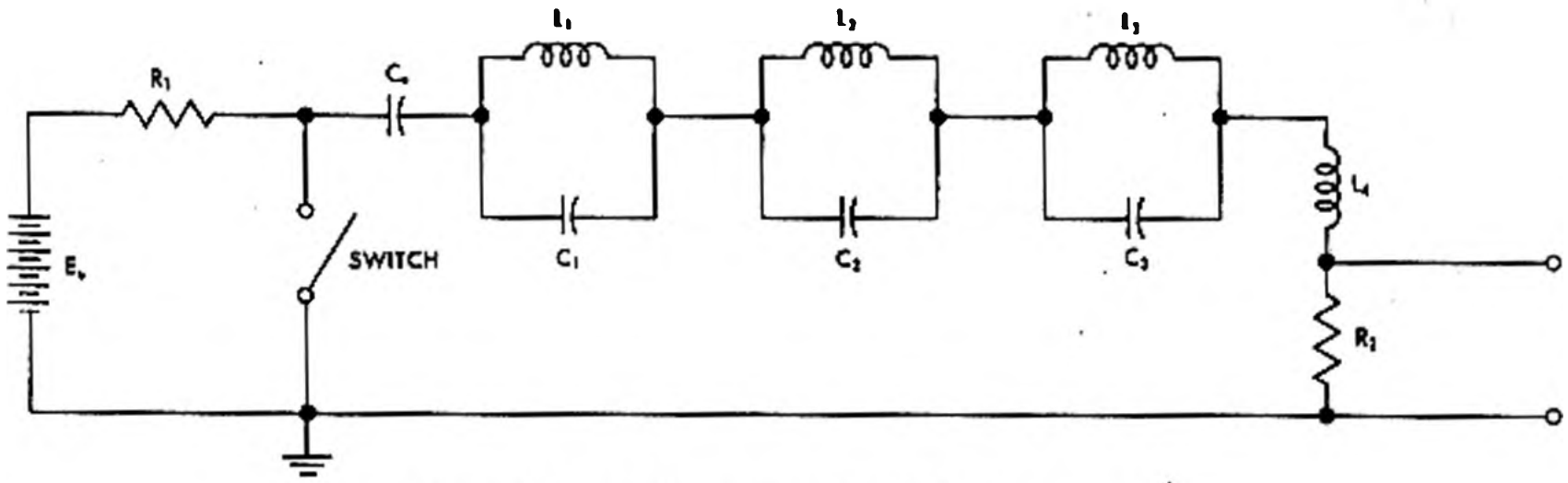


Figure 6-98. Guillemin Line with Mechanical Switch

cal switch. The line is charged through R_1, which could be replaced by a choke. The main storage capacitor of the line, C_0, must withstand the full potential (E_b). Closing the switch causes the line to discharge through R_2. The line is so designed that on discharge a more nearly constant current flows through the terminating impedance than can be obtained with the standard type of line. A rectangular wave with a more perfect flat top is obtained from the Guillemin line, and it may be used to create pulses of high amplitude for direct application to the transmitting tube.

In practical application, the switch is replaced by one or more thyratrons with their cathode-to-plate paths in series so as to divide the high voltage of E_b. The discharge of the line is effected by triggering the grids of the thyratrons. R_2, in this case, would be replaced by the transmitting tube. The impedance of the line, as well as the impedance of the transmitting tube, would be matched to prevent reflections.

There exist many modifications of the foregoing pulse-forming circuits, as well as other devices which are employed to produce pulses. The circuits presented here are basic circuits and are found in general use in microwave radar as used in conjunction with missiles.

Programmers

A flight programmer is a unit into which precalculated flight data are inserted. The programmer supplies these data to the various control circuits of a missile in the sequence necessary to cause the missile to conform to the predetermined pattern of flight. The programmer may include such units as the air log, Veeder counter, map-matching devices, magnetic-tape pickoffs, and stabilized platforms. These units may be found singly or in various combinations in the programmer, depending on the type of control required or the complexity of the flight pattern.

AIR LOG. An air log is used to determine the range of flight. It operates on the principle of an air screw which makes a specific number of revolutions while moving through air for a given distance and at a given velocity. The number of revolutions per unit of distance depends on the pitch of the screw and the density of the air.

Generally, an air log is attached to the outer surface of the nose of the missile and consists of a small 4-bladed propeller, or screw, mounted on a shaft which drives a reduction worm gear with a ratio of 30 to 1. For every 30 revolutions of the air screw, the driven gear makes 1 revolution.

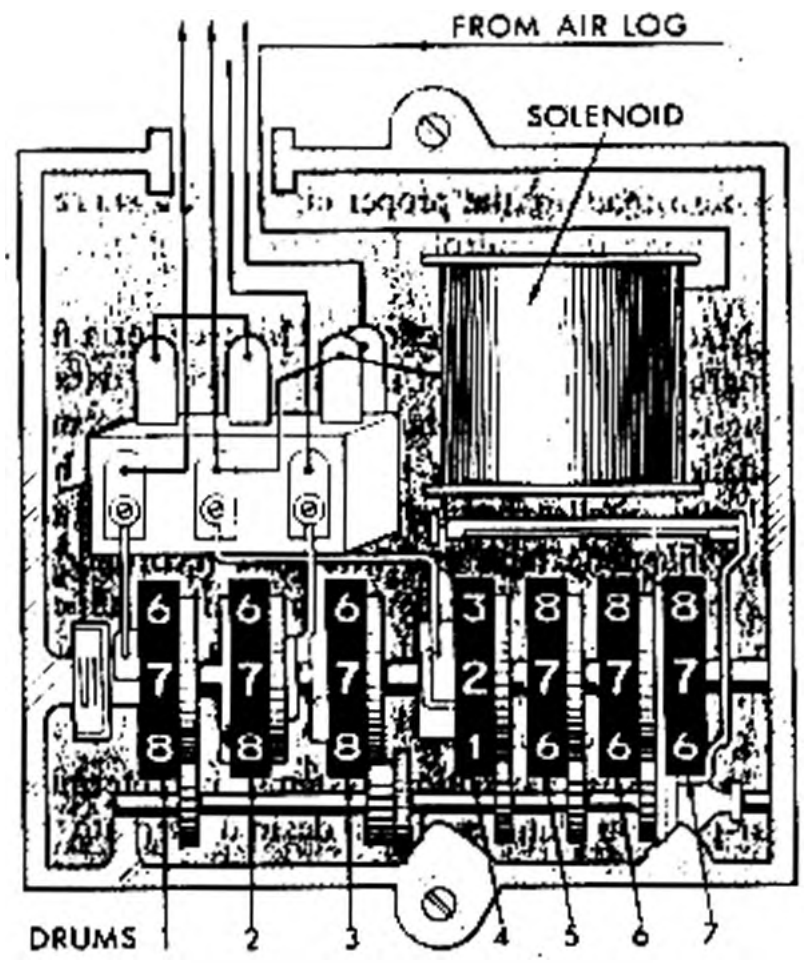


Figure 6-99. Cross-Section of Veeder Counter

The driven gear is machined from hard fibre or some suitable dielectric material, into which is inserted a pair of metallic contact points which make electrical contact with brushes. Two impulses are transmitted for each revolution of the gear. One impulse is transmitted for every 15 revolutions of the air screw.

The contact points and brushes serve to close a magnetic relay circuit in a Veeder counter, once for every 15 revolutions of the air log.

VEEDER COUNTER. A Veeder counter is a device consisting of several calibrated drums as shown in figure 6-99. The drums are driven by a train of gears designed to provide the desired ratio of turns between the drums. The mechanism is similar to that found in the odometer unit of an automobile speedometer. The calibrated drums are rotated and set to the desired length of travel for the missile. Each impulse from the air-log-operated relay represents a specific distance traveled (a specific number of revolutions of the air screw). Each one actuates the Veeder-counter mechanism

so that the calibrated drums rotate from their preset position back toward zero. When the counter reads zero, it indicates that the predetermined dump point and/or arming point has been reached. At zero reading, the proper circuits are energized to position the controls for the terminal dive and to arm the warhead.

MAGNETIC-TAPE PICKOFF. The need for a flight programmer capable of storing sufficient data for long-range flights and having a degree of accuracy sufficient to render it suitable for use in high-velocity missiles, has led to much experimentation with magnetic tape and wire data recording and playback units. These devices make it possible for predetermined flight data to be magnetically recorded on tape or wire and used as a reference in guidance of missiles.

In celestial-navigation systems, predetermined star-position data, such as desired azimuth and elevation angles with respect to the position of the missile at all points or at any specific points along the flight path, are compared with the indicated position data from the star tracker. Any error between the tape data and the tracker data is used to reposition the missile back onto the precalculated course. This necessitates exact recording and playback of data, which requires a timing and synchronizing system with a high degree of accuracy and stability.

When employed in a celestial-navigation system, the tape data specify that at every point along the flight path the angles between the missile and the reference stars should be of some exact value. If the star tracker indicates a discrepancy between the precalculated angle data and the actual position of the missile, the gyros, torquers, and accelerometers on the stabilized platform feed information into the tracking control circuits to bring the missile into the correct angular relationship with the reference stars.

In the trajectory playback system, a time standard and a magnetic-tape playback reproduce (in the missile) information which has been calculated and recorded in advance. This information controls the various circuits of the guidance system according to a predetermined schedule.

Fine steel wire affords a medium for storing a great amount of data on a spool of small size, and it is relatively indestructible. Certain disadvantages make wire recording and playback unsuitable for flight programming. These disadvantages include inaccuracy in sequencing resulting from slippage of the wire on the spool or reel and nonuniform wire speed resulting from variations in the rate of unwinding. As the wire unwinds, the diameter of the storage spool decreases while that of the pulling (capstan) spool increases. This operation gradually increases the rate at which the wire unwinds. No suitable device has been developed for pulling the wire at a constant speed independently of the reels. This disadvantage of variation in wire speed prohibits the use of wire in large spools as would be required for programming long flights.

Positive, uniform recording and playback speeds can be more readily obtained by the use of punched tape or film. Tape and film strips can be designed to provide several channels of data simultaneously. When steel tape or non-shrinkable film is used and more than one set of data are impressed upon the tape, much more data can be stored on tape with less bulk than on wire. Multichannel tape requires the use of a multiple magnetic pickoff head with separate amplifiers and filters for each data channel.

The basic principles involved in magnetic tape and wire recording are the same. The data impressed by electrical impulses are recorded on the tape or wire in the form of magnetized areas. Information is represented by the degree of magnetization in each unit section of tape or wire, or the magnetized areas are coded by spacing or by number with respect to reference points on the tape.

With any flight programmer, a highly accurate timing system must be employed so that the sequencing of the flight-control data shall coincide exactly with the speed and flight position of the aircraft. In some magnetic-tape programmers, a crystal-controlled oscillator is used as a master time-base generator, and its frequency is stepped down through a series of multivibrators to produce the desired time-base frequencies. This method requires a large number of stages which increases the likelihood of error because of tube or component failure or of fluctuations in power-supply voltage, temperature, etc.

Time-base generators or oscillators employing tuning-fork control at low frequencies are more accurate and stable than those of other types.

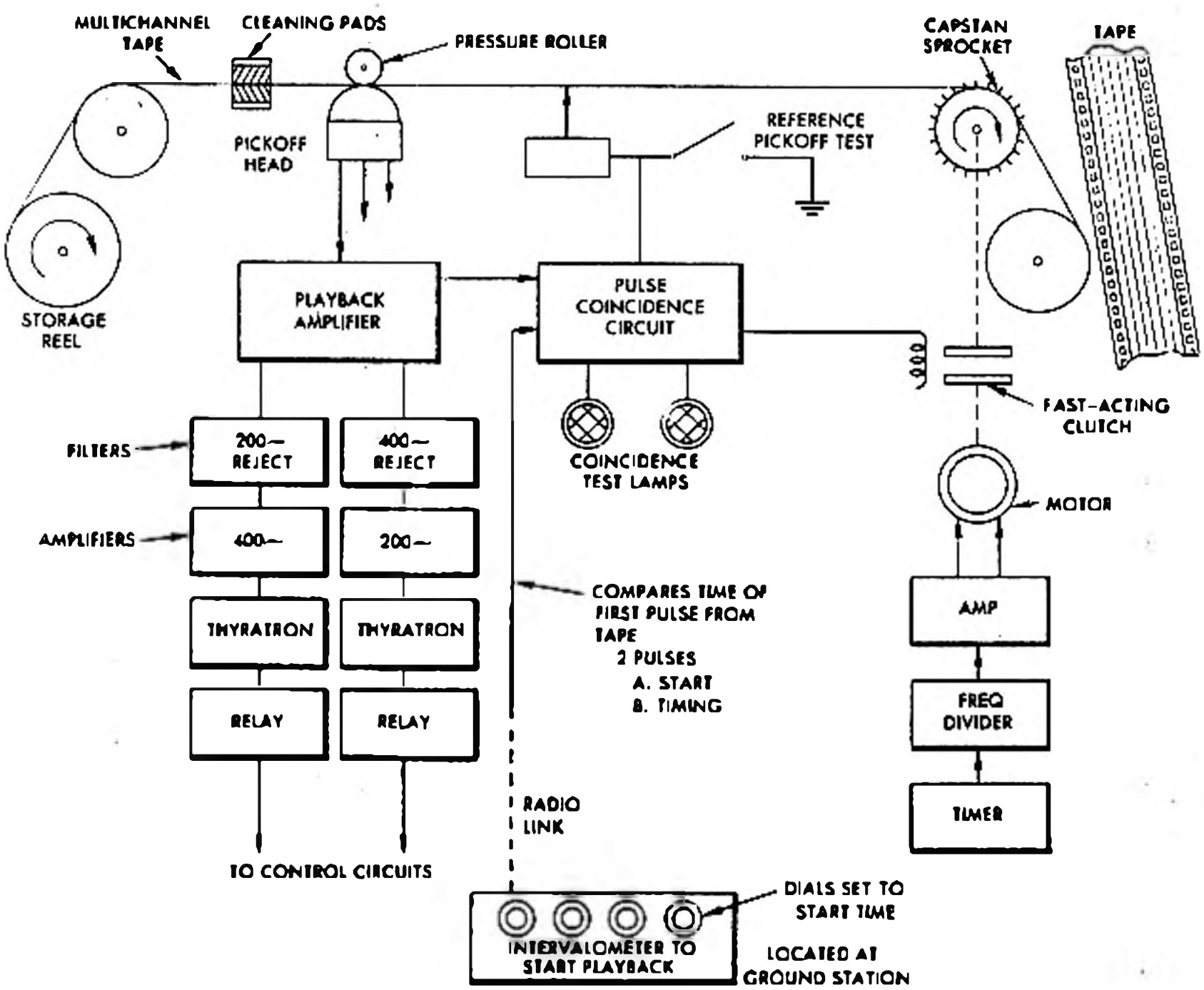


Figure 6-100. Magnetic Tape Programmer

They are becoming more widely used in flight programmers. Figure 6-100 is a functional block diagram of a magnetic-tape programmer which employs a multichannel tape, with each channel capable of carrying two sets of data simultaneously. One set of data is recorded by magnetizing pulses of one frequency, such as 200 Hertz, and the other set is recorded by magnetizing pulses of a different frequency, such as 400 Hertz.

The tape passes through a multiple pickoff playback head of the magnetic or variable-reluctance type with one pickoff for each data channel. Dual frequency outputs from each pickoff are amplified and passed through two filter circuits, one of which rejects the 200 Hertz data and passes the 400 Hertz data, while the other rejects the 400 Hertz data and passes the 200 Hertz data. The data pulses from each channel filter are amplified further and then applied to relay circuits for actuating the controls.

In the block diagram, a thyratron stage is shown in each channel. This stage conserves power by applying current to the relays only when a data signal is present.

The functions of the other blocks are shown on the diagram. The individual circuitry is treated in detail throughout this chapter.

The magnetic pickoff shown in figure 6-101 operates on the variable-reluctance principle, in which the iron core of the pickoff coil is designed with a small air gap to offer a high-reluctance path for magnetic flux. The tape passes through this gap and varies the reluctance of the core by an amount proportional to the degree of mag-

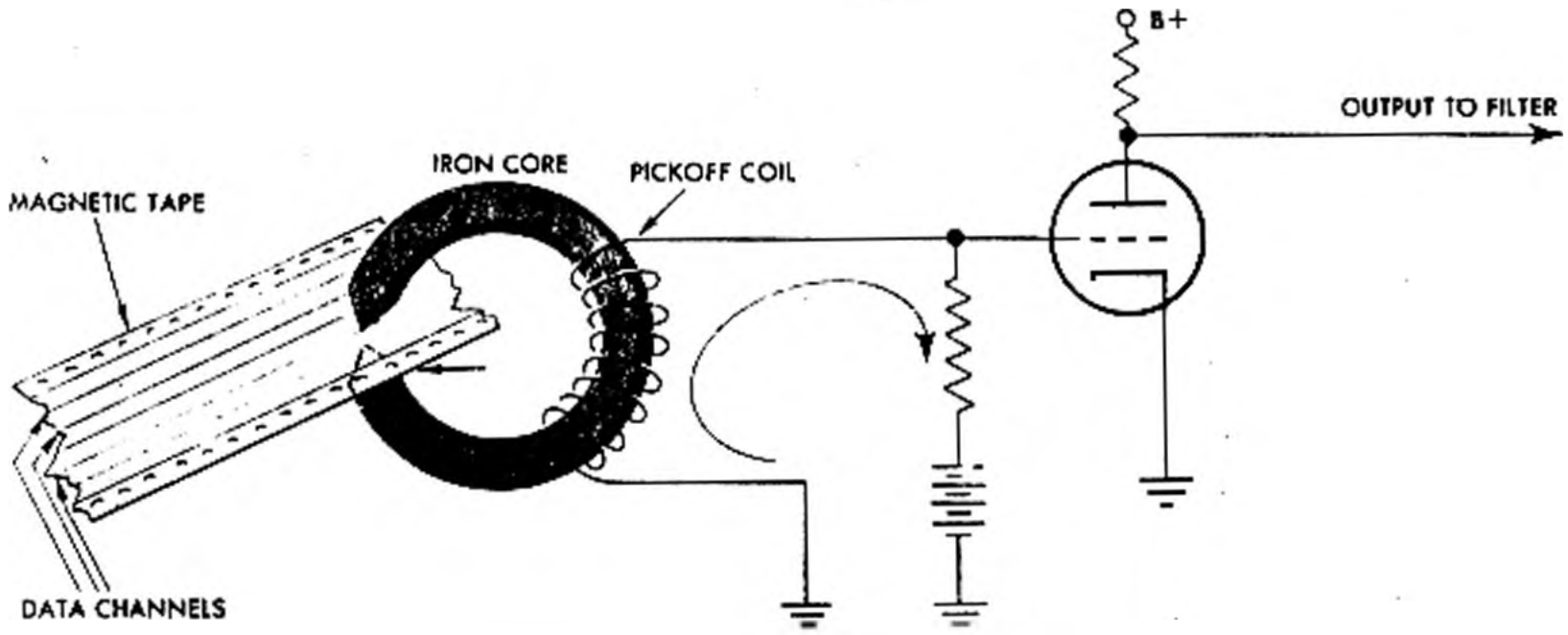


Figure 6-101. Variable Reluctance Magnetic Pickoff

netization of the data channel, or at a rate proportional to the number of magnetized sections passing through the gap in a specific unit of time. When the reluctance of the core is varied, the same effect is produced as would be obtained by varying the inductance of the coil, and the current flowing through the coil will vary accordingly. These variations in current constitute the data or intelligence conveyed from the tape to the controls through the playback circuitry of the programmer.

Map Matching

Radar map matching is another method for providing guidance reference in preset or automatic navigation of missiles. In this system, radar scans the terrain which lies beneath the aircraft course, and the PPI-scope presentation thus obtained is compared with a map or photograph of the same terrain. Any variations between the map and the PPI presentation, after conversion to the same scale, are resolved into error voltages which reposition the aircraft in the manner required to eliminate the error.

Stabilized Platforms

The accuracy of a missile guidance system depends on the accuracy and sensitivity of its individual components and its ability to compensate for the effects of external forces acting on the missile. To maintain a predetermined relationship between the missile and a fixed reference at all points along the flight course requires the resolution of all forces tending to produce any deviation from the desired pattern of flight.

In inertial guidance systems and in automatic celestial navigation systems, an artifical horizon plane always maintains a position perpendicular to the normal flight path. This plane provides an accurate reference for measuring star angles and determining instantaneous position of the missile. This artificial horizon plane incorporated into the aircraft is referred to as a stabilized platform.

A stabilized platform is necessary in all long-range guidance systems. It provides the reference plane for a predetermined path where the path of the missile is adjusted by devices wholly within itself. It makes use of Newton's laws of motion. The platform is independent of outside information except in the celestial type systems where

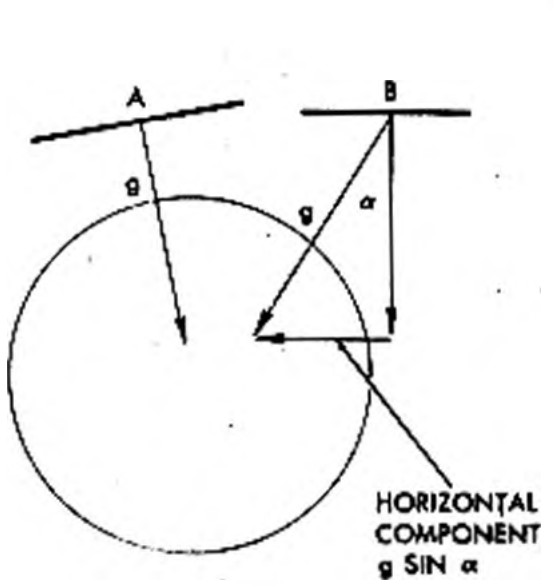


Figure 6-102. Platform Improperly Processed

the telescopes are used to help keep the platform stabilized and oriented. To resolve all of the external forces acting on the missile into either roll, pitch, or yaw errors, the stabilized platform must be kept perpendicular to the direction of the local gravity vertical at all times. The reason for this is that unless the platform is maintained perpendicular to gravity, a horizontal component of gravity will act on the sensing instruments on the platform indicating false accelerations.

This horizontal component is shown as g sin α in figure 6-102. Alpha (α) is the angle between the normal to the platform and the local gravity vertical or the direction of gravity at the geographical position.

In space, a body in motion continues in motion in a straight line with a constant velocity, unless acted upon by some external force. Accurate measurement of the degree of such external forces, with correction factors, would enable a missile to hit any target within its fuel range.

For detection of any deviation of a missile from a predetermined path, the objective or target must be stationary. The target must retain fixed position with respect to the guidance reference at all times or reference is of no value. The path is initially defined by a platform stabilized by gyroscopes around roll, pitch, and yaw axes. These gyros are given a space reference and would take the missile out into space unless corrected. To make the missile follow a path around the earth, the gyros must be torqued so that they precess and orient the platform in its proper position as the missile follows the curvature of the earth as shown in figure 6-103.

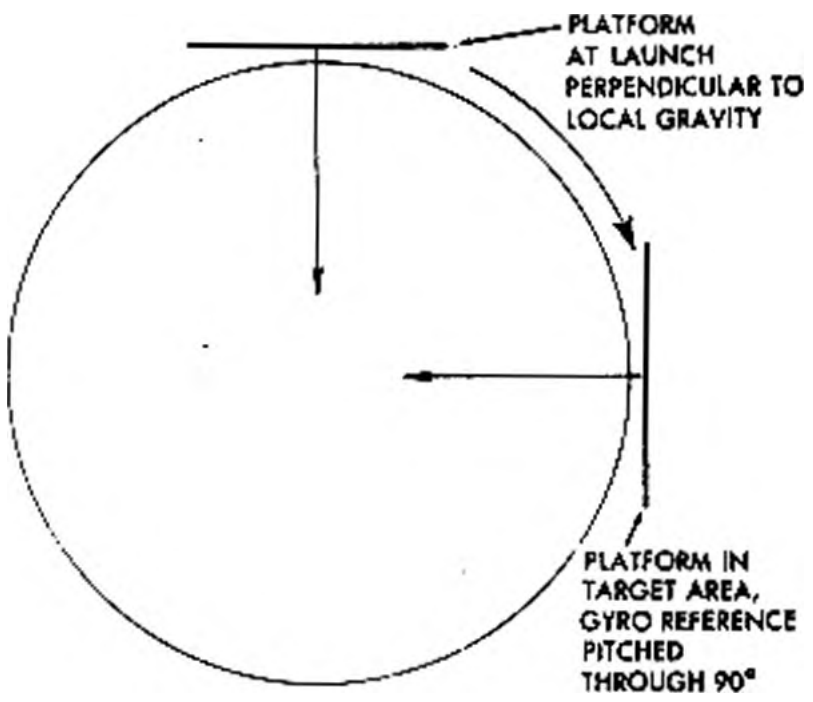


Figure 6-103. Platform Correctly Precessed

If the missile were to fly to a target one-half way around the earth, the platform reference would be pitched through 180° from its original position!

To determine the amount the platform is pitched or precessed, the position of the missile must be known. Position can be determined by knowing the velocity and time of flight. To detect any change in velocity, an acceleration detecting device is added to the platform. The velocity change could be noted by a device whose action is similar to that of a plumb bob. A displacement of the plumb bob would indicate a change in velocity, but in any constant condition it would hang vertically. A change in velocity is acceleration. If this acceleration is integrated and if the velocity and distance outputs are compared with the desired distance and velocity, the position of the missile is easily determined. The functions of distance, velocity, and direction are then set into a program device. This program device specifies a velocity, and any external force acting on the missile produces a change in velocity or an acceleration.

SECTION D
Amplifier, Controller, Actuator, and Feedback Units

This section covers amplifier, controller, actuator, and feedback units of missile guidance systems. These units are just as vital to effective missile guidance as the units already presented at greater length. Coverage of these four units is brief because they are not complex circuits and the units are common to electronic fundamentals, well covered in other texts.

AMPLIFIER UNITS

Certain operations of amplification may appear in any of the units of the block diagram shown in figure 6-1. This explanation covers various types of amplifiers not previously presented. These amplifiers may be solid state circuits using transistors, as well as vacuum tube circuits.

Frequency-Selective Amplifiers

Frequency-selective amplifiers are employed in FM receivers which are used in guided missiles to receive control signals from control transmitters. The control transmitters are located on the ground or in a mother aircraft.

The control signals are frequency modulated by one or more of a number of audio channels. The number of channels varies with the requirements of the system. These signals control servomechanisms, auxiliary functions, and safety equipment in missiles.

The audio control frequencies, after being detected, are routed by selective audio filters to relay energizing circuits which cause relays to close or open, energizing or deenergizing the synchro units associated with the control mechanisms. Each audio channel, comprising a relatively narrow band of frequencies, actuates a specific control in the missile.

The diagram in figure 6-104 illustrates a typical frequency-selective amplifier stage with audio filter and relay control. This is one form of a tuned-grid amplifier.

The filter section is a combination of a series resonant circuit and a parallel resonant circuit with the constants so chosen that a relatively flat and narrow audio bandpass is obtained. A tapped input in the coil of the parallel resonant circuit gives the filter an overall voltage gain at resonance.

Under no-signal conditions, the positive bias on the cathode and the negative bias on the grid keep the tube cut off. The relay in the plate circuit is not energized. When an audio signal is received, it is increased in voltage by the action of the parallel resonant section of the filter. The signal then is applied to the grid of the triode tube. The high signal voltage on the grid of the triode overcomes the bias and causes the tube to conduct heavily. The plate current energizes the solenoid of the relay, causing the relay contacts to close.

A pulsating voltage tends to exist across the relay coil because of tube cutoff during the negative swing of the audio signal cycle and because of the high plate current during the positive swing. This pulsating voltage is smoothed out by the filtering action of the plate bypass condenser. The filtering action makes the current through the relay solenoid relatively steady while the audio signal is being received. Several of these frequency selective amplifiers may operate simultaneously in a well designed receiver.

Other amplifier circuits employed for frequency selection include:

- Tuned-plate frequency-selective.
- Tuned-plate tuned-grid.
- Tuned-cathode.

Each of these circuits is especially suitable for some specific application, but the primary function is the same. The primary function is the selection

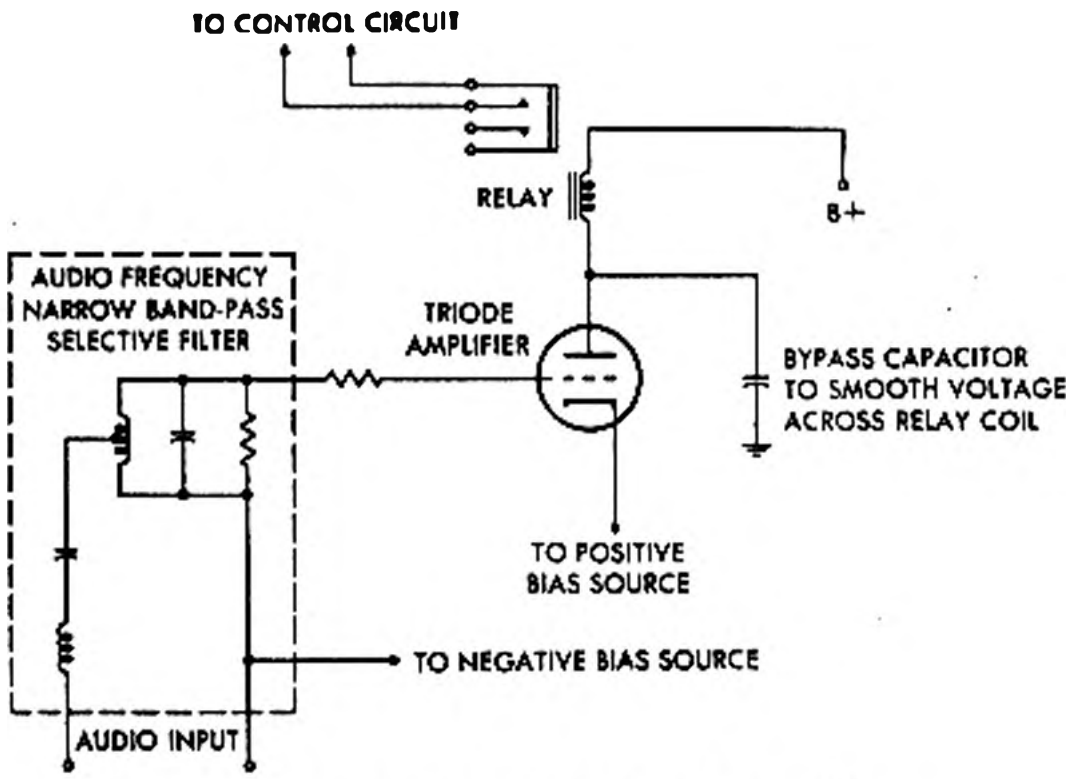


Figure 6-104. Frequency-Selective Amplifier

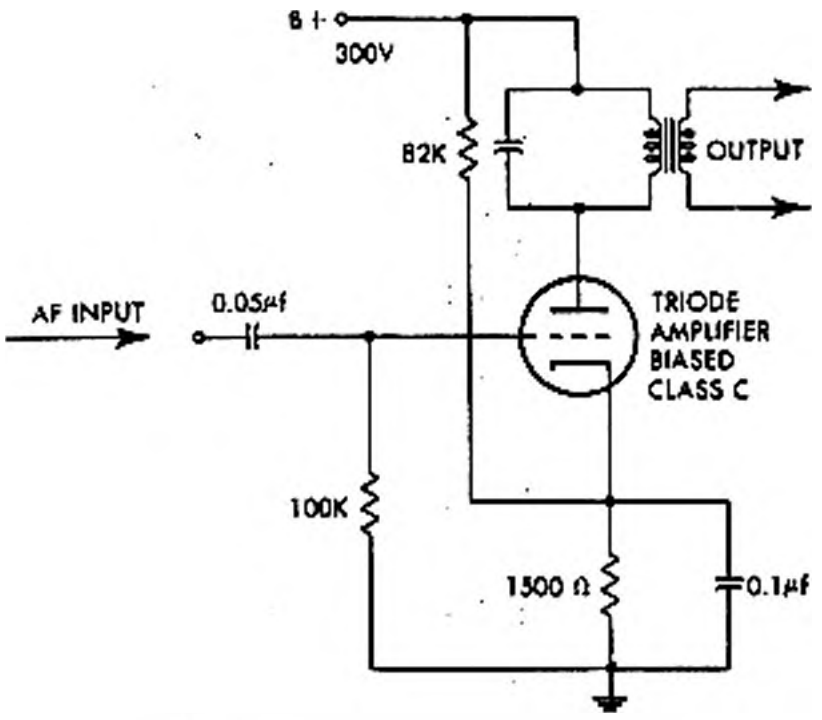


Figure 6-105. Tuned-Plate Frequency-Selective Amplifier

and amplification of a particular band of audio-frequencies.

TUNED-PLATE FREQUENCY-SELECTIVE AMPLIFIER. A tuned-plate frequency-selective amplifier is illustrated in figure 6-105. This circuit represents a class C amplifier with its plate tank circuit tuned to resonance at the frequency desired for amplification. Typical values for components and plate supply voltage are shown in the diagram.

A positive voltage is applied to the cathode, for class C bias, through the voltage divider consisting of the 82K and 1500-ohm resistors. The audio signal is applied to the grid through the 0.05 μf coupling capacitor and developed across the 100K grid-leak resistor.

The output signal is inductively coupled to the next stage. The plate tank circuit, being tuned to resonance at the desired audiofrequency, selects the frequency to be coupled to the succeeding stage. Amplification results from the action of the tuned circuit, the Mu of the tube, and the ratio of the transformer coupling.

TUNED-PLATE TUNED-GRID AMPLIFIER. Figure 6-106 illustrates a frequency-selective amplifier employing a tuned-plate tuned-grid circuit. The tanks are variable capacitors for tuning. The 100K resistors across the tank circuits lower their Q and thus widen their bandpass. The 0.001 μf capacitor in the grid tank circuit keeps the grid below ground potential for automatic volume control (AVC). This circuit is most widely used in IF stages using IF transformers to couple the signal from one stage to the next. The RF energy coupled from the input coil is built up in voltage in the tuned circuit, $L_1C_1R_1$. The voltage that appears across L_1C_1 is applied to the grid of the pentode which further amplifies it. A second resonant circuit, $L_2C_2R_2$, is the load for the plate of the tube and offers high impedance because it is a parallel resonant circuit. The output is inductively coupled to L_3. Most pentode tubes are designed to operate with the screen grid at a lower potential than the plate. A bypass capacitor is used to keep the screen grid at a steady potential by passing to ground any signal that appears on the screen grid.

TUNED-CATHODE AMPLIFIER. Another widely used frequency-selective amplifier is the tuned-cathode amplifier, shown in figure 6-107. This frequency-selective amplifier, which has become widely used in radar and television RF and input IF stages, has the tuned circuit in the cathode and uses a resistive plate load. The signal is applied to the cathode.

In this circuit, the gain of the stage is less than

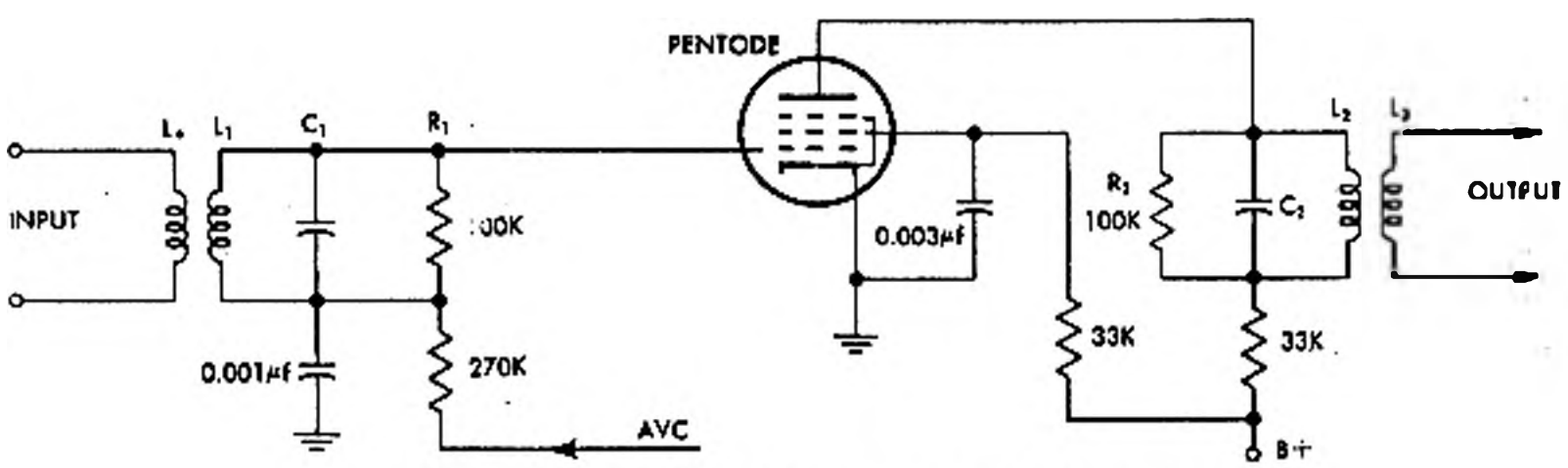


Figure 6-106. Tuned-Grid Tuned-Plate Frequency-Selective Amplifier

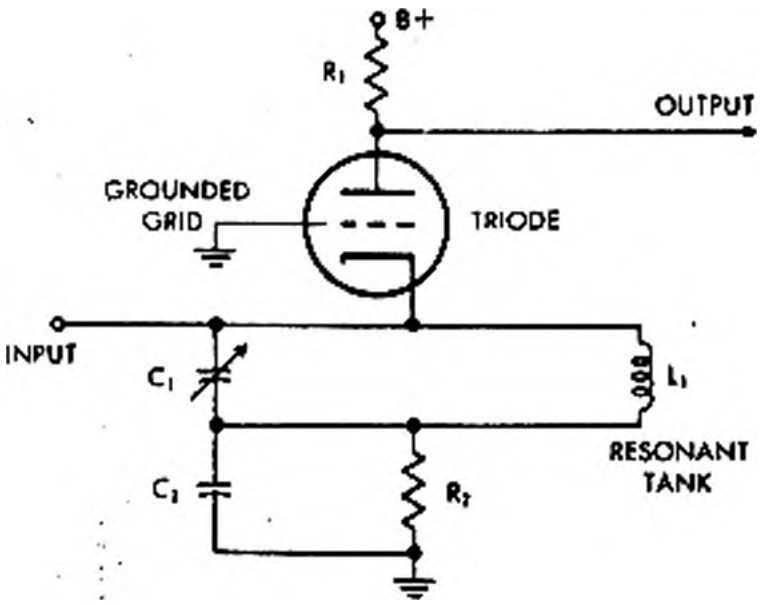


Figure 6-107. Tuned-Cathode Grounded-Grid Amplifier

when the signal is applied to the control grid, but the loss of gain is compensated for by a reduction in noise level so that a less distorted signal is transferred to the following stage.

The network C_1 and L_1 comprises the tuned circuit which amplifies only the frequency band to which it is tuned. The network R_2 and C_2 develops self-bias for the amplifier.

The above frequency-selective amplifiers are found in the amplifying units of many missile guidance systems. An amplifier having high selectivity must be used in the missile to attain precision control.

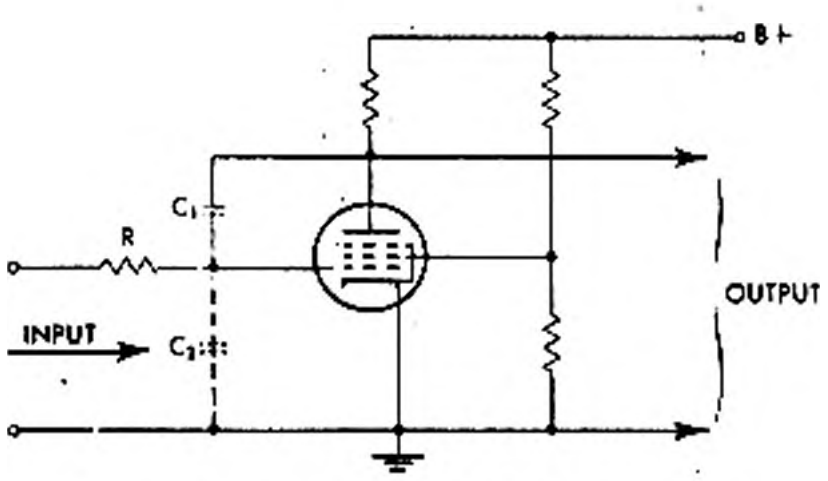


Figure 6-108. Simple Miller-Effect Integrator

Miller Integrators

Miller integrators are found in many radar pulse shaping circuits and in high-fidelity amplifiers using a large value of negative feedback to obtain linearity.

The Miller integrator shown in figure 6-108 makes use of the amplified value of grid to plate capacitance (C_1) to obtain a large effective value of input capacitance (C_2). This capacitance with the input resistance (R) comprises the integrating circuit.

A pulse-shaping phantastron circuit, as shown in figure 6-109, uses the Miller-effect integrator circuitry, thus producing a positive pulse of variable width-to-code pulses of RF energy in pulse-width coded systems.

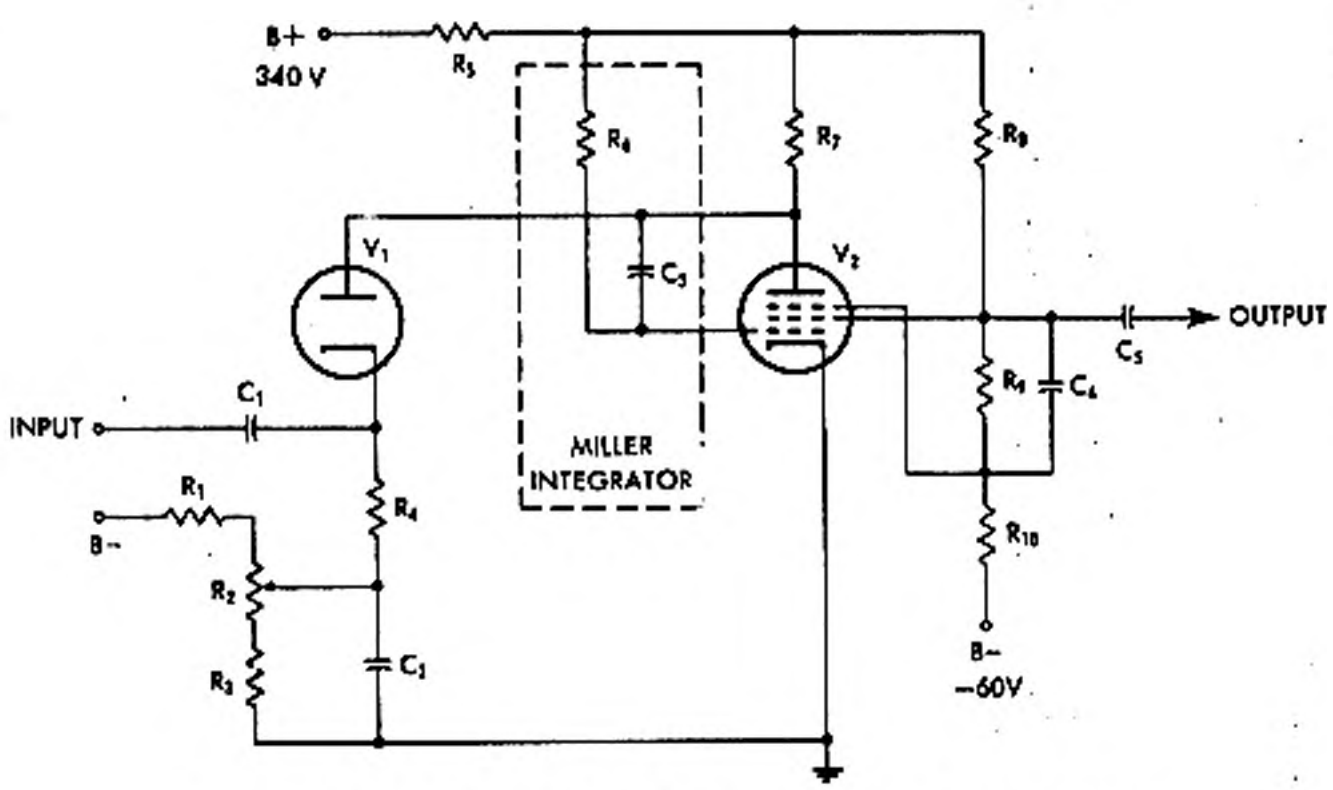


Figure 6-109. Pulse-Shaping Circuit with Miller-Effect Integrator

The width of the output pulse is controlled by the setting of potentiometer R_2. Output is taken from the screen grid of V_2 which is a pentode connected in a phantastron circuit with its plate and grid connected by an RC network to function as a Miller integrator.

Input tube V_1 receives the negative input pulse at its cathode through coupling capacitor C_1. The signal leaves V_1 as a negative trigger pulse and is applied to the grid of V_2 through capacitor C_3. This trigger pulse initiates one cycle of the phantastron action.

The waveforms shown in conjunction with the circuit diagram illustrate how the circuit is capable of converting a narrow negative input pulse into a positive output pulse of symmetrical form and variable width.

Miller integrating networks are also found in many applications, such as DC amplifiers, in which you want to produce a change in the output only after a certain period of time.

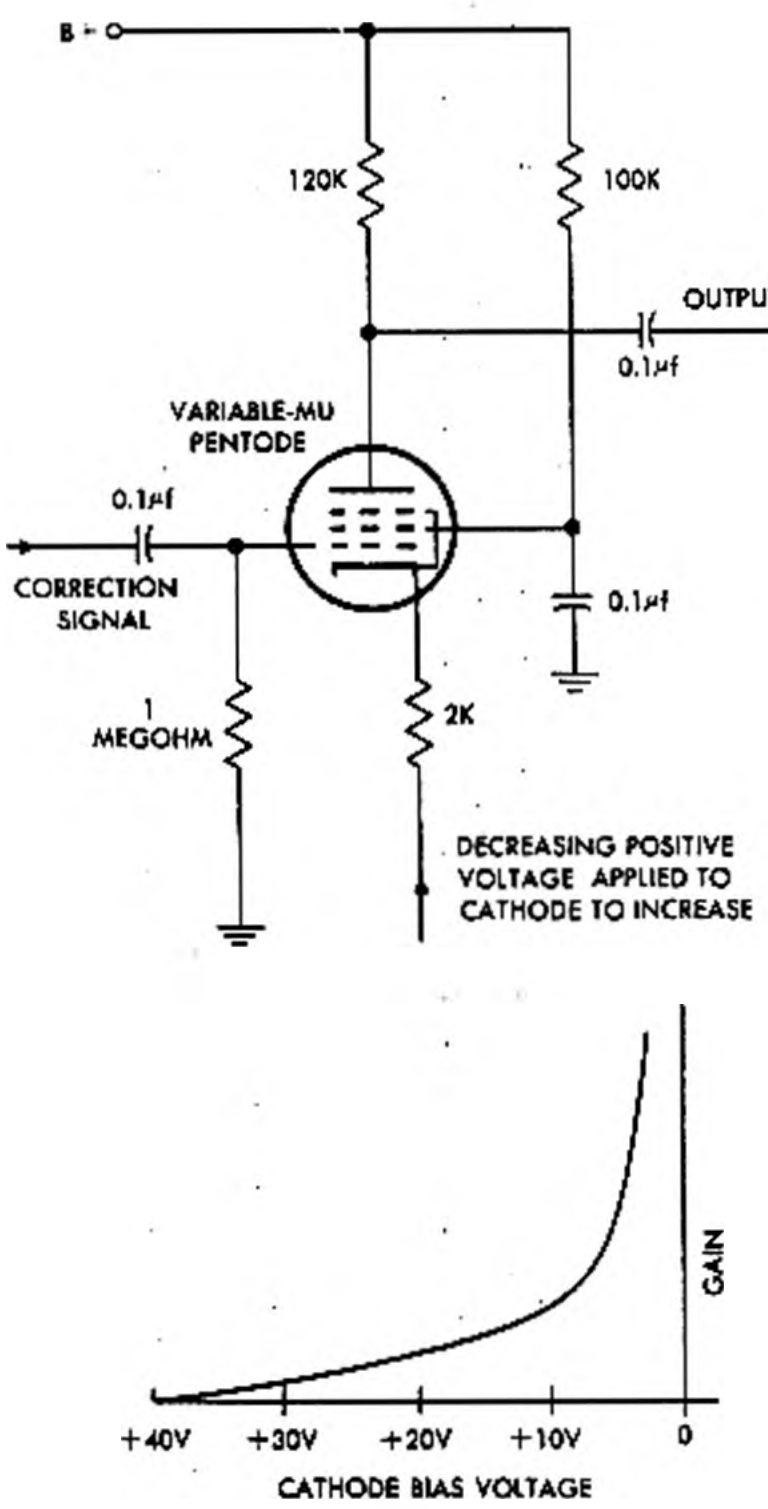


Figure 6-110. Mixing Amplifier

Mixing Amplifiers

Mixing amplifiers are used whenever you desire to impress AC or RF pulses upon a varying DC signal, as in telemetering systems. Mixing amplifiers are also used to control amplification of correction voltages transmitted to the guidance and/or control system of guided missiles when you want to increase sensitivity of the guidance system as the missile approaches the target.

Figure 6-110 illustrates a mixing amplifier circuit in which a decreasing positive voltage is applied to the cathode of a variable Mu pentode. This produces the same effect as applying an increasing positive voltage on the grid. In either case the bias is decreased. Voltage across the tube will increase as the cathode bias is decreased, resulting in higher amplification and greater sensitivity of the circuit to correction signals.

Another mixing amplifier, as used in some telemetering systems to combine from 2 to 16 frequency-modulated audio signals, is shown in figure 6-111.

The 10K potentiometer acts as the master gain control for the circuit. The mixed signals are amplified in two RC coupled stages employing medium Mu triodes V_1 and V_2.

The cathode follower output stage V_3 provides a low impedance output. Negative feedback from the cathode follower output stage is applied to the cathode of V_1 through a 10K resistor and a 0.5 μf capacitor in series. This reduces harmonic distortion to a low level.

An amplifier suitable for combining an RF signal with an AF signal is shown in figure 6-112. Amplifiers of this type are commonly found in AM transmitters and in telemetering systems.

In this circuit, the audiofrequency signal is developed across the 470K resistor and the 500 $\mu\mu$f capacitor. The radiofrequency signal is developed across the 150-ohm resistor. No RF is seen across the 470K resistor since the reactance

of the 500 $\mu\mu f$ capacitor is so low at radio frequencies (10,600 ohms at 30 kHz) that it bypasses the RF to ground.

The combined signal is then applied to the grid of the pentode where it is amplified. A resistor (1K) in the cathode circuit shunted by a 500 $\mu\mu f$

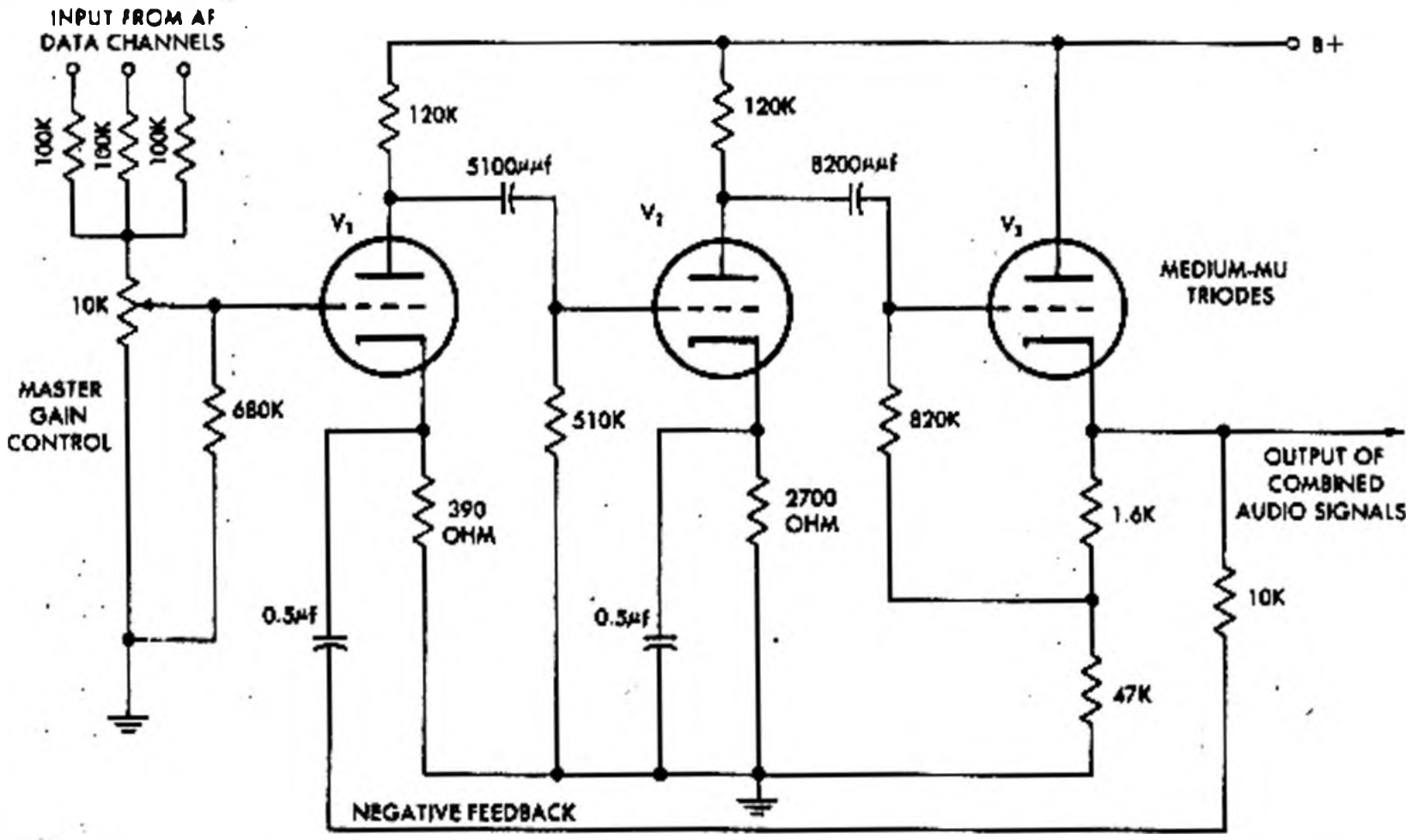


Figure 6-111. Mixing Amplifier used in Telemetry Circuit

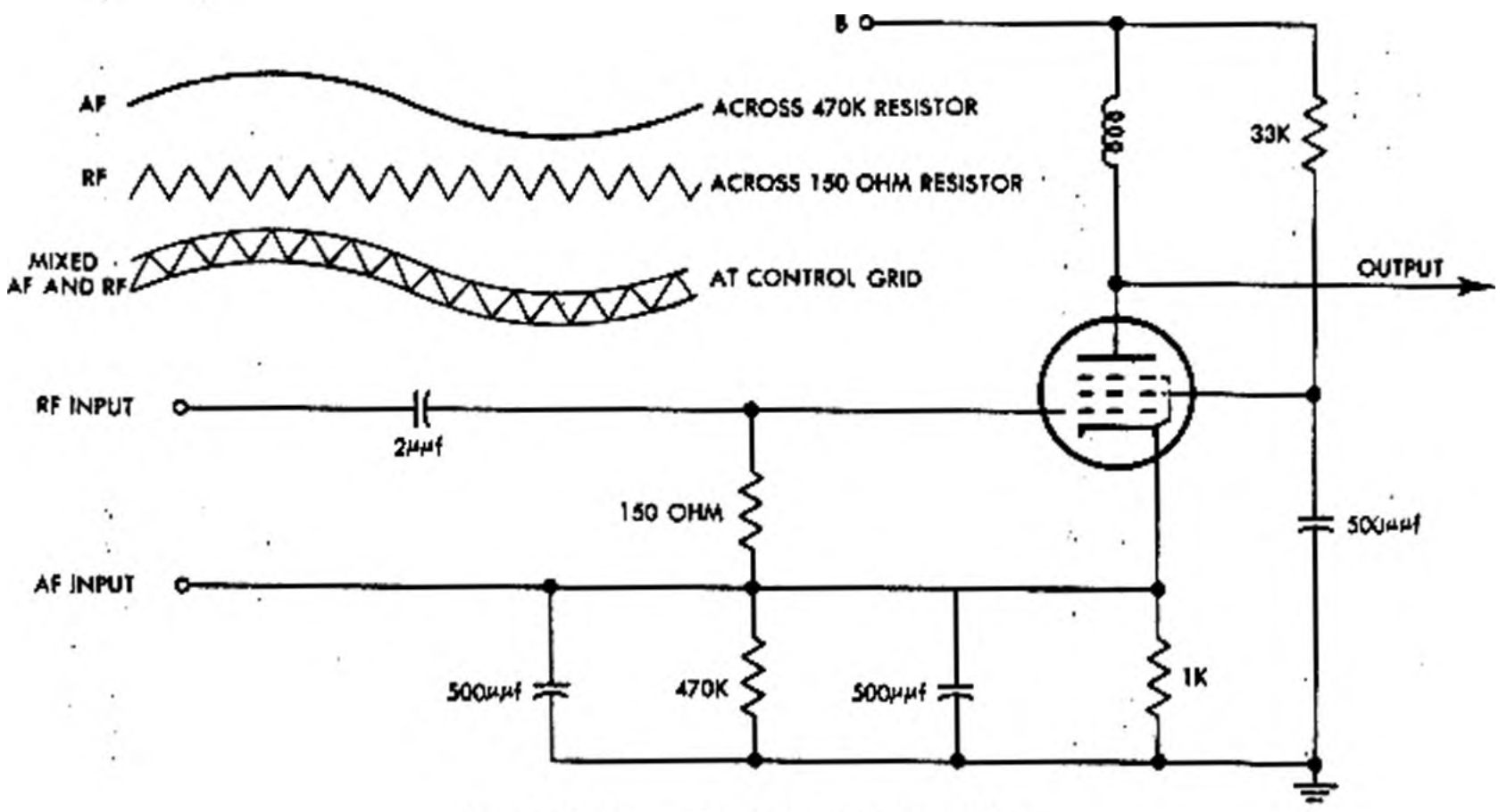


Figure 6-112. RF-AF Mixer and Amplifier

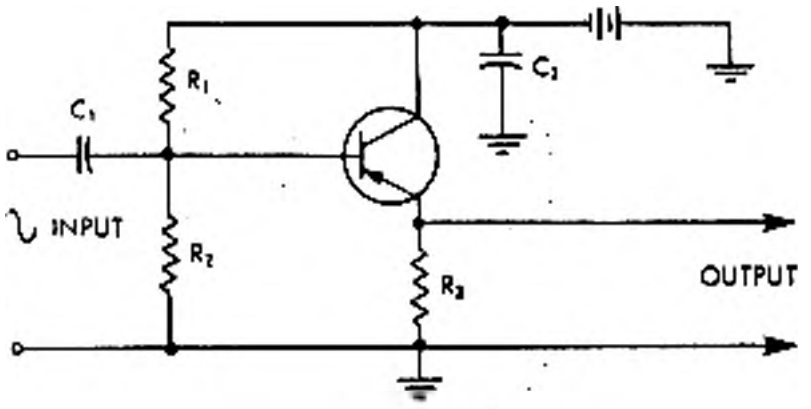


Figure 6-113. Transistor Common-Cathode-Collector Amplifier

capacitor provides self-bias. An inductive load is used in the plate circuit to develop the output signal which is taken from the plate.

Component values shown are typical for the circuit as used in telemetering systems. They will vary in accordance with the range of RF and AF signals which are to be mixed.

The solid state transistor common-collector amplifier may also be used to accomplish cathode follower operation, replacing the electron tube circuit. This circuit is shown in figure 6-113. The input signal is applied to the base and the output signal is taken from the emitter. Since the output emitter signal follows the input base signal in both phase and time, the transistor common-collector amplifier is also referred to as a transistor cathode follower.

The common-collector amplifier, like the common-base and common-emitter amplifiers, can be biased with a single bias battery. The method used in obtaining forward bias for the base-emitter junction in figure 6-113 is a voltage divider network comprising resistors R1 and R2. This arrangement provides the necessary forward bias across the base-emitter junction by making the base negative with respect to ground. If an NPN transistor were to be used instead of the PNP transistor, the polarity of the bias battery would have to be reversed, but the voltage phase relationship would be the same.

The input signal in figure 6-113 is applied between base and ground. Since bypass capacitor C2 effectively places the collector at AC ground, the signal is actually being applied between base and collector.

The output signal is taken across the emitter load resistor, R3, which is between the emitter and ground. The input impedance of the common-collector amplifier is very high, being 50K to 100K ohms. The output impedance is very low, usually less than 100 ohms. The transistor common-collector amplifiers, like the electron-tube cathode follower, has a voltage gain of less than 1. Power gain is fairly good.

CONTROLLER UNITS

Relays find many applications in missile guidance systems as controller units. They provide a method for using relatively weak radio signals to switch on or off circuits in which large values of current are present to actuate control mechanisms.

In conjunction with a vacuum tube circuit, the relay constitutes an electromechanical switch which is timed or delayed with respect to the actuating impulse.

Relays are employed in channel selector circuits, arming circuits, and dive circuits. In these circuits, a certain sequence of operation must be initiated by a pulse or signal received from a command transmitter located at a ground station or from a voltage developed in an automatic navigation system within the missile.

A simple electronic-mechanical plate circuit relay is shown in figure 6-114. The circuit, with modifications to suit timing and power requirements, might be employed in any of the previously mentioned applications.

The triode is biased to cutoff by a positive voltage applied to its cathode. When a signal is received, the grid is driven sufficiently positive to overcome the fixed bias, and the tube conducts.

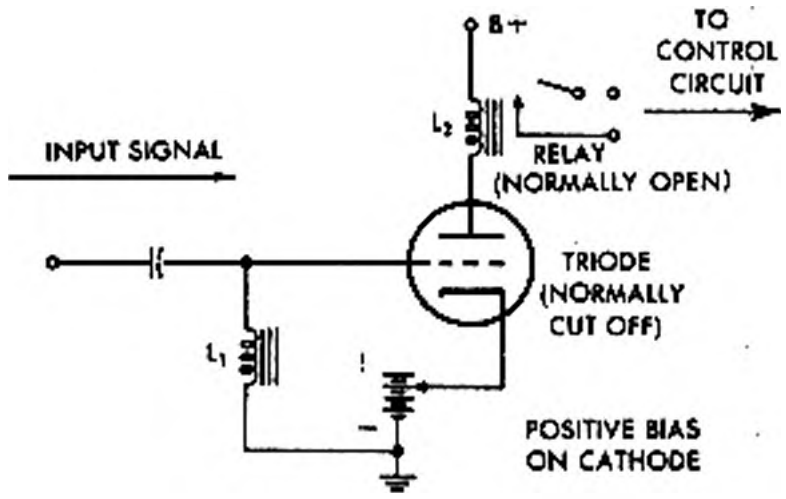


Figure 6-114. Simple Relay-Closing Circuit

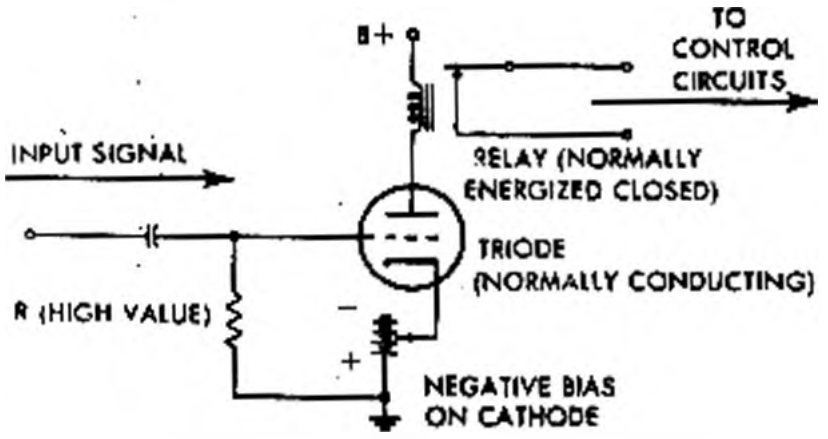


Figure 6-115. Simple Relay-Opening Circuit

The plate current energizes relay solenoid L_1, causing the contacts to close.

If the input signal is of high frequency, the grid circuit of the tube should have low DC resistance so that when the grid goes positive and draws grid current, it will not develop a bias which would aid the fixed bias and prevent the tube from conducting. A choke (L_1) is used in the grid circuit to provide high impedance to the signal and low DC resistance to grid current.

Figure 6-115 illustrates a circuit in which the relay is normally energized and in which a signal serves to cut off the plate current through the tube. This deenergizes the relay, permitting the contacts to open.

In this circuit, the cathode is biased with a negative voltage adjusted to maintain sufficient plate current to keep the relay closed under no-signal condition. When a signal is received, the large resistance in the grid circuit develops a high bias, causing the tube to cease conducting. The relay contacts then open.

Figure 6-116 illustrates a relay circuit in which the opening or closing of the relay is delayed. This is done by means of a capacitor shunted across a high-resistance relay.

In case a delay in the releasing of the relay is desired, a high-resistance relay is shunted with a large value of capacitance, as shown in the figure.

When an input pulse is received, the relay closes, and at the same time the capacitor takes a charge. At the termination of the pulse, the capacitor discharges through the solenoid coil of the relay and keeps the contacts closed for a period of time. The period of time is determined by the resistance of the relay coil and the value of the capacitor.

If it is desired to delay the closing of the relay in the preceding example, a resistance is inserted in series with the pulse input as shown at the bottom of figure 6-116.

When the input pulse is received, a large voltage drop occurs across R. Insufficient current flows through the solenoid coil to close the relay. As the charge on the capacitor approaches the value of the input voltage, current flow increases through the relay coil until it reaches a value sufficient to close the relay. The closing of the relay is affected by the charging time of the RC circuit. If the resistance of the relay coil is low, the effect of the RC circuit is less appreciable.

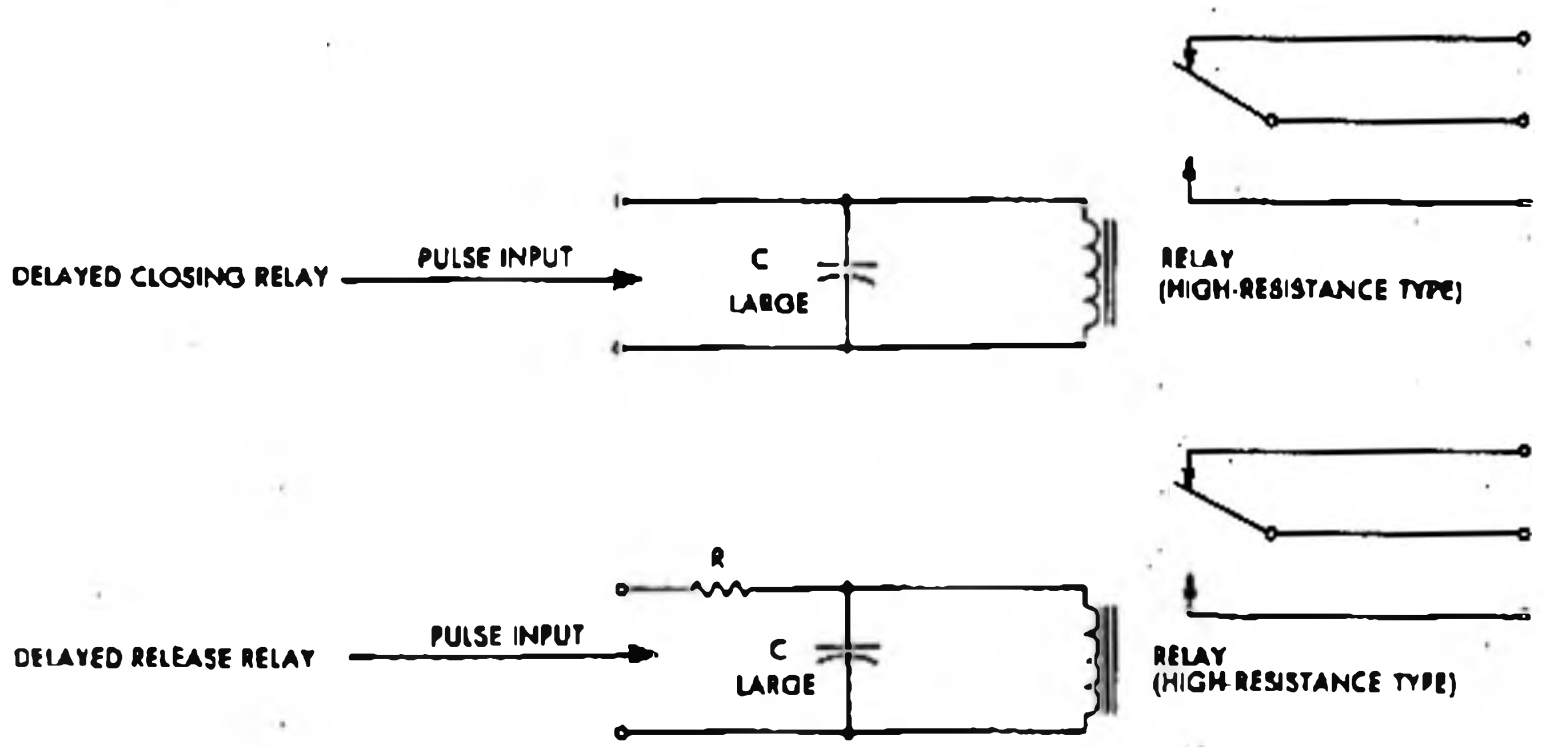


Figure 6-116. Delayed Relay Circuit

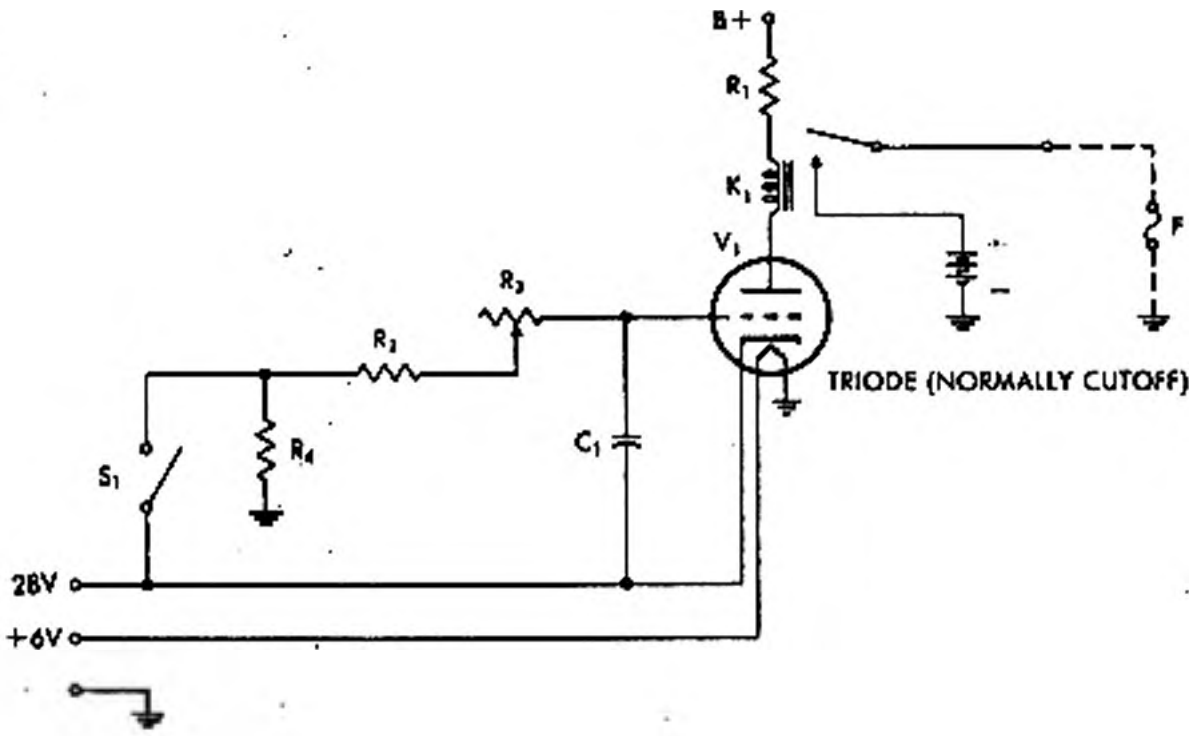


Figure 6-117. Time-Delay Relay Circuit

A high-resistance relay should be used in this application.

Two of these delay-type relays (delay-open and delay-close) are used in parallel for decoding a sequence of pulses.

A relay circuit suitable for application as an arming circuit or gyro-uncaging circuit is shown in figure 6-117.

This circuit consists of an inertially operated set-back switch (S_1) and a triode tube (V_1). It has a normally open relay for applying power to melt the fusible link (F) in the arming mechanism or gyro cage, and it has an RC network and suitable power sources.

The inertial switch (S_1) is set to trip at the desired acceleration. This removes the cutoff bias from triode V_1 after a time determined by the values of the RC components. Once the bias is removed, V_1 conducts, closing relay K_1 in its plate circuit. This completes the arming or uncaging circuit.

Normally V_1 is biased to cutoff by the positive 28 volts on its cathode. As the inertial switch closes, the cathode is connected to the grid through R_2 and R_3, rather than through the 28-volt power supply, removing the bias.

At the same time, R_4 still remains in the circuit between cathode and ground, providing self-bias which prevents excessive plate-current flow through V_1.

A time delay occurs between the closing of S_1 and conduction of the tube. This delay is determined by the charging time of capacitor C_1 which must charge to a specific voltage before the tube can conduct.

The amount of time delay is determined by the adjustment of a variable resistor (R_3) which sets the time constant of the RC combination within the limits of its range of adjustment.

ACTUATOR UNITS

Servomotors are used in missile guidance systems to operate controls such as variable capacitors, inductors, or potentiometers. These controls correct outputs of range, ground speed, and azimuth circuits. The servosystem is designed to provide motor rotation in response to error signals from any two channels. The motor then drives the potentiometer or other correction device until outputs are balanced or until a null point is reached. When a null point is reached, the motor ceases to rotate and controls remain constant until another error signal is received. A simple circuit illustrating this application of a servomotor is shown at the top of figure 6-118.

A servomotor is a reversible, adjustable speed motor. It is controlled in speed and direction of rotation by the phase and magnitude of an error voltage produced by variation or unbalance between the outputs of two system channels. When

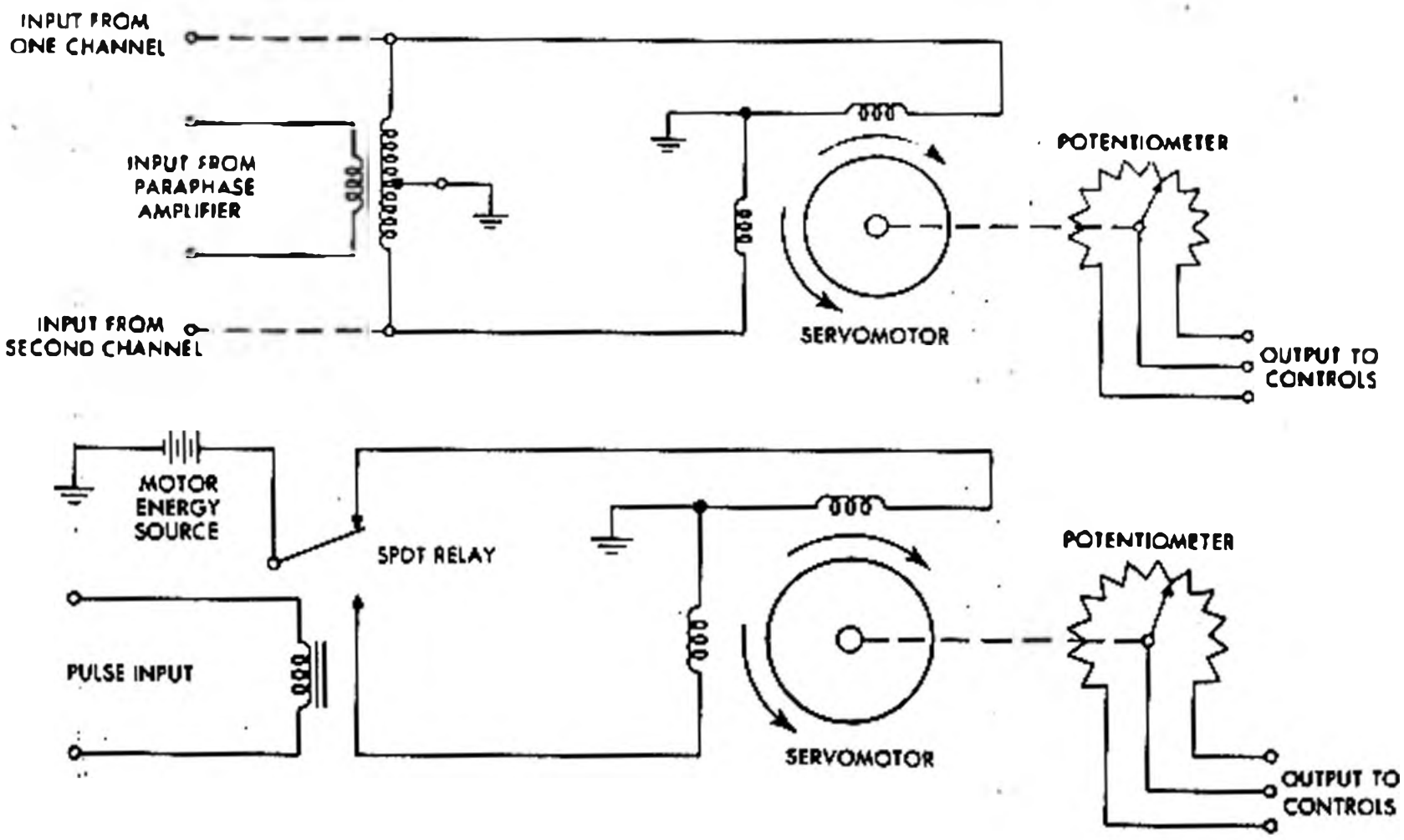


Figure 6-118. Servomotor Activating Units

the missile is following the prescribed flight pattern, the outputs are equal in amplitude and 180° out of phase, cancelling each other or producing a null output. Any variation in the output of either channel produces an output signal or error voltage which causes the motor to rotate in the direction determined by the polarity of the error voltage. The motor moves a potentiometer which regulates the amount of correction voltage applied to the associated control circuit until a null or balanced condition is restored.

An analogous circuit employing a relay, and suitable for application with circuits having a pulse-form output, is illustrated at the bottom of figure 6-118.

When narrow pulses are received, the relay closes for short intervals on the make contact and, because of the long spacing between pulses, closes for longer time intervals on the break contact. If each contact were connected to the windings of a motor so that the make contact caused rotation in one direction (forward) and the break contact caused rotation in the reverse direction, the motor would move in the reverse direction with the first command pulse. With the second command pulse, the motor would move in the forward direction. If the pulses and spacings were equal, the motor would not move.

This circuit is used to decode certain types of command signals and to actuate compensating devices or controls.

FEEDBACK UNIT

Automatic gain control (AGC) is employed in superheterodyne receivers to maintain constant amplification in the RF and IF stages, regardless of variations in strength of the received signal. This constant amplification is accomplished by maintaining the correct bias voltages on each controlled stage.

If the received signal is weak, no bias is applied to the RF stages. This permits the stages to operate at their maximum degree of sensitivity. If the received signal is strong, it tends to overdrive the amplifiers, producing distortion and excessively high levels of noise-to-signal in the output.

If the RF stages were operated at maximum sensitivity at all times, the ratio of noise-to-signal might be so high that the signal would be lost. The AGC system is designed to provide the highest

possible signal-to-noise ratio over the range of operating frequencies.

Deficiencies of most simple AGC systems are their inability to react quickly to rapid fluctuations in signal strength and their lack of stability when noise pulses are present. The deficiencies are particularly noticeable when the signal is in the form of pulses or is pulse-modulated.

In pulse transmission, the changes in signal strength are rapid, and the duration of the pulse is short. The receiver is subject to noise impulses for a long time compared to the duration of the signal. The ratio of noise-to-signal in its output is high. To overcome these deficiencies as much as possible, keyed or gated AGC systems have been devised. In these systems, a synchronizing pulse is applied simultaneously with the received signal to a coincidence tube which gates or energizes the circuit only for the duration of the signal pulse. This excludes the greater portion of the noise, and when the AGC voltage is developed, it is developed to a greater extent by the signal itself and to a lesser degree by noise. The AGC circuit can be designed so that no bias voltage is applied to the RF amplifiers until the signal reaches a certain level. Then the bias will be applied only after the output has reached the desired level.

There are many methods by which AGC voltage is developed and applied, but only one method suitable for use in radar receivers is presented here.

Figure 6-119 illustrates a gating system in which the RF signal is detected in a crystal detector and then applied to the control grid of a coincidence (gating) tube. This tube has sufficient negative bias applied to both its control grid and suppressor grid to prevent the tube from conducting as a result of the presence of a postive pulse on either grid alone. The tube will conduct whenever positive pulses appear simultaneously on both grids. The tube conducts for the duration of a signal if a positive synchronizing pulse, generated at the same PRF as that of the radar transmitter, is applied to the suppressor grid of the coincidence tube (V_1) at the same instant that the signal pulse appears at the control grid. The gated signal is applied to the control grid of an amplifier (V_2). As the cathode current varies with the signal voltage, the voltage across the cathode resistor rises and falls proportionally to the variations of signal amplitude. Suitable AGC voltage is picked off by tapping this resistor so that the desired amount of AGC voltage is obtained by any signal level within the limitations of the circuit. By using potentiometers to set the bias level of the coincidence tube and to regulate the cathode voltage to the amplifier stage, the amount of AGC voltage applied to the RF stages is selected. This produces the desired degree of sensitivity in

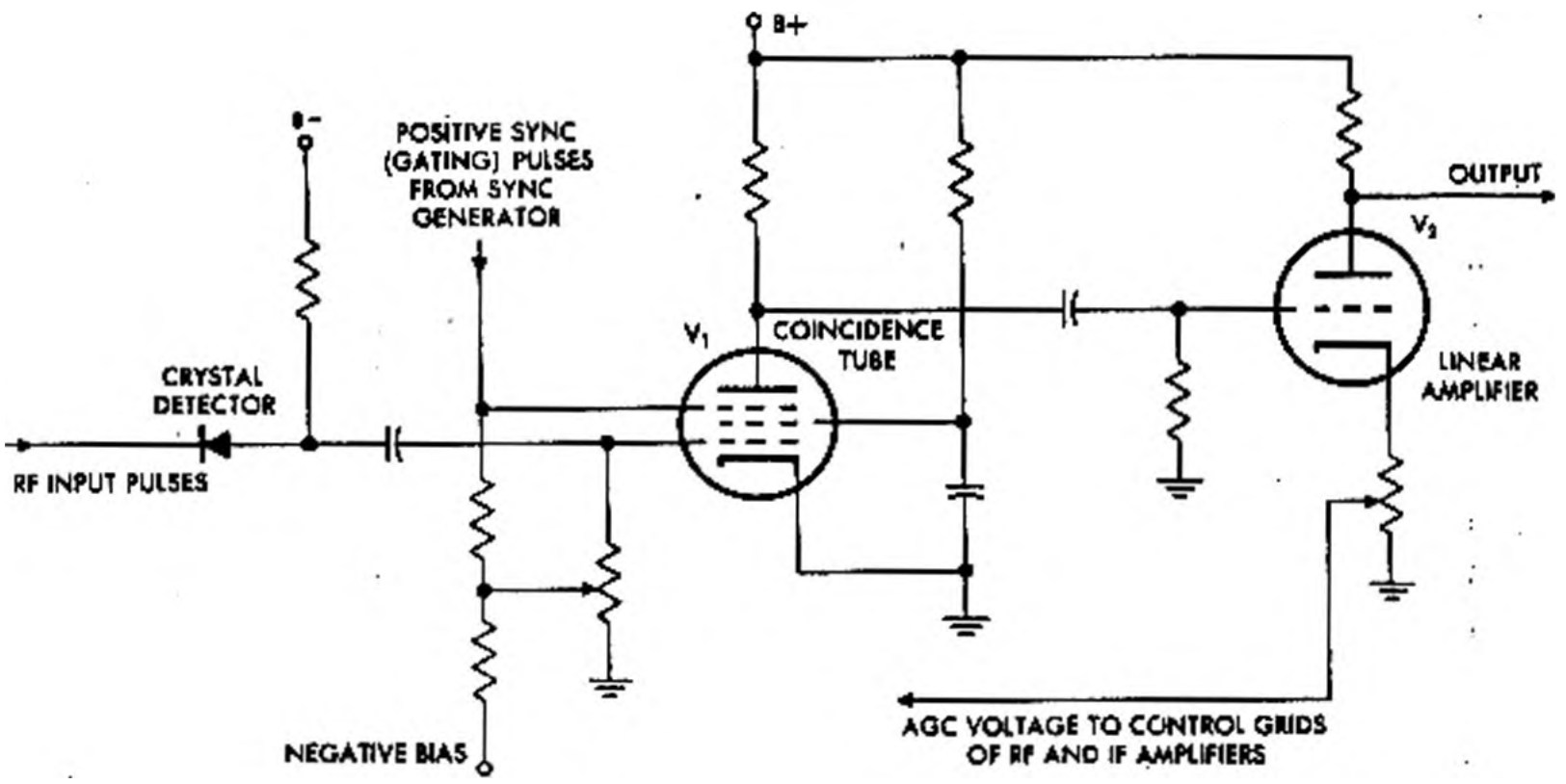


Figure 6-119. Gating Circuit applying AGC to RF Stages for Duration of Signal only

the RF stages when the signal is present and permits a condition of maximum sensitivity between signal pulses.

PRESENTATIONS AHEAD

This brief coverage of amplifier, controller, actuator, and feedback units rounds out the consideration of the missile guidance system components. For explanatory purposes, most of the circuits covered are electron-tube circuits. The state of the art may now include all solid state devices with transistorized circuits, modules, and new integrated circuits.

To this point, the text has emphasized the individual functions of control and guidance system components. The next chapter begins a study of guided missile control systems. And then, after covering several types of control systems, the manual introduces you to missile guidance systems.

CHAPTER 7

Guided Missile Control Systems

This chapter covers different types of control systems which are typical of those in the missile field. No attempt is made to explain every missile control system. Such an attempt would result in much repetition since all control systems have many similarities.

After studying this chapter you should be able to take any control system with which you come in contact and closely relate it to some system in this manual. This will enable you to analyze the operation of a new system with a minimum of time and effort.

A knowledge of the operation of a system with which you are involved will help you to understand not only *how* a certain procedure is performed but *why*. Such knowledge will increase your confidence and interest in your missile work and will enable you to learn other jobs more rapidly.

Numerous operations on missile equipment are required from the time a missile is manufactured until it is fired. Assembly, test, and firing procedures for the control, propulsion, guidance, and launching systems must be closely coordinated. Knowledge of systems operation will enable you to understand what part each job plays in the launching of a missile.

In this chapter, four common types of control systems—*electric*, *hydraulic-electric*, *pneumatic*, *and pneumatic-electric*—are presented. Details of individual components are contained in chapter 5.

A CONTROL SYSTEM DEFINED

A control system consists of interconnected components which control the missile so that it will maintain a selected heading and attitude in flight. A clear cut distinction is difficult to make between flight control and guidance of a missile, since both components are actually part of an integrated system for automatically directing the missile along its intended flight path. Generally, a control system is intended to provide smooth, *stable flight*, but it does not know where to direct the missile. The guidance system provides the *location information* necessary to direct the missile to the desired target. The guidance system functions through the control system to direct the missile on a desired course. This necessitates close tie-ins between the two systems. The same control system may be used with different types of guidance systems. This fact makes a division between control and guidance logical for most types of missiles.

REQUIREMENTS OF A CONTROL SYSTEM

A missile control system does a complex job. The control function maintains a missile in a stable, smooth flight condition regardless of outside disturbances. In a piloted aircraft this job is normally performed by a human pilot. The problems attendant on stable flight are explained in chapter 2, dealing with aerodynamics.

Many human activities are highly complicated. Little is known today as to exactly what happens in the human mind and body when people are engaged in activities such as flying an aircraft, driving a car, or even picking up a pencil. Experts agree that our mental and physical processes are highly complex. The duplication of certain functions by means of automatic systems is also complicated.

Flying an aircraft is a complex human activity, and it certainly requires an extensive design technique to create a system which will replace a human under all flight conditions. Systems for controlling aircraft were developed before the need came about for systems for controlling missiles. Aircraft autopilot components were modified and adapted for use as control systems in some early missiles. As missile speeds and accelerations increased, control systems with faster response times were also developed.

To analyze what a control system must do is not difficult. First, it must have a means of sensing when the missile is deviating from the desired flight conditions. The deviation may be in attitude, altitude, or speed.

Second, the system must be able to distinguish in what manner the missile is deviating from the desired flight conditions. A measure of the deviation, called the "error," is sensed and measured by one of the systems components, a sensor. The sensor determines the magnitude (amount) and sense of the error. Sense as used here means the sensor provides an output as to whether the error is to the right or left, up or down, plus or minus, etc., from the desired flight condition. The output of the sensor is referred to as the "error signal."

After the error has been detected, it must be relayed to the controls and a correction must be performed. This means the system must operate the control surface or device.

To operate the control device properly, modification of the error signal may be necessary to represent additional information. The signal must also be amplified so it has enough power to operate the control device.

The control system must produce a correction signal without a long delay after the error signal is generated. The period of delay is known as the *response time* of the system.

After the control devices begin to move, the control system still has not completed its job. During the entire period of missile correction, the system must continually detect instantaneous missile errors and make constant corrections so the missile can maintain a smooth flight. In most systems the speed, or rate, of deviation is also detected and taken into consideration when corrections are made.

TYPES OF CONTROL SYSTEMS

There are four methods commonly used for aerodynamic control of missiles. They are *aerodynamic control surfaces, jet vanes, movable jets, and fixed steering jets.* These controls are explained in detail in chapter 2.

When using control surfaces, jet vanes, and movable jets, the control system deflects these devices to change the effects of the forces acting on the missile. With fixed steering jets, the missile is controlled by varying the thrust from each jet. The rate or time of burning can be controlled by valves.

Control systems are not classified by the type of control device, but rather by the method used to move these devices. The four methods used for producing this motion are (1) hydraulic, (2) pneumatic, (3) electrical, or (4) mechanical. The type of control system, then, is determined by whichever method or combination of methods is used to actuate the control surfaces or devices.

The selection of the control system depends on factors such as speed, size, altitude, range, and weight of the missile. Incidental factors such as the preference of the engineering people designing the system and the preference of the user may also influence the selection. The economy of using off-the-shelf items as opposed to the cost of engineering new equipment also influences the selection of a type of control system.

CONTROL-SYSTEM POWER SOURCES

Energy is required to operate missile control devices whether they are aerodynamic surfaces, movable jets, or fixed steering jets. The energy to operate a control system must come from within the missile and ordinarily is supplied by one of three sources: the missile engine, an auxiliary engine, or a battery.

Energy from Missile Engine

One common source of energy is the missile propulsion system. Two methods of deriving energy from this source are illustrated in figure 7-1. If a turbojet or gas turbine engine is used to propel the missile, a generator may be mechanically connected to the turbine to provide elec-

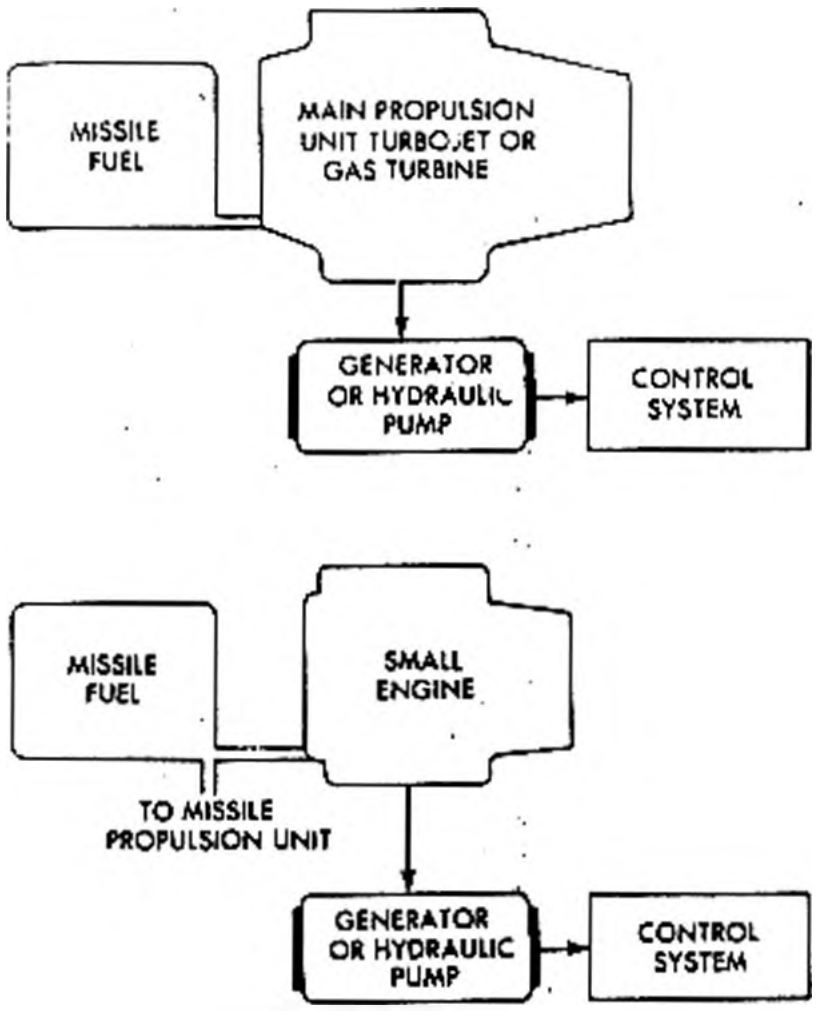


Figure 7-1. Two Methods of Obtaining Power from Energy of Fuel

trical power. Similarly, a hydraulic pump may be connected to the turbine to provide hydraulic pressure. This is a very practical method of developing control power and uses part of the missile thrust power and fuel. This system can produce a comparatively large amount of power for the control system.

Energy From Auxiliary Engine

A small auxiliary engine, using the missile propellant fuel, may be used to drive a generator or hydraulic pump directly.

A variation of this system often used on large missiles is the gas generator. This generator uses its own fuel, which may be a chemical such as hydrogen peroxide. A gas generator is ordinarily a high output, short duration device. The time of operation is limited by the amount of fuel carried. Such a power generator is normally used on missiles having a relatively short flight duration.

Stored Energy

Storage batteries are a possible source of electrical energy where the requirement for large amounts of power are of short duration.

Another source of stored energy is compressed air or gas stored in containers on board. These gases usually provide the power indirectly by driving a small turbine and function like the gas generator mentioned previously.

Now consider the specific types of control systems. The following section covers hydraulic-electric systems.

SECTION A
Hydraulic-Electric Control Systems

The hydraulic-electric system is the most commonly used control system in guided missiles. Everything in the hydraulic-electric system is electrical except the controller and actuator. The actuators are moved by hydraulic fluid; and since fluid is nearly incompressible, the hydraulic actuator has a very fast reaction time.

A completely hydraulic system is not feasible; in fact, sensing units which operate hydraulically are nonexistent. The use of a computer which operates by hydraulics would be limited by lack of flexibility in performance and design.

In a hydraulic-electric system, hydraulic pressure is maintained by a continuously operated pump. As the fluid is used over and over again, the operating time of the system is unlimited. This makes the system well suited for use in long range missiles.

Gyros are used as an attitude reference for the missile. Electrical pickoffs sense any deviations of the missile in relation to the gyro reference. The signals are computed and amplified electrically. The amplified signals operate the controller which is a hydraulic transfer valve. The controller regulates the amount and direction of fluid flow to the actuator.

ADVANTAGES AND DISADVANTAGES

The most important advantages of a hydraulic-electric control system are the rapid response to error signals and the large amount of force available when using hydraulic actuators.

A sensor and computer which operate electronically can be designed to produce an output of

the correct form for any error which the missile may acquire in flight. A signal of correct form is one which, at all times, is of the proper amplitude and phase to produce actuator movement which results in the required correction. Couple this electrical signal to a rapid actuator, which responds to signals the correct amount regardless of impeding forces on the control surfaces, and the result is nearly an ideal system.

A disadvantage of this system is the need for two power sources: electric and hydraulic. This requires maintenance, assembly, and checkout personnel trained for electrical and electronic systems in addition to personnel trained in hydraulic systems. There is also the problem of supplying not only electrical parts but also hydraulic parts and test equipment. Of course, these disadvantages exist in any combination type control system, and they are, in reality, minor disadvantages when compared to the advantages.

One must view the guided missile field with a broad outlook, keeping in mind that a system which is ideal for one missile may not be satisfactory for another.

CONTROL SYSTEM CHANNELS

There are three distinct channels of control in hydraulic-electric systems for providing control of the missile about the three axes. These channels are referred to as roll, pitch, and yaw. The following paragraphs consider each of these systems separately.

Roll Control

A hydraulic-electric system used to correct for missile roll is shown in figure 7-2. This system is simple and is used on a missile that is aerodynamically stable in the roll axis and does not have a tendency to roll excessively.

The system illustrates proportional control only. It reacts to information which tells the amount of deviation of the missile from level flight. To do this, the signal is proportional to the deviation and is called a displacement signal. In this particular system, *rate control* is not included. Rate action is explained in a later section.

Note that the components can be closely related to the basic control block diagram presented in chapter 5. The gyro acts as a reference for the system. The selsyn is connected to the gyro and senses any roll of the missile. If roll is detected, the selsyn produces an error signal proportional to the amount of roll. This error signal applied to the servoamplifier is the difference between the followup signal and the gyro signal. This error signal is amplified to increase its power so that it can operate the controller. The controller is a hydraulic transfer valve which regulates the flow of fluid to the actuator piston which moves the control surface.

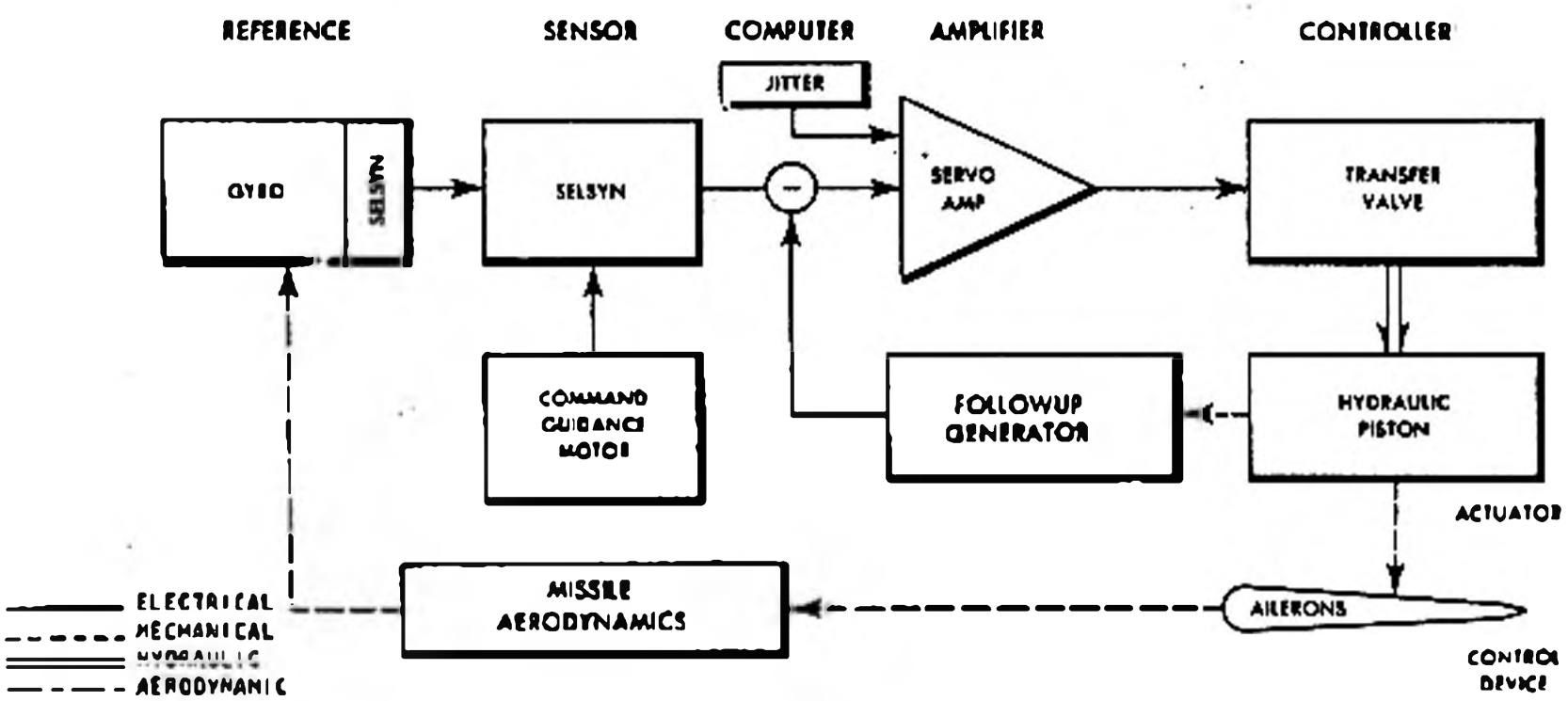


Figure 7-2. Hydraulic-Electric System for Roll Control

Any difference between the actual and desired position of the missile in relation to the displacement gyro produces the required control surface movement. This action produces a corrective movement of the missile which affects the relationship of the reference units and the missile. This action is represented by the block, Missile Aerodynamics, in the illustration, and the dotted connecting line which completes the servo loop.

Now is a good time to explain some commonly used symbols which you will see from time to time in illustrations in this and following chapters. These symbols, representing certain operations in the system, are used to explain system functions.

The circle enclosing an "X" indicates a connection or mixing circuit for several signals. Signals are added if the circle contains a plus sign. The output signal is the sum of several inputs. If the circle contains a minus sign, the resulting signal is the difference between the input signals. A triangle amplifier symbol forms an arrowhead which always points in the direction of the output. A followup generator may be any pickoff which detects position such as a selsyn, potentiometer, or reluctance pickoff.

The block labeled *jitter* represents a generator or oscillator which provides a small AC signal of approximately 25 Hertz. This jitter signal keeps the transfer valve constantly vibrating as an aid to overcome friction and to make movement of the controls smoother.

Pitch Control

A hydraulic-electric system for pitch control is illustrated in figure 7-3. In this system a gyro and selsyn are again used as the reference and sensor units for changes in pitch attitude. The error signal is put through several preamplifiers to strengthen a weak selsyn signal before it is fed to the demodulator or servoamplifier input. Part of the proportional error is fed to the computer, and the rest is fed directly to the servoamplifier. The servoamplifier increases the power of the selsyn signal to the point at which the signal can operate the transfer valve solenoids.

The transfer valve controls hydraulic fluid to the actuator which moves the control surface. The control surface affects the attitude and changes the relationship of the gyro and missile as indicated by the dotted airframe dynamics feedback loop.

The dynamic loop, when combined with the followup signal, completes the proportional control. This particular system includes rate control. Rather than using a separate rate gyro, the displacement signal is modified by the computer which changes it to a rate-of-change signal. The displacement signal and the rate-of-change signal are combined at the input of the servo-amplifier.

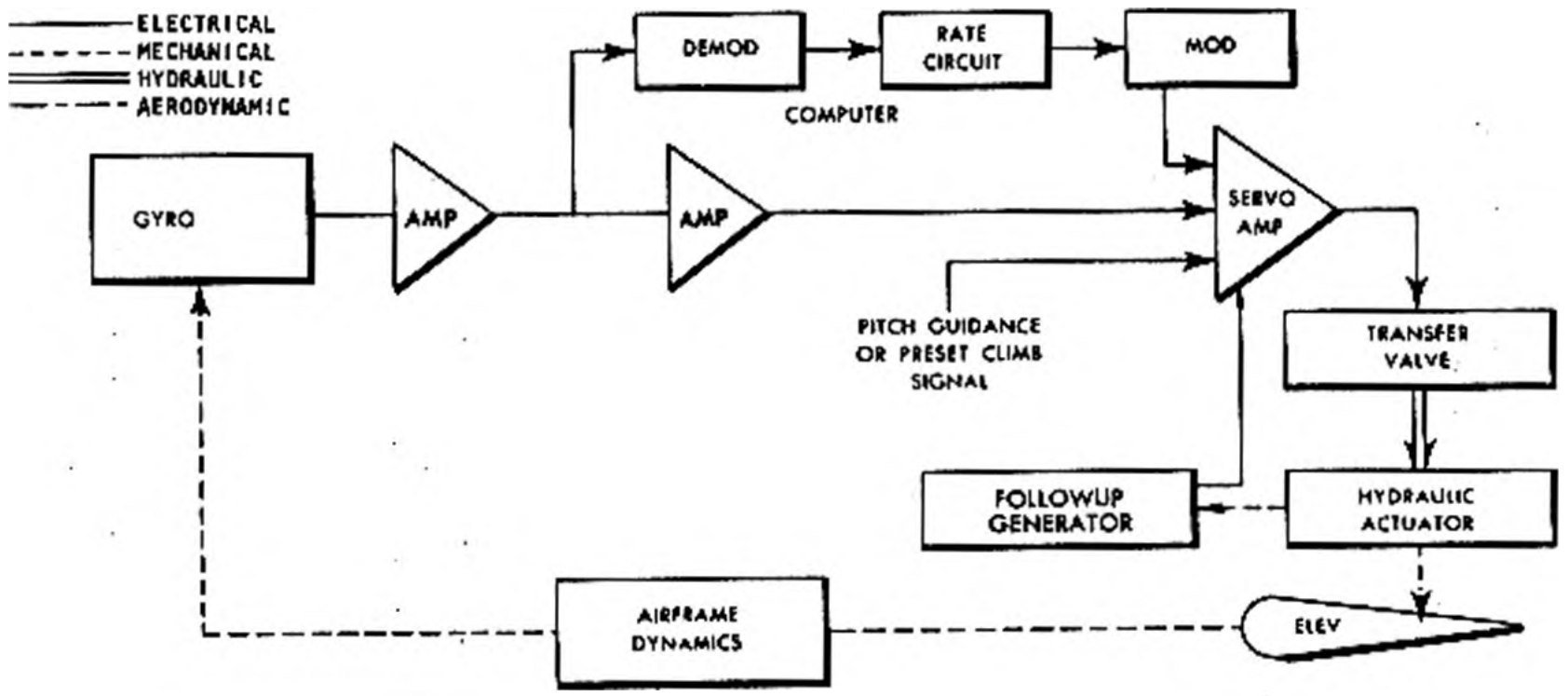


Figure 7-3. Pitch Hydraulic-Electric System

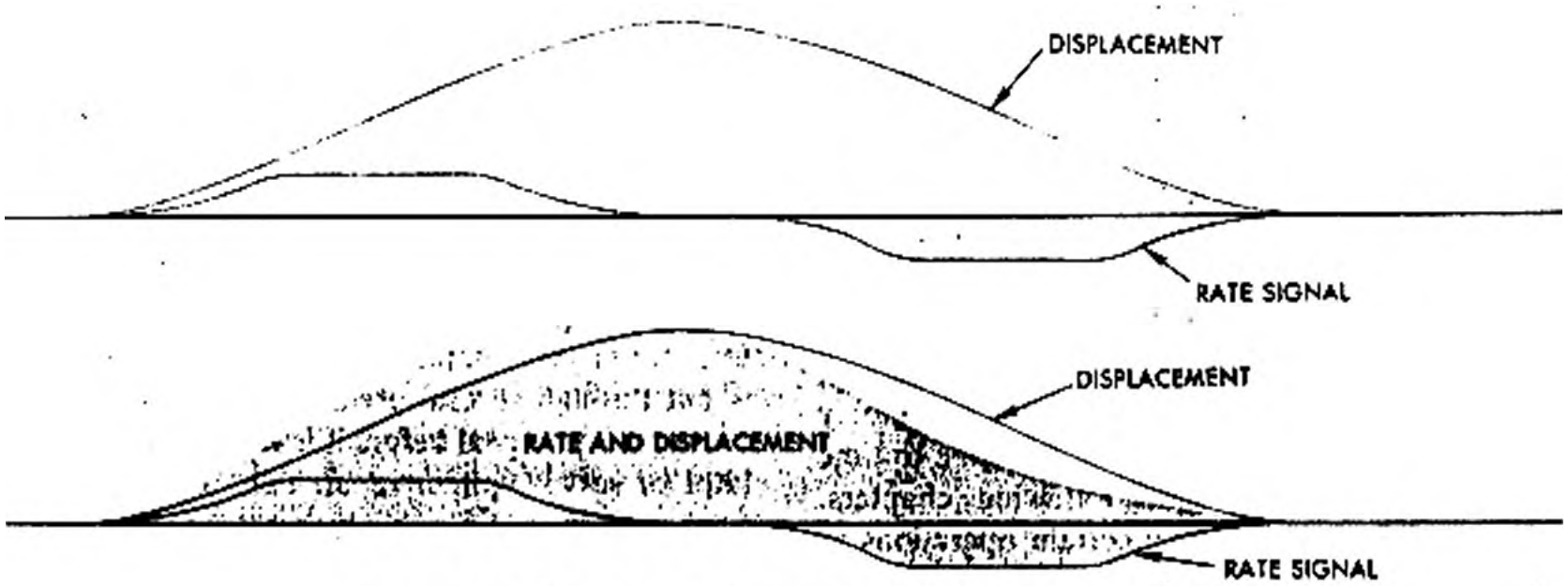


Figure 7-4. Effect of Combining Rate and Displacement Signals

Figure 7-4 illustrates the effect of combining rate and displacement signals.

Assume the missile nose begins to raise. The displacement gyro senses the movement and the selsyn generates an error or displacement signal as illustrated in the top part of figure 7-4. Part of this signal is routed to the rate circuit (indicated by the white area in figure 7-4 top). These two signals are combined at the servo-amplifier. As shown in the bottom drawing of figure 7-4, you can see that the rate signal supports the displacement signal at this time, increasing the corrective action taken by the system. When the missile has deviated as far as it is going to, you will note that the rate signal returns to zero and the total rate and displacement signals are now equal to the displacement signal only.

Now consider the missile nose swinging down again toward normal level flight attitude. The displacement signal is decreasing and the rate-of-change signal from the computer has changed phase. This signal is now opposing the error signal, causing a decrease in the resulting corrective signal. Referring to the bottom of figure 7-4, you can see that the combined rate and displacement signals are now less than the displacement signal alone. As the missile nose approaches level, the two signals cancel out and cause the missile smoothly to assume level flight. Without the rate signal, the nose would continue to drop until the opposite going error signal initiated corrective action and caused the nose to raise again. This action would continue through several oscillations, being referred to as *hunting*.

One servoamplifier input in figure 7-3 provides for injection of another signal into the control system. This signal may be a pitch guidance signal or a preset signal which causes the missile to climb immediately after the launch phase.

A guidance signal is necessary in all types of missiles where directional information leading to target interception is required. These signals can originate from one of many types of guidance systems such as command, beam rider, inertial, or homing. The gyro and guidance signals operate together in the control system. The gyro gives stability while the guidance signal provides direction or corrections to the gyro heading.

Yaw Control

Figure 7-5 illustrates the yaw control loop of a hydraulic-electric control system. The yaw stabilization loop is similar to the other two loops except for the addition of integral control. A reluctance pickoff produces a signal proportional to the angular displacement of the missile with respect to a directional gyro sensor. This signal has three paths to the servoamplifier. The top path provides the proportional signal; the middle path, rate control; and the lower path, integral control. These signals combine at the servoamplifier to provide the desired correction signal no matter how far, how fast, or how long the missile deviates from the desired heading. Guidance signals may be injected at either of the two points indicated on figure 7-5.

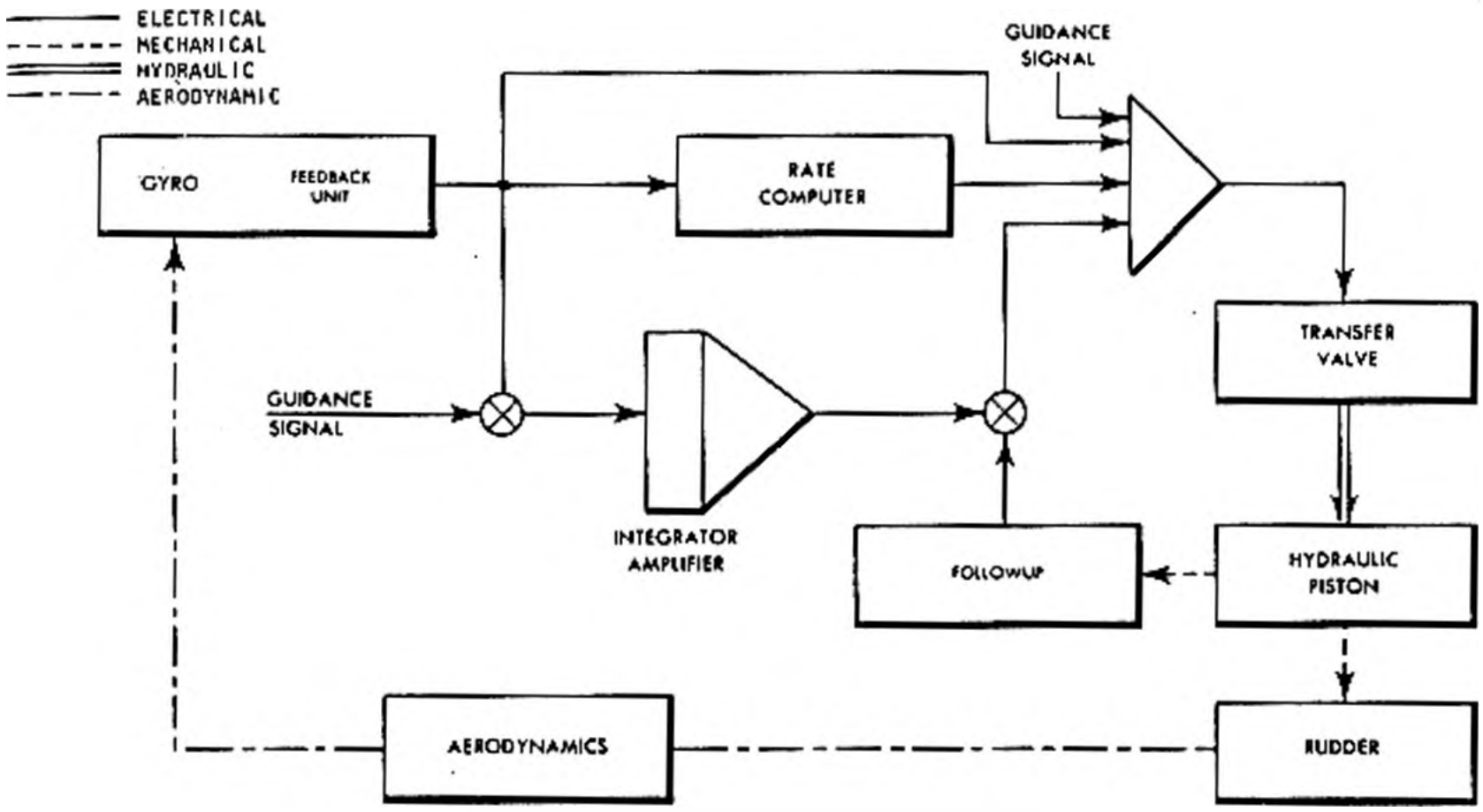


Figure 7-5. Yaw Hydraulic-Electric System

INTEGRATOR ACTION

The purpose of the integrator circuit is to detect an error of a certain phase that has existed for an extended period of time. This detecting is done by reproducing an output which is proportional both to the amplitude of the error and to the length of time that the error exists. The integrator actually accumulates the error signal. This integrated signal is mixed with the other error signals at the servoamplifier.

In figure 7-5 you can see that the signals entering the yaw servoamplifier are the displacement, rate-of-change; and the accumulated or integrated error signals. The integrator functions in the system are as follows: Assume that the missile is blown off course by a strong crosswind and the error is so great that the control system response is insufficient for complete corrective action. The integrator responds to this error and after a period of time produces a signal large enough to cause complete correction of the yaw error. The integrator signal aids the proportional displacement signal to overcome an error that exists for a period of time. The integrator circuit does not respond to rapidly changing error signals.

CHANNEL INTERCONNECTION

Most control systems have some type of channel interconnection. Interconnection of channels is always necessary when the control device must control the missile in more than one axis of movement. An example of this is the use of ruddervators which control both pitch and yaw. So far, references have been made to aerodynamic control surfaces. System principles are much the same regardless of whether control surfaces, movable jets, or fixed steering jets are used. Movable jet control is most likely to use channel interconnection and is used here as an example.

To obtain full control in all three axes of movement, at least two movable (gimballed) jets must be used. Such a system is illustrated in figure 7-6. As shown here, control in all three axes is possible. The jets move in any direction and each jet responds to signals from any of the three channels. Consider a system having four movable jets. Each of these jets moves back and forth in only one plane. Two jets are used for pitch and the other two for yaw. All four jets are used to provide roll control. The roll channel must be connected to both the yaw and pitch channels, a condition which introduces channel interconnection.

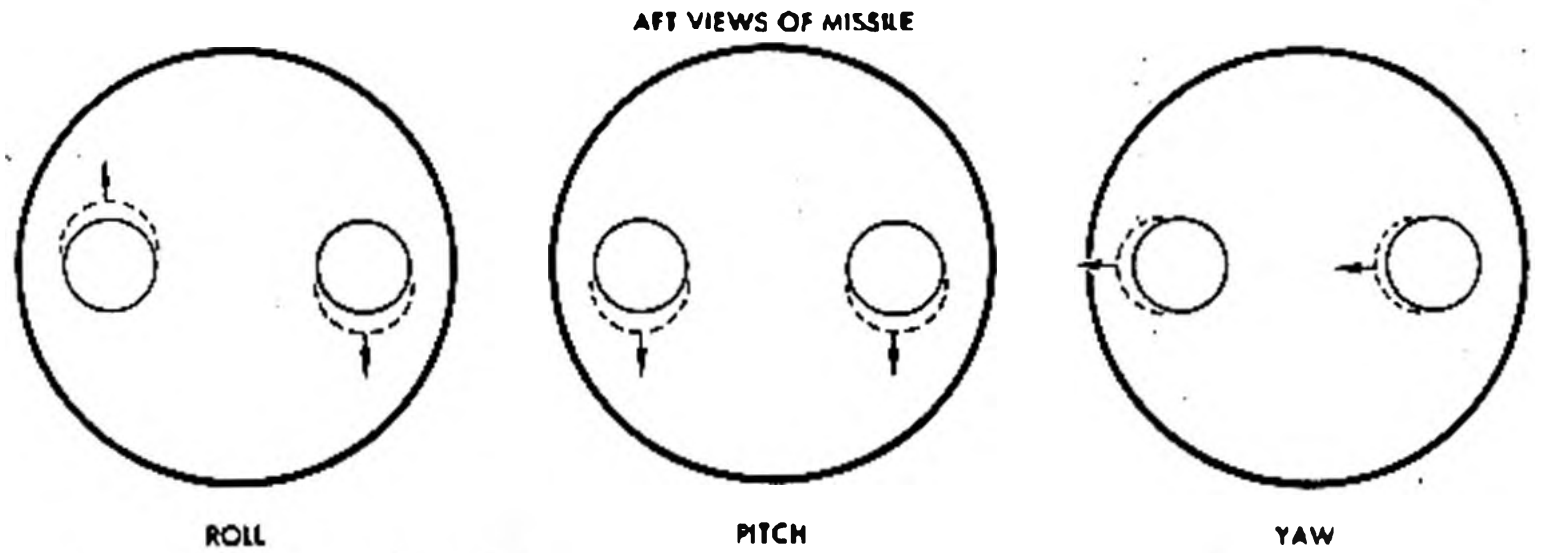


Figure 7-6. Attitude Control using Two Movable Jets

Hydraulic Interconnection

Two methods of interconnecting the roll channel with the yaw and pitch channels are considered; one hydraulic and the other electrical. The hydraulic method, shown in figure 7-7, uses three normal control channels up to the hydraulic actuators. At this point the three transfer valves are used to move four jets. Differential movement of the jets is accomplished by interconnecting four hydraulic actuators.

In the figure, a certain fluid output from each transfer valve is assumed. The cylinders contain fluid on both sides of the pistons. To simplify the explanation, the side representing high pressure is shaded in color. Note that the cylinders can also move. The arrows indicate either fluid, cylinder movement, or jet movement depending on the location of the arrow.

The interconnection of the system can be understood by assuming displacement of each transfer valve individually and then following the resulting flow of hydraulic fluid through the cylinders to determine the jet movement. Movement of the jet linkage to the right produces a clockwise movement of the jet nozzles as shown in the drawing representing a rear view of the missile.

In the pitch channel the fluid flow will result in movement of the nozzle of jet 4 clockwise and jet 2 counterclockwise. This action tilts both jets upward and produces a nose-up movement of the missile. As both cylinders are the same size, they have equal displacement and both nozzles move the same distance.

In the yaw channel, fluid flow results in movement of jet 1 clockwise and jet 3 counterclockwise. This causes both jets to move to the right, producing a turn to the right.

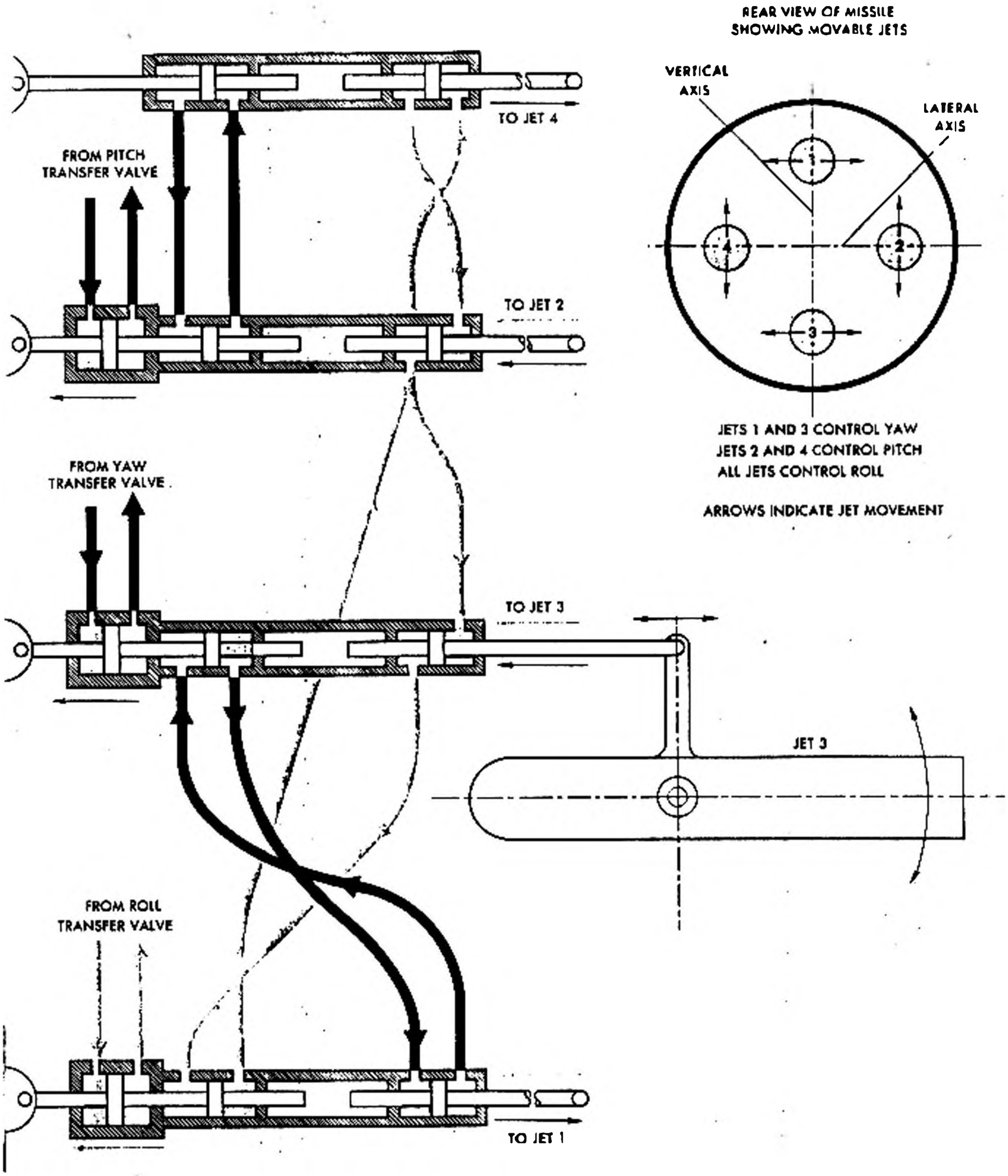
Movement of the roll master cylinder produces movement of all the jet pistons. In this case differential action is produced as one of the pitch jets moves down and the other up, and one of the yaw jets moves to the right and the other to the left. The missile will roll clockwise. There is no pitch or yaw movement, yet each jet is aiding the roll movement of the missile.

While roll action is occurring, there is nothing preventing the pitch and yaw channels from producing corrections at the same time. The yaw and pitch channels can either aid or counteract the roll action of each cylinder to produce the required jet positions to cause changes in the yaw and pitch attitude of the missile.

With this type of channel interconnection, followup signals must indicate the relative displacement of each channel piston rather than the nozzle displacement. If the followup signal were connected to the actuator linkage, it would indicate the result of two channels simultaneously.

Electrical Interconnection

The second method of interconnecting channels is by electrical means. As shown in figure 7-8, the interconnections are made just prior to the servoamplifiers. The electrical distribution is made at the output of the three channel amplifiers. These amplifiers are double-ended and produce two outputs 180° out of phase. The pitch signal again affects jets 4 and 2 by feeding into the respective servoamplifiers. If you assume that a signal of a certain phase causes clockwise movement of all



REAR VIEW OF MISSILE SHOWING MOVABLE JETS

JETS 1 AND 3 CONTROL YAW
JETS 2 AND 4 CONTROL PITCH
ALL JETS CONTROL ROLL

ARROWS INDICATE JET MOVEMENT

Figure 7-7. Hydraulic Interconnection of Channels

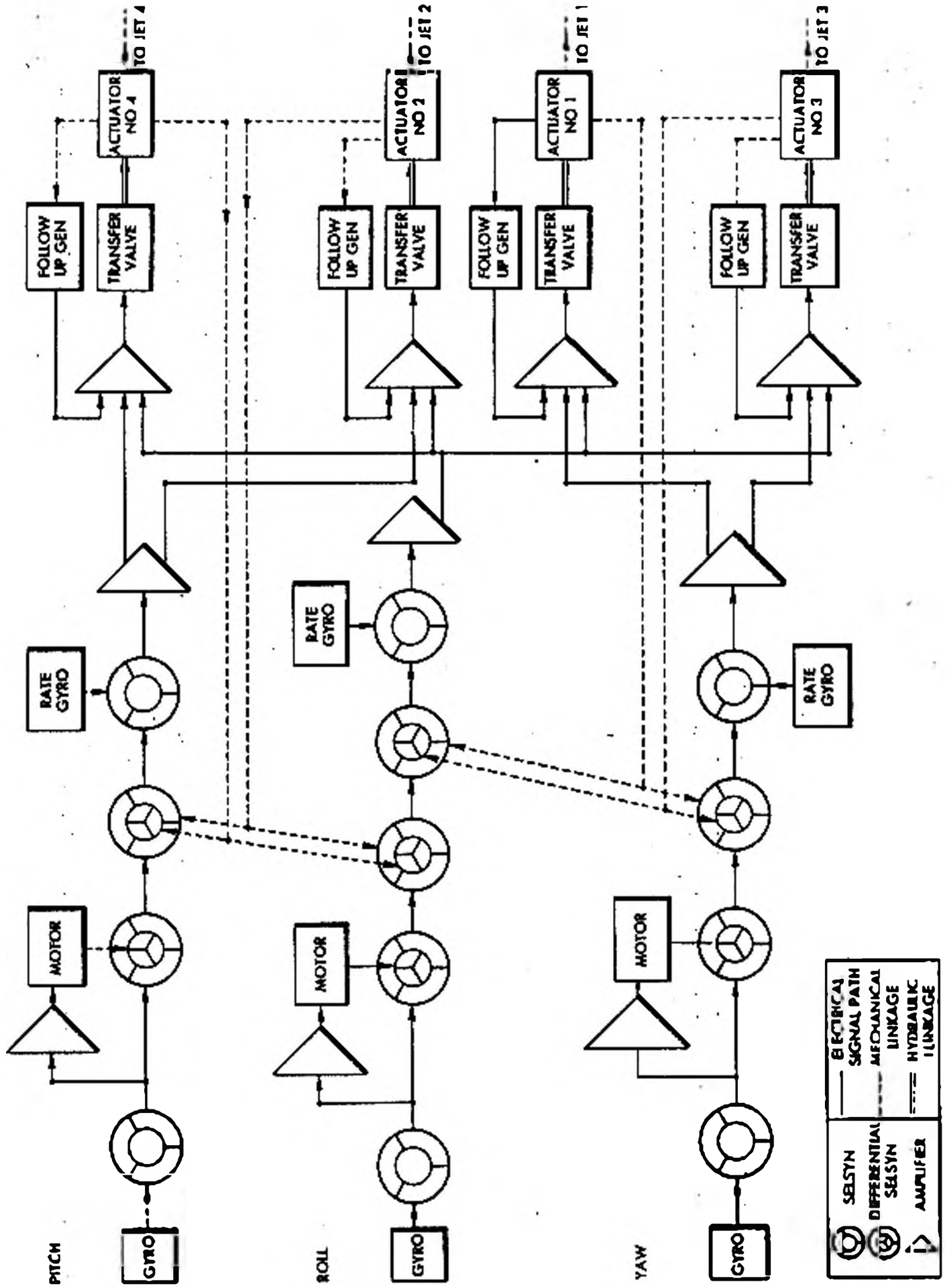


Figure 7-8. Electrical Interconnection of Channels

the jets, then the signal to jets 4 and 2 must be of opposite phase for pitch control. The double-ended amplifiers produce the out-of-phase signals. The yaw signal feeds into the other two servoamplifiers in the same way. The roll signal feeds into all four servoamplifiers, since it must operate all four actuators.

Rate and integral control are included in this system. You can see on the figure that the rate signals are obtained from gyros, and integral control is obtained from variable speed integrators fed by the displacement gyros. Differential selsyns are used to mix information from the displacement, rate, integral, and followup selsyn generators.

The system has several followup loops. In each case, actuator position information is fed back to the input of the respective servoamplifier. The followup signal produces jet movement that is proportional to the servoamplifier input. Actuator position information is fed back to two of the three control channels. This feedback is necessary because each actuator has an effect on the missile in two axes. The roll channel has followup signals, since each actuator affects the missile in roll. The combination of actuator signals to any channel produces a signal which represents the true followup for that channel.

SECTION B
Pneumatic and Pneumatic-Electrical Systems

There have been several missiles that used pneumatic or a combination of pneumatic-electric control systems. The systems are very similar to the hydraulic operated control systems. The major difference is the use of air pressure instead of fluid as the activating medium. Some of the systems even use air to drive the gyros and for pickoff and relay functions.

Figure 7-9 is a block diagram of a pneumatic control system. You will notice that the reference unit, as in the hydraulic system, is a displacement gyro. In place of selsyns, this system uses air pickoffs as sensors. Air relays are used to control the servoamplifiers which are also air activated. The servos control the control actuators which are operated by air pressure and move the control surfaces.

Referring to figure 7-10, you can associate the pictorial representation of the gyros, pickoffs, relays, and controller-actuator units that are depicted in the block diagram. There is no electrical circuitry associated with this system.

A combination of pneumatics and electronics is illustrated in the block diagram of figure 7-11. The actuators and controllers are operated by air pressure while the reference, sensor, and amplifier sections are electronic.

The pneumatic and pneumatic-electric control system has some decided disadvantages. The range of the missile is limited by the amount of air that can be stored and carried by the missile. This fact restricts the missile to a relatively short range. The response time of a pneumatic system is not as rapid as with a hydraulic system. This is because of the compressibility of air. Because of these disadvantages, pneumatic control systems have not had the success of electrical and hydraulic systems, and present day missiles all use either electrical, hydraulic, or a combination of the two systems.

SECTION C
Ballistic Missile Control System

A ballistic missile is essentially a large cylinder with a thrust device on one end. Such a device is inherently unstable. Uncompensated disturbing factors cause unrestricted rotational motion. Aerodynamically, the center of pressure lies ahead of the center of gravity. To complicate this situation, a ballistic missile is ordinarily subject to disturbing forces during flight, such as propellant slosh, missile airframe flexing, and vibrations caused by engine gimballing. The result is an unstable object which is agitated by flight conditions and, therefore, presents a serious control problem.

Various methods are employed to stabilize a ballistic missile. Spin stabilization can be achieved by giving the missile a large axial angular momentum, or external fins can be used for aerodynamic stability.

You may think of a ballistic missile control system as having two parts: one which measures missile attitude, compares it with an established reference, and computes signals for proper thrust direction; a second which changes the thrust vector in accordance with the computed signals. The first

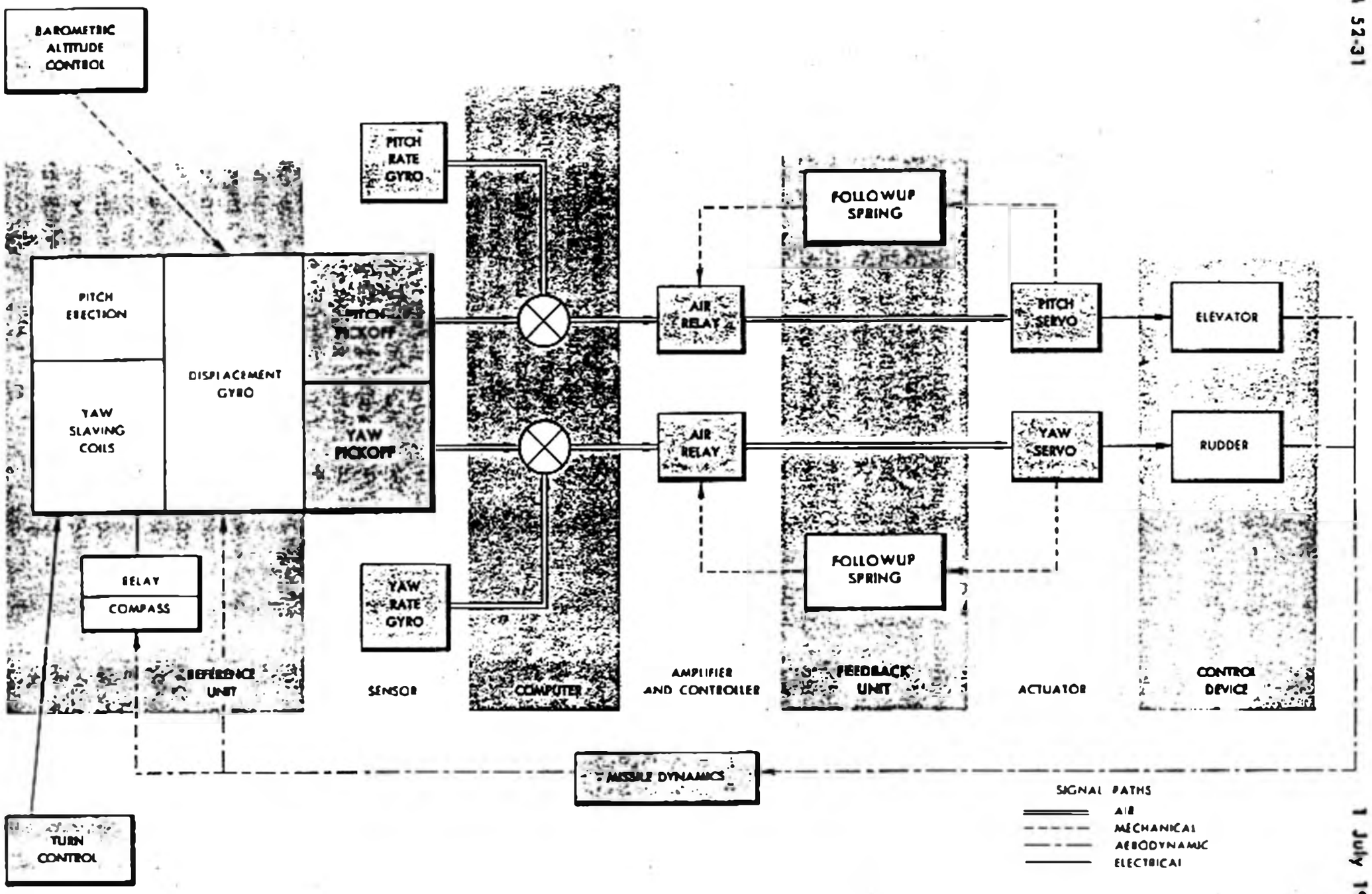


Figure 7-9. Operation of Pneumatic Control System

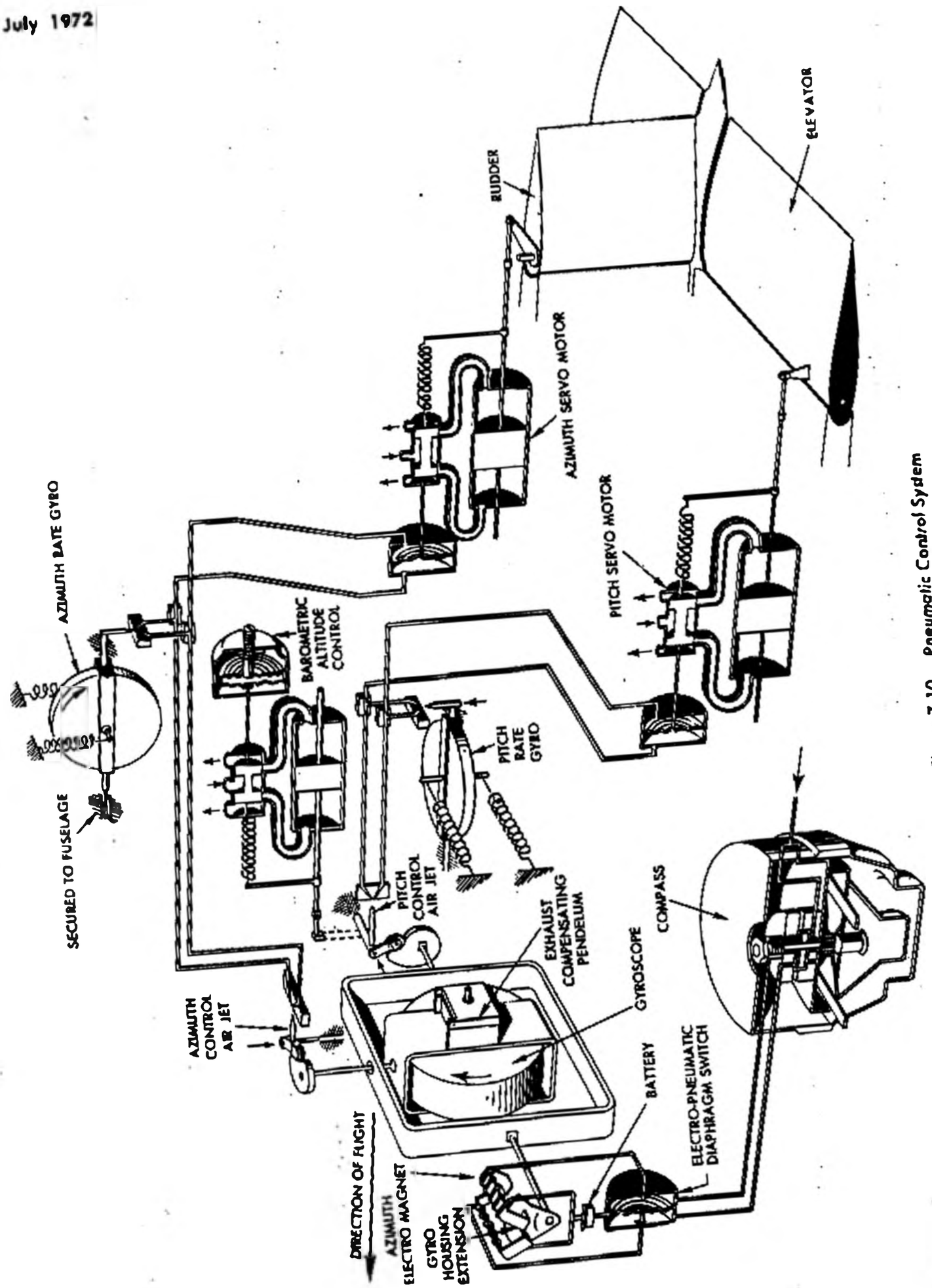


Figure 7-10. Pneumatic Control System

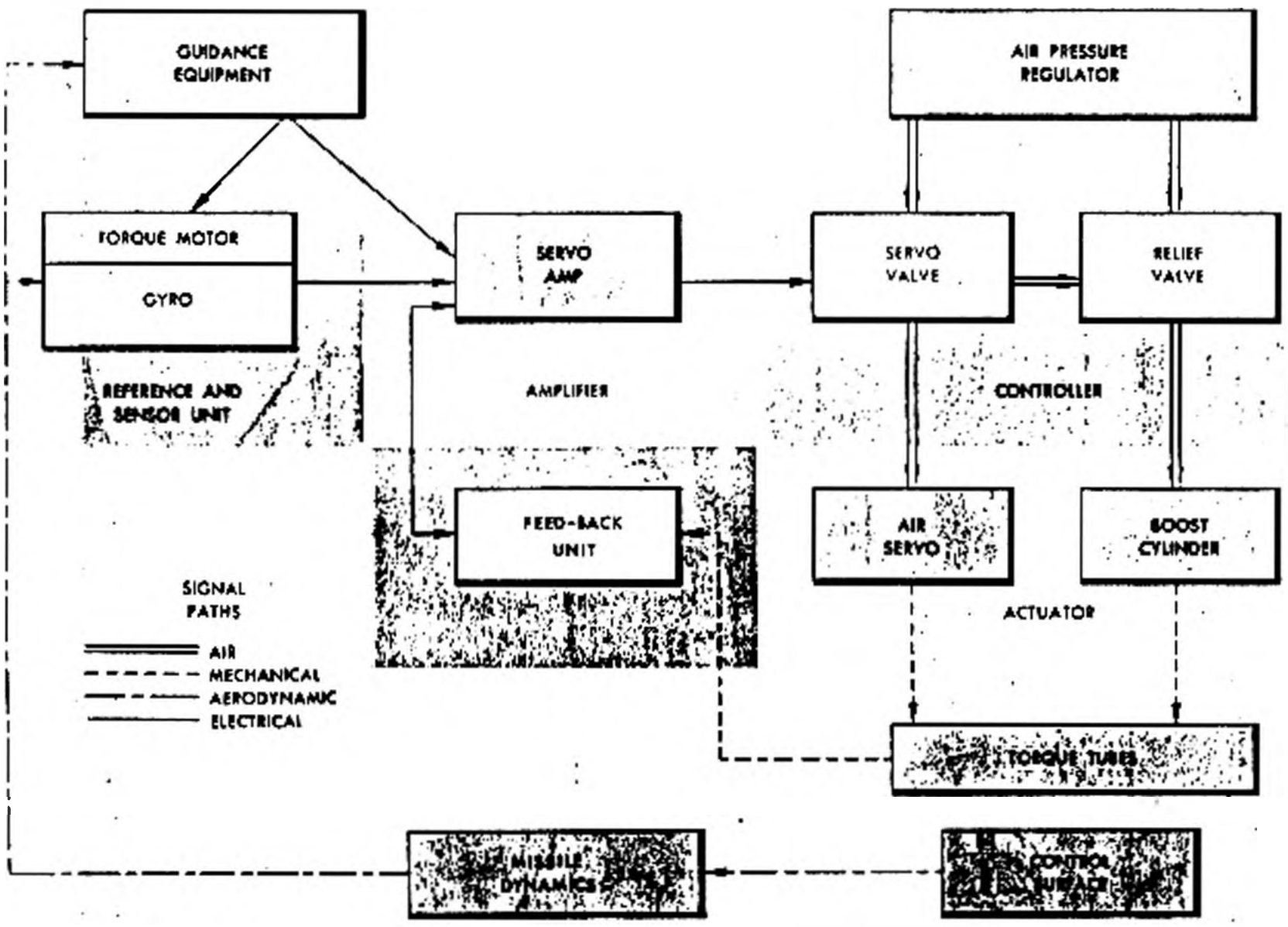


Figure 7-11. Pneumatic-Electric Control System

group of functions is the autopilot and the second the thrust vector control. Figure 7-12 is a block diagram of a typical ballistic missile control system.

A control system has several functions. During the initial portion of flight, its programmer feeds attitude reference signals into the system. These inputs are preplanned and based on studies of the dynamics of the missile, the aerodynamic forces expected, and the desired course the missile is to fly during the first part of its flight. Measurements of actual missile attitude made by the rate and displacement gyros are compared with the attitude references from the programmer. The difference between the two is sent as a corrective signal to the actuation system. This changes the direction of the missile thrust by gimballing the engine, providing a compensating, stabilizing torque. This is a closed loop and the true deflection of the engine is compared with the deflection called for, and any error is detected and sent back through the system for further corrective action. On most ballistic missiles, the actuation system consists of hydraulic valves and actuators.

In addition to providing stability, the control system provides the missile with a type of programmed guidance during the initial period of flight. By holding the missile attitude to established references, the control system can maneuver the missile by changing the references according to a preplanned schedule. These maneuvers, which may include a roll program, a transition turn, and a gravity turn, are achieved by establishing appropriate reference attitudes as a function of time in the programmer. The programmer provides sequenced signals to the gyro torquers, and these determine the desired reference attitudes for the desired maneuvers.

Finally, the control system provides responses to guidance commands and, in the later portions of its flight, enables the missile to follow the commands initiated by the guidance system.

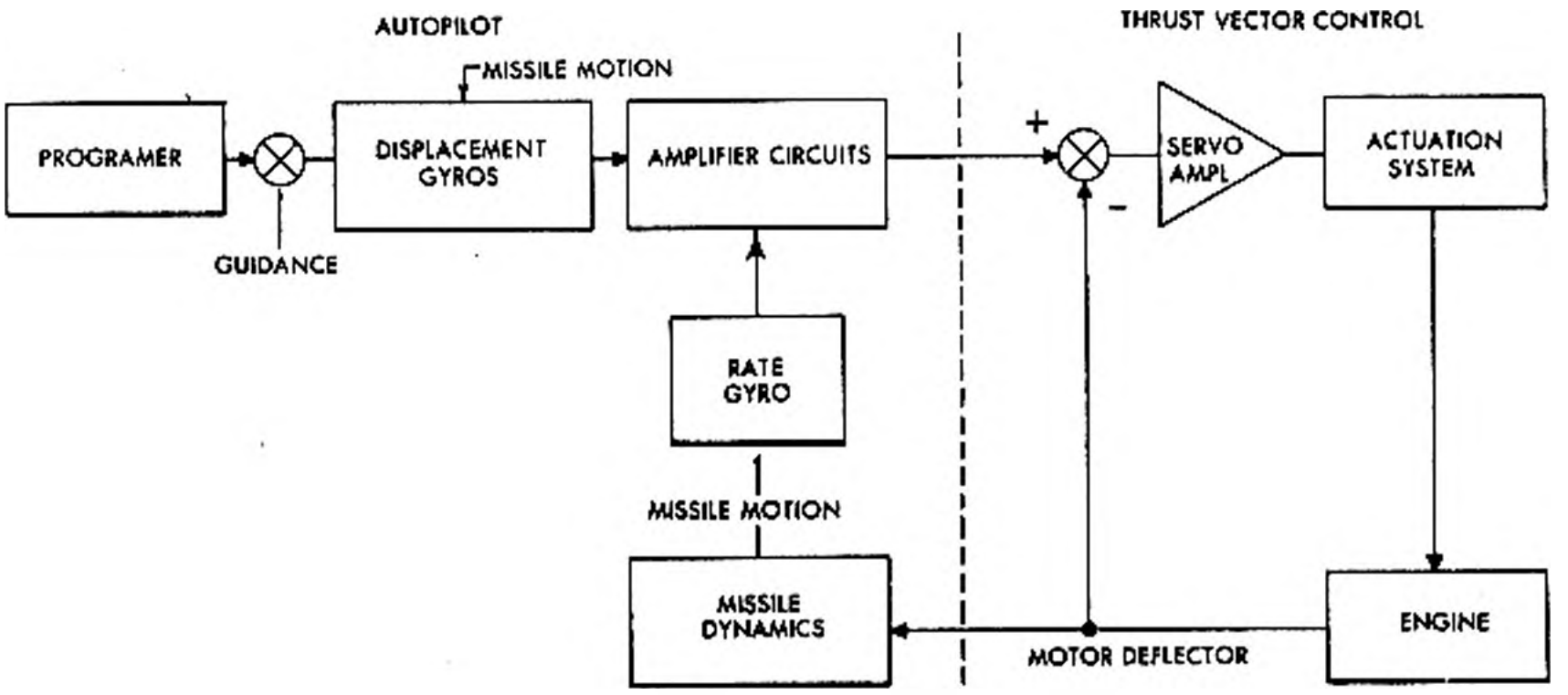


Figure 7-12. Typical Ballistic Missile Control System

We begin examination of the control system by considering the part of it to which the programmed and computed signals are sent—the thrust vector control.

THRUST VECTOR CONTROL

The thrust vector of a missile can be controlled by (a) changing its direction and (b) cutting it off. Liquid-propelled rockets present entirely different problems from solid propellant rockets in achieving thrust control.

Thrust Direction of Liquid Propellant Engines

The most common method of deflecting the thrust vector of a liquid propellant rocket is by swiveling the thrust chamber. There are many problems inherent in this type of control. The thrust chamber is mounted in gimbals and the propellants must be routed through high pressure hoses or the gimbal bearings. Either method introduces the possibility of fuel leaks.

To provide thrust vector control around all three axes of rotation, at least two thrust chambers are necessary. Figure 7-13 illlustrates a ballistic missile using the two booster engine thrust chambers for thrust vector control during the launch and boost phase of the flight. During the later part of the flight while the missile is in the extreme high altitude, the small vernier engines on the sides of the missile provide attitude control.

Figure 7-14 is a closer view of the thrust chambers and illustrates the complexity of the piping necessary to feed propellants to gimballed chambers.

Thrust Direction of Solid Propellant Engines

In most solid propellant engines, the thrust chamber is a large percentage of the mass of the missile which makes gimballing the chamber impractical. Some solid propellant missiles use a swiveling nozzle for thrust vector control.

An early approach to vector control of solid propellant missiles was the use of control surfaces (called jet vanes) placed in the rocket exhaust stream.

Jet vanes have not proven to be an effective method of control. Vanes in the exhaust stream cause perturbances and result in loss of thrust. They do not have linear characteristics at large deflection angles. As new solid propellants have been developed, new compounds, such as aluminum, have been added. These additives are not compatible with jet vanes and cause buildup of slags or erosion of the vanes, either one of which changes their control characteristics.

Another means of control that has been used is external control surfaces or fins. These have not been effective immediately after liftoff nor are they

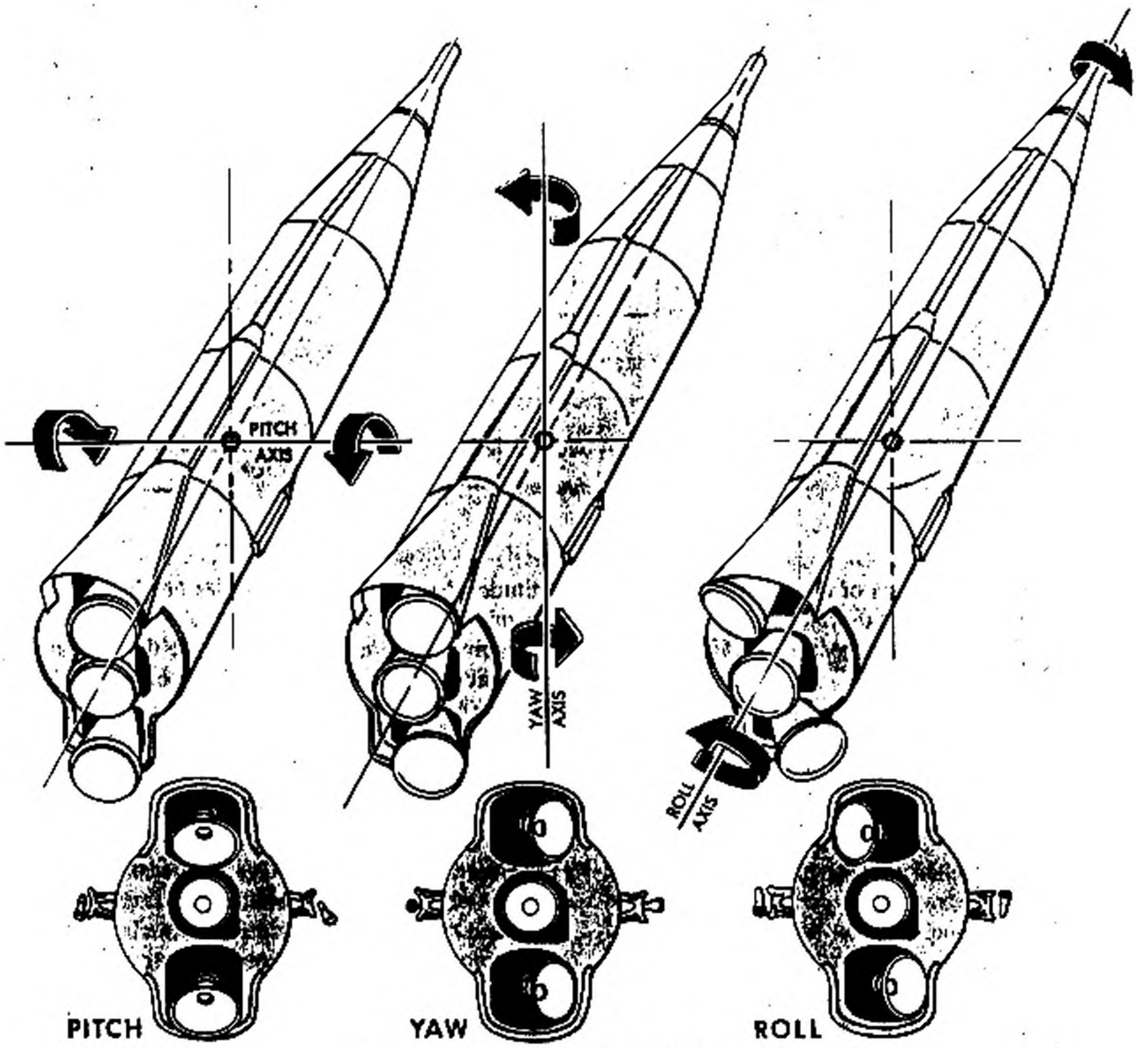


Figure 7-13. Action of Gimbaled Thrust Chamber

useful under atmospheric conditions which approach a vacuum. Some other type of control must be used in conjunction with fins to provide adequate response under various flight conditions.

Some of the latest developments in the control of large ballistic missiles are exemplified in the LGM-30 Minuteman. The Minuteman is a 3-stage solid propellant missile. Stages I and III use control systems with only slight differences. Stage II uses a system entirely different from the other two.

Attitude control of stages I and III is accomplished by swiveling the exhaust nozzles (four on each stage). An electrohydraulic package is used to swivel the nozzles to provide pitch, yaw, and roll movement about the missile axes. An angular accelerometer unit senses accelerations about the pitch and yaw axes of the missile. These movements are measured as pitch and yaw angular rates and transmitted as signals to a computer. The computer analyzes and compares these signals with data stored in its memory in order to apply limits to the angular velocity of the missile during stages I and III firing periods. This action prevents the missile from exceeding accelerations that could

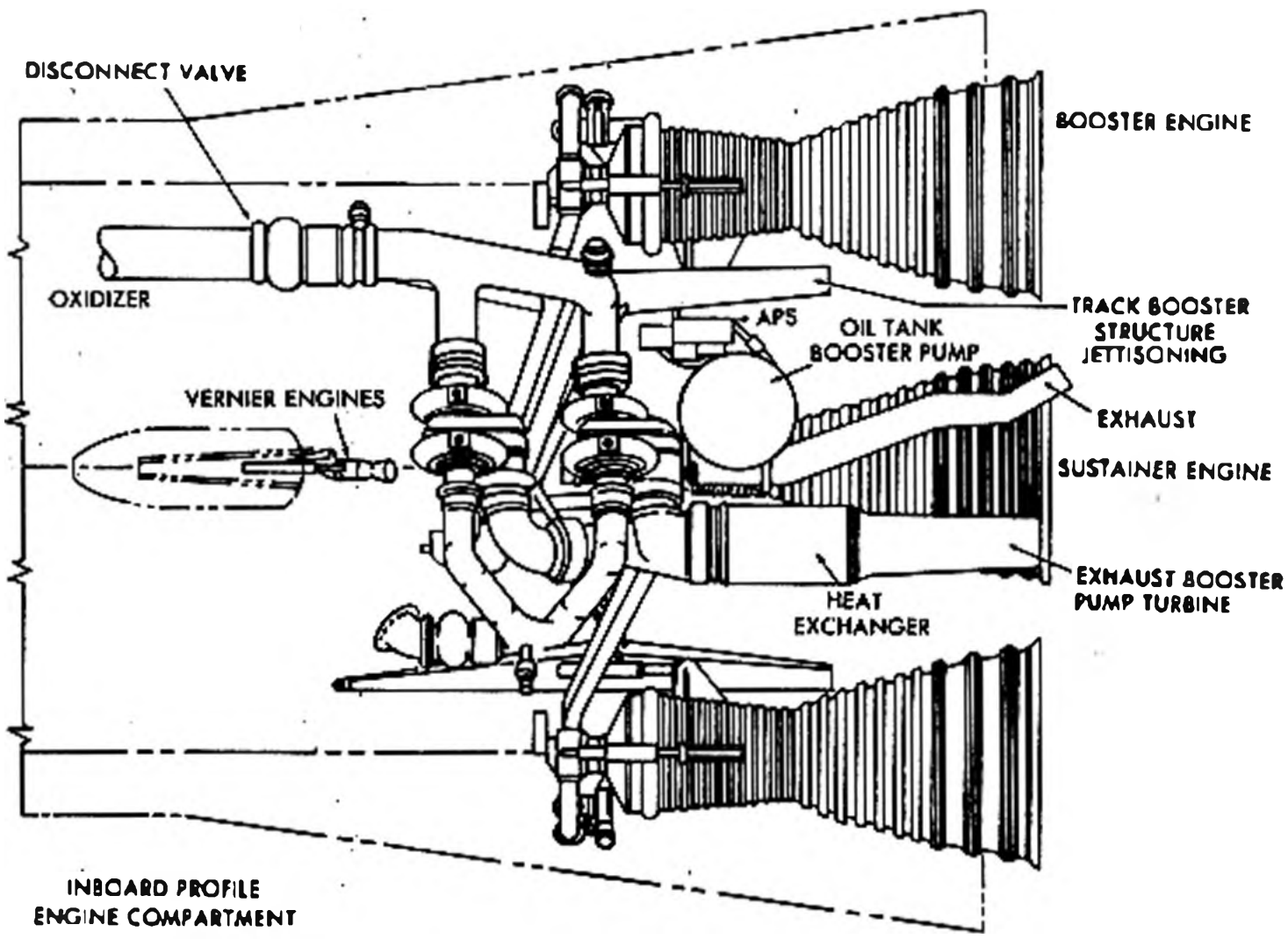


Figure 7-14. Propellant Feed to Gimbaled Thrust Chamber

cause structural failure. The output of the computer is applied to the exhaust nozzle control servoamplifier. The four computer correction commands are pitch plus roll, pitch minus roll, yaw plus roll, and yaw minus roll. Thrust vector control of the four nozzles is performed by an amplifier assembly and an exhaust nozzle control servocontrol loop. The nozzles are moved to the required positions by four hydraulic servocylinders, each of which is attached through linkage to a nozzle. When the nozzle control receives a correction command from the computer, the servocylinder moves a discrete distance, dependent on the degree of correction, thereby moving the nozzle. This movement provides a change in the thrust vector causing the missile to change its attitude position.

Stage III is controlled in the same manner as stage I. This stage is much smaller than stage I and does not depend on its control functions until the missile has reached an extremely high altitude. The control system of stage III is much smaller than the system for stage I. As an example, stage III hydraulic pressure is only 1500 psi as compared to 3000 psi for stage I.

Thrust vector control of stage II is completely different than for the other two stages. Stage II has only one nozzle and control is accomplished by the release of liquid freon into the exhaust nozzle through four freon injector actuators (two yaw and two pitch injectors) which are mounted radially around the motor exhaust cone. This system is known as Liquid Injection Thrust Vector Control (LITV). Figure 7-15 illustrates the stage II nozzle and shows the location of the injector ports and the freon tank. The injector actuator responds to commands from the guidance computer and amplifier.

When the attitude control injector receives a command (yaw for example), the yaw freon injector actuator unit lifts a pintle valve allowing liquid freon under pressure to flow into the exhaust gases and causes a change in thrust direction that causes the missile to correct its yaw attitude position. The direction of correction required determines which of the two yaw freon injectors func-

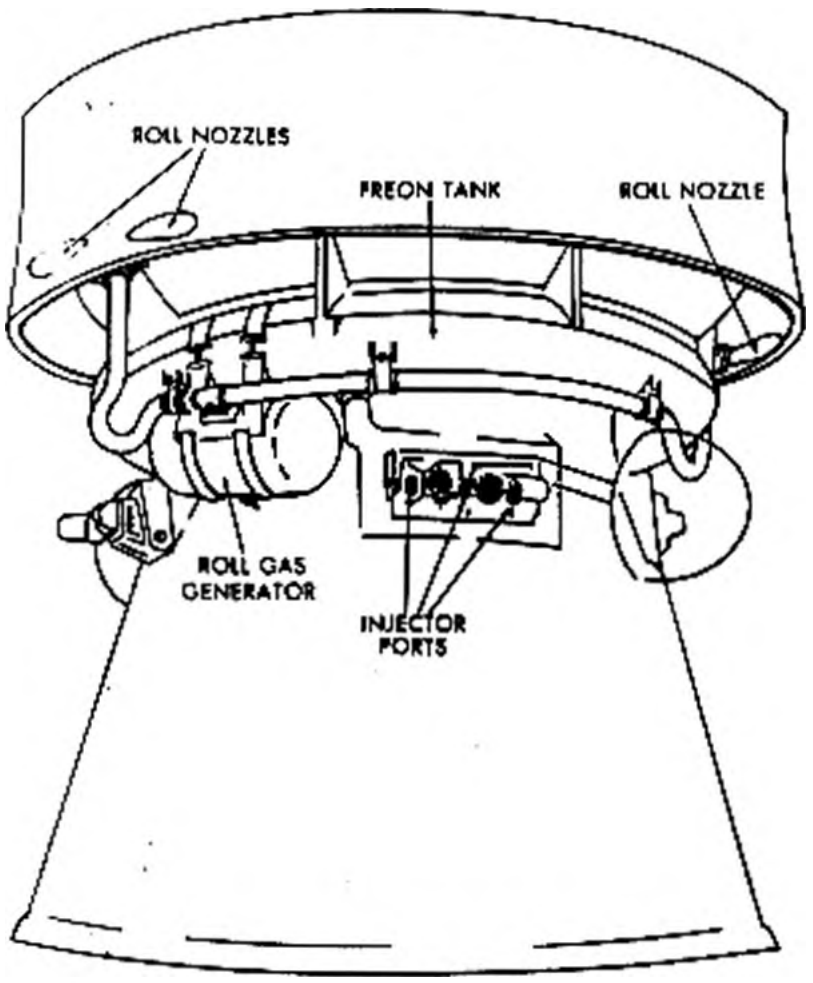


Figure 7-15. Nozzle Configuration (UTV control)

tions. A pitch error will cause a like operation in the pitch channel and will activate one of the pitch freon injectors to cause missile attitude correction. Figure 7-16 shows the positions of the freon injector actuators around the nozzle and the location of the freon pressure generator.

Figure 7-16 shows a roll gas generator and two roll control exhaust ports on each side of the missile. It is through this system that roll control of stage II is accomplished. When the stage II is first activated, exhaust gases from the roll gas generator are exhausted through the four gas jets. The jets are in neutral or null position and thrust is generated in equal and opposite directions. When missile roll is required, commands from the computer are fed through an amplifier assembly to the roll control assembly. Roll control solenoids move the gas jets from their null position and position them in opposite directions, causing unequal roll thrust to be generated. When the roll correction has been made, the solenoids are deactivated and the nozzles return to the null position.

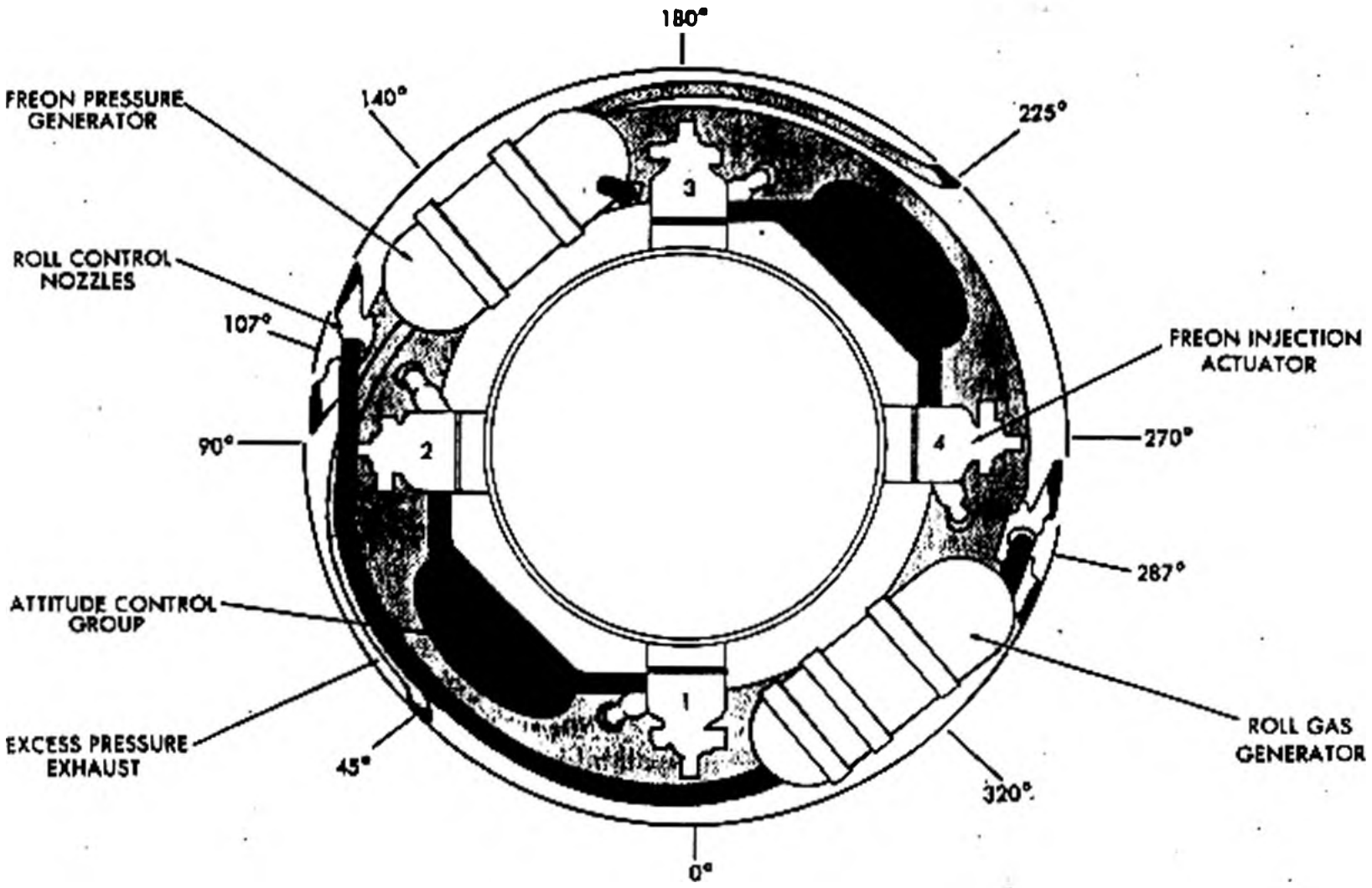


Figure 7-16. Attitude Control Injector Orientation

THRUST CUTOFF CONTROL

In both liquid and solid propellant rockets, precise thrust cutoff control is important to the accuracy of the mission. Very small errors in the final velocity of the missile will result in large inaccuracies at impact. This problem is complicated by the fact that there is a period of thrust decay after engine cutoff is signaled. This results in a gradual decay of engine impulse rather than a sharp cutoff.

Because of the uncertainty about the extent of the impulse after the main engine cutoff signal, verniers are normally used to adjust final velocity. Verniers may be small rocket engines or they may be gas jets. The impulse of vernier engines is much smaller than the thrust of the main engine and very discrete thrust velocity adjustments can be made with them. Verniers are sometimes used for controlling missile roll during main engine operation as is the case with the second stage of the Minuteman missile. They are always used for roll, pitch, and yaw adjustments when used after main engine cutoff.

Liquid Propellant Engines

Thrust cutoff on liquid propellant engines is accomplished by closing valves to stop the flow of fuel and oxidizer into the combustion chamber. The accurate timing of the operation of the fuel and oxidizer cutoff valves is very important to the accuracy of the missile impact. In the case of a satellite or space probe, precise cutoff of the engine, is necessary for the space vehicle to attain its programmed orbit. Ordinarily the engine cutoff signal is initiated from the computer.

During the time of flight of the missile, the guidance system continuously feeds course data to the flight control system and the computer stores and compares these data with the programmed information pertaining to missile location in respect to target location. At the proper point in time and space, the computer initiates a cutoff signal. This signal is amplified and energizes a solenoid that closes the fuel and oxidizer cutoff valves. As the main engine shuts down, the computer calculates the exact location and attitude of the missile and generates signals necessary to make the fine adjustments necessary to correct the attitude and speed of the missile. In the case of a ballistic weapon, these corrections are usually made immediately prior to the release of the warhead so that the warhead is released at the precise point for it to complete a ballistic path to the impact point.

Solid Propellant Engines

Controlling the thrust of solid propellant rocket engines is a factor of propellant burning rate. Burning rates and thrust developed by various solid fuels are explained in detail in chapter 3. Burning rate is a function of pressure, and for solid propellants there is some minimum pressure below which efficient burning will not occur. The thrust of a solid propellant rocket may be terminated by rapidly lowering the chamber pressure. There are several ways of lowering chamber pressure. One way is by blowing off the nozzle. This causes a severe momentary thrust increase which is undesirable. Another method of lowering pressure is by blowing plugs out of the side of the chamber. This method is not entirely satisfactory because of the side thrusts that are created.

The most satisfactory and effective means devised for thrust termination of a solid propellant rocket is that used on the Minuteman stage III motor. Figure 7-17 is a cutaway illustration of this motor showing the thrust termination ports in detail. The small detail sketch of the port shows that a small amount of the propellant covers the ports. After ignition, the propellant must burn several seconds before the activation of the termination port can be effected. Any time after this initial burning time, the forward thrust of the main engine can be terminated by a computer signal that will release the hatches from the ports by means of an explosive bolt or squib. Notice that the exhaust of the ports face forward. Not only do the ports lower the main engine thrust chamber pressure but the exhaust of gases through the ports provides a reverse thrust to the missile. This actually causes stage III to back away from the reentry vehicle and warhead which is the final stage of the Minuteman. By providing precise separation, this action prevents the usual decay of the rocket shutdown from affecting the programmed ballistic flight of the reentry vehicle.

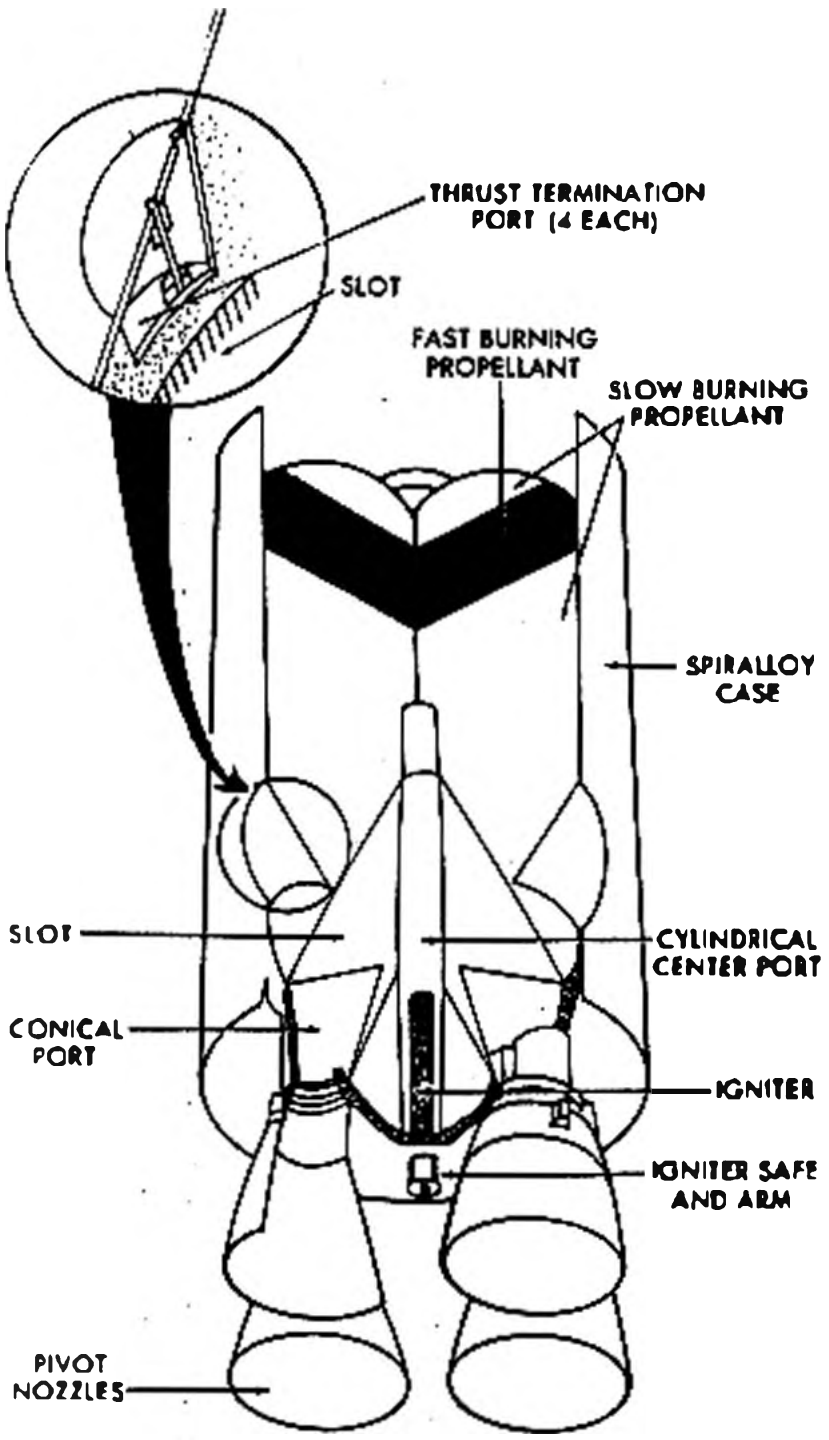


Figure 7-17. Minuteman Stage III Motor

THE AUTOPILOT

Now consider the second half of the control system, the autopilot. The autopilot consists of the displacement gyros, rate gyros, torque amplifiers, and a reference axes programmer.

Displacement Gyros

The displacement gyros are three mutually perpendicular, single degree-of-movement gyros mounted on the missile airframe. These gyros establish angular references along the three axes of the missile. Deviations from a prescribed flight path are measured as deviations from the established axes references. The missile may be regarded as a stable platform whose flight control system maintains it at an attitude in which the displacement gyros are at a null position.

Programmed deviations are effected by changing the angular references of the displacement gyros. The angular references for pitch and roll are changed by inputs from the programmer to a torquer which rotates the gyro gimbals, changing the reference angles. An optimum powered flight is determined for each missile. During this powered flight portion of the missile trajectory, changes in angular pitch reference are programmed. Roll reference is established by the azimuth location of the target. If the missile uses a stabilized guidance platform, this is aligned with the azimuth of the target prior to launch. The missile is put into a programmed roll maneuver during the initial seconds of flight so that the missile pitch plane is coincident with the stable platform.

Rate Gyros

The attitude signal from the displacement gyros in itself is not sufficient to provide complete stabilization of the missile. Rate gyros are added to provide damping to the system. The rate of deviation of the missile is provided by three rate gyros (yaw, pitch, and roll). Outputs from the rate gyros are fed to the amplifier where they are combined with the outputs of the displacement gyros to give the total angular displacement. The resulting signal is applied to the servoamplifier and then transmitted to the thrust vector control.

Amplifier Circuits

This section of the autopilot is simply a series of circuits which receive input signals and demodulate, amplify, filter, and mix them. The output of the amplifier provides lead compensation and power amplification necessary to operate servo valves in such a way that proper signals are sent to the thrust vector control.

Reference Axes Programmer

The reference axes programmer is a memory device. Its primary purpose is to furnish roll and pitch command signals during the initial part of the flight. Most programmers employ a timing motor which drives tapped potentiometers or cam and switch mechanisms. If integrating gyros are used, the commands are in the form of rate signals.

The following section presents electrical control systems. Although these systems have some disadvantages, they also eliminate some of the problems found in combination systems.

SECTION D
Electrical and Auxiliary Control Systems

This section completes the coverage of guided missile control systems. Most of the section is devoted to flight stabilization control. The last part of the section includes auxiliary control systems.

ELECTRIC CONTROL SYSTEMS

An electric control system consists of components all of which are powered electrically or electronically. There is no pneumatic or hydraulic power-transfer system required.

Except for the controller and actuator, the components used are similar to those used in the hydraulic-electric system. Heading reference is established by an electrically driven gyro and sensed by electric components such as a selsyn or reluctance pickoff device. Any computation performed on the error signal is done by an electrical or electronic computer. Rate signals are developed from electrically driven rate gyros or from an electronic rate circuit operating on the displacement gyro signal. An integral signal is obtained from a motor driven selsyn or from an electronic integrator circuit. Voltage amplification is provided by an electronic amplifier. Power amplification can be obtained by electronic means or by using a motor generator such as a dynamotor or amplydine.

The controller is a device that varies electrical power in proportion to an input signal and causes the control surfaces to move a proportionate amount and direction. In the case of an all-electric system, the type of controller used depends on the type of actuation it must control. More details pertaining to controllers are presented in later paragraphs.

Electric power can produce mechanical motion through magnetic force. This means that either a solenoid or motor may be used as an electric actuator. As solenoids do not produce enough force to move an airfoil, motors are normally used.

An electrical motor could be connected directly to a control surface, using the motor shaft as the pivot of the control surface. To apply enough torque to move an airfoil sufficiently, such a motor would be too large and heavy to be practical.

A small motor running at high speed has the same power potential as the large motor. A small motor may be connected to the control surface through a gear train. The mechanical advantage of a gear train results in a larger torque exerted on the control surface pivot. The motor may be either a *constant-speed* motor, operating through a clutch, or a *variable-speed* motor.

The required fast rotation of a small motor introduces a major disadvantage to an electrical system. Sudden deviations of the missile require rapid reaction of the control surface. The rotor of a small motor must run at a high speed to produce a rapid control response. A large amount of inertia exists when trying to reach a high rotor and gear-train speed from a standstill. This inertia opposes a change in velocity and causes an undesirable lag in control surface response. This lag results in insufficient sensitivity of the control system.

A larger motor could be used but it would still possess inertia due to its greater mass. A partial solution to this problem is the use of a continuously operating motor and clutch assembly. Response can be improved by the use of small high-torque motors.

Figure 7-18 is an illustration of a variable-speed motor used to actuate a control surface. The motor rotates in either direction, depending on the phase of the input signal. The speed of the motor is proportional to the strength of the input signal. The motor is coupled to the control surface through a reduction gear train and the movement of the control surface is proportional to the speed of the motor.

The controller for this motor must have a high output power capability to provide enough power to drive the motor. The variable-speed motor can be driven directly by the output of an electronic power amplifier.

The controller shown as part of the system in figure 7-19 converts the power of the controller drive motor to 3-channel variable-speed motors. If the pitch amplifier develops an output, the magnetic field of the pitch generator increases. The voltage output of the generator feeds power to the

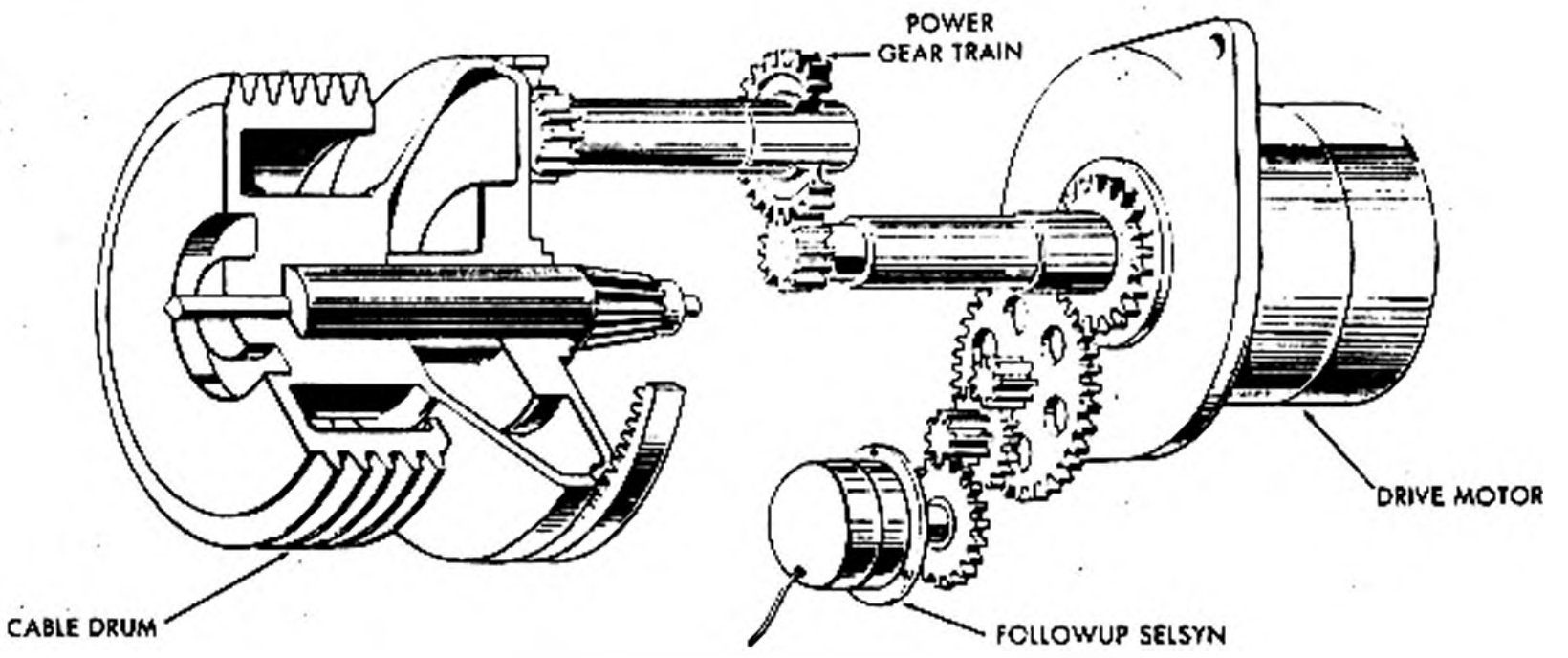


Figure 7-18. Electric Actuator (variable speed)

variable-speed motor. The controller drive motor must be capable of maintaining a constant speed regardless of load. If the pitch output decreases, the speed of the motor and, consequently, the output of the roll and yaw generators are affected. This speed variation would cause undesirable cross-coupling between channels.

The effects of inertia when starting and stopping a variable-speed motor can be eliminated by using a drive motor which runs continuously at a constant speed. This type of motor is connected to the control surface through a clutch assembly which acts as the controller. The clutch controls the power transmission from the motor to the control surface. Using two clutches and a gear train allows control actuation in both directions.

A control system using a constant-speed motor and clutch assembly is illustrated in figure 7-20. The illustration also shows a cutaway drawing of a clutch assembly. The clutch disc is operated by

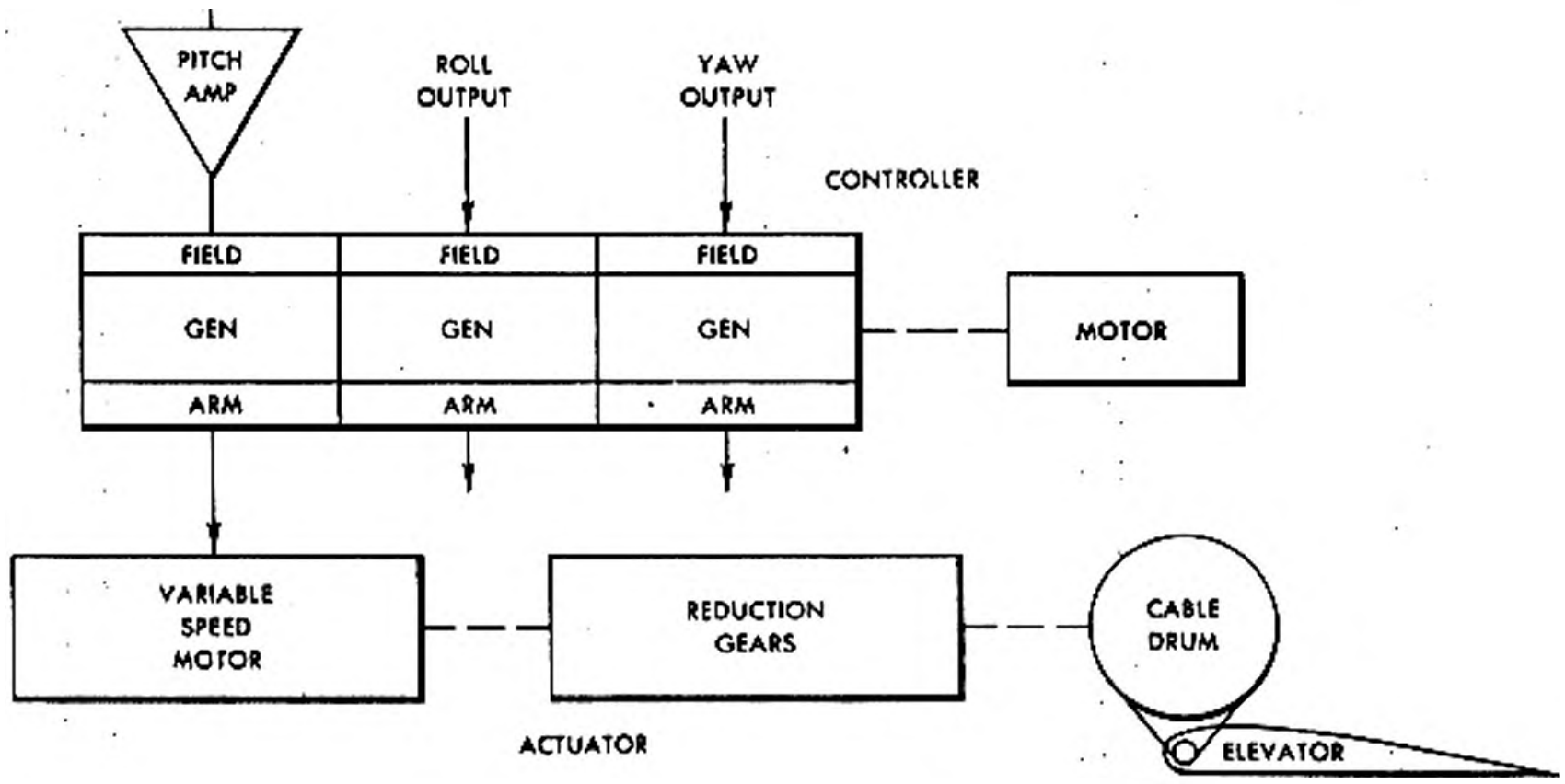
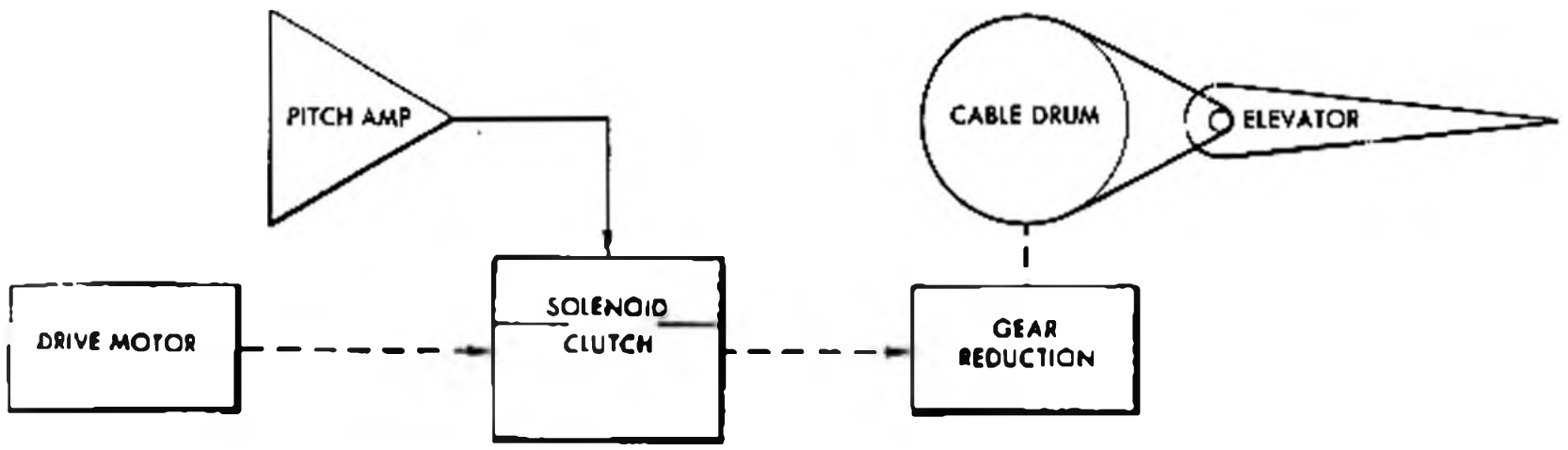


Figure 7-19. Part of Pitch Electrical System

a solenoid receiving power from the channel power amplifier.

An all electric control system is most easily adaptable to cruise type guided missiles that use conventional type control surfaces. Ballistic missiles travel at such high speeds and altitudes that conventional controls are ineffective. The electric-hydraulic system is normally used in most of these missiles.

The next paragraphs present control systems which do not have flight stabilization as a primary function but which are still of equal importance to the overall mission of the missile.



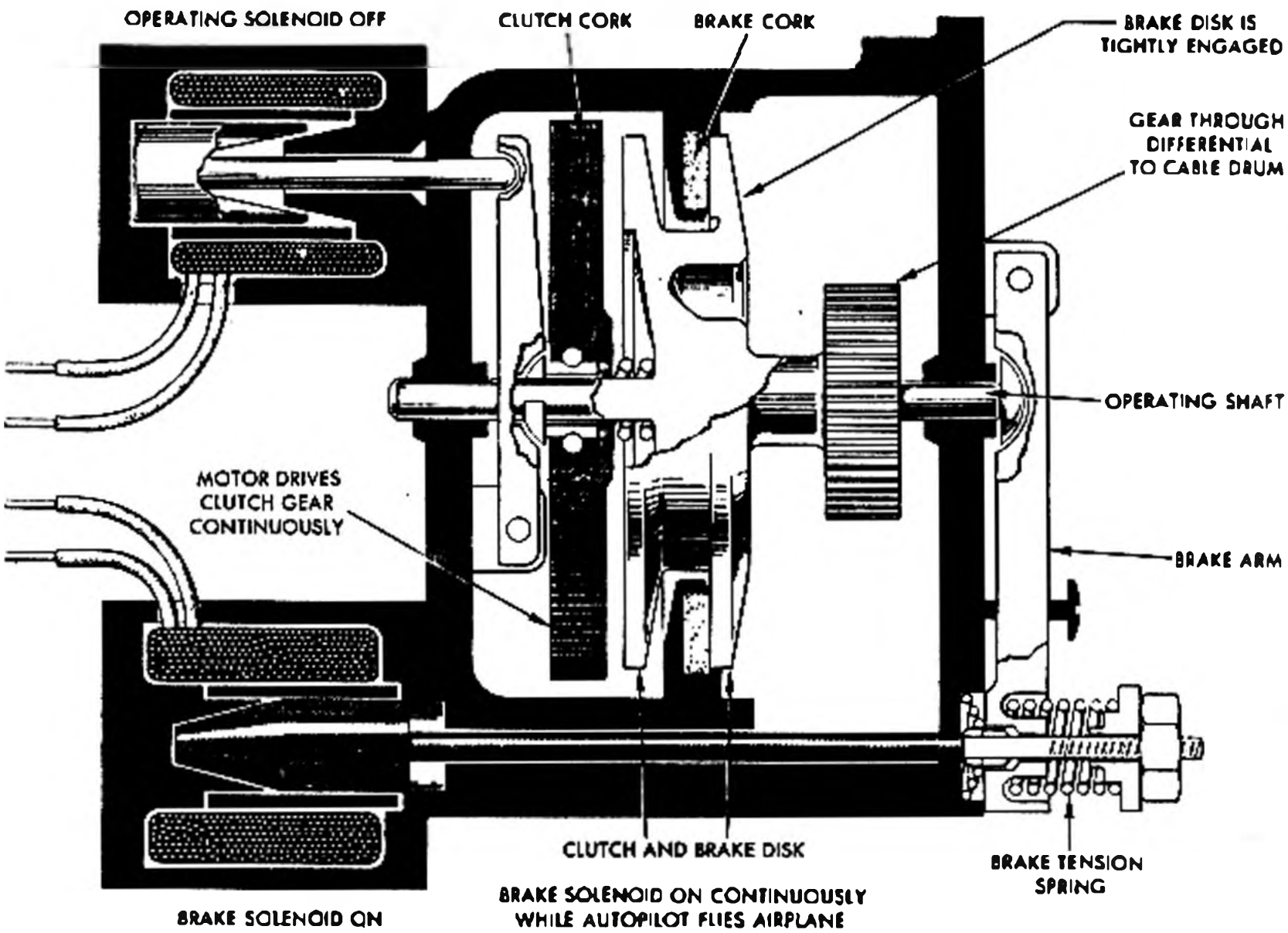


Figure 7-20. Electrical System using Clutches and Motor-Clutch Control of Autopilot

AUXILIARY CONTROL SYSTEMS

There are control functions other than flight stabilization that must be performed during a missile flight. These functions are performed by separate systems that may or may not be inter-connected with the flight stabilization control systems. The functions of these auxiliary systems are just as important as flight stabilization. Although the output power requirements are normally not as great, the accuracy requirements may be even greater.

Some guidance systems require that data be sent or received by the missile to a certain point by means of a very narrow radio beam. The continuous aiming of the missile antenna requires an accurate auxiliary control system. Any small antenna error caused by missile deviation could interrupt the radio link. Another antenna requiring an auxiliary control system is a search radar antenna used with some guidance systems. It also must be aimed accurately. Gyro slaving systems can also be considered as auxiliary control systems.

The operation of many of these auxiliary systems is similar to the operation of stabilization systems. The auxiliary system, for example, which maintains the stability of a stabilized reference platform requires a reference unit and sensor unit to produce an error signal. The error signal indicates a comparison between existing conditions. The error signal is then changed in form and amplified to control an actuator. The actuator produces the correction in the stable position. This type of stabilized platform is used to keep the guidance system sensing components such as telescopes or accelerometers in a steady angular position in space.

The auxiliary system must be extremely sensitive and accurate to keep the movement of the stable platform to a minimum and to maintain it at a definite angular position. The platform cannot tilt with deviations of the missile airframe. This would cause errors in the guidance sensing system and could even make it inoperative.

Some auxiliary control systems do not operate on the same principle as stabilization systems. Stabilization systems operate on the servomechanism principle, which is a *closed-loop system*. The closed-loop system senses the results produced by its output and continually makes corrections. Many functions to be performed by auxiliary control systems are simple, one-time operations and do not require further corrections. These functions can be performed by open-loop systems.

An *open-loop system* is usually used to change some situation on the missile after it is launched. Examples would be the jettisoning of boosters and jettisoning of skirting prior to ignition of the next stage engine. Some cruise type missiles have fairings on doors that close after booster rockets have been released. Figure 7-21 shows a control system used for this function. The difference between this system and previous systems is the absence of a followup connection between the device being controlled and the input of the system. The action starts at the timer and ends at the door, with no feedback path to complete the system loop.

Supporting systems requiring high power and fast response, like the flight stabilization system, normally use hydraulic power. Systems not having

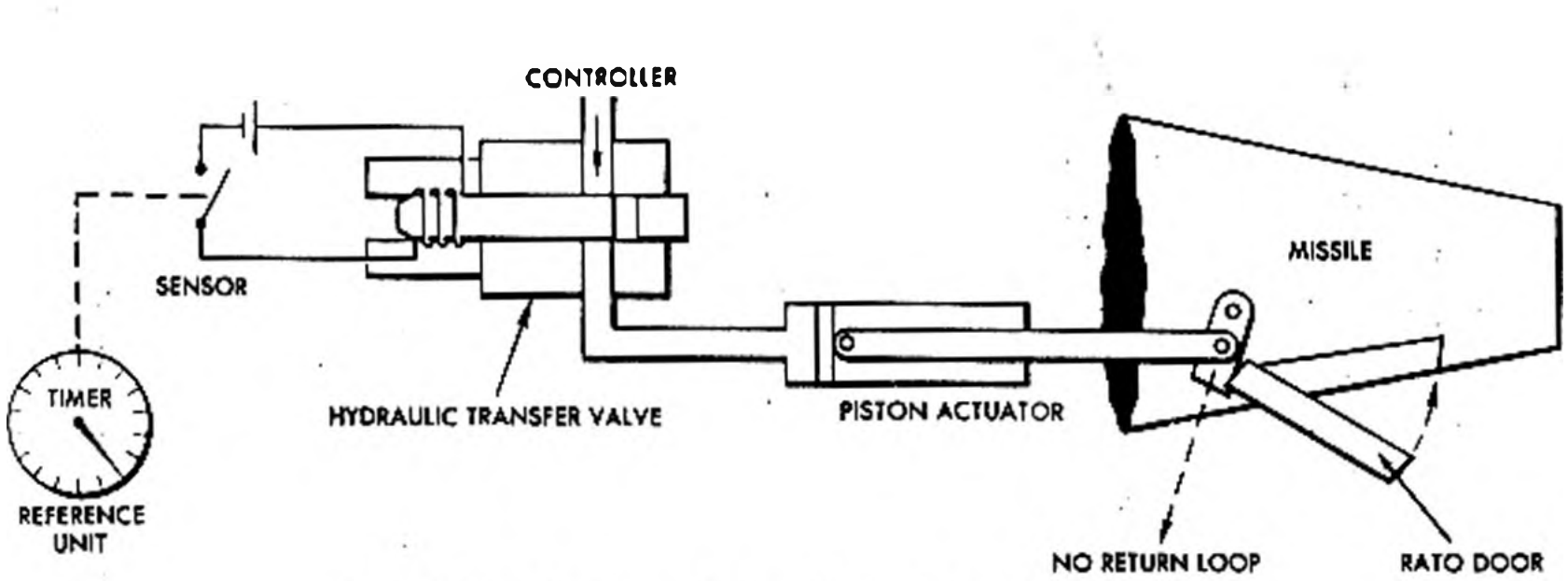


Figure 7-21. Open-Loop System: Hydraulic RATO Door Closing

high performance requirements may use small electric motors. Some systems such as booster jettisoning systems may use explosive bolts or squibs.

These auxiliary systems are usually checked out by the same people responsible for checkout of the flight control or stabilization system unless the auxiliary system is related to the guidance section. Checkout procedures which are similar to flight stabilization checkout procedures are presented in the final section of this chapter.

SECTION E

General Procedures for Checkout of Control Systems

Contained within a guided missile are many electronic and mechanical components that must function correctly at the proper time during the flight of the missile. If a part fails, there is no one aboard to compensate for or to repair the part, and the failure can result in loss of the missile or at least cause the missile to miss its target. To reduce the possibility of failure during flight, the missile is checked at scheduled intervals.

The final prediction of reliability depends on a preflight checkout. These checks are designed to be as reliable and complete as possible. To be reliable, the checks must be made in the correct sequence. Checks made incorrectly or not in the proper sequence result in inconclusive results and possible damage. Technical orders provide the mechanic with the directions for performing checkouts properly.

An important step in any checkout procedure is an accurate record of results. These records consist of checklists that are annotated as each step of the checkout is performed, as well as records showing what corrective action has been taken to repair a malfunction. Such records actually have many uses. Consumption of spare parts is tabulated to establish stock levels for supply. Trends of repeated malfunctions can be calculated to establish replacement or inspection cycles. These records are sometimes used to establish product improvement studies to develop better, more reliable components and better, more efficient maintenance techniques.

For large missiles, the checkout job is divided into major sections such as power system, guidance system, control system, and airframe and propulsion system. Different teams of technicians and mechanics are assigned to perform checks within their job specialties.

The development of extremely reliable equipment has resulted in the increase of time periods between system checks. In the case of some ballistic missiles such as the LGM-30F, scheduled periodic maintenance has been all but eliminated. The missile is assembled and checked out in the missile support area. It is then transported to the launch site and installed in an underground launch silo. After the missile has been connected into the launch facility electrically, the condition of the missile systems are monitored by the launch crew. The launch crew is located in the launch control facility which is a minimum of three miles from any of the ten launch sites which it controls.

The LGM-30 is designed to be held on alert status for an extended period of time. During this time, maintenance is performed only if a system malfunction in one of the missiles is indicated by the monitoring equipment at the launch control facility. In the event of a malfunction indication, the specific missile is put on a standby status and a maintenance crew is dispatched to the launch site to check out and to repair or replace the faulty component. The missile is not removed from the launch tube unless the extent of the malfunction is beyond the possibility of on-site repair.

Military tactics and strategy demand that preparation for launch of a missile be rapid. Speed and certainty are the important objectives in preflight checkout of a missile. The time between the decision to fire and the actual launching must be an absolute minimum. For this reason preflight checks are designed to be as short and simple as possible. A maximum amount of the work, such as component and systems testing, is done in the assembly area. Most large missiles are "counted down" to a hold condition that is just short of actual launch (usually not exceeding 2 minutes). Under these conditions all the various systems, with the obvious exception of the propulsion and arming systems, are operating and being monitored. On command to fire, a minimum number of final checks are made and the engines ignite, launching the missile in a very short period of time.

In the case of the LGM-30 Minuteman missile there are no "hold" conditions in operational countdown. When launch is commanded by the crew, the airborne computer counts all systems down to stage I ignition, and missile away.

The possibilities for human error in the checkout of these sophisticated weapon systems have been greatly minimized by the use of automatic computer-operated checkout equipment. By the use of prepunched cards, the equipment will sequence the complete checkout of the missile system. In the event of a malfunction, some of these computers will even print out the correct repair procedure. Of course, a technician must make the actual repair. This requires a man who *knows what he is doing*.

CLASSIFICATION OF TESTS

With the wide variety of missiles in existence, many different tests and procedures exist. This makes it difficult to analyze checkout procedures from a general standpoint. The following is a listing of one possible breakdown of all control systems checks on a representative missile.

1. Physical inspection.
2. Checks on the power supply system.
3. Electrical point-to-point checks.
4. Zeroing and balancing adjustments.
5. Servo static error simulation.
6. Servo dynamic response checks.

These checks may be made through the use of common test equipment or they may be performed automatically by specialized computer programmed equipment.

Although these categories are listed in a logical order, the checks are not necessarily performed in that order nor are all checks under any one category performed at once.

Equipment included under the *power supply system* are those components providing power to the control system and the actuators. In most cases, this includes electrical and hydraulic power.

CHECKING THE POWER SUPPLY SYSTEM

Tests on the power supplies are normally performed before checks on the rest of the systems. This is logical since proper operation of all systems depends on amplitude and frequency of electrical power and adequate hydraulic pressure.

After making certain that the power supplies are functioning correctly, check the control systems components.

Electric Actuation Checkout

The folowing general procedures are accomplished:

1. *Battery check.* Check the battery for full charge. Check voltage regulation by measuring the terminal voltage under maximum load conditions. Check the power wiring by measuring voltage at the load. High-resistance wire or connections create an abnormal voltage drop when current is drawn.

Some missiles use batteries that are not activated until launch. These batteries have a separation between the battery and its electrolite. The electrolite may be introduced into the battery by gas or air pressure or by use of a squib. In any case, the potential of the battery cannot be checked during normal preflight checks. Special procedures are used for checkout of such a power system and these procedures are spelled out in specific system technical orders and checklists.

2. *Generator and voltage regulator test.* The generator is normally driven by the propulsion system. On a missile using a conventional jet engine, the generator checks can be made during engine tests. A gas-driven generator cannot be fired to perform output checks. They require specialized test procedures.

3. *Controller check.* Checks on an electrical controller are made for electrical balance and power gain using specific input signals.

4. *Actuator check.* The major requirement of an actuator is that it move fast enough and in the correct direction with a given input signal.

Hydraulic Actuation Checkout

The following are general procedures for checkout of the hydraulic system:

1. *Filling and bleeding the hydraulic system.* Procedures for filling and bleeding the system includes precautions for cleanliness, use of proper type of fluid, and the proper torque to be used when tightening fittings. Bleeding is a process of removing all air bubbles from the system. This

procedure insures precise operation of the actuators. Any air in the system will cause a "soft" reaction in the cylinders like the reaction you may have experienced with hydraulic automobile brakes. Bleeding the hydraulic control system is usually accomplished with the same test unit.

2. *Pressure test.* After the system is filled, the test unit pressurizes the system to check that pressure is maintained within the prescribed limits.

3. *Leakage test.* The system is made to reach a pressure above its normal operating pressure and all fittings and components are checked for a period of time to ascertain that no leaks exist. The increase over normal pressure insures that the system has an operational safety margin.

4. *Operating test.* During the pressure and leakage tests, the hydraulic test unit actually cycles the system components to assure complete circulation of fluid, as well as to determine that all actuators function throughout their limits of travel without binding or erratic oscillations.

PHYSICAL INSPECTION

At some time during the control system checkout, the physical condition of the equipment should be checked. Actually, a physical inspection should be done before power is applied to a system. Following is a list of things which should be checked.

1. Frayed insulation and damaged wires.
2. Poor contact of plugs in jacks.
3. Secure suspension of electronic chassis.
4. Cables in correct locations and secured.
5. Breaks, cuts, or deterioration of rubber tubing.
6. Corrosion or dirt in equipment.
7. Leaks in metal or rubber tubing and fittings.

Some procedures include specialized checks for physical conditions such as:

1. Caged or uncaged condition of gyros.
2. Energized or deenergized condition of relays.
3. Position of switches.

ELECTRICAL POINT-TO-POINT CHECKS

Point-to-point electrical checks are continuity, voltage, and resistance measurements made between two points in a circuit. Some of the measurements that are made are supply voltages, plate and bias voltages, and amplifier input and output voltages. On most present day missiles, point-to-point checks are made only at the time of assembly of the missile stages or during maintenance procedures. After the missile is installed in the launch facility and counted down to a hold position, the continuity of electrical wiring is monitored continuously by the launch control equipment.

The old method of performing point-to-point check was with a multimeter. This method is now outmoded due to the complexity and sensitivity of missile systems.

Most missile systems now have specialized test consoles that are designed to perform continuity and voltage checks. These sophisticated testers are automated and are programmed to follow a sequence of tests. As long as the circuitry is sound and voltages within the proper tolerances, the checker continues to step from check to check. In the event of a short or open circuit or an out of tolerance reading, the checker stops and indicates a malfunction. In some of these testers a trouble-shooting capability is built in. The tester will perform a series of checks and by means of a light display or printed tape will indicate the malfunction correction procedures. As systems become more and more sophisticated, the chances of human error are minimized or eliminated.

A major problem exists with these test consoles. They are complicated in themselves and are susceptible to malfunctions. Many of the adjustments in the consoles are critical, and maladjustment leads to erroneous test results. Separate maintenance and check procedures for the consoles must be established, and highly specialized personnel must be trained to perform these checks. To carry the job of testing to still a higher level of efficiency, test units are designed to test the missile test console.

ZEROING AND BALANCING ADJUSTMENTS

An important part of preparing a missile for launch is the zeroing and balancing of control equipment. This must be done prior to the checks which determine the operation of a complete system. Most zeroing and balancing that is required is in the amplifiers and pickoffs. Many of

the amplifiers used in servo systems have double-ended outputs. These outputs must be balanced if the gain is to be equal for each section.

Pickoffs must be zeroed for a condition that correctly indicates error signals. Pickoffs in gyros are zeroed during the manufacturing process, so the missile technician has only the problem of aligning the gyro case with the missile airframe. There are also other pickoffs in the system. These include followup generators, airspeed transducers, and altitude detectors. These all must be adjusted to the proper reference position. Selsyns are adjusted to produce a null or zero output with the rotor at a desired reference position.

Closely related with the zeroing of pickoffs is checking the linearity of the pickoffs. Normally, the output of a pickoff is proportional to the displacement of the control device. If a resistance bridge is used as a pickoff, a linear pot must be used. There are many other miscellaneous adjustments which must be made on the missile such as timers and amplifier gain. Most adjustments that have been presented here are monitored on the test consoles and if some adjustment is not within tolerance it shows up as a no-go indication on the console.

STATIC SERVO RESPONSE CHECKS

After a manned aircraft has been preflighted, a test flight determines that all the systems are functioning properly. During an actual flight, the intelligence and ability of the pilot analyzes and corrects for any inaccuracies that may show up.

Missiles cannot go on a test flight. Missiles are designed for *one flight—one way*. There is no pilot in a missile to correct for any instability or deviation from its course. Dependability and integrity of mechanical and electrical equipment must be depended upon for stability under all flight conditions. Since the missile cannot be test flown, the next best thing is simulated flight. There are three possible ways to do this.

- If the missile is small such as an AIM-4 Falcon, it may be mounted on a special test stand. By positioning the missile at a certain angular attitude, the reference and sensor create an error signal, and the control system applies corrective action to the control surfaces. The surfaces move to a certain angle and remain in that position as long as the same error signal is applied to the missile. In a system using proportional control, the followup signal is exactly canceling the attitude error signal. With a test error signal, the control surface must deflect to within a given angle to assure proper flight control. If the angle is not within tolerance, adjustments must be made. The test signal can be reversed in phase to check control surface movement in both directions.
- A second method of simulated flight testing involves moving the stabilized platform a given amount and checking the effect on the control devices. This method is usually used with large missiles. The test is performed with a test set having the capability of slaving the stable platform and tilting it a known amount and direction. The output error signal from the gyro is then applied through the control system and causes the control devices to react in the amount and direction determined by the strength and the phase of the error signal.
- The third method of simulating missile attitude errors is by connecting an artificial error voltage into the sensor. The voltage must be of a prescribed amplitude and phase to be equivalent to the output of the sensor when the missile is deviating at a certain angle. The action of the control system can then be measured and the gain can be determined by the ratio of angular movement over voltage input. A disadvantage of this system is that the sensor and reference units themselves are not checked in the process.

The voltage that is injected into the system simulates an error signal *from* the sensor. It is usually an AC voltage with a phase and frequency relationship to the reference voltage of the control system.

Completion of checks by one of these three methods determines whether or not the control surfaces or jets move the proper distance for a given constant error signal.

Additional checks must be made on most control systems to determine if they will react correctly to guidance signals. The output of the guidance system is normally electrical; therefore, an electrical signal can be inserted into the control system at the proper point to simulate a guidance error. The guidance error indicates that the missile has deviated from its planned course. The

control surface movement must be within a certain range for a given guidance signal.

DYNAMIC SERVO RESPONSE CHECKS

The checks listed so far do not provide a complete check of a control system. Constant voltage errors were simulated and the results measured under static testing. Additional checks are needed to predict the stability of the missile in actual flight. Actually, during flight, attitude errors normally are not constant voltages but vary irregularly. The error signal is constantly changing in amplitude at different rates depending on wind, thrust, and missile inertia.

Most systems are not given dynamic checks in the field. Either the checks are performed at the factory or the static and component checks are relied on for prediction of stability. Component checks may include a dynamic response check on some part which has a great influence on the response of the entire system. Such components may include rate circuits, hydraulic transfer valves, and actuators.

The closer a checkout comes to simulating actual flight conditions, the more thorough and decisive are the test results. Dynamic response checks are used to test a missile under simulated operating conditions. Dynamic means moving or operating, while response means reaction or result. The purpose of performing dynamic response checks on a missile control system is to:

1. Determine the system reaction to known simulated error signals which vary at rates which could occur during flight.
2. Determine if the system reaction is within the prescribed tolerances to insure flight stability.

The fact that movement is attained with a constant signal does not insure that proper movement will be produced with an error signal that is changing in amplitude at a varying rate.

The difference in checking a control system with a constant error and checking one with a changing error is comparable to measuring the voltage of a generator or battery when they were not producing any current, which would be under a static condition, and checking the terminal voltage under a load condition, which would be a dynamic check.

Constant Versus Changing Error Signal

The results of a changing error signal are different from the results of a constant error signal. A constant error signal causes a given control surface deflection which is also constant. A changing error signal results in varying deflections of the control surface, which are affected by other factors within the system. These factors affect the response time of the control surfaces.

Control surface or jet movement lags the error signal due to mechanical and hydraulic reaction time. Any action requires time. The electrical, mechanical, and hydraulic reaction times present in control systems cause control movements to lag behind the error signal. Any bending or expansion and contraction as well as slack in connections causes mechanical lag in a system. Hydraulic lag is caused by expansion and compression of hydraulic lines and also by the time it takes fluid to flow through transfer valves to the actuators. This compressibility and time of flow of hydraulic fluid is very small, but for precise control it must be taken into consideration. In the electrical part of a system, any components which affect the phase of an AC signal voltage have some effect on the response time of the signal. All coupling and filter capacitors and inductors contribute to a time difference between the input and output.

These lags in a control system are objectionable because they affect the capability of the missile to correct rapidly and smoothly. A means of reducing the effect of this control lag is to use rate gyros or rate circuits, sometimes referred to as lead circuits. The output of a rate circuit is combined with the proportional signal. If the rate signal is great enough, it will overcome the lag in the system and sometimes will produce a lead rather than a lag.

Meaning of Lead

The term *lead* does not mean that a correction occurs before the error begins. This would not be possible. Suppose a changing error signal for making dynamic checks is simply a sudden application of or removal of a constant voltage. With a sudden change of input, the system reacts an instant later to produce a new output condition. If a *lead circuit* exists in the system, the lag of the reaction is minimized.

If a simulated error signal, with the amplitude carying as a sine wave, is applied to the system, the control unit oscillates to each side of the neutral position. By using this type of signal, a lead or lag can be measured. This type of changing signal is used most often for making dynamic tests. The frequency of the amplitude variation is low (between 1 and 15 cycles per second). To complete the dynamic tests, several lead or lag determinations are made for each channel. Each check is made at a different frequency to simulate different rates of change of error and these tests are usually referred to as *frequency response* checks.

The deflection of the control unit also represents a sine wave. Whether this deflection leads or lags, the signal depends upon the particular control system. The lead or lag is determined by comparing the *phase* of the control movement to the input signal after several cycles have been observed. The lead or lag must be within a certain range for a particular system and is specified in the system checkout procedures.

Signal Simulation

Most control circuits are designed for AC error signals of a certain frequency, usually 400 cycles. The input signal must be modulated at this frequency. A common method of simulating an AC signal which will increase and decrease with a sine wave function is by rotating an autosyn at a rate equal to the frequency at which it is desired to move the control unit. A constant AC voltage with a frequency equal to the frequency for which the control surface is designed is applied to the stator. As the rotor passes through a null position, the output drops to practically zero and changes phase. It then increases with the opposite phase. The average amplitude approaches a sine wave function as illustrated in figure 7-22. The speed of rotation is adjusted to provide the modulating frequency.

Another method of simulating a changing signal is comparable to that used in simulating static error signals. The gyros, or the entire missile if it is a small one, can be rotated back and forth about an axis at a given rate.

Analysis of Results

After the simulated error signals have been applied, the final job is to measure the results and determine if they are within the prescribed limits. The control unit deflection is best determined by making a connection to a pickoff unit at the actuator linkage.

The output voltage of the pickoff is compared in *phase* and *amplitude* to the input signal to determine the response of the system. These signal comparisons can be made by one or more of the following methods:

1. Visual observation on an oscilloscope.
2. Reading of meters or oscilloscopes connected to special phase and amplitude comparator circuits.
3. Analysis of paper tape record produced by pen-and-ink recording oscillograph.
4. Analysis of record produced by photo-oscillograph.

The most modern and commonest method is by the use of the automatic checkout equipment. This equipment includes one or more oscillograph recorders which reproduce the results of tests, including frequency response checks, which can be analyzed and compared with the technical data for the system.

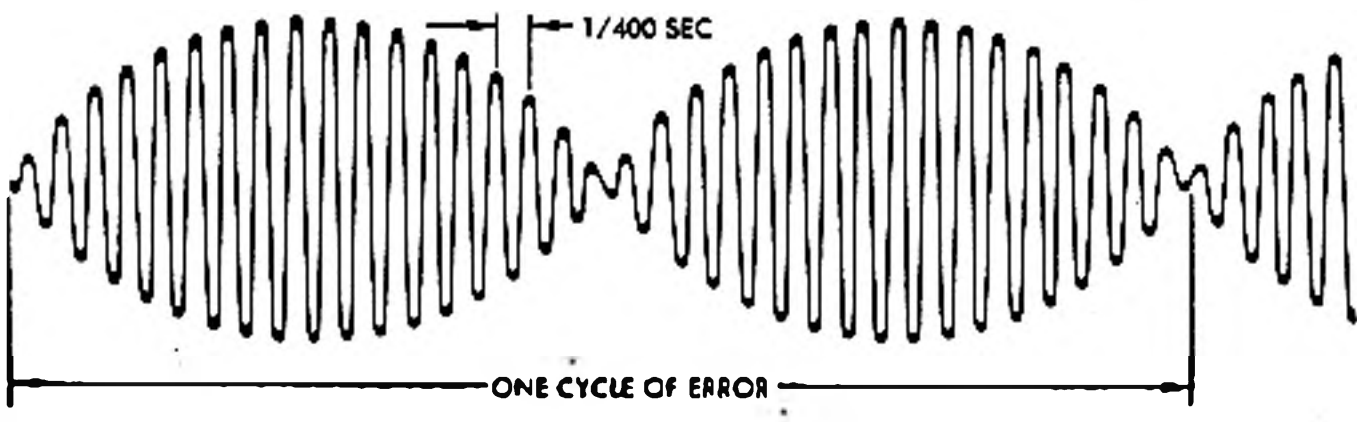


Figure 7-22. Simulated Error Signal for Dynamic Response Checks

MISSILE CONTROL SYSTEMS

This study of missile control systems has included the basic components. the control system as a whole, and general checkout procedures.

All guided missile control systems are similar in operation. Their purpose is to compensate for changing winds, equipment unbalances, and airframe dynamics so as to keep the missile stable in flight. The system is also capable of accepting location and heading errors from the guidance system and making corresponding heading changes. The system must be able to respond rapidly enough to sustain a smooth, straight flight. For this reason the hydraulic-electric system is the most commonly used system.

While control systems are very similar, guidance systems differ greatly depending upon the missile type and the mission requirements. Comprehensive presentation of various guidance principles are covered in other chapters of this manual.

CHAPTER 8

Trajectory Considerations

In this chapter, the path taken by a guided missile is considered. This path is commonly referred to as the trajectory of the missile. Missiles curve through space on a track which can be analyzed by advanced geometric methods. In the early stages of developing a missile, the possible trajectories that can be followed are calculated so that the airframe and control characteristics can be tailored to the optimum performance of the missile under the expected conditions of its operation. A study of missile trajectories helps your understanding of missile guidance requirements. For your use, it is not necessary to obtain more than an understanding of terms and basic principles—you need not study analytic geometry.

REFERENCE CONCEPTS

The concept of reference axes is basic to all missile systems. To detect the direction and magnitude of an error, there must be a standard reference point for up-down and right-left measurements so the control system can be commanded to give the proper corrections. This standard becomes more complicated as the complexity of the guidance system increases, especially for systems for long-range missiles.

A reference system can be based on an actual arrangement of some physical components, or it can be quite arbitrary. For arrangement of physical components, a homing-type missile such as the AIM-4 Falcon is an example. The reference for such a system is the physical and corresponding electrical orientation of the antenna. In a homing missile with an antenna that looks directly forward, the arrangement of the antenna is easily analyzed. Figure 8-1 illustrates the effect of attitude on missile reference. What is *up* to the antenna is *up* to the missile. Whether the missile is inverted or not, the command resulting from the error signal detected at the antenna directs the missile to the target.

For the initial planning of a flight, the primary reference of the trajectory coordinates must be exact. The selection of the coordinates for an inertial system is an arbitrary decision. Since the earth is pursuing its orbit about the sun and rotating about its own axis, the requirement for the reference to be stable or fixed in respect to the guidance system is not directly attainable. Because a gyroscope wheel maintains rigidity in space or, more scientifically, in inertial space, this inertial space becomes the reference. In respect to inertial space, the reference is a particular direction, usually with the axis of the gyro either parallel or perpendicular to the force of gravity at the time and place of launching the missile.

The remainder of this chapter presents the variations and combinations of coordinate systems required by missiles.

GUIDANCE COORDINATES

The simplest concept of guidance coordinates is that mentioned previously in which a seeker head is fixed to a missile so that its viewing axis is only to the front along the line of flight of the missile. The seeker is orientated so that signals from the detector to the control system initiate control movements that place the missile back on a path directed toward the target. When the missile rolls or changes attitude, the attitude of the

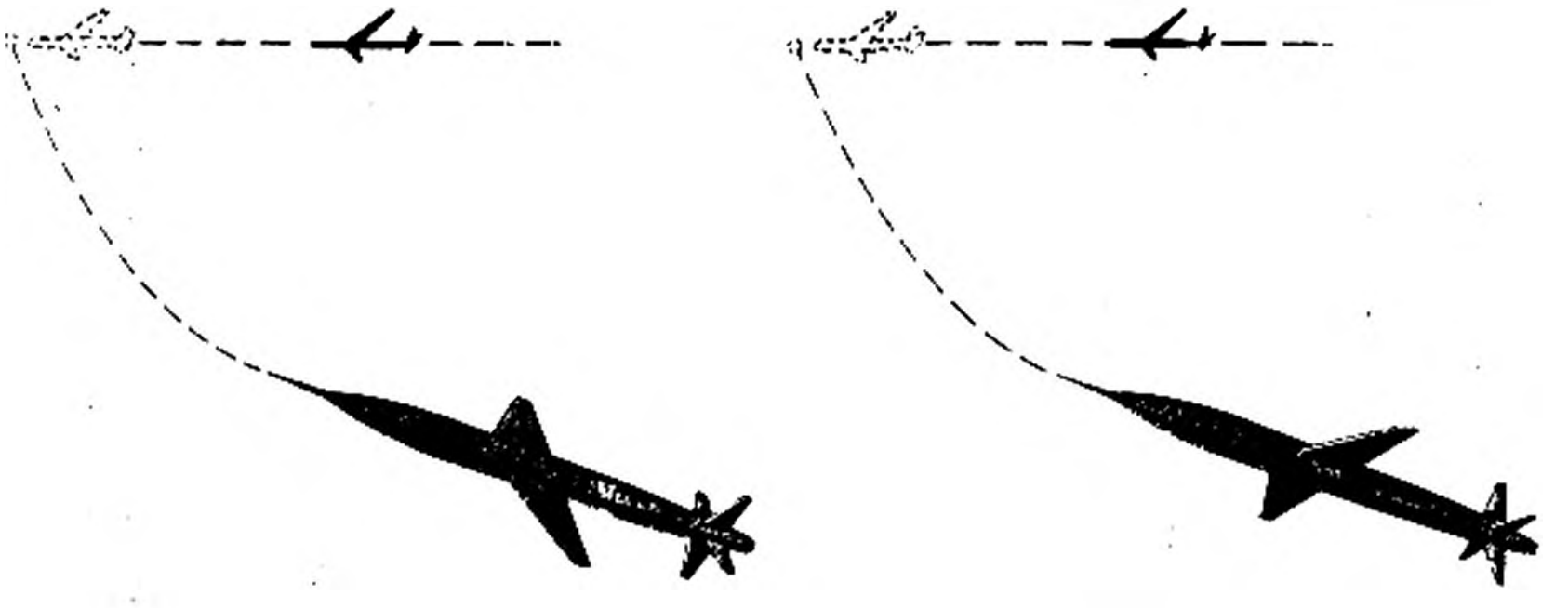


Figure 8-1. Identical Missiles in Different Attitudes making Proper Corrections for Targets above and to the side

guidance coordinates change accordingly. The guidance reference is fixed to the missile as shown in figure 8-2, and the guidance reference moves with the missile so there is no confusion as to the movement required to correct any error signal.

This is a two-dimensional reference system since only up-down and right-left signals are detected. The range dimension may or may not be a function of guidance. In some missiles, range is an ordnance function and is only necessary for warhead safe-and-arm and detonation. In such cases, range has no direct use for guidance.

In most applications of this reference system, the seeker head is mounted in the missile by a universal joint (gimballed) arrangement. With such an arrangement, the seeker can look at a moving target while the missile course is projected at a lead angle or collision course, as illustrated in figure 8-3. The course is plotted to converge on the target with a minimum of maneuvering by the

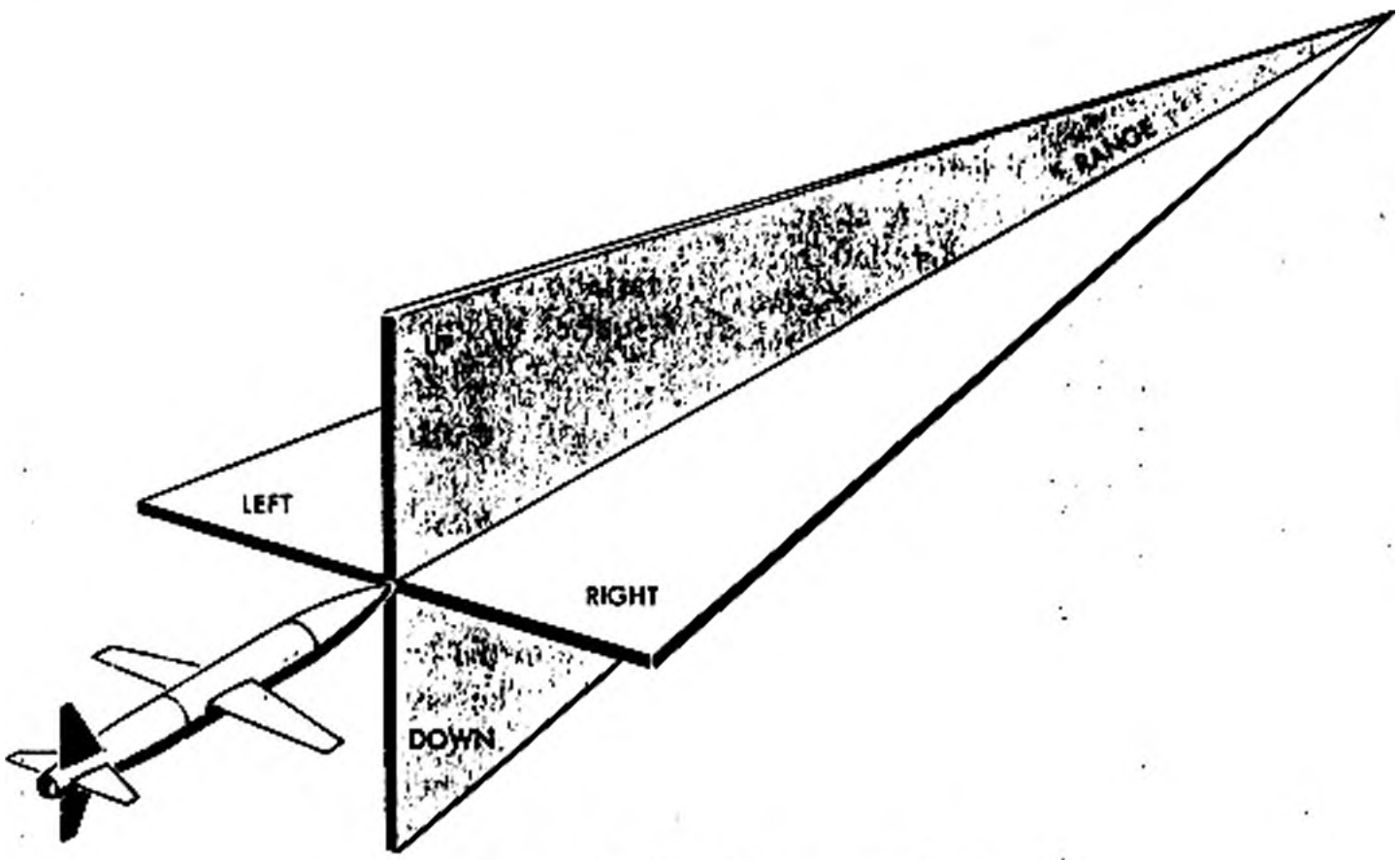


Figure 8-2. Simple Guidance Coordinates

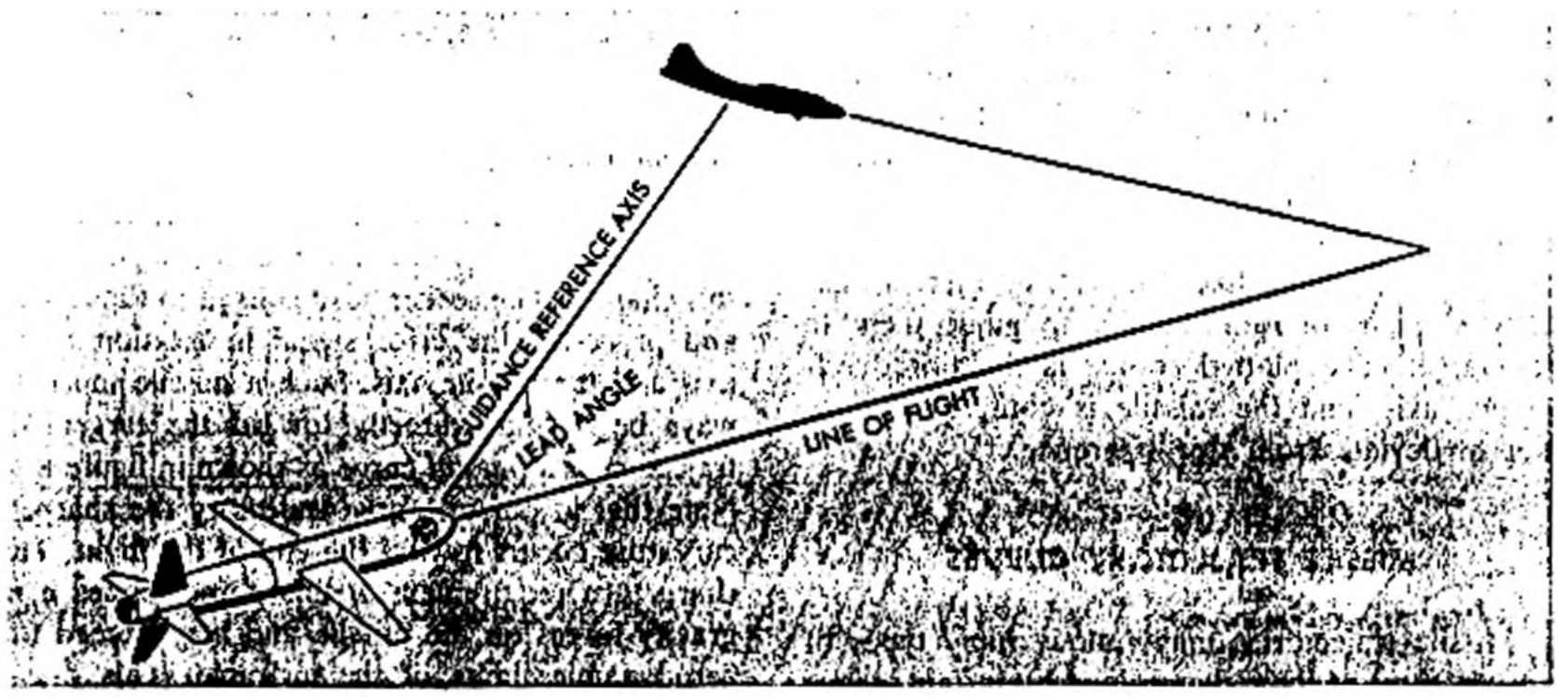


Figure 8-3. Variable Guidance References

missile. The missile "looks over its shoulder," and error signals are interpreted by the control system so that a correct collision course is maintained.

Geometrically, two intersecting lines determine a plane, which is a dimensional surface. If the lines were ideally straight, the problem would be two dimensional; however, the problem is complicated by the lead angle involved. The error signal must be modified according to trigonometric functions. A common way to obtain trigonometric operations on an electrical signal is through the use of selsyns. The reference is still fixed to the missile, but it is transformed or interpreted from the antenna signal. The antenna signal is in the guidance reference to the control system signal which actuates the missile controls.

TRANSFORMATION OF AXES

Most missile guidance systems perform the transformation of the reference system to the axes of the control system. In some systems there are several of these transformations. In the system referred to in the previous section, the reference axes are rotated through an angle to make them correspond to the control reference. Rotation of axes is a method of transformation.

An important point to remember is that when a transformation is performed on a system of axes or a coordinate system, the effect is on all three axes. Transformation of a reference system can involve both rotation and translation. Moving a set of axes for comparison with a standard is called *translation of the axes*. The performance of these operations is not often obvious from reference to schematic diagrams, but such functions are the reason for many complications in the missile electronics equipment. The missile must receive and interpret guidance errors in terms of its reference frame, and compute control movements in terms of the aerodynamics axes to give the proper corrections. These interpretation and computation steps require complex electronic circuitry.

Some missiles do not perform error signal transformation to develop the control command. Some guidance systems that do not require these complex operations are chosen for some missiles. This type of system detects the direction of the error and applies it to the control system directly after a correction has been made for the amount of deviation. Computation of the amount of error for use by the control system is not always a direct result of the concept of reference coordinates, but such a concept does enter into most computations. Almost all systems now make some attempt to distinguish the magnitude of error so as to develop a proportionate amount of control action.

Missiles controlled by external means require a further consideration. These missiles have an ideal trajectory plotted for them and are then commanded to stay on that trajectory. The flight

path is actually controlled by some means external to the missile. Remote radio control is an example of external control. The tracking information for remote control may be by visual, optical, or electronic methods. Both the beam-rider type of missile and the long-range electronic navigation system, like the Loran system, use charted paths of radio energy to guide them to the target. The plotted course is the prime reference axis, and the missile is controlled so as not to deviate from this reference.

MISSILE TRAJECTORY CURVES

Missile trajectories may follow many types of curves, most of them determined by instantaneous position of the missile relative to the target. An exactly predicted path is the hyperbolic course laid out by the Loran-type system.

Hyperbolic System

A hyperbolic system is shown in figure 8-4. Two radio transmitters (a *master* and a *slave*) are located some distance apart, and each transmits an RF pulse at specific time intervals. As the pulses are transmitted, a family of hyperbolic curves is generated.

A line joining the two transmitters is the base line of the hyperbolic curves. The curves are symmetrical about this base line, and at each point on the base line the sequence of signals arriving from the transmitters has a different time separation. At any particular point on the base line, a singular time separation between the signals is apparent. If this time separation is kept constant, movement away from the base line traces a hyperbolic curve, except at the exact center of the base line. Here a location of all points of the same time difference traces a straight line which is the perpendicular bisector of the base line. As each end of the base line is approached, the hyperbolas become more curved. The best course to follow is very nearly at the center of the system because it will be the closest to a straight course.

By adding a second slave station, a second group of hyperbolas is generated. This is represented by the darker lines on the diagram. The two systems are set up to intersect at definite points from which an exact position is plotted. Chapter 9 covers applications of this type of coordinate or reference systems.

Pursuit Curve

The principle of the fixed seeker head in the homing missile was introduced at the beginning of this chapter. The seeker head is fixed to the missile and processes the error signal in relation to its axis and the missile axis. Such a missile must always be heading directly toward the target. Its trajectory is a pursuit curve as shown in figure 8-5. Note that in pursuit curve trajectory the sharpest curvature occurs toward the end of the flight. The sharp turn requirement at this point applies high gravity forces on the missile and may exceed the aerodynamic limits of the airframe. This drawback of the pursuit curve homing system can be overcome by gimballing the seeker head in its mount and modifying the system for a lead-angle course.

Lead-Angle Course

By gimballing, the seeker head can be made to look in directions other than that in which the missile is flying. The trajectory flown by the missile is then on a collision or lead-angle course. Note in figure 8-6 that this course approximates a straight line, causing the missile to do a minimum of maneuvering. If the angle between the seeker axis and missile axis remains fixed, the missile will intercept the target. The fact that this reference angle stays constant when both the target and the missile are in motion compensates for any difference in speed between the two. The representation of this problem can be shown by vectors where the magnitude of the speed is taken into account. This vector development is illustrated in figure 8-7.

Beam-Rider Course

The beam-rider type of missile is confronted with the same problems as that of the homing missile. The beam-rider course is a result of the beam movements because the beam-rider stays within the radar beam controlling it. For the same reason that the pursuit curve attack was eliminated, the controlling of the missile by the tracking radar has been discontinued. A controlling radar is now used so that target position is predicted and the missile flies a collision course to the target.

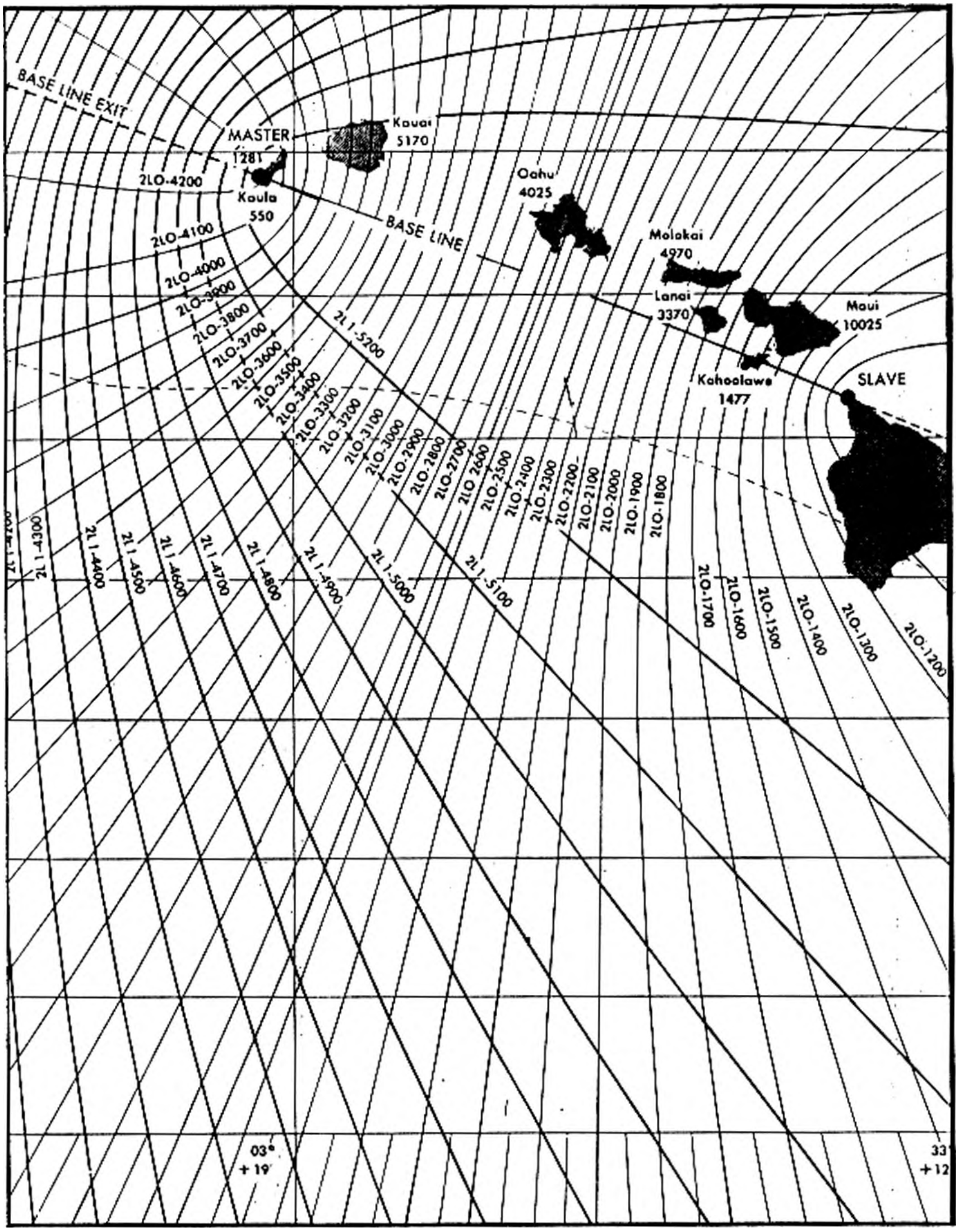


Figure 8-4. Hyperbolic Lines of Position Generated by Synchronized Stations

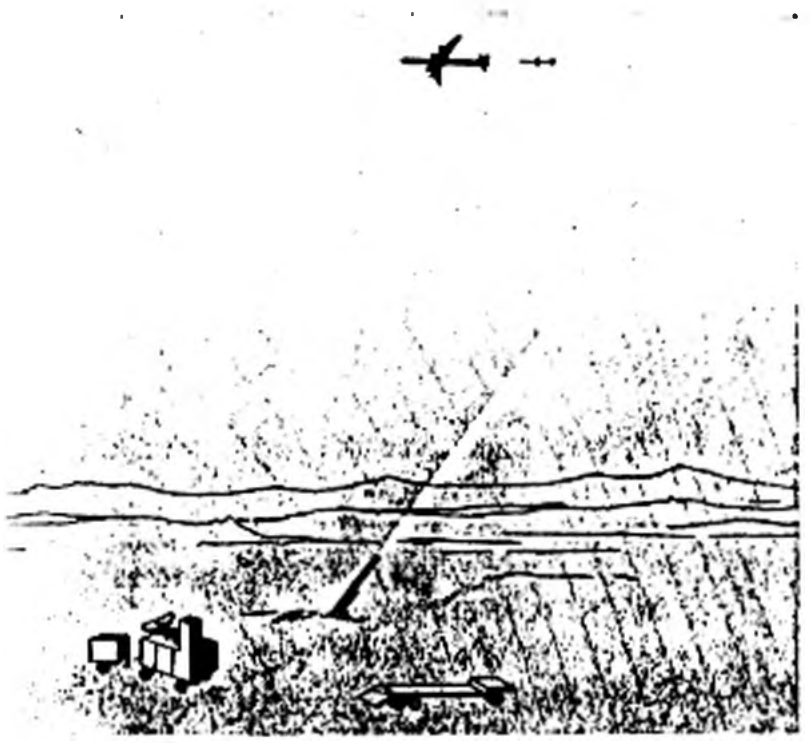


Figure 8-5. Pursuit Curve of Surface-to-Air Missile

Ballistic-Type Trajectory

A parabolic trajectory is the classic curve accepted for a surface-to-surface missile in free flight. If this same free flight is extended until the missile no longer returns to earth but becomes a satellite traveling about the earth, the trajectory becomes an ellipse. At some point the path of the missile undergoes a change of characteristic and the transition from the parabolic to the elliptical curve takes place. However, the missile has been flying under conditions caused by the same forces whether it returns to earth or remains in orbit. Except for air friction, the actual curve that the missile follows is the same in both cases.

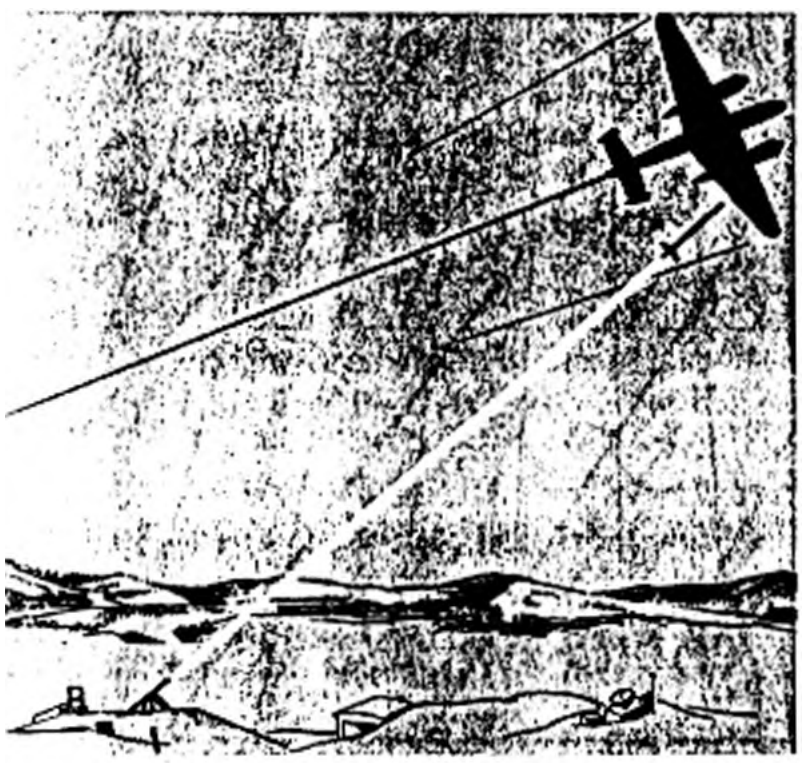


Figure 8-6. Lead-Angle Interception

In the theoretical consideration of the ballistic curve, an erroneous assumption led to the belief that the path was a parabola. This assumption was that effectively the earth was flat and gravity acted parallel to itself. The truth is that the earth is a globe and gravity acts toward its center. The conditions in analytic geometry of the equation parameters that define the curve of a parabola allow but a single shape for the parabola. The conditions for elliptical curves include an infinite number of shapes that are determined by their eccentricity. By using actual conditions of earth shape and gravity direction and analyzing a free-fall path, the equation parameters of a free-fall path are found to be elliptical rather than parabolic. The curve of any ballistic trajectory, therefore, is actually elliptical.

There are several trajectory possibilities for surface-to-surface missiles. The choice of trajectory is made before the missile is designed. Once it is built, a missile is limited to flying the trajectory for which it is designed.

The word "trajectory" is commonly associated with the ballistic trajectory of a rifle or artillery piece. The artillery-type trajectory usually has launching and attack angles of less than 45°. The trajectory curve is illustrated in figure 8-8. Examples of missiles using this type of trajectory are the Army Little John and Honest John.

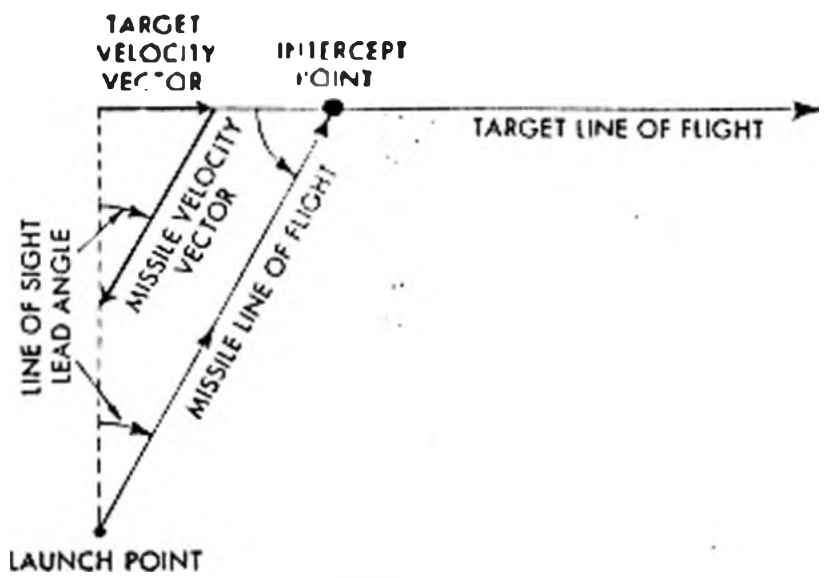


Figure 8-7. Vector Diagram of Lead-Angle Development

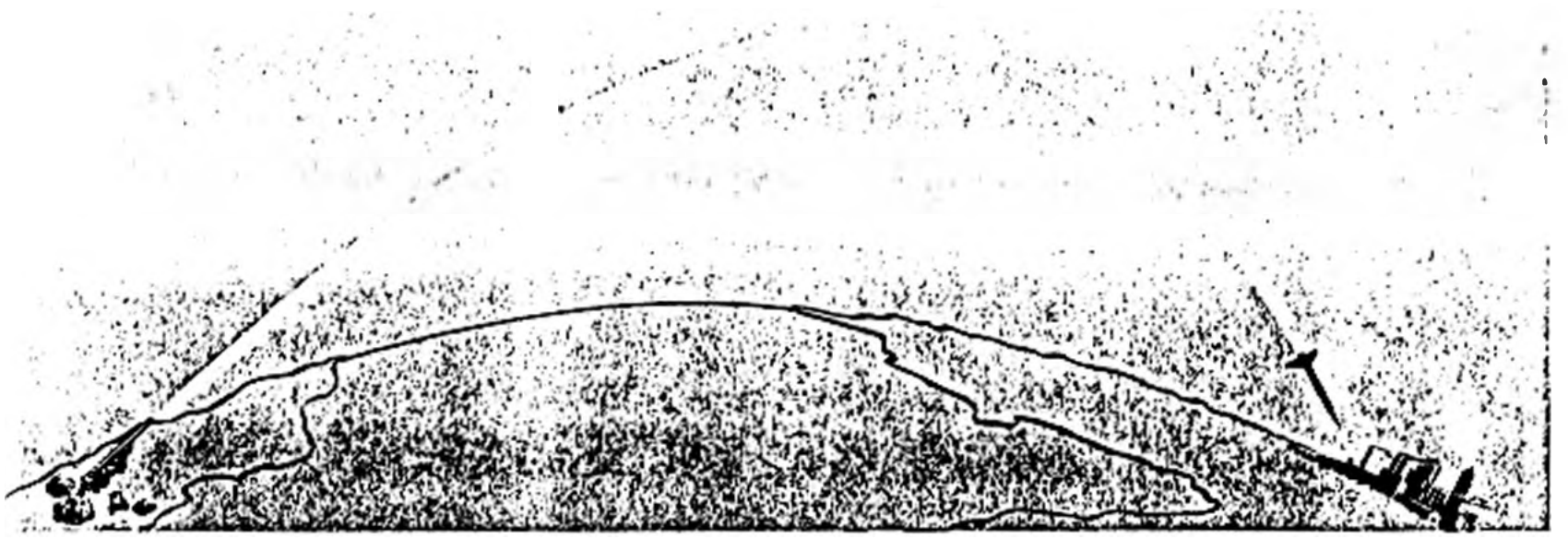


Figure 8-8. Ballistic (artillery-type) Trajectory

Flat Trajectory

A flat trajectory is employed by some missiles. These are usually the cruise type missile such as the Air Force Hound Dog and the Navy Talos. The mid-course altitude may be maintained constant or it may be varied for tactical reasons. The altitude chosen is at the level best suited to the propulsion system and the tactical situations. A representation of such a flight is shown in figure 8-9.

Rocket Trajectories

Rocket trajectories are similar to artillery trajectories. The major difference lies in the launch angle of the missile. Note in figure 8-8 the missile was launched at an angle to the horizontal. In figure 8-10, the missile is launched from a vertical position. This attitude is maintained until the missile reaches a predetermined altitude at which time it pitches over into a programmed turn. Propellant cutoff occurs when the correct velocity (magnitude and direction) is attained for a hit on the target. From this point, the missile is in free flight and on a ballistic trajectory.

The altitude reached by a missile of this type is very high. Ballistic missiles have a reentry vehicle that separates from the rest of the missile shortly after propellant cutoff. Only the reentry vehicle containing the warhead reenters the atmosphere. The missile propulsion system, having performed its function, will burn up in the atmosphere while the reentry vehicle continues on to the target. To prevent the reentry vehicle from burning up in the atmosphere, it is protected from the heat by some form of shielding which absorbs the heat.

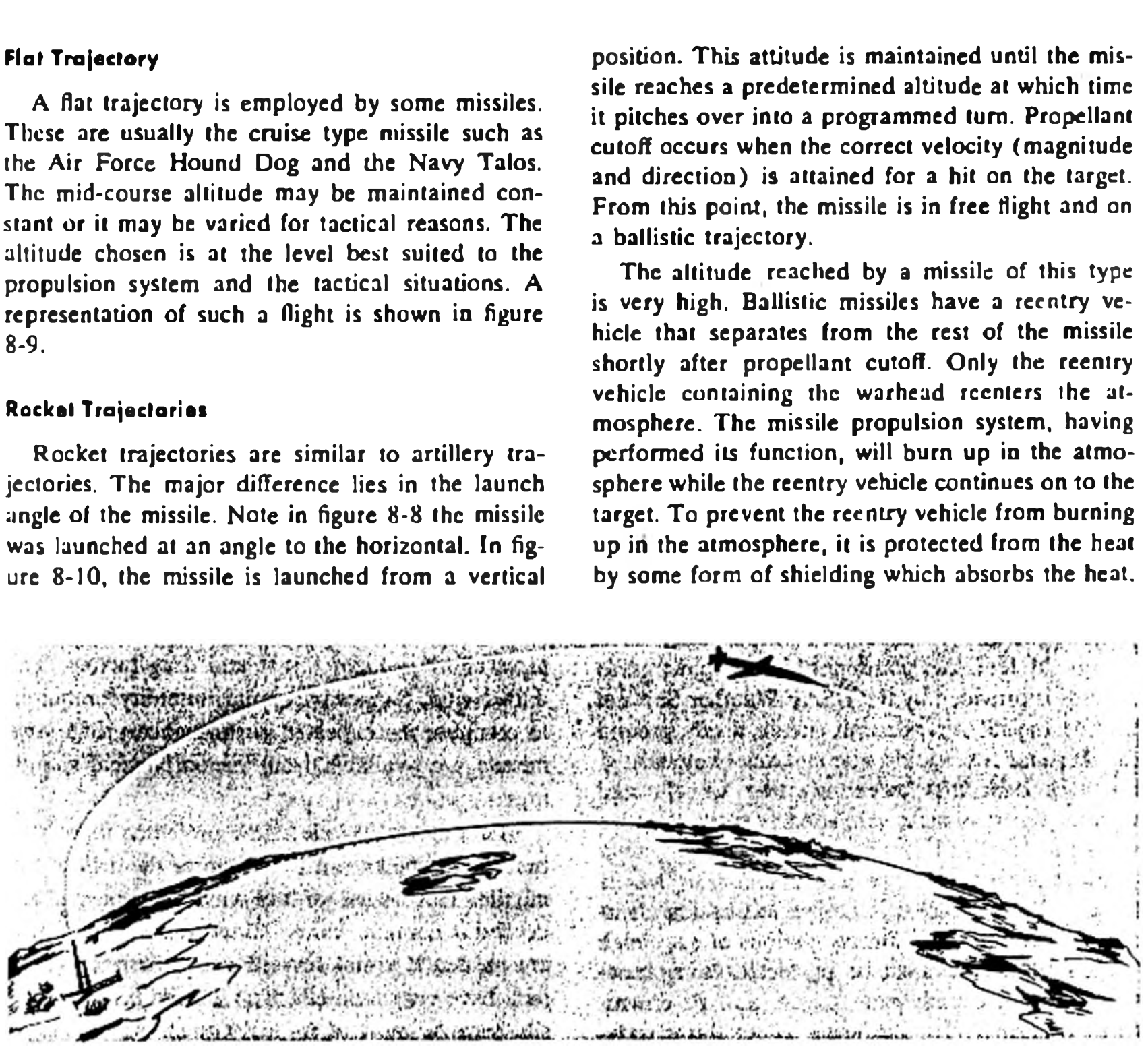


Figure 8-9. Constant Altitude Trajectory

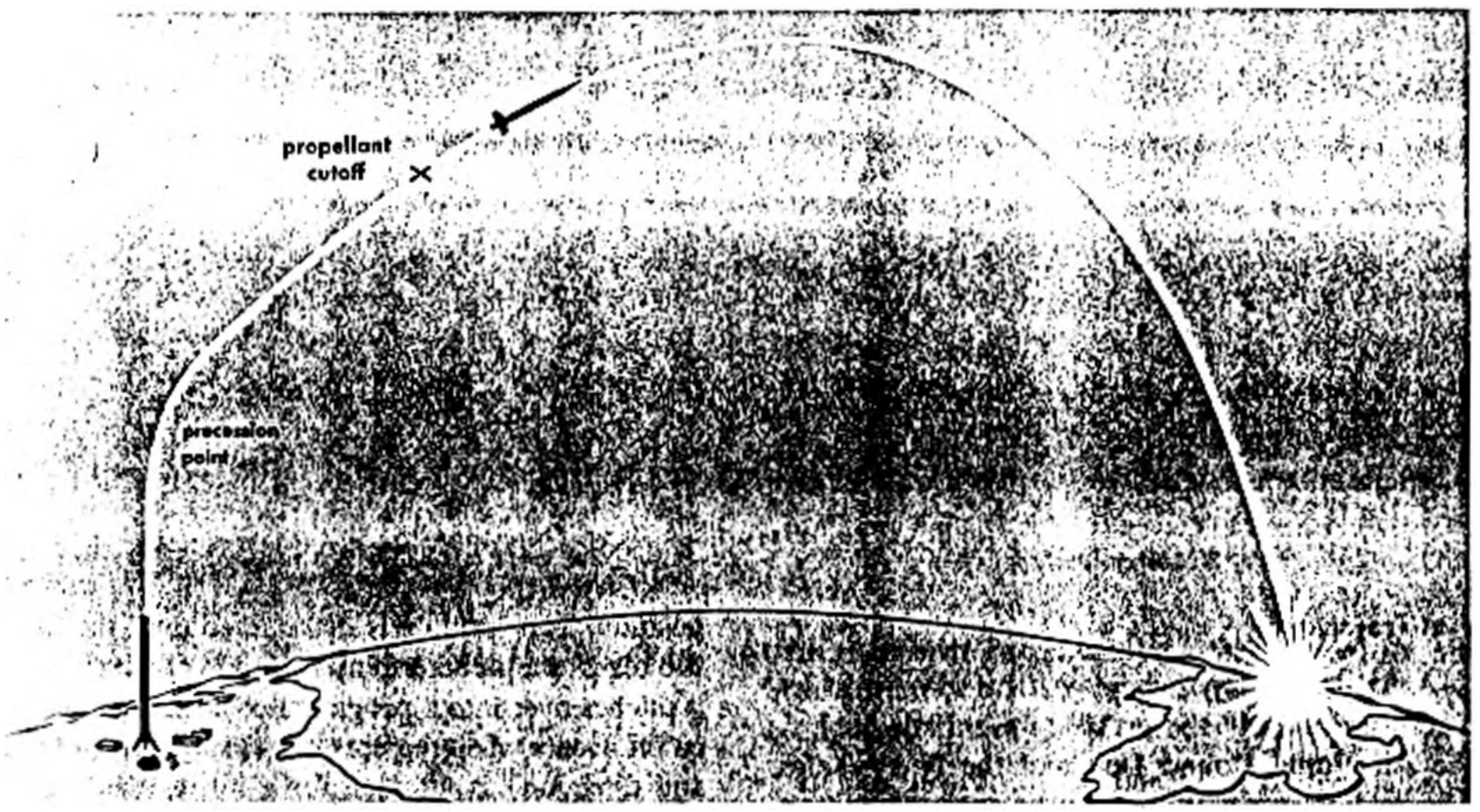


Figure 8-10. High-Angle Rocket Trajectory

NATURAL FACTORS INFLUENCING TRAJECTORIES

So far, no consideration has been given to factors influencing trajectory other than those affected by the missile itself. As missile flights are made in or through the atmosphere, some thought should be given to the effects on a trajectory by meteorological phenomena and terrestrial phenomena.

Influence of Meteorological Conditions

All missile systems are affected by weather elements under both operational and training conditions. Meteorological support for guided missile systems is provided by the USAF Weather Service. Meteorological observations made from ground weather stations, radiosonde stations, meteorological rocket facilities, ocean weather vessels, and weather reconnaissance are used in providing this support.

On the Ground. The larger and more fragile the missile, the more it is subject to physical damage by surface winds. During periods of very high winds, the missile must be protected from structural damage by adequate covers and tie-downs. Frost, freezing rain, and snow affect the aerodynamics of the missile. During periods of thunderstorm activities, lightning can cause damage to electrical circuitry and components.

In Flight. Immediately after launch, surface winds exert considerable influence on the performance of a missile. Some missiles have critical wind component factors above which the missile cannot be launched. Initial guidance may be affected by high surface winds. Cloud cover and visibility data are needed to select optimum conditions for photographic or television tracking. The humidity distribution in the atmosphere affects the accuracy of ground radar tracking. Air temperatures affect in-flight thrust and drag factors. The data (wind, temperature, and humidity) required to compute the expected performance profile of a missile are available from meteorological soundings.

Mid Flight. Winds at the projected flight altitude should be known to determine the drift of the missile. Head-wind or tail-wind components are needed to compute range. Cloud and visibility data are needed if visual surveillance is desired. These data have very little effect on a large ballistic missile because most of its flight is above the atmosphere.

IMPACT. Weather conditions at the impact area are important considerations if a missile is expected to be recovered as in some test operations. If impact data are to be evaluated optically, cloud and visibility information is necessary. For recovery operation in an ocean area, the conditions of the sea must be known.

Effects of Gravity on Missile Flight

A spirit level resting on a surface is horizontal when the bubble of the level is centered. Similarly, a plumb bob defines vertical in relation to the surface of the earth. Gravity is the force acting on these two items to produce the results described. In physics, gravity is defined as the force of attraction between the center of two masses. A little study will lead to a paradoxial situation in connection with this definition.

The earth is not a true sphere, but rather a figure known as an ellipsoid. This is like a sphere pushed in at the poles. As you study figure 8-11, you see that a vertical to a tangent at any point on the surface, except at the equator and the poles, does not pass through the exact center of mass. It appears that either the definition of gravity is wrong or that there is error in the method of measuring the direction of the force of gravity.

Actually, the definition of gravity given previously is not complete. The force that causes the plumb bob to align to the surface of the earth is known as *apparent gravity*. Apparent gravity is a resultant of true gravity plus centrifugal force of the rotation of the earth. The earth rotates about its axis and as a result centrifugal force acts in a direction perpendicular to this axis. The amount of centrifugal force applied to an object on the surfact of the earth is proportional to its distance from the axis. Thus, centrifugal force is greatest at the equator and decreases to zero over the poles.

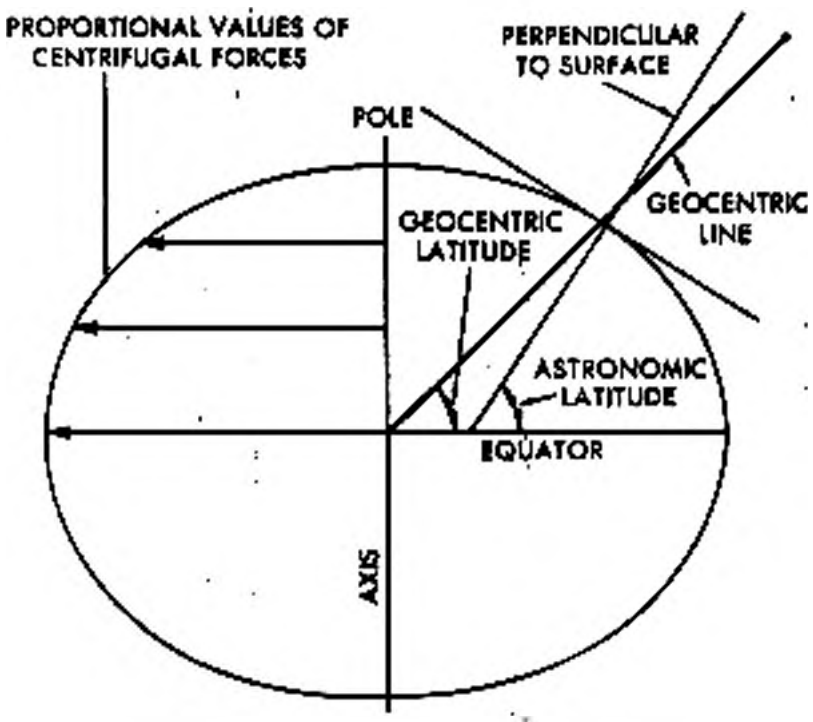


Figure 8-11. Exaggerated Elliptical Cross-Section of Earth

Apparent gravity acts perpendicular to the surface of the earth except in certain local areas where large mountains or particularly dense underground deposits deflect the apparent gravity.

Except for these local variations in gravity, the force of gravity is considered as acting perpendicular to the earth's surface on a body at rest on the surface. Any body in motion is contributing an additional force which alters the apparent gravity. Because centrifugal force contributes a substantial amount to apparent gravity, any motion of an object on or over the surface of the earth alters the effect of centrifugal force upon the object and thus alters the apparent gravity acting on that object.

Effect of Coriolis Force

A missile moving over the surface of the earth tends to turn toward the right in the Northern Hemisphere and toward the left in the Southern Hemisphere. This deflection toward the right or left is the effect of two motions—the rotation of the earth and the movement of the missile relative to the surface of the earth. This deflective force is known as *Coriolis acceleration*. It amounts to the difference between the speed of the missile relative to a non-rotating coordinate system and a coordinate system which rotates, such as the earth's surface.

Consider a missile flying toward the east in the Northern Hemisphere. The earth rotates in an easterly direction. Since the missile is moving in an easterly direction, it experiences the effect of an even greater centrifugal force. The increased force is straight out from the earth's axis and results in a motion to the right of the easterly movement. Study figure 8-12 in which P represents a missile moving into the page and around the earth's surface.

If P were relatively fixed over the earth's surface, the only force acting on it would be apparent

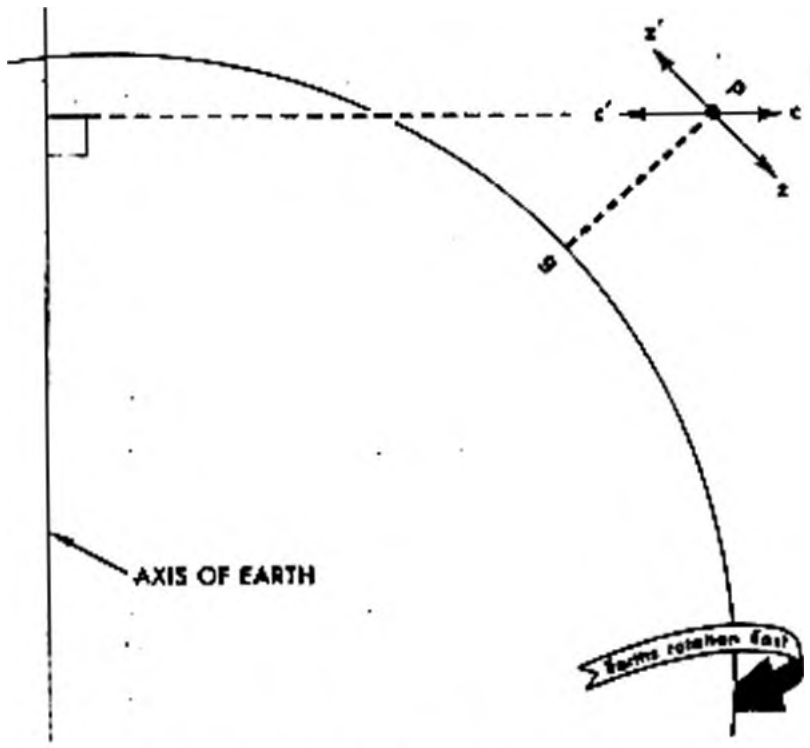


Figure 8-12. Action of Coriolis Force

gravity along the line Pg. As P moves eastward faster than the surface of the earth, greater centrifugal force is exerted on it because of its increased angular velocity about the axis of the earth. The centrifugal force acts perpendicular to the earth's surface through point P in the direction c away from the earth. The component of vector Pc that P experiences is in the horizontal direction toward the equator or to the right of the path of actual motion. This component is represented by Pz.

If the missile were moving to the west, the effect of centrifugal force would be decreased in the direction Pc'. The horizontal component would then be Pz', again to the right of the direction of travel.

Coriolis acceleration is tabulated for use by navigators in the Air Almanac. It is necessary to compensate for this Coriolis effect when programming a flight over the earth's surface.

Coriolis effect can also be recognized as the force responsible for deviation of the winds to the right in the Northern Hemisphere and to the left in the Southern Hemisphere.

Effect of the Earth Magnetic Field

The earth is a magnetic body with poles of unlike magnetism situated within the Arctic Circle and the Antarctic Circle. A magnetic field surrounds the earth, and this field influences the magnetic compass. The earth is not magnetized symmetrically with respect to the geographical poles. Generally speaking, a compass needle does not point in the direction of the geographical pole and this error varies from point to point. The angle of error is known as magnetic variation or declination.

MAGNETIC VARIATIONS. These are magnetic variations between places on the earth, as well as daily, monthly, and annual variations. These variations are minor compared to a progressive variation which has been consistent over the centuries. These variations have an effect on calculations involving exact and precise magnetic headings and coordinates.

DIP. A magnetic needle freely suspended in the earth's magnetic field has a vertical movement in aligning itself with the earth's total field. This vertical component of the earth's magnetism is known as *dip*. Dip does not occur at the magnetic equator which lies close to the geographical equator. Moving away from the magnetic equator, the dip increases until it becomes vertical at the magnetic pole.

INTENSITY. The direction in which the suspended magnetic needle is aligned indicates the line of force of the magnetic field. In this position the needle is parallel to the maximum magnetic intensity of the earth. The intensity is the number of lines of magnetic flux in a standard cross-section area. This intensity can be measured by using equipment operating on the same principle as the flux valve of the gyrosyn compass.

MAGNETIC STORMS. The earth's magnetic field is subject to fluctuations called *magnetic storms*. These fluctuations last from brief periods to several days. They result from sudden changes in electric currents which circulate within the earth and in the region surrounding the earth. The fluctuations occur at random and apparently have some correlation with sunspot activity. They may occur simultaneously over the whole earth or may be restricted to certain regions.

The range of the effect of magnetic storms upon a compass does not often exceed half a degree in the lower latitudes, but it is of great concern in the higher latitudes.

Local disturbances in magnetism occur in regions where mineral substances within the earth

possess magnetic properties. The amount of this effect is determined by actual survey.

PRACTICAL NAVIGATION

Surface-to-surface navigation of missiles presents more stringent accuracy requirements than the conventional air navigation as performed by the navigator in an aircraft. With precision radio navigation aids at the destination of a flight and the exact placement or location from pilotage, dead reckoning and celestial navigation need be accurate to within only a few miles. A missile does not have radio-beacon homing equipment, and it has no pilotage reference, so its unattended navigation system must have pinpoint accuracy.

Target Location on Maps

The location of the target is the primary problem. The target must be accurately plotted in respect to the launching point. To do this requires reference to a map or chart. Maps are developed for information as to the character of the earth's surface and location of man-made features.

In developing and drawing a map, a certain amount of accuracy must be sacrificed. Some smaller features must be portrayed larger than scale if they are even to show in a scale reduction of an area. For instance, a line representing a road may be almost a mile wide according to the scale. The earth's surface is curved, and to lie flat, a map must distort to some extent the area it represents. The method by which the surface is "stretched" to make a flat map is referred to as *projection*. Another basic difficulty is the inaccuracy of the original information from which a map is developed. This is not apparent to those who are acquainted only with maps of the United States and America, but there are areas in the world where unknown errors actually exist.

Radio Propagation

There are some difficulties in exact placement of a point for those guidance systems using radio waves. The propagation of radio waves is affected by the conditions of the earth over which they travel and the condition of the medium through which they travel. Air has an effect on radio waves because of several factors. Figure 8-13 illustrates the ionospheric skip effect on short waves.

Magnetic storms cause considerable difficulty with radio reception. As the frequency of the radio

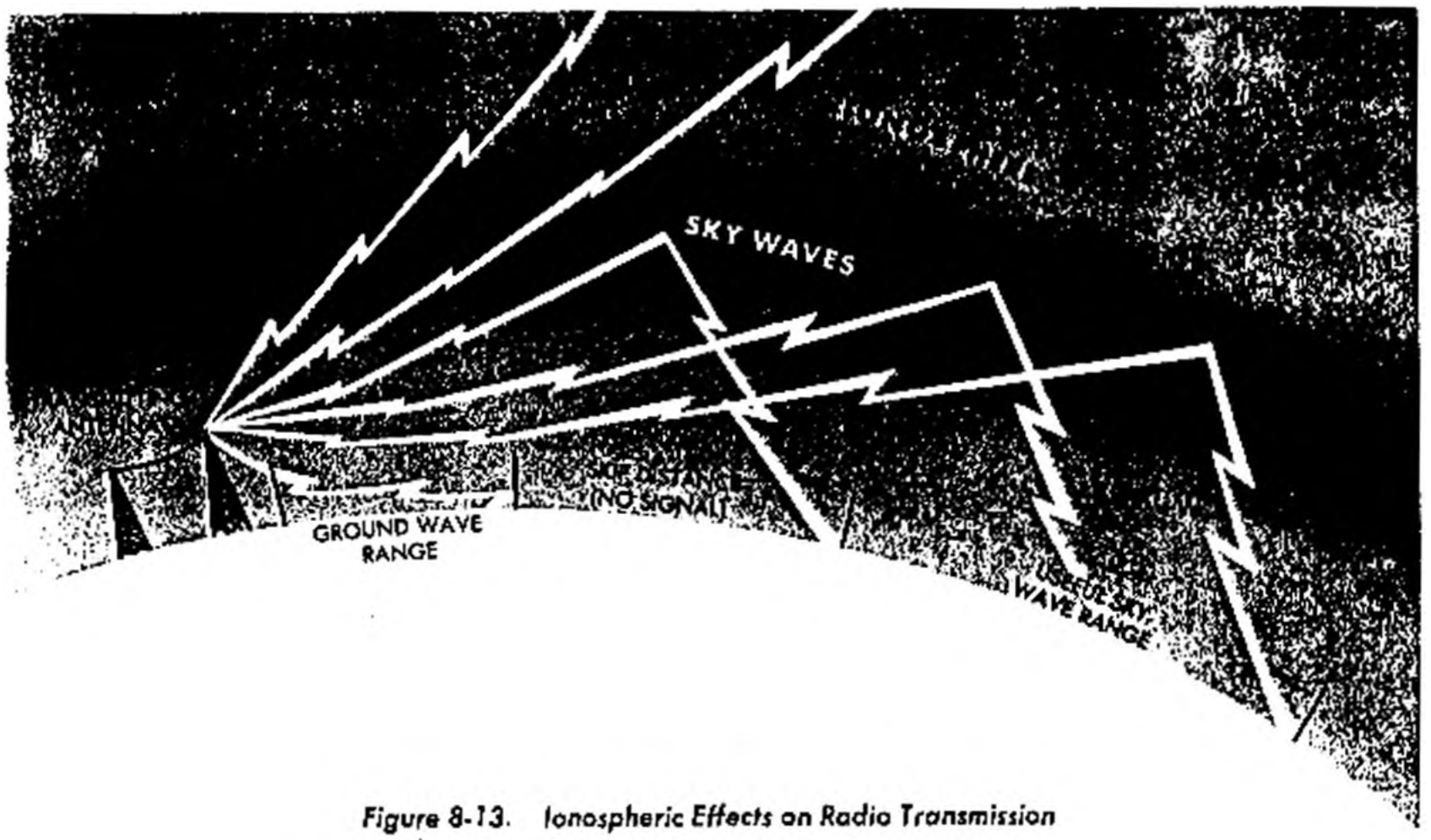


Figure 8-13. Ionospheric Effects on Radio Transmission

waves is increased to VHF and above, even the density of the air has an effect on transmissions.

You will recall that in the explanation of weather, the density of the air is measured in terms of pressure it applies. The refraction of waves passing through media of different densities was explained in the chapter on physical principles. Refraction also occurs in radio waves passing through varying air densities.

A condition of temperature inversion, when warm air overlays cold air, is desirable for long range communications on VHF bands. Note in figure 8-14 that warm air, being less dense, refracts the radio waves back toward the cooler air which is at the earth's surface. You can see then that there are conditions in which radio waves do not travel a line-of-sight route and cannot always be depended upon to do so.

Line of Position (LOP)

If you are on a line which has a definite bearing to a known point, you are on a line of position (LOP). A line of position is the plot of all positions on a line of a particular bearing. LOP may be derived from a compass or radio bearing, Loran sites, or celestial sights. The exact position is not known until a second LOP is obtained which intersects with the first LOP. The point of intersection is known as a fix. A line traced by a series of fixes is a course or track. An accurate fix may also be obtained when the course crosses a known landmark.

Course Terms

To maintain a course, the missile track must be headed along that course. The heading is most often taken from a magnetic reference and is known as a magnetic heading. Because of magnetic variations, a true heading is often used, giving the actual geographic direction of flight. A true heading may not result in a course of the proper direction because of drift. The greatest drift is caused by crosswinds through which the missile flies. This is illustrated by the diagram in figure 8-15. A true heading is usually chosen that is correct for a course between two checkpoints for a long flight, or between the launch point and target point for a short flight. The true heading course is called a rhumb line. A rhumb line is a straight line that intersects all meridians of longitude it crosses at the same angle and is therefore a single-heading course. A rhumb line is not the shortest distance between two points on the earth's surface.

Great Circle

The shortest distance between two points on the surface of the earth is measured along a great-circle course. Figure 8-16 shows a great circle. A great circle connects two points on a line that a flat plane would make on the earth's surface if it passed through the two points and the center of the earth. The great circle is a circumference of the earth, and it has the least amount of bending and thereby is the shortest distance between two points. Most guidance systems are programmed to use the great circle route for navigation.

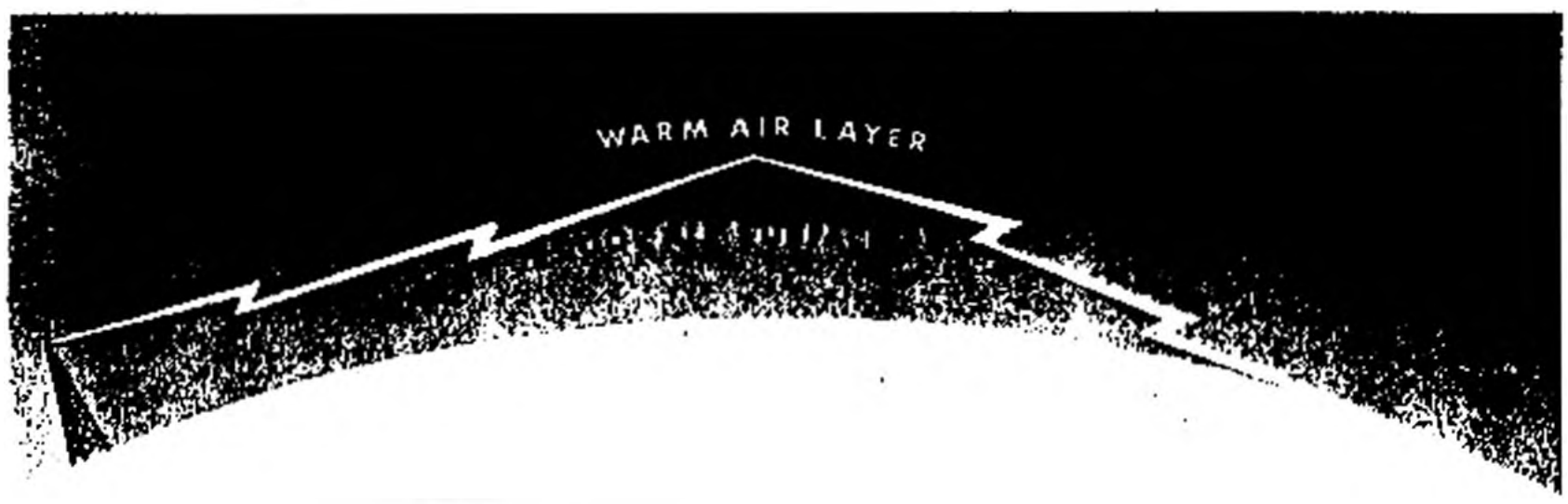


Figure 8-14. Refraction of Radio Waves by Temperature Inversion

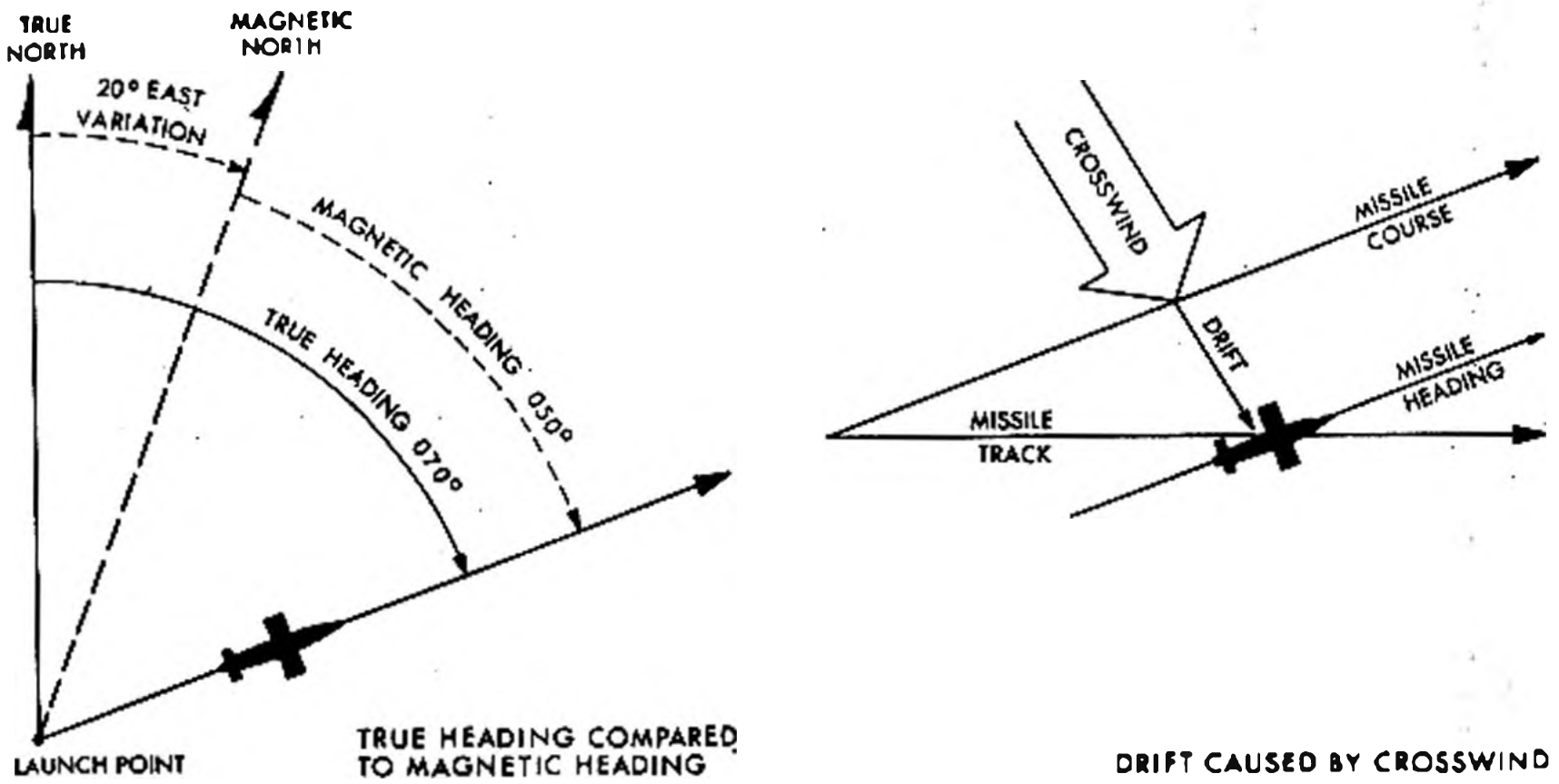


Figure 8-15. Navigation Terms

CELESTIAL NAVIGATION

To simplify the concept of the heavens for celestial navigation, it is assumed that the stars are on the inner surface of a large sphere which rotates about the earth. See figure 8-17. This is not a modern concept. It was originally proposed by Aristotle and his followers. The stars will rotate together, but the sun, moon, and planets have eccentric orbits, and are treated separately. The celestial sphere apparently rotates on an axis that is an extension of the earth's axis. At any instant, a point or star on the celestial sphere is over a corresponding point on the earth because this geocentric system consists of two concentric spheres. By using a sight measuring device, such as a sex-

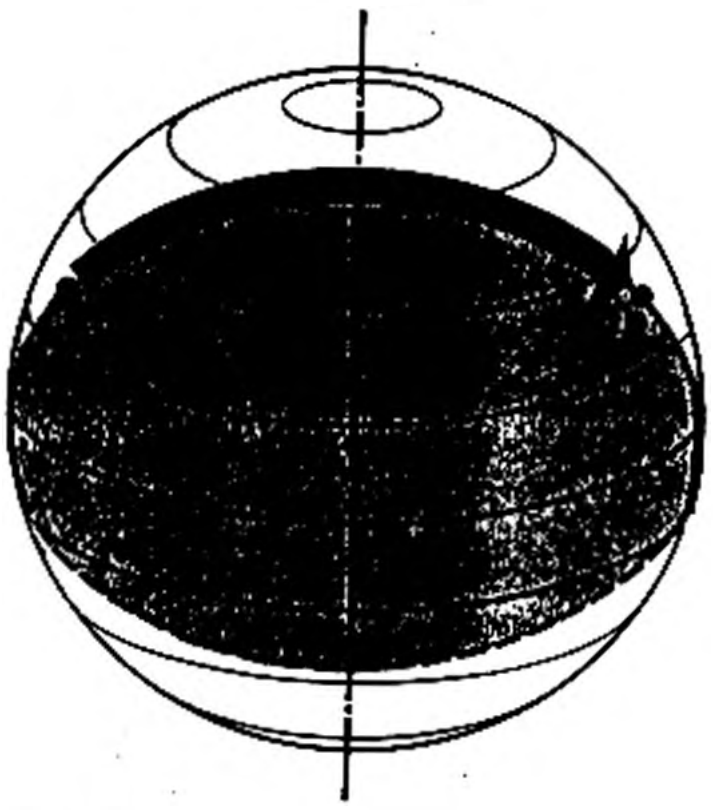


Figure 8-16. Plane of Great Circle Route from A to B

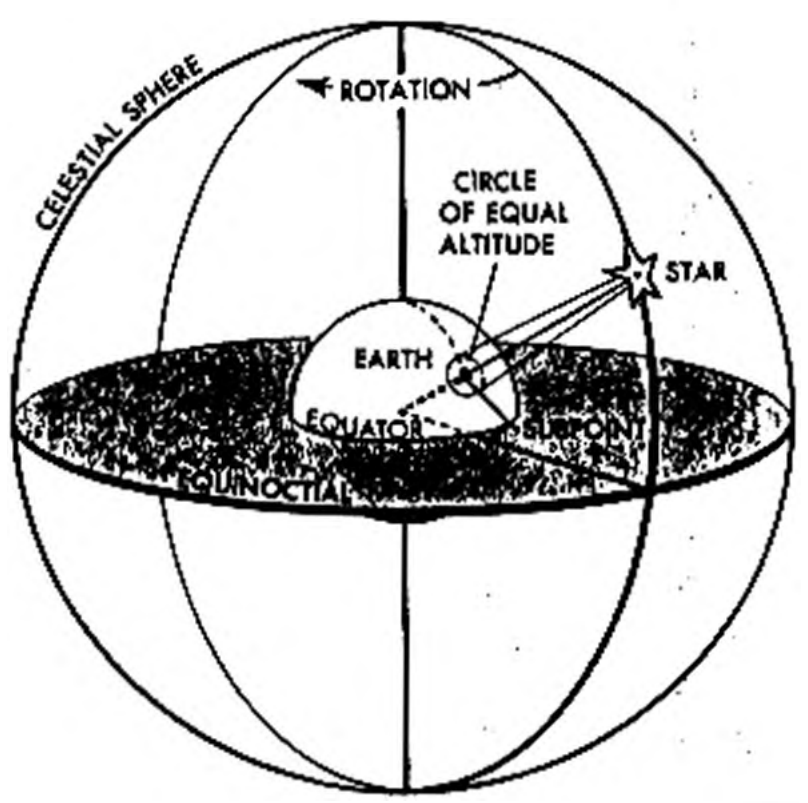


Figure 8-17. Celestial Concepts

tant, and trigonometry, an observer's position can be determined. A navigator, knowing a star's geographical position on the earth, uses that position as the center from which to measure an arc, the radius of which is ascertained by his observation and calculation. This arc is part of the circle of equal altitude, so called because at all points on this circumference the star will have the same angle above the horizon at that instant. The circle of equal altitude is a celestial LOP. The intersection of two celestial LOPs gives a position or fix.

Stars As Navigational References

Although stars give the least amount of light of all celestial bodies, they are the most desirable as navigation references. One of the reasons for this is that they are point-sources of light and can be accurately sighted on. They are at such great distances from the earth that light rays coming from them are essentially parallel. This is not true of the divergent light source from other celestial bodies.

To an observer on earth there is no relative movement among the stars, while there is considerable motion among the planets, sun, and moon. As the earth rotates in its orbit about the sun, so do all the planets. Their distance from the sun and their orbits are different from those of the earth. This combination of motions when viewed from the earth accounts for the apparent erratic movements of the planets. For purposes of navigation, stars are considered as fixed to their position on the celestial sphere.

Stars are classified on a brightness scale. This scale is known as the *magnitude scale*, and the lower the number, the brighter the star. Those stars visible to the naked eye on a clear night are 6th magnitude or brighter. Telescopes used for celestial navigation of guided missiles have been developed with the capability of "seeing" and locking of stars that are not visible to the naked eye. Such telescopes can be effectively used during daylight hours.

Stars also give off light of different colors (wavelengths), an important fact when using an electronic device to track them. Certain phototubes are more sensitive to light of a particular color. Color-corrected magnitudes are used and the tracking device is able to identify a particular star by its magnitude and color wavelength.

Celestial Coordinates

The celestial sphere has certain features corresponding to those of earth. Because the earth rotates within the celestial sphere, latitude and longitude cannot be used to determine the position of a star. The celestial sphere uses declination and sidereal angle for locating coordinates. Declination is the celestial latitude in respect to the equinoctial or celestial equator. The sidereal hour angle is measured westward from an arbitrary meridian known as the *first point of Aries*.

The first point of Aries is the point at which the sun crosses the equinoctial line at the time of the vernal equinox and corresponds to the prime meridian of earth as a starting point for longitude. The sun's path on the celestial sphere does not follow the equinoctial line but rather follows a curve called the ecliptic, because the earth is tilted with respect to its orbit about the sun. The ecliptic reaches its northernmost point at the summer solstice and crosses the equinoctial line during the equinoxes. The sun's declination varies from 23°27′ north to 23°27′ south during the course of a year.

Latitudes

Certain small discrepancies develop because of the ellipticity of the earth. Three different latitudes may be developed for the same spot because the world is not a perfect sphere. These are the geocentric, astronomic, and geodetic latitudes. If you refer back to figure 8-11, you will see that the geocentric latitude is that angle which a line through a point to the center of the earth would make with the equatorial plane. The astronomic latitude is the angle with the equatorial plane that a line perpendicular to the earth's surface will make. Geodetic latitude is the same as the astronomic latitude with corrections in the direction of gravity for local conditions.

The ellipticity of the earth is exaggerated in the illustration. Actually, the ellipticity at latitude 45° causes only 11 minutes difference between the two latitudes, geocentric and astronomic. Even this small difference, if not compensated for, has a considerable effect on the accuracy of a long-range missile.

As mentioned earlier in this chapter, the greatest problem is predicting the exact location of the

target at the time of arrival of the missile. In the case of the interceptor missile, the problem is obvious. The missile is fired at a moving aircraft. The aircraft is moving in relation to the missile and is capable of taking evasive action. The missile must have superior flight characteristics and computer system to out-maneuver its target and score a hit.

Although not as obvious, the same problem exists for a long-range missile flying toward a target on the earth's surface. The target is on the earth which is rotating on its axis while revolving around the sun. The target is actually moving with reference to inertial space and describes a complicated trajectory through space.

The missile does not fly toward the actual target, but rather to a point in space at which the target is predicted to be at the instant the missile arrives. The target is fixed and does not perform evasive action. Many forces acting on the missile, however, can have the same effect. A very small uncalculated force is sufficient to deflect the missile many miles off target on a long flight. The winds, Coriolis force, and equipment tolerances are some of the forces.

The problem of a missile and its path to the target is solved through an application of mathematics called *trajectory analysis.* The positions of the missile and target are located precisely, and a course for the missile is then plotted.

GEOMETRIC COORDINATE SYSTEMS

Because a missile flies in 3-dimensional space, a 3-dimensional coordinate system must be used. You are familiar with coordinate systems under several names; for instance, latitude and longitude are coordinates.

A coordinate system uses a reference point as its center. Normally, this point is the place from which the missile is launched. This reference point must be precisely located, for it is the source of all other measurements. Measurements from this reference to the target are made by several methods. The method of making the measurements determines the name of the coordinate system. Of the variety of systems, you should become familiar with the rectangular or orthogonal, cylindrical, and spherical systems.

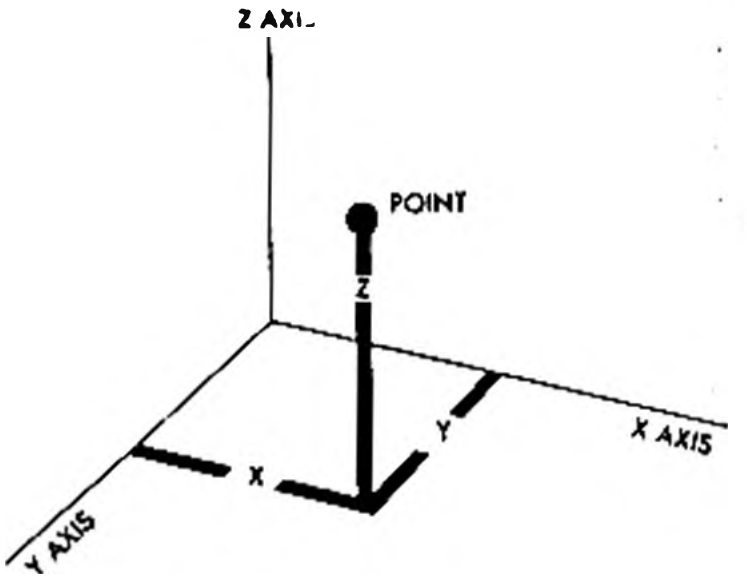


Figure 8-18. Rectangular Coordinates

Rectangular Coordinates

This system consists of three axes which are mutually perpendicular to each other at a common point as shown in figure 8-18. The point of intersection is the origin of the system. A point in space is fixed by the three distances along x, y, and z from the planes formed by each pair of lines.

Cylindrical Coordinates

In a cylindrical system the position of a point is determined by polar coordinates in a horizontal plane and the distance along the vertical axis as illustrated in figure 8-19. Polar coordinates are easy to understand when they are used in speaking

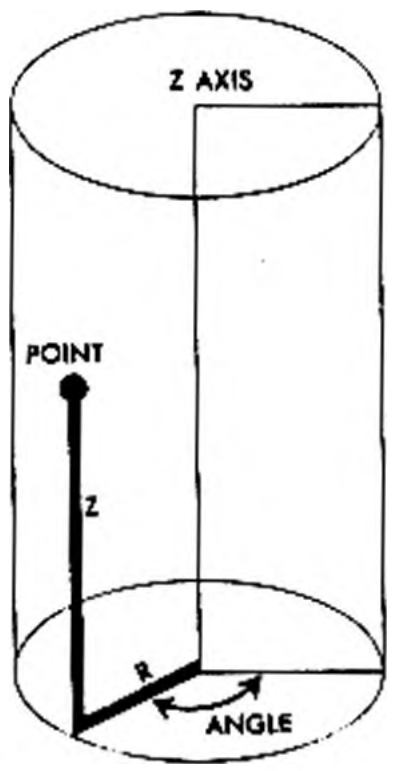


Figure 8-19. Cylindrical Coordinates

of a location as being so many miles NE of a position. Polar coordinates are a direction (angle) and a distance (radius) from a point of origin.

Spherical Coordinates

The spherical system is better known as the geographic system of coordinates. As illustrated in figure 8-20, the positive location of a point is determined by its distance from the origin (the radius), an angle in the horizontal plane known as longitude, and the angle from the vertical axis called co-latitude. In the sphere, the orgin lies at the intersection of the axis with the horizontal plane.

Coordinate Applications

In an analytical concept of a trajectory problem, the target may be considered the origin of a coordinate system. The missile would then be a point in space; the objective being to bring the coordinate values to zero, resulting in the missile hitting the target.

In the case of long-range missiles the most predictable reference point is the center of the earth. The target and missile are both represented in terms of this coordinate reference or possibly some other reference point—for instance, a point fixed in space at the predicted intercept point. The equations of motion are calculated, using the terms which define any force affecting the motion of either the missile, the target, or both. The terms are then transcribed to give the equation parameters in values that the missile guidance and control system can effectively use.

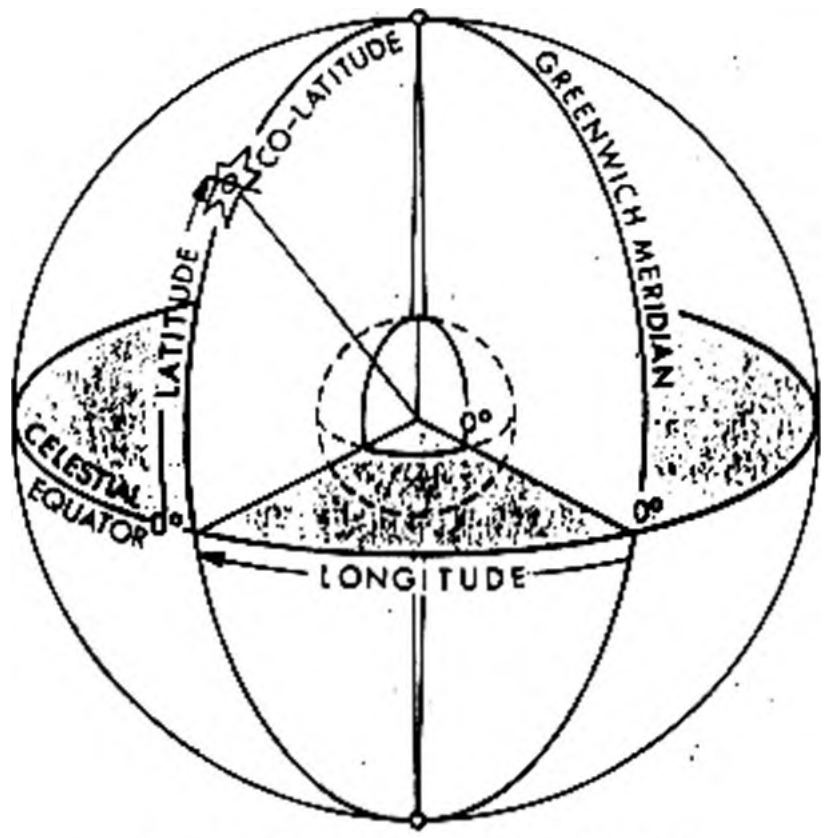


Figure 8-20. Spherical Coordinates

This chapter does not attempt to present a complete study of missile trajectory. The purpose of this chapter is to present the principles involved in missile trajectory computations and to acquaint you with terms with which you should be familiar.

CHAPTER 9

Guidance Systems

Factors to consider in selecting a guidance system for use in a missile are accuracy, protection against countermeasures, distance over which the system is usable, dependability and durability, physical size of the system, simplicity of operation and adjustment, and economy in manufacture.

In the past, emphasis was too often placed on the quantity of projectiles a given fire-power could deliver. As a result, accuracy suffered. Artillery used several rounds to bracket a target, correcting until a hit was made. The advent of missiles brought the necessity for a "one-shot" barrage per target. The expense and complexity of a missile led to the requirement for a *very high degree of accuracy*. The first shot must be right on center. If it is not, the element of surprise is lost and an expensive missile is wasted. This means that not only must the guidance be accurate but also the pre-launch planning, calculations, and adjustments must be precise. Little tolerance from calculated requirements is permissible. Accurate fire in volume is acceptable, but volume of fire as a substitute for accuracy must be rejected.

An ideal guidance in all classes of guidance systems is one for which there can be *no countermeasures*. Prior to countermeasure must come detection. A missile that emanates a radiation or uses one generated by the launching vehicle is detectable at the target at ranges equal to or greater than that at which the missile itself can see the target. The target can then take measures to evade or neutralize its attacker. A missile that uses a signal or radiation generated by a friendly source for guidance overcomes such countermeasures by using extremely high speed or by short time of flight. The target defense then does not have time to react after it detects the missile.

The *distance* over which a guidance system can function dictates the use to which it is put. A system of limited range is best used for close targets, as in the case of aircraft interception. The longer range systems are used as offensive bombers.

The *field of view* of a system is important. If the missile has a large error at the beginning of a flight and can still determine and strike its target, it is more valuable than a missile which is limited to a narrow zone of approach.

Physical and electrical factors become important in the tactical movement of a missile. Its size, the need for minimum maintenance and replacement, ease of storage and operation, and economical manufacture are all considered in the design of new equipment.

As the field of missile development progresses, the concept of range keeps pace. At the start, short range was classified as that which was within visual limits. This was later extended to a missile that flies less than 25 miles. The definition of long range has gradually expanded from that of several hundred miles to the global range of 6000 miles or more.

In the next few chapters, some systems are covered for the historical background only. As systems evolve, the old ones become obsolete. Even so, the parent systems are worthy of consideration in the interest of understanding the reasons for trends in present systems.

This does not mean that equipment will be increasingly complicated as time goes on. By combining principles and making accurate components, certain refinements in present systems will no longer be necessary. These refinements only exist to compensate for possible signal errors in

the first place. In the end an uncomplicated, accurate system is hoped for and anticipated.

A type of guidance is not necessarily used for any one missile. An inertial system, for example, is used in a number of missiles. The guidance unit in an air-to-air missile could also be used as the terminal guidance in a surface-to-surface missile. Certain guidance systems, such as radio command, are primarily used during research and development to check on flight characteristics of the missile. Other systems, which also employ a radio link between missile and controller, are used in testing and instrumentation.

The use of a combination of guidance systems makes possible the division of a missile flight into phases. In the case of a missile that contains more than one type of guidance equipment or in which the guidance equipment operates for only a portion of the flight, the flight can be divided into phases which are defined by the type of guidance equipment operating.

A standard flight breakdown into phases which would fit all types of situations is not possible. Some missiles have a flight consisting of but one or two phases, while other more complex systems have a flight consisting of four phases, with some of the phases subdivided into periods.

All missiles have an *initial or launch phase*. During this phase, the missile becomes airborne. Often, because of the acceleration due to propulsion thrust during launching, no guidance system is working. Reasons for this are covered later. In specially designed systems the guidance equipment can operate during the launch phase. The launch phase is then combined with a later phase of flight.

In the more complex systems, a *calibration phase* follows the initial phase of flight. During the calibration phase, the midcourse guidance system is supervised by an external system and corrected to the proper trajectory. The calibration phase lasts until the missile system is aligned, with the minimum error, to the desired trajectory.

The *midcourse guidance phase* follows the initial phase of flight in most medium- and long-range systems. Midcourse guidance is the primary system in these missiles because it must guide the missile through the major portion of its operation. It must be extremely accurate because of the great distance over which it functions.

As a result of the required guidance accuracy, the midcourse guidance system contains the greatest refinements in circuitry and components and is the most difficult to understand. At periods when midcourse guidance is not able to function, there may be provisions for replacing the guidance signal with a standby system.

Terminal guidance is the final phase of controlling the flight of a missile. It is the phase in which the missile makes contact with its target. A short-range guidance system that allows the missile to home on its target with minimum error is used. It may also be that the terminal guidance is an inertial system whose accuracy depends upon the accuracy of the midcourse system. The information available to a terminal inertial system is the error information supplied by the midcourse guidance, up to the time when the terminal system took over.

The division of phases of a flight into periods is somewhat more arbitrary than the division of the flight into phases. This is true because of different requirements and principles of operation of various guidance systems. Although a flight is separated into phases, no hard and fast rule can be used to determine this separation. Allowances must be made for the ingenuity of guidance system design in which the functions, as outlined by the several phases, may be combined into a common time or phase of operation.

SECTION A
Short-Range Guidance Systems

Short-range guidance systems can be divided into two general types: preset and command. A preset system of guidance falls into the category of a self-contained system. The intelligence necessary for flying the midcourse and terminal phases of the flight is entered into the missile before it is launched. The usual definition of preset guidance includes only a simple system that defines heading, altitude, time or length of flight, and programmed turns.

In a command guidance system, the missile receives correction signals from an external source. A correction signal is a command that activates the controls for a definite time in a specific direction. Do not confuse it with an error signal

which is the detected discrepancy from the required course, altitude, attitude, or speed. Ordinarily, the guidance system detects the missile position error and then develops the correction signal. After the correction signal is properly formed, it is fed into the control system.

PRESET GUIDANCE SYSTEMS

A preset system of guidance is completely self-contained. A simple preset system can be separated into two functions: one function is to determine the heading and the other is to determine the distance of flight. This portion of the explanation is divided into two corresponding sections. Various means of accomplishing each of the functions are considered.

The most obvious form of heading reference is that using the earth's magnetic field. For many years the magnetic compass and its refinements have served as a heading reference. A control arrangement, using the gyrosyn compass, is an example of a preset reference for heading. The yaw gyro of a control system is another typical example of a preset reference. If a yaw gyro is not capable of being precessed by an external source of intelligence while in flight, the heading set in prior to lauch controls the path of the missile.

The length of flight can also be determined in a simple fashion. The distance a missile has flown can be determined by the use of an air log or by integrating the average velocity of flight with the time of the flight. An air log consists of a small propeller of known pitch, turning a Veeder-root counter. Through a precision gear box connecting the propeller and counter, the counter can be made to show miles, feet, kilometers, or an arbitrary measure of distance.

Air pressure developed by missile movement can be used to make an airspeed meter give an indication of velocity. By integrating (summing the average) airspeed over the time of the flight, a continuous check of the distance flown is obtained. This method has the same accuracy as the air log, and the only reason for choosing one over the other is the availability of components.

The latter method makes use of standard equipment but usually requires an electronic installation. The air log equipment is all mechanical, merely energizing a switch or relay at the dump point.

The simplest form of a preset system, as used in almost all missiles, is known as launch bias. Because of the tremendous acceleration occurring during the launch of a missile, the mechanical and electromechanical components do not always function properly. To take the missile through this initial stage of its flight, fixed control and throttle setting are made before the launching takes place. These control settings are calculated for a stable climbout during the initial acceleration, until the regular guidance and controls can take over.

In addition to heading and distance, a third aspect of flight exists. This aspect consists of altitude reference. An altitude transducer is used to control the height of the missile flight. An altitude transducer is an altimeter with an electric output. The simplicity and accuracy of the altitude transducer have resulted in its remaining as the primary altitude control or reference in even the latest guidance systems in aerodynamic missiles.

The preset system was the earliest of the methods of guidance, and in some form will undoubtedly be the longest in use. The settings for launching missiles are always necessary whether they are made manually as with the surface-to-surface missiles or automatically as set in by gunsight computers in air-launched missiles, and whether they last for many minutes as in the World War II missiles or for mere seconds as in many present systems.

COMMAND GUIDANCE SYSTEMS

A command guidance system is one in which the flight path of the missile is controlled from a source outside the airframe of the missile. This definition necessitates certain basic operations. First, some tracking device must determine the positions of both the missile and the target, either continuously or at designated intervals. Some missions may require that only the missile be tracked, not the target. This would be true in the case of an air-to-surface missile used against a stationary target. Second, some operation must be performed to interpret the tracking in-

formation and compute the guidance signals to be transmitted to the missile. Third, some method must be provided for transmitting the command signals to the missile. Usually, the command signals are relayed by some form of radio link. Finally, the missile must be equipped to receive, interpret, and use the command signals so desired changes in the flight path may be produced. There are numerous methods of accomplishing these operations.

Uses of Command Guidance

Command guidance systems have many uses. Besides its use as the primary system, the command guidance system is used in some form in most missiles as an auxiliary or secondary system. Only the very short-range missiles use the command system as the primary guidance function. For example, some of the VB, GB, and JB series used during World War II were guided by command signals until impact. The Azon was a standard general-purpose bomb equipped with a special tail assembly. This tail assembly contained a small radio receiver, a control servosystem, and movable vertical fins or rudders. The receiver circuits received command signals transmitted from the mother aircraft and converted these signals into an error voltage. The error voltage was coupled to the servosystem which caused the rudders to move in a direction determined by the polarity of the error voltage. The bombardier controlled the Azon in azimuth, from drop to impact, visually tracking it with respect to the target.

The range over which a command system can be successfully operated is limited, since command guided missiles must always be tracked either visually or by optical or radar equipment. If optical tracking methods are employed, the range is further reduced by darkness, clouds, and other visibility-limiting factors. Instead of being guided completely by command signals, missiles frequently use a common system in conjunction with some other system of guidance. In combination systems, the command unit performs certain specialized operations. Probably the most frequent use of a command link is to guide the missile from the launcher to a point where it is in a position for some other type of guidance system to take control. For example, command guidance is used in conjunction with a beam-rider system to guide the missile into the radar beam. Once in the beam, the beam-rider equipment could take control and guide the missile to the target. A second example is the use of a command system in conjunction with a stellar-supervised guidance system. After launching, the missile is guided to a previously calculated point in its flight path by some form of command guidance. After the missile reaches this point, the star-tracking telescopes locate and lock onto the previously selected stars and guide the missile on the midcourse phase of the mission.

An important application of command guidance is its use as a safety feature during missile test flights. All missiles launched from test ranges within the continental United States are required to have a fully operational command guidance system in addition to the guidance system being tested. The command system serves as a secondary means of controlling the missile if the primary guidance system fails to operate satisfactorily.

It is possible that a failure of the primary guidance system could cause a missile to fly into populated areas. In the event that tracking data shows a missile to be out of control, the command system is used to take control and guide the missile back to the range or, as a last resort, destroy it. Destruction is achieved through a command destruct channel in the missile equipment which, when actuated, detonates an explosive charge. This explosive charge will blow off the wings or rupture the fuselage. The objective is to destroy the aerodynamic characteristics and stability in some way and cause the missile to come to earth in an unoccupied area.

One of the main disadvantages of a radio command system is that it limits the number of missiles that can be launched within a given interval of time. This may be very important in a tactical operation. Each missile must be tracked and controlled separately from the control station. If several missiles were launched simultaneously, signals intended for one particular missile might easily be received by and change the path of some other missile. By use of different command signal frequencies, two or more missiles may be launched and controlled simultaneously from the same control station. But even this technique does not permit an unlimited rate of fire.

A further disadvantage of these systems is their susceptibility to jamming. Countermeasures designed to jam a command guidance system are most likely encountered near the target area. From the standpoint of defense, such countermeasures require that the signals transmitted to the missile receiver be capable of overriding or jamming the signals from the command guidance system. This requires a knowledge of the frequency used in the guidance system of the attacking missile and the time required to initiate jamming countermeasures.

In general, for command guidance systems, a wide range of frequencies must be made available. Allowance must be made for control frequency changes on short notice, and for the use of coding systems to hinder the possibility of jamming the command system.

Command Guidance Techniques

There have been many approaches, some quite ingenious, to the problem of command guidance. Perhaps the most straightforward of the early approaches was a multiple transmitter and receiver installation. In this arrangement the controlled craft had a separate receiver for each command function to be performed. The transmitter was either multiple or capable of transmitting its carrier on any of the several radio frequencies for which the receiver was tuned. This system had an advantage in that no channel interaction was experienced, but it had the disadvantage of being bulky and requiring too wide an RF spectrum. Because of the number of receivers, only the most simple and, therefore, insensitive receivers could be used.

The natural refinement to this type of system was the use of tone channels. By modulating the transmitter with various audio frequencies, it is possible to use one RF carrier and receiver. After demodulation of the carrier in the receiver, the audio frequencies are used to excite the tone channel, which is resonant to the particular frequency for the function they perform. In the all-RF system, the number of functions capable of being accomplished was limited to about four, because of space and channel requirements. By using a tone-modulated system, 10 audio channels can be used without undue complications. Then by proper combinations of the audio channels, 18 or more functions can be performed. In new "miniaturized" equipment, 20 primary channels are available, giving a possibility of many more functions.

The tone-modulated system developed new interference problems which become apparent especially in the running of development tests where accuracy is a necessity. Interference from other sources containing the frequency of a tone used for a control function causes a missile to react. Often a harmonic or sideband of a voice-modulated signal contained frequency components sufficient to upset a whole operation. The change to frequency-modulated systems eliminated much of this interference. Manmade interference of an FM nature still could create difficulties. To overcome this, coding combinations of the tone channels were devised. Certain operations could not take place in the craft unless several selected tones appeared at the receiver simultaneously. The chance that an interfering source could duplicate this combination was negligible.

COMMAND TRANSMITTERS. The RF channel of the command transmitter is not unlike any other FM or PM transmitter, as evidenced by figure 9-1. A quartz crystal is used for stability and accuracy of the carrier output. The modulation of the carrier must necessarily follow the oscillator if stability is to be retained. After modulation, the necessary stages of RF multiplication raise the carrier frequency to an output in the VHF band. The power amplifier furnishes the RF power necessary to send information through space to the receiver.

Notice in the diagram that in the lower audio channel the peculiarity of command equipment is evident. A group of audio tone generators are present. Each generator operates only when it is actuated by the circuits in the keyer, which may be manually controlled or computer controlled. The individual outputs of the tone generators are mixed together into a composite audio signal and applied through the audio pre-emphasis network to the modulator.

The pre-emphasis network maintains the audio signal-to-noise ratio at optimum operating con-

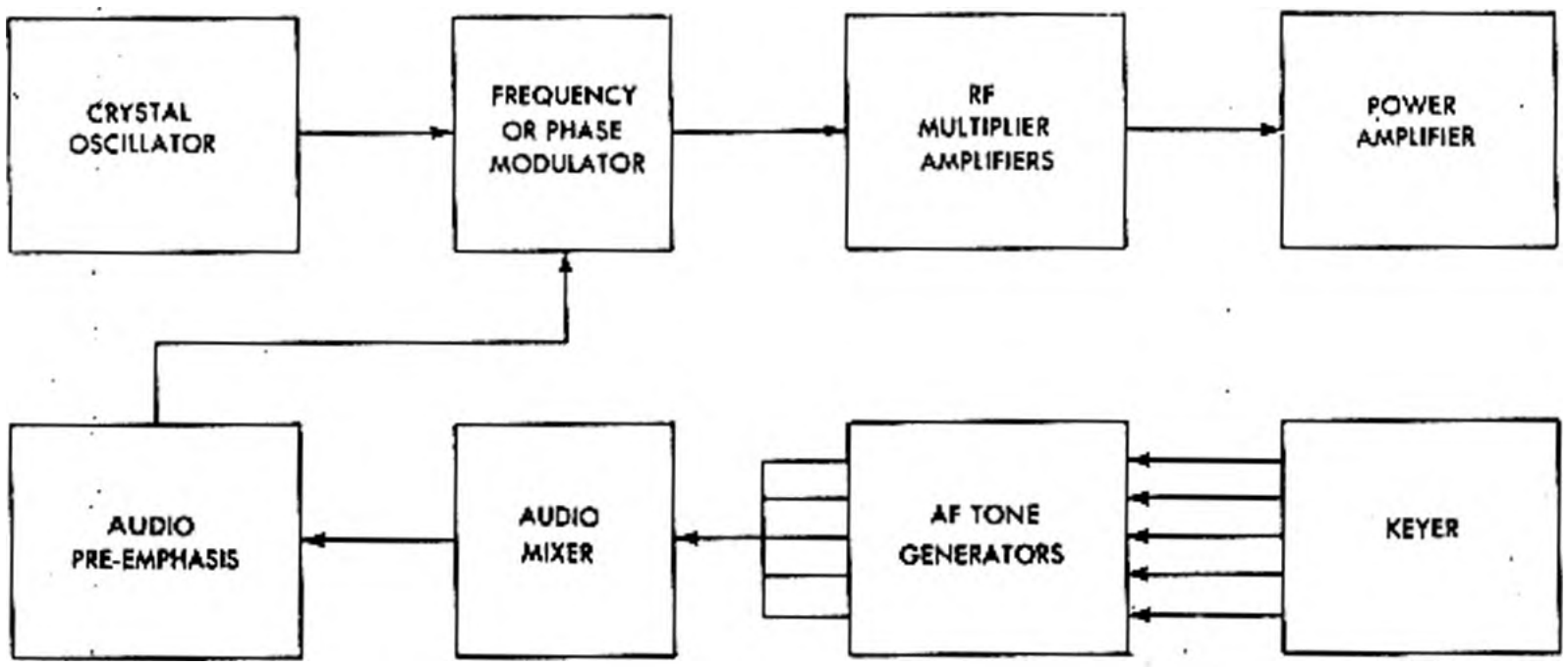


Figure 9-1. Command Transmitter

ditions for the overall system (including the receiver). The function of this network is to emphasize the higher audio-frequency tones which are later deemphasized in a network of opposite characteristics in the receiver. This operation causes the signal-to-noise ratio to remain more constant throughout the audio range. Because the noise appearing with the signal consists of high frequency components, the pre-emphasis and deemphasis on the audio frequencies need be performed only in this range.

COMMAND RECEIVER. The circuits of the receiver shown in figure 9-2 are those of a standard FM communications receiver with some added refinements. Ordinary superheterodyne connections are used through the second detector. At the limiter stage preceding the detector, the limiter grid current, which is proportional to the strength of the incoming carrier, operates a carrier fail relay in case the carrier should get too weak. The transmitter carrier is usually left on, even if no modulation is being sent, so the receiver will have a means of determining whether it is continuing to receive a signal.

After amplification, the discriminator (detector) output is applied to the selector channels. These selector channels and their operation make the receiver unique in its application. There is one selector channel for every tone that the transmitter may send. The selector channel, a breakdown of which is presented in figure 9-3, is usually an amplifier with a bandpass filter at the input and a relay in the output circuit. The selector tube is held below cutoff bias until the tone to which the input filter is tuned appears. This tone is passed by the filter and appears on the grid, causing plate current to flow. The plate current flowing in the coil of the relay energizes the relay, completing the circuit for performing the desired function.

Figure 9-4 shows a variation of the preceding selector channel. The tuned feedback is adjusted

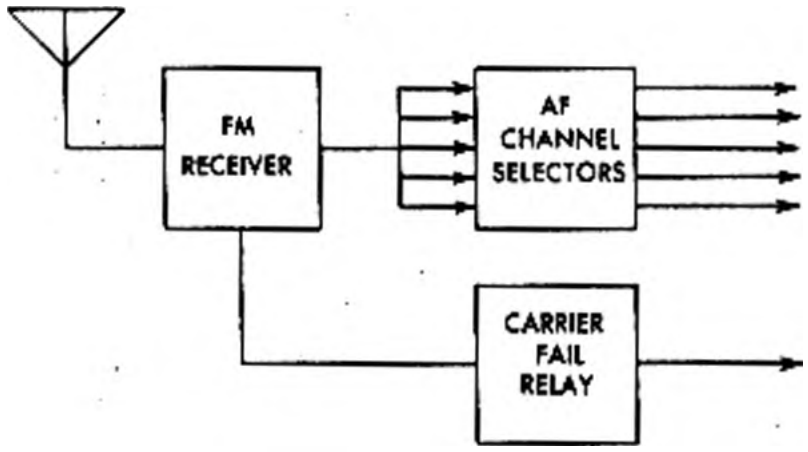


Figure 9-2. Command Receiver

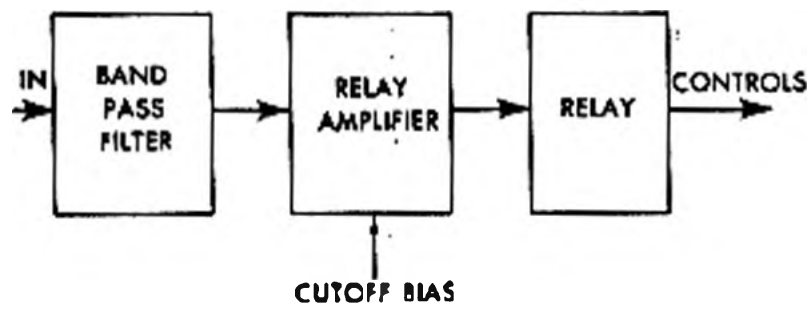


Figure 9-3. Selector Channel using Bandpass Filter

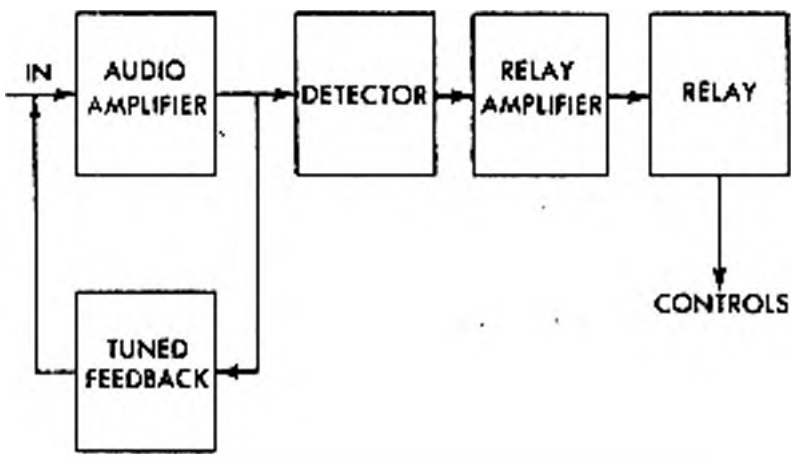


Figure 9-4. Selector Channel using Tuned Feedback

to aid the audio amplifier in amplifying only the desired frequency signals. Detection of the signal is then made before applying it to the relay amplifier.

A rather unique combination for issuing command signals is the use of tracking radar and command receivers. To extend tracking ranges on the relatively small missiles which offer poor radar return, a radar beacon is installed in the missile. This radar beacon transmits a much stronger echo pulse than the echo that would be returned on skin track (without the beacon).

The radar beacon is a small receiver and transmitter operating in the tracking radar band. The receiver has circuitry that only accepts a radar signal with a definite pulse separation. The receiver responds only to a certain coded interrogation. If the proper pulse combination is received, the beacon transmitter sends out a pulse of RF.

The transmitter operates on a slightly different frequency than the exact frequency of the radar. The radar receiver is tuned to this slightly different frequency so it can detect the beacon signal without having it confused by the ordinary return of the radar transmitter.

The beacon receiver can also be arranged to accept command signals. An additional set of coded pulses is added to the transmitted signal from the radar, arranged to give some intelligence such as a command tone. The beacon set accepts these signals and channels them into a different function, which decodes them for the intelligence they contain.

This intelligence is impressed on the radar signal as pulse position modulation. This modulation gives the autopilot its commands just the same as a human controller voices his commands to a pilot over the radio communications circuit. Pulse position modulation requires but two pulses per sampling cycle to give one intelligence channel. It is possible to include other intelligence channels in one sampling cycle by tone modulation, as was explained earlier. The beacon interrogation pulses are also in the pulse train received from the radar. The position of the modulated pulse is varied back and forth about its rest position at an audio rate. The number of intelligence channels is limited because of the limitation of audio modulation which can be accomplished by the sampling frequency and because of the fact that the radar PRF (which is the sampling rate) is dictated by the maximum required range of the radar set.

Examine the diagram in figure 9-5. Note how it is possible to use the beacon as a command receiver.

Trace the path of a signal through the diagram to show the transition from the pulsed RF input to the sine wave audio output. The receiver converts the RF signal to a set of video pulses which are separated by the pulse selector. The first pulse is used to start the multivibrator and the second modulated pulse is used to stop it, thus producing a square wave. This gives a square wave of varying width. The varying width follows the excursions of the modulated pulse. An integrator forms a saw-tooth wave whose amplitude varies with the width of the square wave. The pulse position or time difference in the received signal is converted to the amplitude difference of a sawtooth wave with a frequency the same as the radar PRF. The sawtooth waves of varying amplitude are fed into a frequency selective amplifier, and all components except the fundamental modulating frequency (which is evidenced by the amplitude variations) are removed. The resulting sine wave output then excites the relay amplifier and institutes the desired control function.

The criterion for selection of tones to be used in any command system requires that there be no possibility of a false signal being created through interaction of other signals. The audio tones must be kept below a certain frequency to limit the sidebands. Any harmonic of any tone, or beat note of any combination of tones, must not be able to actuate a receiver channel which is not

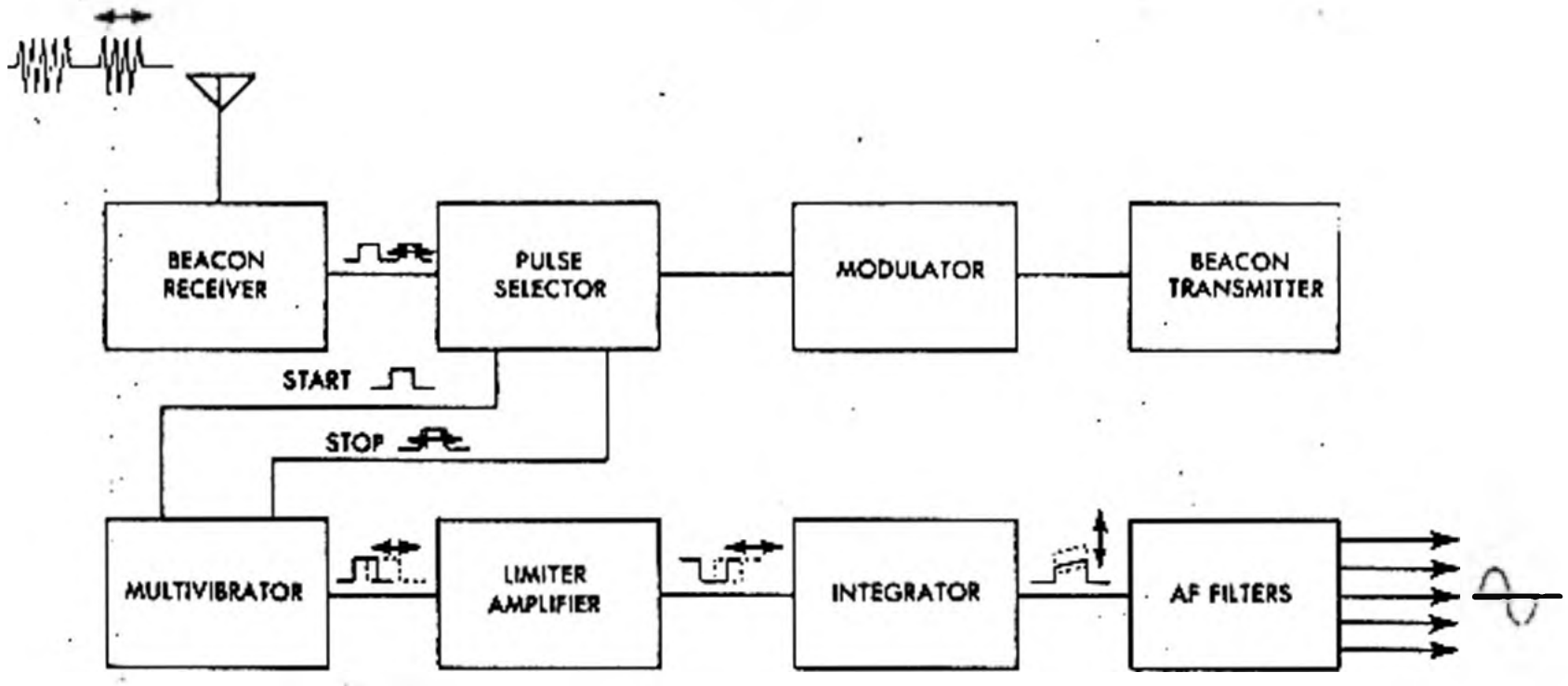


Figure 9-5. Combination Radar Beacon and Command Receiver

desired. The practical bandpass characteristics of the selective circuits make it difficult for them to be high "Q"; therefore, the separation between tones must be even greater than if high high "Q" circuits could be used. A set of frequencies to satisfy the above conditions requires quite a bit of study.

Inasmuch as the commands are so closely tied in with the control section, the control capabilities or requirements usually dictate the method of controlling the modulation of the command link.

In devising the control unit for keying the transmitter, the ultimate in development is proportional control. This means that a signal is sent which is just large enough to correct the particular error and no more.

A contrasting method of control is an on-off system, which means full-strength control applied for the period of the command signal. An example of the on-off system is control by a stepper box. The stepper box is a control box whose main component is a four-way toggle switch mounted so that its movement simulates that of an aircraft control stick. The operator merely holds the stick in the proper position to bring about the desired change.

It is not too difficult to realize what the normal outcome of this type of control could be. A pilot controller commands a function until it is evident to him what is taking place. At this time, the controls are well advanced so that there is a greater movement of the craft than desired. A countercommand has to be sent to correct this overcontrol. The missile then tends to follow an oscillating course which, if not properly damped, could become uncontrollable.

The closest approach to remote proportional control is a method by which the controller is varied in small increments to obtain smooth action and to allow the aircraft to receive the control necessary to produce an exact amount of change. The equipment in the missile follows these incremental moves exactly. This is done with a pulsing arrangement, much like the selector actuated by a telephone dial. The command operator elects the number of pulses to be sent. Each pulse produces an incremental change in the desired missile control unit. The unit would be a selsyn or some other type of pickoff. The pulse repetition rate is determined by the pulsing equipment, just as it is in the free rotation of a telephone dial.

In order not to lose the reference position, the operator may automatically step the missile unit to the position by pressing an index button. This indexing arrangement checks for lost or extra pulses. The index usually requires the use of another channel in the command equipment. Actuation of the indexing channel causes the selector in the missile to return to a reference point.

RADAR COMMAND GUIDANCE. The variation of command guidance that employs the use of

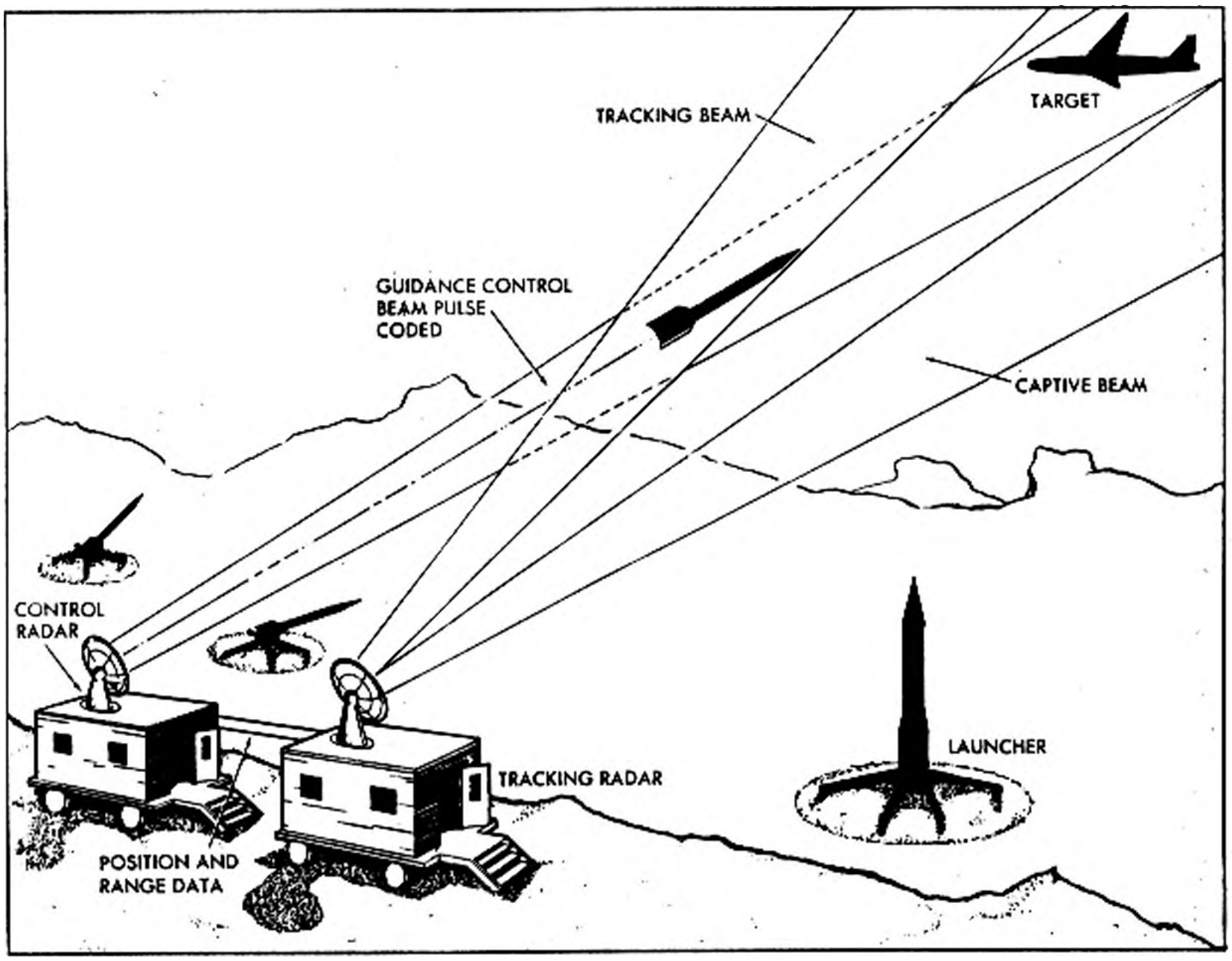


Figure 9-6. Radar Command Guidance

radar equipment to perform both the tracking and directing operations is illustrated in figure 9-6. This method is especially applicable to surface-to-air missiles, where both the missile and the target must be tracked to obtain command signals to guide the missile. In figure 9-6, one radar tracks the target while another radar tracks and sends commands to the missile.

The target-tracking radar obtains a continuous flow of information which gives the slant range, elevation angle, and azimuth angle of the target. At the same time, the missile-tracking radar is keeping a continuous account of the slant range, elevation angle, and azimuth angle of the missile. Figure 9-7 shows the relationship of these three factors. The information obtained by the two tracking radars is fed to a computer unit which compares the data and issues the corrections to the missile-tracking command radar. The computer controls changes in the missile flight path; when the slant range of the missile is equal to that of the target, the missile will also have the same elevation angle and azimuth angle as the target. When these conditions are satisfied, a collision of

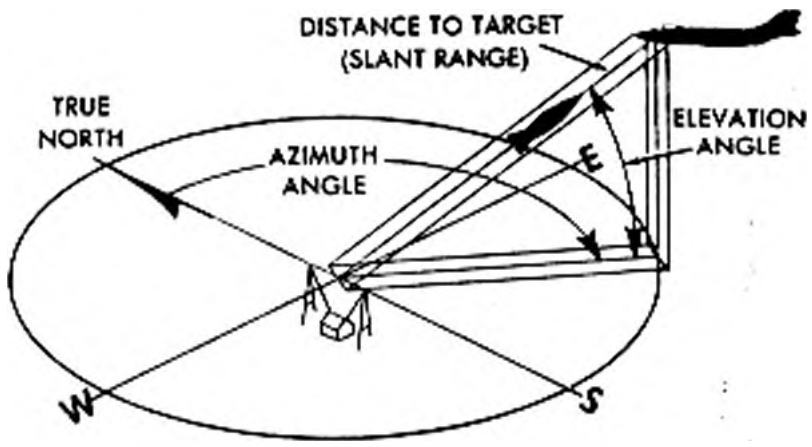


Figure 9-7. Slant Range, Elevation Angle, and Azimuth Angle

the missile and target will result. The missile uses some type of proximity fuze to detonate the warhead when the missile and target come within a calculated distance of each other.

The use of one radar system for both tracking and transmitting command signals forms a rather unique combination. Such a combination can be designed to extend the tracking range for relatively small missiles which might ordinarily offer a poor radar return. This problem can be solved by installing a radar beacon in the missile. When properly triggered, the beacon transmitter will transmit a much stronger pulse back to the ground-based radar receiver than the echo that would normally be returned because of the small size of the missile.

The beacon in the missile is a small receiver and transmitter operating in the ground-based tracking radar band. The receiver section of the beacon has circuitry that will accept only those signals from the tracking radar that have a definite pulse separation. The beacon transmitter does not operate on exactly the same frequency as the tracking radar, but on a slightly different frequency. The tracking radar receiver is also tuned to this frequency. This permits the tracking receiver to distinguish the signal transmitted by the radar beacon from the reflected return of the missile-tracking radar.

In addition to its functions relative to tracking, the radar beacon receiver in the missile may be designed to accept command signals from the missile-tracking radar. To accomplish guidance, an additional set of coded pulses is added to the signal transmitted by the ground-based tracking radar. The arrangement of these pulses determines the information contained in the command signal. The radar beacon receiver in the missile accepts these coded command signals and routes them through a network which decodes them for the guidance information they contain. This information is impressed on the signal from the ground radar as pulse position modulation which gives the missile its guidance commands.

Pulse position modulation requires two pulses per sampling cycle to give one intelligence channel. The output of the ground radar contains both the missile beacon interrogation pulses and the command signal pulses. The position of the modulated pulse is varied back and forth about its rest position at an audio rate. Consequently, the number of intelligence channels is limited by the audio modulation which can be accomplished by the sampling frequency. The sampling rate depends on the operating range of the radar set.

A one-radar system might also be used where the target is stationary. Knowing the geographic coordinates of a target and the capabilities of the missile, a desired trajectory can be computed prior to launching. After launching, one radar is used to track the missile and compare its actual path with the desired path. When deviations of the missile from the desired trajectory are detected, command signals are transmitted to the missile to produce the necessary corrections.

Short-Range Hyperbolic Navigation

The range of hyperbolic navigation depends upon the radio frequency used as the carrier. If a short-range navigation system of good accuracy is desired, hyperbolic principles can be applied to accomplish this. By using microwaves, a small, highly directive antenna can be contained in the missile without interfering with its aerodynamic characteristics. The main problem is to eliminate errors in the received signal by discriminating against skywaves. The directional characteristics of the antenna are narrow in the vertical plane and fairly wide in the horizontal plane. An important effect of the directional receiver antenna is that it has gain. It actually does the same job as an RF amplifier connected to a standard dipole antenna. Also, its directional characteristic decreases the possibility of countermeasures by jamming. Most of the skywave discrimination exists because transmitted energy can be compressed in the vertical pattern, and the line-of-sight transmission to the missile affords little opportunity for interference from multipath signals.

This discrimination against skywave interference is necessary in the synchronization of the ground stations. The synchronizing pulse must be transmitted via a direct, constant path to be accurate. There should be no variable factors such as skip effect which would alter the transmission of synchronizing signals. Establishing a condition in which these variable factors do not alter the transmission poses a greater problem with UHF

systems using a base line longer than the line-of-sight distance. A means which does not introduce any varying or unpredictable delays is used to relay the synchronizing signal. The transmitter stations have precision timing signal generators whose outputs are used to modulate RF transmitters that are typical for the band employed. They use the same sort of circuitry as would a radar transmitter in the same band. The slave ground station also requires, for the synchronizing signals, a receiver that is quite similar to the reception equipment used in the missile. The transmitters must have high power output to give a high signal-to-noise ratio at long ranges. The separation between ground stations is normally limited to less than 100 miles because of the limitation of line-of-sight distances. This short base line would result in a very short-range system if a three-station (one master and two slaves) system were used. The lines of position (LOP) in such a setup would cross at an extremely obtuse angle and would make an accurate fix difficult for any point located at any distance from the base line.

A four-station system has the advantages of both the line-of-sight transmissions and long base line systems. Two pairs of stations are used; each pair consists of one master and one slave operating on the same frequency, and they are properly synchronized. The pairs of stations are separated by some distance so that the resulting hyperbolic grid system has the LOP more nearly normal to one another in the desired target areas. Figure 8-4 in the preceding chapter shows a two-station grid at nearly right angles, the target points would be more positively located because of the LOP crossing at angles closer to a right angle.

One pair of guidance base stations is used to give the azimuth guidance hyperbola. One time difference line of this pair (a hyperbola) is chosen that will cross the target area and thus serve as the desired track of the missile. The guidance equipment within the missile can determine when the received signal pulses have the proper time separation and show the on-course condition. If the received guidance signals do not have the desired time difference, the guidance equipment can determine whether the difference is greater or less than the on-course condition, thus determining whether the missile is right or left of the desired course. The error signal from the guidance section is then sent to the control system for processing into the proper control correction in yaw.

The other pair of guidance base stations is used for range determination. A particular time difference line of this system is calculated to pass through the correct point for instituting the terminal phase of the flight. The intersection of this range line and the course line gives a fix at the predump point. The missile guidance equipment develops a voltage output proportional to the distance from the preselected target point. This varying output voltage approaches zero at a rate proportional to the velocity of the missile.

This output voltage is converted to a voltage proportional to the rate of change of the output voltage. This proportional voltage is then a measure of the missile groundspeed. Circuits are necessarily incorporated to correct for the hyperbolic expansion, because adjacent lines become farther apart as the distance from the base stations increases. The groundspeed signal is necessary in the terminal phase to alter the dive path so the missile will not overshoot or undershoot its target as a result of airspeed error from the preselected speed.

The output of the range portion of the guidance equipment decreases to zero volts, at which time the terminal guidance takes over. The missile then makes an approach, contacting the target area under terminal guidance.

Figure 9-8 illustrates the operation of the guidance equipment which might be located in a missile. The signals are received at the missile by a directional microwave antenna which is located so as to receive signals from the rear. A cover is placed over the end of the antenna, designed to hold air pressure around the antenna and transmitter at sea-level pressure no matter what the altitude of flight. The cover also prevents entry of moisture.

The two receivers form the beginning of the range and azimuth channels, as shown in the figure. The single local oscillator frequency is midway between the range and azimuth carrier frequencies. The resultant IF is then the same for each channel. The mixer is a crystal, located in the tuned-cavity coaxial connectors coupling the mixer output to the respective IF amplifier.

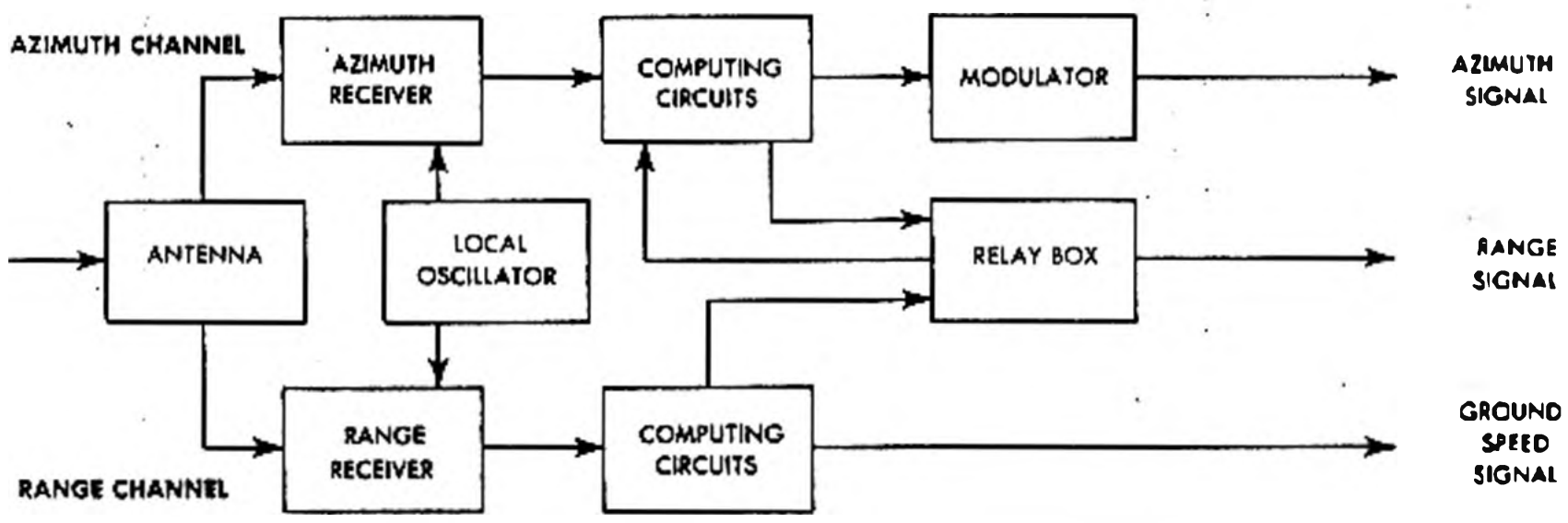


Figure 9-8. Short-Range Hyperbolic Missile Guidance

The IF amplifiers are stagger tuned to provide a bandwidth wide enough to preserve the shape of the short signal pulses. The output of the IF stages is demodulated by a crystal detector. The video section which follows then brings the signal level up to that required by the computing circuits.

The computing circuits first identify the master and slave pulses and then compare the time difference between them to the desired time difference. In the azimuth channel, the desired time difference represents an on-course condition. In the range channel, the desired time difference represents the beginning of the terminal dive phase.

Identification of the pulses is necessary for two reasons. First, the slave signal must be distinguished from the master signal so that the proper signal becomes the reference. Second, false signals caused by noise or enemy jamming must be prevented from controlling the equipment. Each signal must possess some peculiar characteristics so that it may be positively identified. This positive identification is done by sending a pulse train of two or more pulses in place of a single pulse from each transmitter. The pulses may be coded by using certain amplitude differences, pulse width, pulse separation, number of pulses, or some combination of these methods.

When time difference coding is used, the two slave and the two master signals contain two pulses each. In either channel, the slave signal has a different pulse separation from the master signal. Two decoders are used in each channel to identify the master and the slave signals.

The output of the range channel computer determines the groundspeed. The computing circuits are constantly measuring the distance from the missile location to a preselected (zero) hyperbola near the target area. The output of the computer is a constant DC voltage until the missile comes within a given range in the range hyperbolic grid. At this given range, the DC varies in proportion to the range from the preselected zero hyperbola. This constantly decreasing voltage can now be differentiated to indicate the rate of change of range time differences (crossing of range hyperbolas) as the zero hyperbola is approached. The operation becomes missile velocity.

The changing DC output of the computing circuits is also applied to another circuit where the signal amplitude at any given moment is an indication of the range time difference (range-to-go). The phase of the wave denotes approach to the zero hyperbola or passage beyond this point. At a preset voltage level following passing of the zero hyperbola, a relay is activated, causing the control system of the missile to take over. The controls then determine the dump or dive point after precise automatic corrections of the range-to-go voltage.

Beam-Rider Guidance

Beam-rider guidance has its primary use in air-to-air or surface-to-air missiles. It requires that the launching vehicle and guidance radar remain fixed on the target until the interception is complete.

The beam rider was one of the first of the interception or antiaircraft automatic guidance systems. It was a logical development from the automatic tracking radar for gun direction which was developed to a high level during World War II. Although highly successful against conventional aircraft of that period, such a system is limited by the range of the antiaircraft artillery and is not sufficiently accurate or flexible for effective use against modern aircraft and guided missiles. An interceptor missile launched on a fixed trajectory might miss the target because of evasive action by the target or pointing errors at launching, even though the tracking radar could easily maintain its track of the target. If the missile could be made to follow or "ride" the radar beam, the possibility of a hit would be greatly increased, and so the idea of a beam-rider guidance system was conceived.

Beam-rider guidance is a system in which the missile senses its position in the radar beam and corrects itself to remain in the center of the beam. The original plan was to have the ground radar, which was tracking the target aircraft, control the interceptor missile. Because this radar beam was moving continuously, the missile was subjected to sideward accelerations which tended to strain its aerodynamic capabilities.

A two-radar system was then devised, consisting of a tracking radar and a control radar. The output of the tracking radar is fed into a director computer which points the control radar at a predicted target position. The missile is launched into the control radar beam and rides it toward the predicted target position. Continuous corrections for this position must be made, but the beam movements and the missile movements are very small as compared to those of the tracking radar.

A one-radar system is used in the air-to-air version of the beam rider. This means that the launcher aircraft, the missile, and the target aircraft have to remain in line at all times. It is possible to fire missiles in sequence with a half second or so separation, controlling more than one missile at a time.

Consider a missile which is intended for use with an associated pulse-modulated radar system. The target in such a system is detected by directing a beam of pulsed high-frequency radio energy in a preset pattern over the space to be searched. When the beam strikes an object, energy is reflected, and a small portion of the reflected energy is returned to and detected by the radar system. The beam is narrow, so the direction of the target is known. Since the energy is pulsed, the elapsed time between the transmission of the pulse and the receipt of its echo gives a measure of the range.

The projected beam is not fixed with respect to the axis of the antenna reflector but is caused to trace a cone in space by nutation or rotation of the central antenna feed. Rotation is the spinning of the antenna dipole, slightly offset from the reflector focal point, to produce a conical pattern. Nutation moves the antenna in circular position about the focal point without varying its polarization; that is, the antenna remains in a horizontal or vertical plane throughout its cycle of nutation. Nutation of radar antennas is a much more efficient method of conical scanning because the returned signal always has the same polorization as the antenna. In rotating feed, the polarization of the returned signal always slightly lags the antenna position and thus gives a weaker signal than if the antenna were properly polarized.

A target on the axis of the reflector always has the same echo amplitude for all positions of the beam. If the target moves away from the beam axis, the echo signal varies approximately sinusoidally with the rotation of the beam. The reflected signal can thus identify the target displacement from the reflector axis in direction. For small target deviations, the error signal amplitude indicates the amount of displacement. The axis of the conical trace of the radar beam tracking the target provides a line-of-sight route. The problem concerns the method of controlling the missile to fly along an automatically maintained path or to ride the beam. To accomplish this, the missile must be able to sense when it is on the beam axis. If displaced, the missile must know the direction and degree of displacement.

The missile has no internal phase reference as does the radar tracking system. The modulated signal of the beam must furnish the necessary guidance information. In addition to other functions, the beam is modulated so that it can either furnish the missile with the sensing information in a one-radar system or command the missile in a two-

radar system. The basic idea of radar command control has already been covered.

In a single radar system the spinning beam of the director radar impresses signals on the beam to identify the position of the missile about the beam axis. These identifying signals may be in the form of coded pulses transmitted while the director radar antenna is in a specific quadrant. This coding could be in the form of pulse-time modulation applied to every other tracking pulse and used in a system as shown in figure 9-9. This pulse-time modulation, when detected in the pulse position detector of the missile receiver, would direct the error sine wave to the position channel at the conical spin or scan frequency, which is usually about 30 to 60 cycles per second.

This reference frequency is compared in the position channel phase demodulator to the sine wave resulting from the amplitude modulation on the received pulses. The reference frequency is then phase-detected to give the direction of error. Correction is obtained to return the missile to the center of the beam in the up-down axis and right-left axis.

In a system using quadrant identification signals, the signals are employed in the missile control system. The tracking pulses are not used by the missile. With such an arrangement, jamming is more difficult since the coding of the pulses by the time spacing between them is the only information desired. The quadrant signals can be thought of as turn commands—UP, DOWN, RIGHT, LEFT. If the missile is on the axis of the cone, all turn signals balance, and the missile continues its flight without correction. If the missile is displaced from the cone axis, one signal overbalances the other. The missile obeys the stronger signal to turn back toward the beam axis.

Beam-rider guidance data received by the missile antennas is detected and sent as coded pulses to the guidance receiver. After sufficient video amplification has been accomplished, the signals are decoded in suitable coincidence circuits and then filtered and transmitted to the control system as up-down and right-left error signals. These signals appear as electrical voltages proportional to the missile deviation from the beam. The control system then institutes appropriate action to correct any off-course error.

The beam-riding missile can be distinguished by the fact that the antenna or antennas for guidance are located on the aft section, or rearward looking from the missile.

To pick up the echo signal would be detrimental to the guidance system within the missile itself. Because of the dispersion and scintillation of the returned signal there would be confusion as to the quadrant identification. The guidance antenna

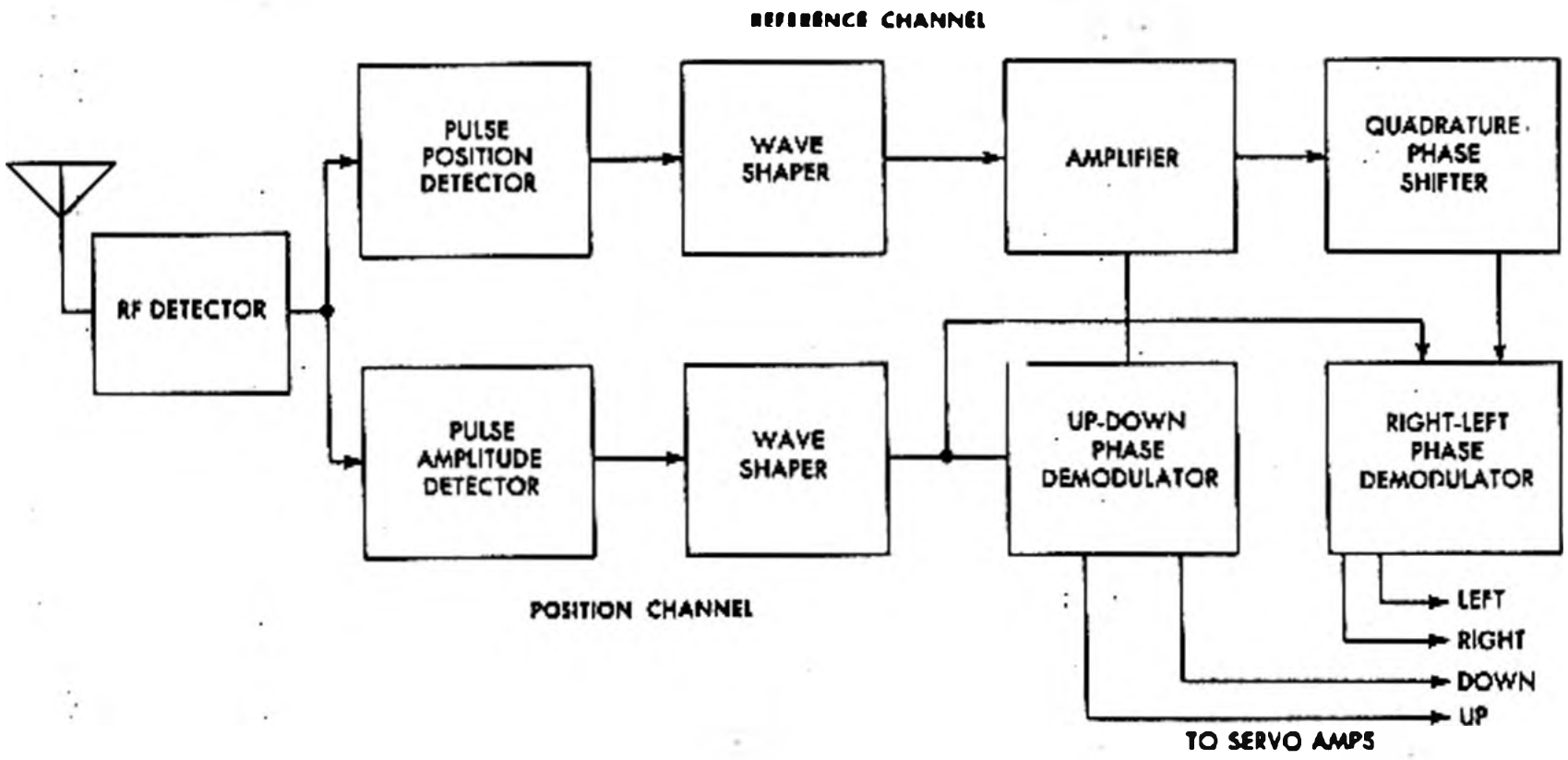


Figure 9-9. Beam-Rider Receiver using Pulse Position Reference

system of a beam rider, therefore, has its pickup lobes to the rear of the missile.

Because of the low traffic-handling capabilities of a beam-rider system, it is of limited use.

SECTION B
Long-Range Guidance Systems

This section presents various methods of missile guidance for long-range missiles. These guidance systems include inertial, celestial navigation, hyperbolic, and terrestrial reference and magnetic systems.

INERTIAL GUIDANCE SYSTEMS

Of the many types of automatic guidance systems, the latest and most important developments pertain to the inertial system. In its simplest terms, inertial guidance can be described as a type of guidance which is complete within itself. It needs no exterior energy or radiation source to determine its course. It emits no signal, and it does not depend on ground equipment to operate it once the missile is launched.

Inertial guidance is especially advantageous for missiles. Since it sends no signal and receives no signal, it cannot be jammed. Also, it is almost impossible to detect or intercept. It is not influenced by weather conditions. Missiles can be launched and guided accurately to the target with all corrections for winds, atmospheric conditions, and other factors automatically made in flight. The inertial system is presently considered the best guidance system for use against *stationary* targets.

During flight, the system computes its present position, altitude, and velocity; and it applies various compensations to its computer circuits. These corrections minimize the errors introduced into the system by gravity, Coriolis, gyro unbalances, and the non-spherical shape of the earth.

An inertial guidance system makes use of Newton's second law of motion which states: "An unbalanced force acting on a body causes the body to accelerate in the direction of the force, and the acceleration is directly proportional to the unbalanced force and inversely proportional to the mass of the body."

Acceleration is the rate at which velocity is increasing. The law also applies to deceleration, which is the rate at which velocity is decreasing. The acceleration of a body is proportional to the outside force exerted upon it. In an inertial system, missile acceleration is measured by means of the inertial force of acceleration upon a mass. Since this acceleration has an effect on flight, the inertial guidance system changes the flight path to compensate for undesirable changes in missile accelerations.

Inertial guidance may be used with an atmospheric missile, either air- or surface-launched, that flies a level flight path within the atmosphere, or it may be used with one that flies a ballistic path.

The ballistic trajectory is often used to describe the flight path of a missile. An operational ballistic missile may acquire speeds up to 15,000 mph or better at heights of several hundred miles. At the present time the speeds of such craft are considered unlimited, being governed only by available materials and design knowledge. Ballistic missiles are the type least likely to be intercepted. Usually inertial guidance is used for a ballistic trajectory only during the very early part of the flight (up to fuel cutoff) to establish proper velocity for a hit by free fall.

Inertial guidance can be used on both short- and long-range missiles. Regardless of the trajectory or flight path, complete control by inertial guidance requires detection of accelerations along the three axes of movement: (1) movement to right or left along the pitch axis, known as *lateral motion*; (2) movement forward or aft along the roll axis, called *longitudinal motion*; and (3) movement along the yaw axis, producing an *altitude change*. Note that the term "along" an axis rather than "about" or "around" an axis is used. Along an axis means parallel to the axis.

Basic System Operation

All inertial guidance systems are similar in basic operations. In contrast, different missiles may use different methods and electronic circuitry to accomplish the same basic functions. Discussion on inertial guidance in this manual covers one specific and proven operational system.

The three basic elements of any inertial system relating to a specific flight problem are acceler-

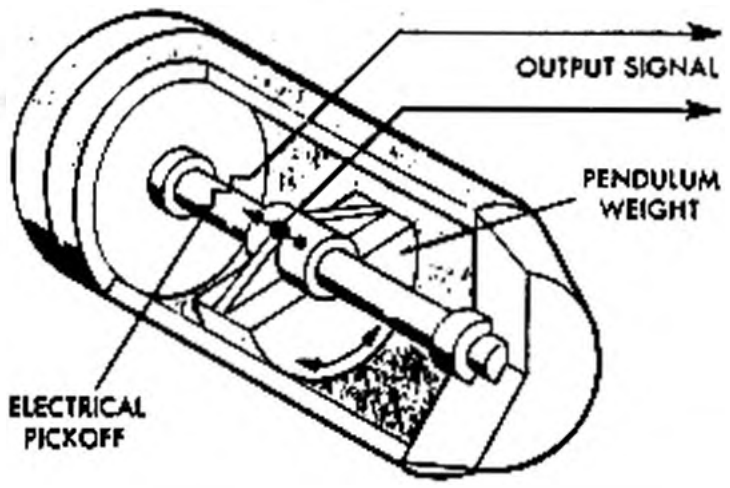


Figure 9-10. Measuring Pendulous Movement

ometers, memory devices, and gyroscopes. Even the most sophisticated of inertial guidance systems today have some counterpart to these three basic elements.

ACCELEROMETERS. Two accelerometers are used in a typical inertial system. One accelerometer measures speed changes along the desired flight course, while the other measures lateral speed changes; that is, those to the left or right of course. A simplified drawing of an accelerometer is shown in figure 9-10. The direction of the pendulum motion indicates either acceleration or deceleration, and the extent of speed change is reflected by the amount of angular deflection of the weight. The pendulum weight (mass) rotates when any speed change occurs. Attached to the pendulum is a contact arm which rotates with the pendulum. The angular rotation of the entire inner element produces an electrical voltage, proportional to the rotation, which is picked off by the wiper arm. When no speed change is taking place, no electrical signal is generated. The polarity of the electrical signal depends on whether acceleration or deceleration has taken place.

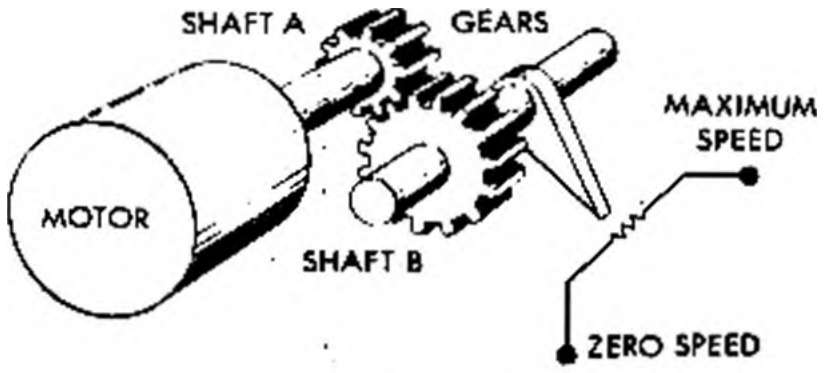


Figure 9-11. Integrating Device

The second basic element in an inertial guidance system is a memory or integrating device, which memorizes the speed changes that have taken place. The simplest memory device is an electric motor. The direction of rotation depends on the polarity of the applied voltage. The larger the voltage applied to a motor, the faster the motor runs. The faster the motor runs, the more motor shaft rotations result in a specific length of time.

A gear train can be attached to the motor to provide a reduction in the amount of speed and movement of the wiper arm (figure 9-11). Shaft A, coupled to the motor, drives shaft B through a gear train arrangement. Attached to shaft B is an electrical contact arm which rotates and develops an electrical signal proportionate to the motor speed. With the wiper arm starting at the bottom of the resistance element (zero speed), the amount of movement toward maximum speed will indicate the speed of the missile.

To obtain an airborne speedometer, the pendulum accelerometer and the simple motor memory device must be combined. See figure 9-12. Since the pendulum registers zero in the prelaunch position, the speed memory is stationary. As the missile starts to gain speed, the acceleration produces an electrical signal which energizes the

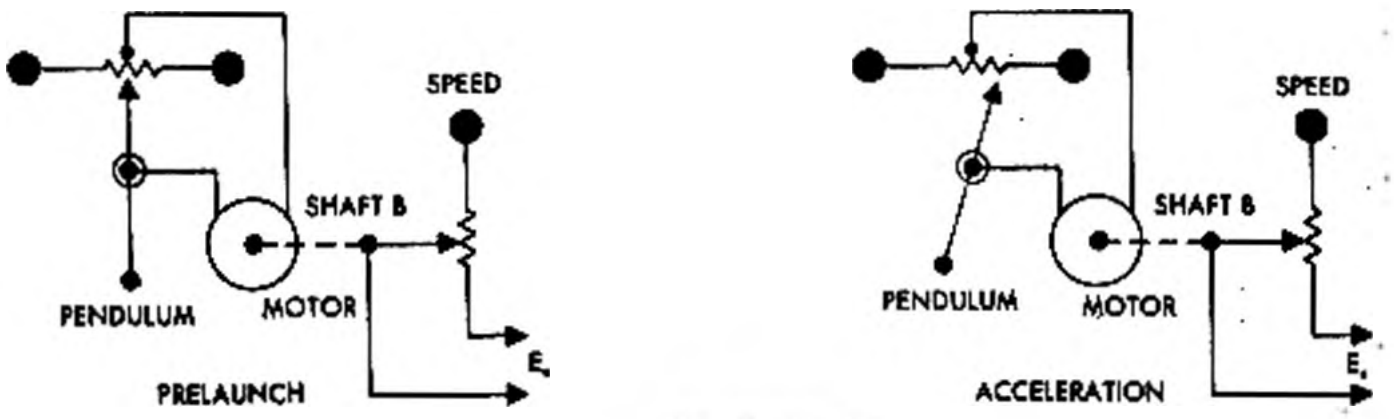


Figure 9-12. Combining Accelerometer and Integrator

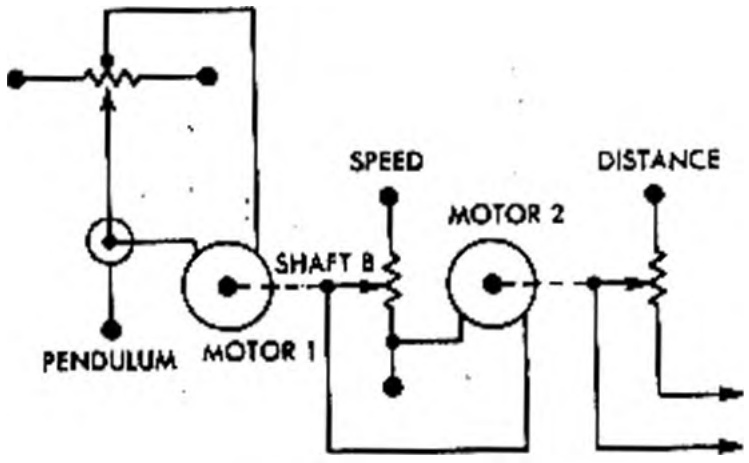


Figure 9-13. Distance Traveled Indicator

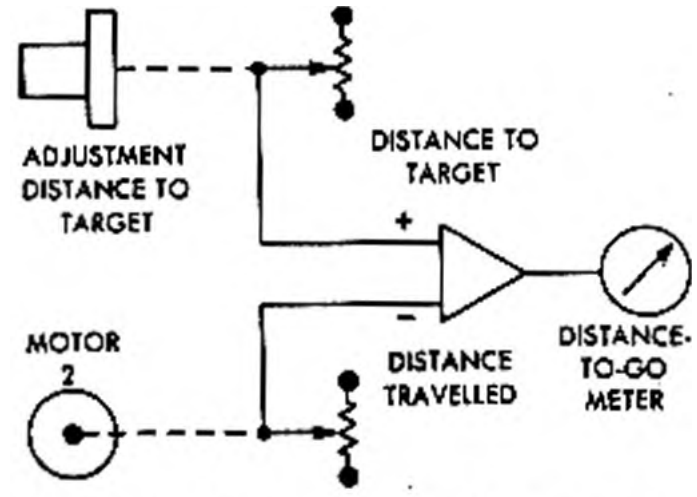


Figure 9-14. Distance Comparison Circuit

motor, which drives the wiper arm of shaft B along the potentiometer. When the missile stops accelerating, the pendulum returns to the zero position, stopping the motor. The wiper arm will remain in this position, memorizing (recording) the missile speed as long as there is no acceleration or deceleration.

To obtain the distance the missile has traveled, the use of an additional memory device is required. This consists of a second motor connected electrically to the wiper arm driven by shaft B (figure 9-13). As soon as voltage registers on the first potentiometer, the second motor begins to rotate, driving a wiper arm along another potentiometer. When the missile has attained a constant speed, the pendulum will return to vertical, motor 1 will stop, and motor 2 will continue to run at a constant speed. Any change in acceleration or deceleration will cause the cycle to be repeated, changing the speed of motor 2.

Since the distance from the launch point to the target is known, distance to go can be computed using another circuit—distance to target. If the signals from distance traveled and distance to target are connected to an amplifier circuit (figure 9-14), a continuous indication of distance to go can be computed. The signals from the two circuits are subtracted in the amplifier, and when

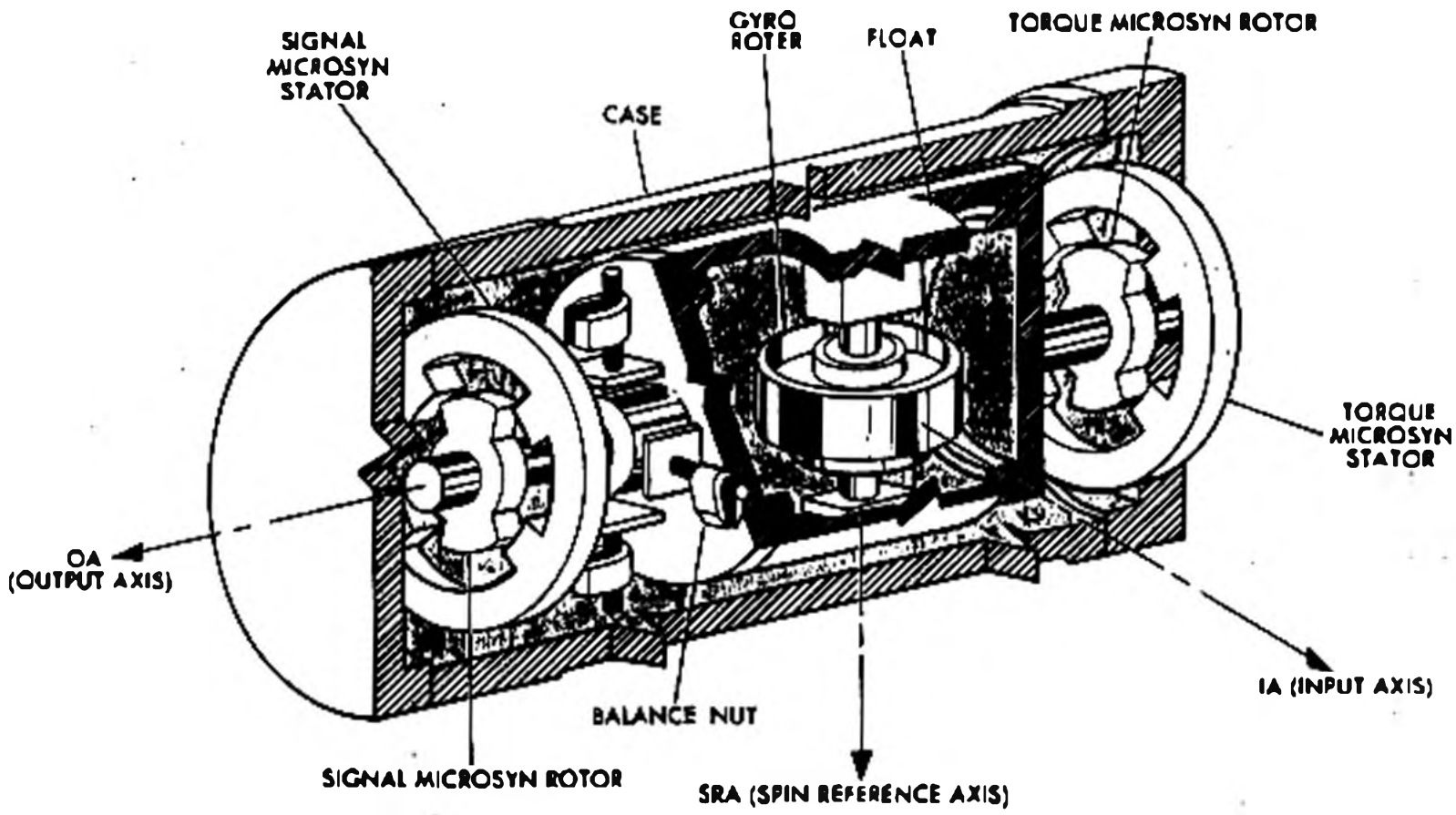


Figure 9-15. Gyro Cutaway

the resultant is zero, the missile will be on target as shown by the display meter.

An accelerometer and two memory devices provide the speed and distance traveled in the direction in which the accelerometer is pointed. In the normal flight of a missile, external forces can cause it to roll, pitch, and/or yaw. For the accelerometers to function properly, a means of keeping them properly oriented, regardless of missile angular motions, is required. In other words, the accelerometers must be stabilized. This is accomplished by mounting three gyroscopes on the same mechanical platform as the accelerometers.

GYROSCOPE. To use the gyroscopic principles in an inertial guidance system, the case (figure 9-15) is attached to a stable platform. Any rotation of this platform in the direction of axis IA will precess the gyro around axis OA. From an electrical pickoff system, a signal proportional to the precession (deviation of the platform from a given orientation) can be measured. The gyro is attached to a support frame (gimbal ring) that holds the platform. When the gyro spin rotor axis is normal to the stabilized platform, the wiper arm is in the center of the electrical pick-off. Any rotation about the input axis will result in the precession of the gyro. The signal from the pick-off is sent through the amplifier to the gimbal stabilizing motor, which turns the platform (gimbal) until the pick-off signal is nulled or zero. This action returns the platform to its original oriented axis.

The gyro serves the purpose of memorizing the original position of the platform on which it is mounted. In actual operations, the desired platform orientation (established before missile launch) automatically centers the gyro rotor for zero output. Since any platform rotation will precess the gyro and drive the gimbal motor to reposition the platform for a null signal, the stabilized platform will retain its oriented position.

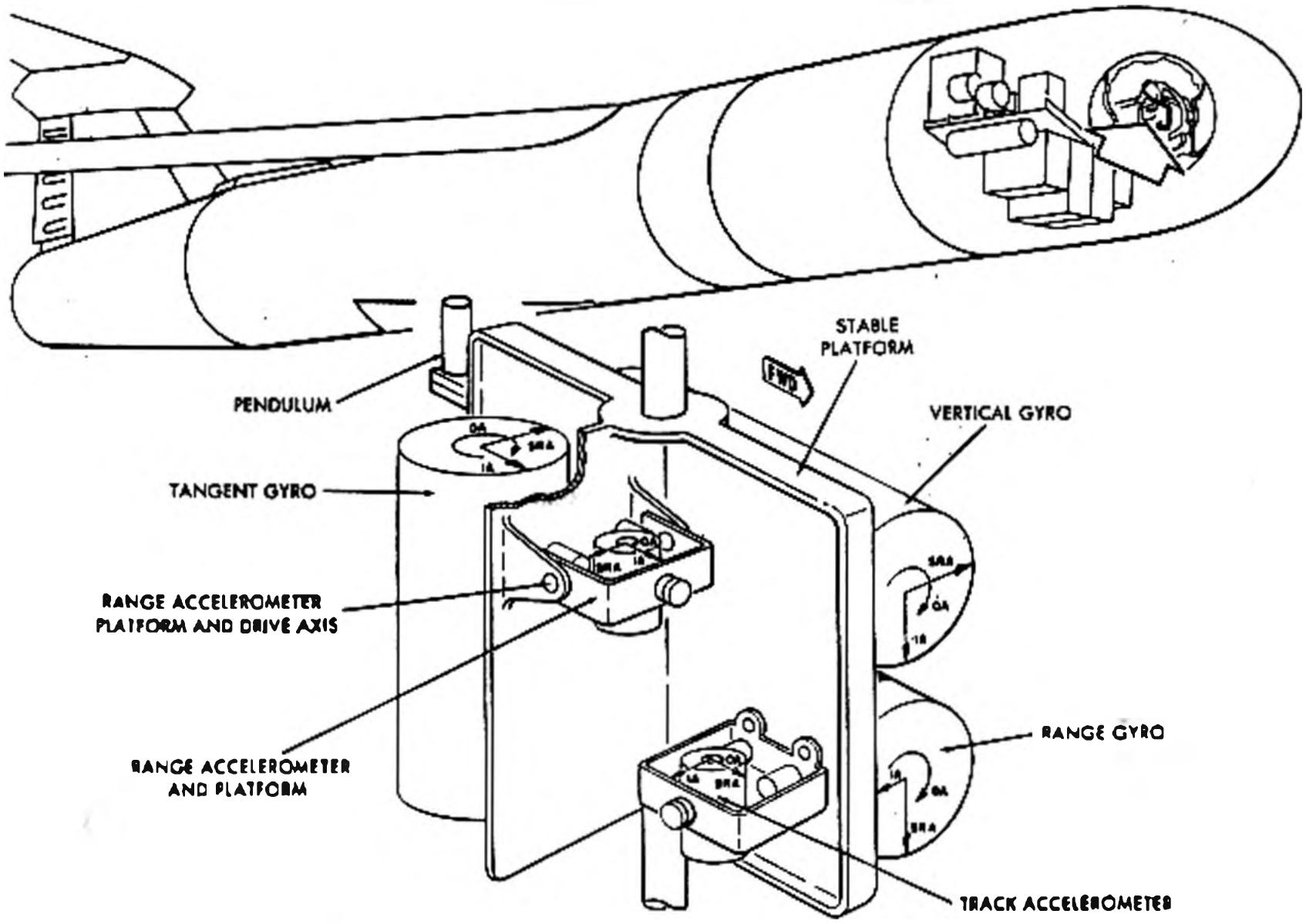


Figure 9-16. Stable Platform

A missile may perform three different angular motions; i.e., roll, pitch, and/or yaw, at the same time; therefore, it is necessary to use three gyroscopes. Each gyro memorizes a single axis in space and must be mounted on a common platform, which allows the missile to go through its normal angular motions with a minimum of disturbance to the gyro platform.

STABLE PLATFORM. For the missile guidance equipment to compute missile position, measurements must be performed with respect to a physical reference which is related to the guidance geometry. This physical reference is the stable platform. Before flight, the stable platform is oriented to the launch point vertical and the target azimuth, which jointly define the reference flight path. (See figure 9-16.)

The two accelerometers, range and track, are also mounted on the stable platform with their input axis aligned such that one senses accelerations in the range direction (along the flight path) and the other senses accelerations in the track direction (left or right of the flight path). The track accelerometer is mounted directly to the stable platform. The range accelerometer is mounted on the range accelerometer platform which is mounted to the stable platform through trunnions. The range accelerometer platform is driven in such a manner that the range accelerometer input axis remains perpendicular to the local vertical. Drive elements for the range accelerometer platform are also mounted on the stable platform.

Other functional assemblies on the stable platform are a two-degree-of-freedom pendulum, which provides erection signals, and an alignment mirror and drive, which are used during erection to orient the stable platform about its yaw axis.

MECHANICAL FILTER. The mechanical filter, shown in figure 9-17, is used to filter transient and quadrature voltages. Mechanical filters are used in both the range and track second integration loops. The mechanical filter is a heavily viscous-damped device, using a low-inertia, 2-phase motor. The motor drives a shaft which positions the rotor or a resolver.

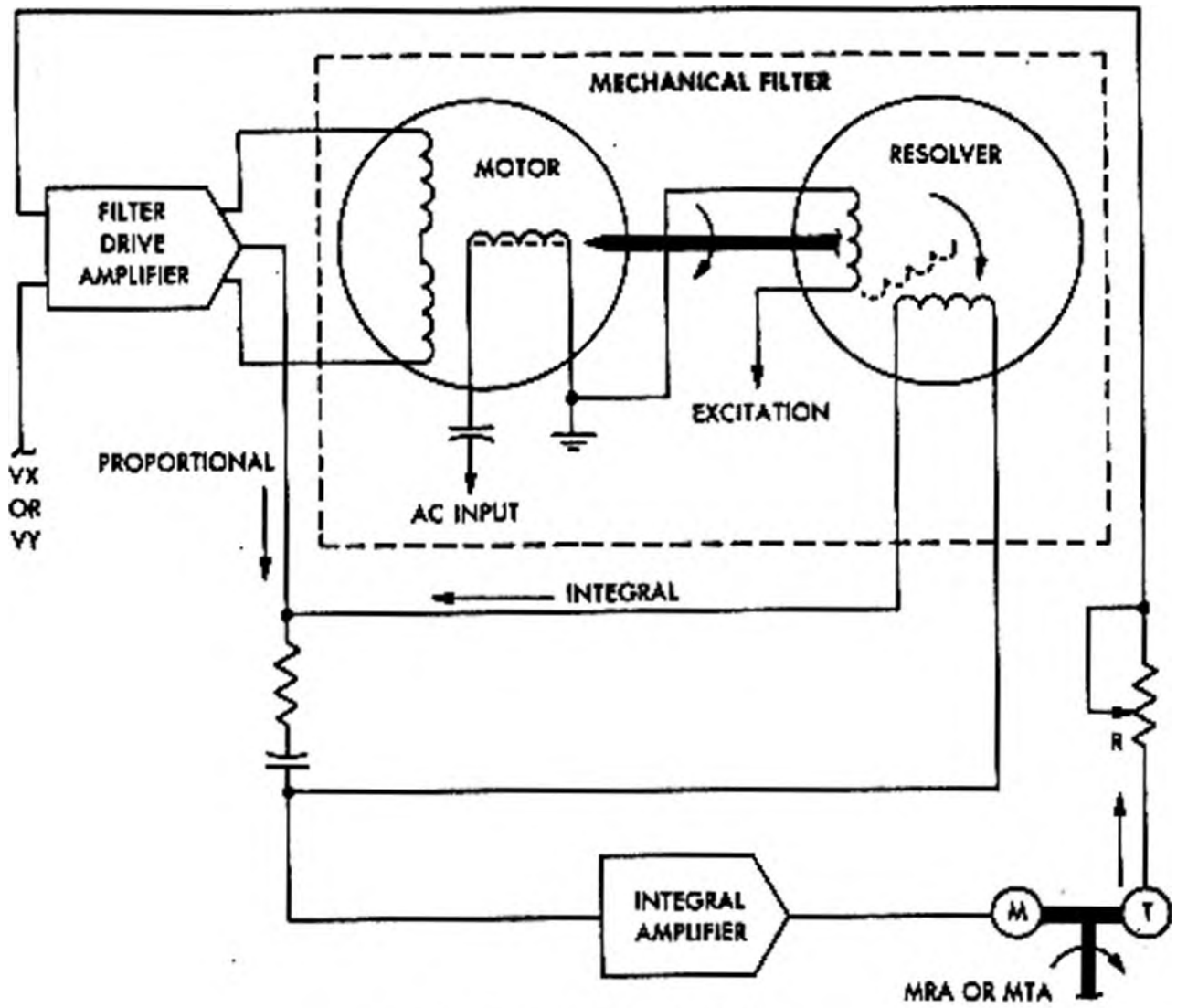


Figure 9-17. Mechanical Filter

When the velocity signal to the filter drive amplifier changes there are two outputs from the filter drive amplifier. One output is sent to the proportional channel and the other goes to the integral channel.

The proportional channel signal is applied across an RC network to the integral amplifier. The signal is then amplified and applied to the motor of a motor-tachometer group. The motor starts changing the rate of MTA (missile track angle) or MRA (missile range angle) shaft rotation, providing fast response to a velocity signal change.

The integral channel signal is applied to the motor of the mechanical filter. The motor is a 2-phase motor, which operates only when the input signal is 90° from the motor excitation. This eliminates quadrature voltages since they are either in phase or 180° out of phase with the motor excitation. The output of the motor is mechanically applied to the resolver by a shaft which positions the rotor of the resolver. The output of the resolver is applied to an RC network to compensate for the phase shift caused by the inductance of the resolver. The integral signal is first applied to the integral amplifier and then to the motor of the motor-tachometer group, resulting in a corrected MTA or MRA signal. Tachometer feedback is applied to the filter drive amplifier, canceling the input signal.

BRAKE-CLUTCH. The brake-clutch (figure 9-18) allows the memory potentiometer to be positioned during gyro erection and holds the potentiometer stationary after the missile is launched.

The input shaft of the brake-clutch is attached to a clutch plate. A second clutch plate and brake plate are mounted on the output shaft which passes through a retainer wall. Attached to the wall is a stationary brake plate. One end of a spring fits in a racetrack in the wall, and the other end is connected to the output shaft. A part of the output shaft is made of magnetic material which acts as a plunger in a coil. External to the brake-clutch assembly is a memory potentiometer which is connected to the output shaft.

During gyro erection, a DC voltage is applied to the coil, causing the output shaft to move to the left. This movement compresses the spring and engages the clutch plates. Rotation of the input shaft is then transmitted through the clutch plates to the memory potentiometer.

SCHULER TUNING. For the accelerometer to sense only horizontal accelerations, it is necessary at all times to have an exact indication of the local vertical or gravity vector. If not, a component of gravity is sensed as well as accelerations, and an error develops. To prevent errors in the computation of MRA caused by this gravity component, the range accelerometer platform must be driven through an angle equal to MRA and at a rate equal to the angular velocity of the missile over the surface of the earth. To accomplish this, the range accelerometer platform drive loop must have a definite response time. The response time of the loop must be equal to one-quarter of the time of a cycle of a pendulum having a length equal to the radius of the earth. A

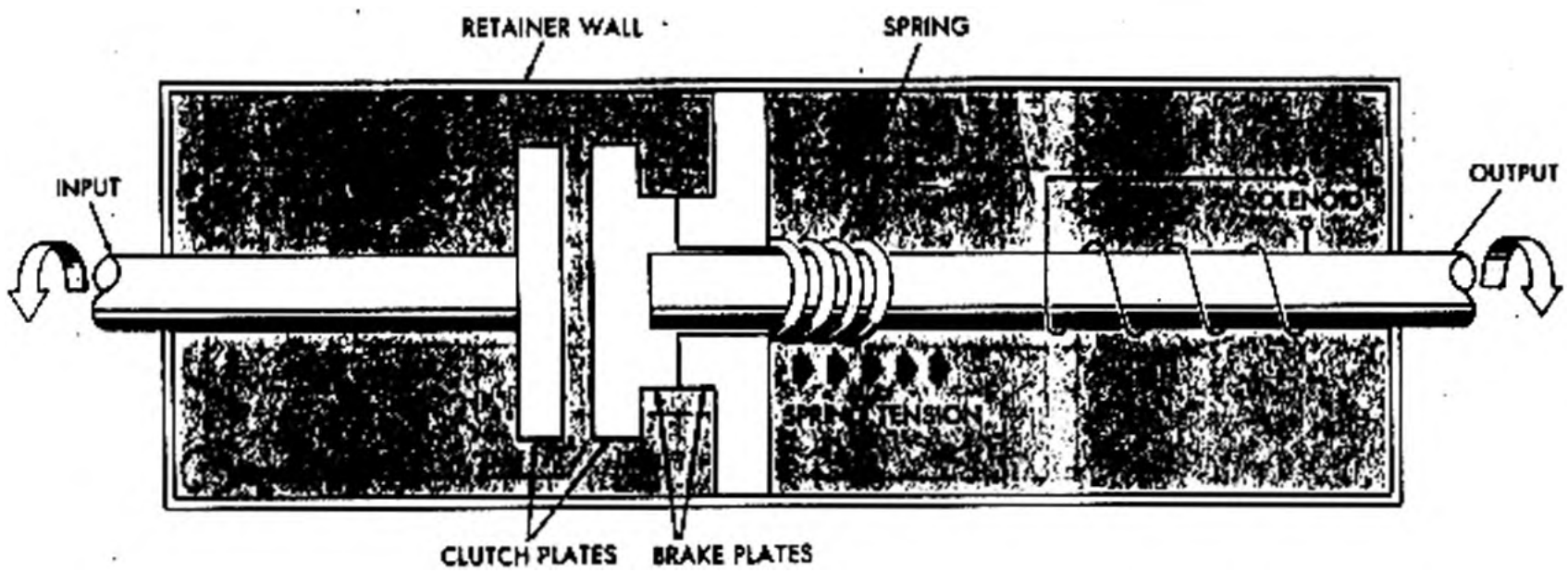


Figure 9-18. Brake Clutch

pendulum of this length will have a period of 84.4 minutes, or it would make one complete oscillation in 84.4 minutes. This system of tuning the loop to a definite response time is called "Schuler Tuning" in honor of Dr. M. Schuler, who discovered the principle.

GYRO-STABILIZED PLATFORM. The gyro-stabilized platform provides the reference from which all missile movement is measured. It provides a fixed launch point coordinate reference for the missile guidance set. The platform is erected and aligned prior to launch.

The operations related to the stable platform are alignment of the platform, memorization of the platform position, compensations, and sensing of deviations with the associated corrections. These operations are performed in the stabilized platform by its five types of functional loops and a mirror drive loop. A vertical sensing element and mirror drive loop are used to establish the orientation of the platform prior to launch. The stabilization loops maintain this orientation as a reference and the accelerometer loops measure horizontal accelerations of the missile with respect to this reference.

The stable platform consists of three gimbaled rings. The outer ring or pitch gimbal is mounted to the support gimbal which physically engages the housing of a measurement unit and is rigid with the frame of the missile. The middle ring or roll gimbal is mounted to the pitch ring. The inner ring or yaw gimbal is mounted to the roll gimbal. Mounted rigidly to the inner ring are three stabilization gyros and two accelerometer gyros.

At the end of each axis, mounted between pairs of rings, are data boxes which supply electrical information proportional to the angular relationship of the rings. At the other end of the axes, mounted between pairs of rings, are torque motors which control the angular relationship.

SIGNAL RESOLUTION. During missile flight, the stabilization gyros might not sense the full missile movement because of the misalignment of the gimbal and missile axes. Signal resolution resolves or corrects for gyro signals which do not represent a pure error for the particular missile movement. By resolving these gyro signals, proper platform stabilization is assured. The block diagram shown in figure 9-19 represents a typical inertial system and shows the various signal paths to control the missile in flight.

The range gyro senses missile movements about an axis perpendicular to the range plane. The tangent gyro senses missile movements about an axis tangent to the range plane at the launch point. The vertical gyro senses missile movements about an axis which is vertical or coincident to the launch point vertical. The outer gimbal mounting for the stable platform is the pitch gimbal, the middle mounting is the roll gimbal, and the inner mounting (or stable platform) is the yaw gimbal. Since all three stabilization gyros are mounted on the stable platform, a specific gyro input axis could become misaligned with respect to the platform gimbal axis and hence the missile axis.

The vertical gyro, being mounted on the yaw gimbal, cannot have its input axis misaligned with respect to the platform gimbal axis. No resolution is required for an error signal from this gyro. The tangent gyro axis can only be misaligned with the roll gimbal axis by a movement of the yaw gimbal. Signal resolution is required about or with respect to the yaw angle for an error signal from the tangent gyro. The range gyro axis can be misaligned with the pitch gimbal by a movement of either the roll or yaw gimbal. Signal resolution is required about or with respect to both the roll angle and the yaw angle for an error signal from the range gyro.

One way to determine when resolution is required is to remember the platform gimbal orientation. The pitch gimbal is about the roll and yaw gimbals, so pitch (range gyro error signal) must be resolved about both roll and yaw. The roll gimbal is about the yaw gimbal only, so roll (tangent gyro error signal) must be resolved about yaw. The yaw gimbal is about no other gimbal, so no resolution is required for yaw.

In determining what the output will be from one of the gyros whose input axis can be displaced from the platform gimbal axis, it must be understood that the input axis will be displaced by an angular movement. When the input axes of the gyros are not displaced, they will each sense a particular movement of the missile; i.e., range gyro-pitch; tangent gyro-roll; and vertical gyro-yaw. For example, if a missile yaws 90°, the range gyro will sense no pitch but all of the roll, while the tangent gyro will sense no roll but all of the

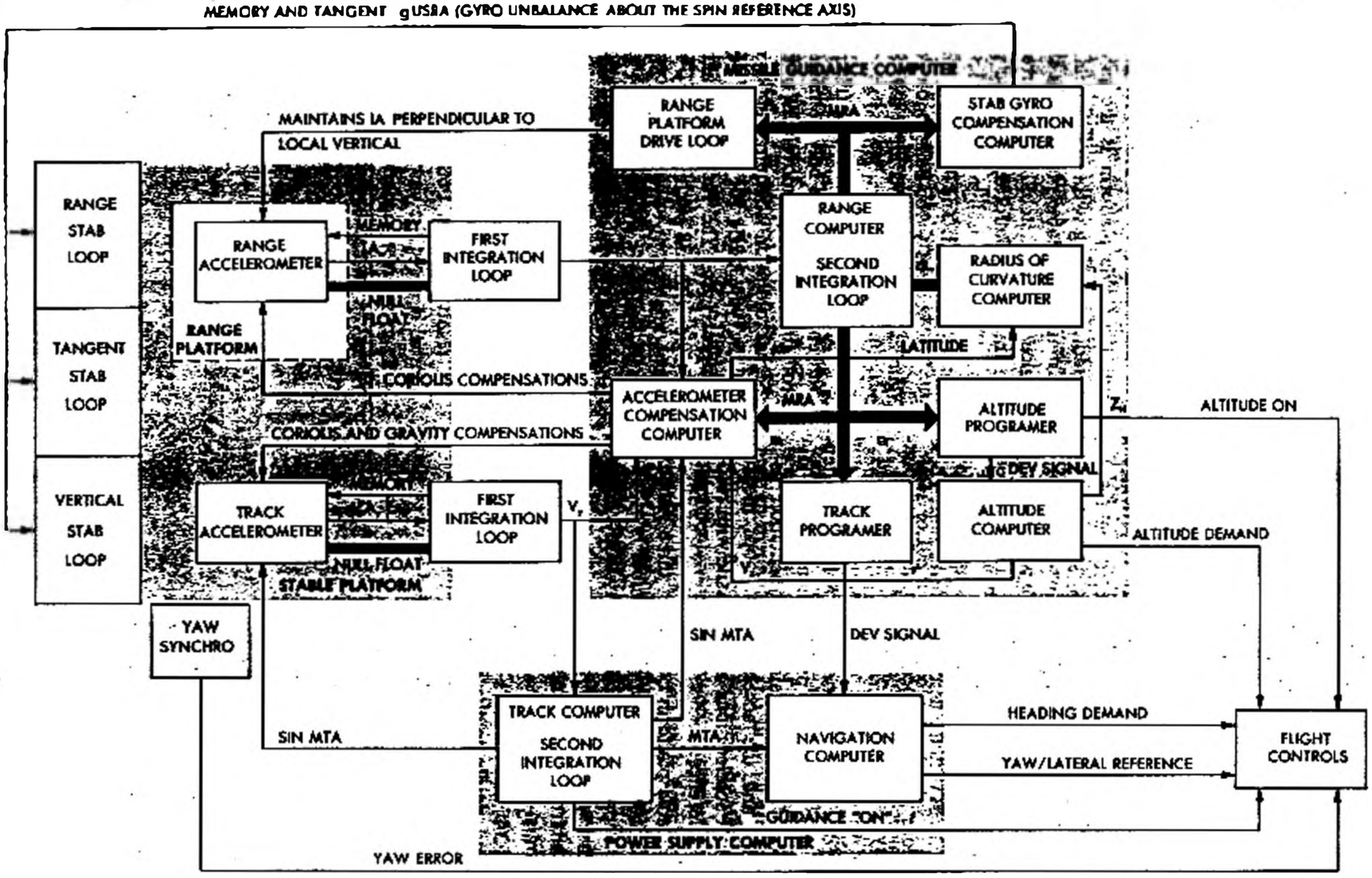


Figure 9-19. Block Diagram-Inertial Guidance System

pitch. Two rules for resolution about yaw may be formulated:

- The gyro which would normally sense the error will sense (or give as an output) the cosine of the yaw angle times the amount of error. Example: When a yaw angle exists and the missile pitches, the output from the range gyro will be pitch times the cosine of the yaw angle.
- The gyro which would not normally sense the error will sense (or give as an output) the sine of the yaw angle times the amount of error. Example: When a yaw angle exists and the missile pitches, the output from the tangent gyro will be pitch times the sine of the yaw angle.

Resolution of pitch or roll about yaw may be developed as follows (figure 9-20):

The output of the range gyro when a yaw angle exists is

Pitch × cos yaw angle (1)

The output of the tangent gyro when a yaw angle exists is

Pitch × sin yaw angle (2)

The signal induced from P1 to S1 is

Pitch × cos yaw angle × cos yaw angle
= pitch × $\cos^2$ yaw angle (3)

The signal induced from P2 to S1 is

Pitch × sin yaw angle × sin yaw angle
= pitch × $\sin^2$ yaw angle (4)

Adding (3) and (4)

Signal from S1 = (pitch × $\cos^2$ yaw angle)
+ (pitch × $\sin^2$ yaw angle)
= pitch ($\cos^2$ yaw angle
+ $\sin^2$ yaw angle)

From trigonometry: $\cos^2 + \sin^2 = 1$

Therefore, the signal from S1 to the pitch gimbal torquer motor = pitch × 1 or the full pitch error.

The signal induced from P2 to S2 is

Pitch × sin yaw angle × cos yaw angle (5)

The signal induced from P1 to S2 is

−Pitch × cos yaw angle × sin yaw angle
(negative value) (6)

Adding (5) and (6)

Signal from S2 = (pitch × sin yaw angle
× cos yaw angle) (7)
+ (−pitch × cos yaw angle
× sin yaw angle) (8)
= 0 (9)

Therefore, the signal from S2 to the roll gimbal torquer motor is zero.

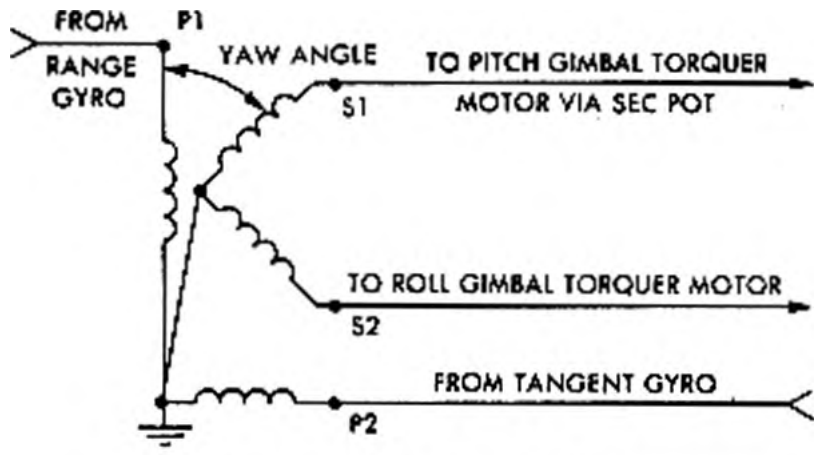


Figure 9-20. Yaw Resolver

To illustrate the resolution of pitch about yaw, assume the following conditions: (a) yaw angle = 30° and (b) the missile pitches with a force = 10v.

The output of the range gyro is

Pitch × cos yaw angle = 10 × cos 30° (1)

The output of the tangent gyro is

Pitch × sin yaw angle = 10 × sin 30° (2)

After resolution

Output = pitch × $\sin^2$ yaw angle + pitch ·
× $\cos^2$ yaw angle (3)
= 10($\sin^2$ 30° + $\cos^2$ 30°)
= 10($0.5^2 + 0.866^2$) (4)
= 10(0.25 + 0.75) (5)
= 10v (6)

Therefore, the output of the resolver is the full amount of pitch. Resolution of roll about yaw can be done in the same manner.

The resolution of pitch about roll is performed by the secant pot which is positioned by the platform roll gimbal. If the platform roll gimbal becomes displaced because of a roll angle, the range gyro will not sense the full amount of the pitch movement; or

Output of the range gyro, when a roll angle exists, is

Pitch × cos roll angle (1)

This signal is multiplied in the secant pot; or

Pitch × cos roll angle × sec roll angle (2)

From trigonometry: The secant of an angle is the reciprocal of the cosine, or

$$\text{Output} = \text{pitch} \times \cos \text{ roll angle} \times \frac{1}{\cos \text{ roll angle}} = \text{pitch} \quad (3)$$

As a result, the pure pitch error signal is used to reposition the pitch gimbal, keeping the stable platform at the launch point orientation.

By resolving the outputs of the range and tangent gyros, when necessary, it is assured that the stable platform will retain the launch point orientation.

The basic rules of signal resolution can be summarized as follows:

- Pitch is resolved about roll and yaw.
- Roll is resolved about yaw.
- Yaw is not resolved.
- Phase output of a given pitch will remain the same for either direction of yaw.
- Phase output of a given roll will remain the same for either direction of yaw.

ACCELEROMETER SERVO LOOPS. The outputs of the range and track accelerometers provide the signals for inputs to the range and track computers and pulling loops. These computers use these inputs to determine such parameters as missile position, heading, velocity, and distance.

Since the range and track accelerometer servo loops are similar in operation, only the range loop is shown (figure 9-21). The accelerometer

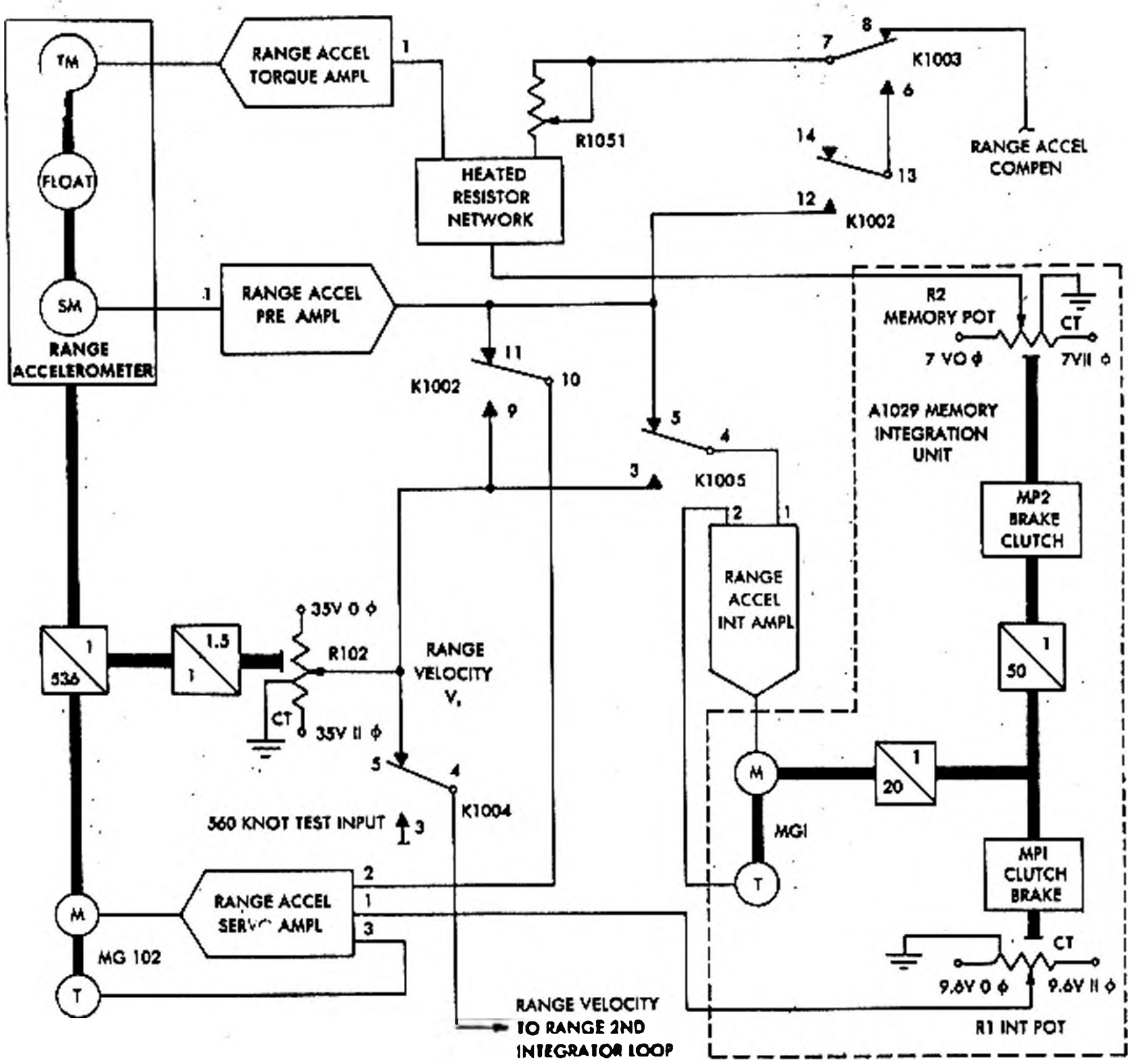


Figure 9-21. Range Accelerometer Servo Loop

servo loops, known as the first integration loops, have three functions:

1. They memorize the drift of the accelerometers.

2. They integrate acceleration, Ax or Ay, with respect to time to compute velocity, Vx or Vy.

3. They generate a signal which will null the float of the accelerometers.

Memorization is accomplished during checkout. Since there is no acceleration on the accelerometer, any output of the signal microsyn will be caused by drifting of the gyro. The output, after preamplification, is sent to the range accelerometer integral amplifier. This output will drive the motor tachometer in the memory integration unit. The motor will position the memory pot through the brake-clutch (which is acting as a clutch). The output of the memory pot will be sent through the heated resistor network and the range accelerometer torque microsyn. This will cause the gyro to precess an amount equal to the drift rate of the gyro. Memory is completed which deactivates the brake-clutch and locks the memory pot at its last position, applying a constant voltage for use as memory.

The *integration* of Ax to get Vx is accomplished by the simultaneous operation of two channels: proportional and integral. The proportional channel provides the fast response necessary to assure accurate measurement of acceleration. The integral channel provides the accuracy and maintains the null float signal. (The integral channel senses the changes in acceleration to be summed and stored so that the acceleration can be remembered even though the float is nulled.) The proportional channel is the straight output from the pre-amplifier to the range accelerometer servo-amplifier. It provides immediate action for any output from the accelerometer which would indicate a change in the acceleration. The range accelerometer servoamplifier and the associated motor tachometer is the point at which the actual integration of Ax to Vx occurs. The proportional input to this integration enables the circuit to respond faster and thus have a more accurate Vx. The mechanical output from the motor is converted to an electrical Vx by the shaft position of R102 which is sent to the range computer or the second integral loop. The integral (summing) channel uses part of the same circuit that was used for memorization. The output from the preamplifier, besides being sent as a proportional signal, is sent to the range accelerometer integral amplifier, causing the motor tachometer to drive the integral pot through the clutch-brake which acts as a clutch during flight. The output of the integral pot is applied to the servoamplifier along with any proportional signal present.

Nulling the float or turning the case to precess the accelerometer to null is accomplished by the motor tachometer and shaft which computed Vx. This shaft will turn the accelerometer case end-over-end about the input axis which gyroscopically precesses the accelerometer to null. The accelerometer must always be nulled after sensing acceleration because the force of acceleration felt at an angle; that is, other than at 90° (nulled), will not be a true force of acceleration. When the acceleration is changing, there will be a proportional signal aiding the integral (summing) signal. The integral channel sums Ax, allowing it to become constant and has no proportional signal output.

Refer again to the range accelerometer servo loop shown in figure 9-12. Accelerations displace the float of the range accelerometer. This float displacement develops a signal in the signal microsyn. For a positive acceleration, the output is zero phase. For a negative acceleration, the output is pi-phase. The output of the signal microsyn is applied to the range accelerometer preamplifier. The output of the preamplifier is applied at two places: (1) to the proportional channel for fast dynamic response to accelerations and (2) to the integral channel for the summation of acceleration changes.

When subjected to a positive constant acceleration, the action of the proportional channel is as follows: The proportional channel signal is applied through relay K1002 and to the range accelerometer servoamplifier. The output of the servoamplifier is applied to the motor-generator set MG102. Tachometer feedback is applied to the servoamplifier for stability. The motor drives a shaft that performs two functions: (1) it positions the velocity potentiometer R102 and (2) it turns the accelerometer case end-over-end about the input axis. This rotation of the accelerometer applies a force to the gyro which precesses the wheel mass and the float toward the null condi-

tion. As the float approaches null, the output from the signal microsyn decreases. The output of the preamplifier decreases proportionally. MG102 drives at a slower rate, causing the accelerometer cases to turn end-over-end at a reduced rate. At this point, the float is displaced away from the null condition once again, and the cycle of proportional channel servo loop action is repeated.

The action of the integral channel for the conditions of positive constant acceleration is as follows: The output of the preamplifier is applied to the integral channel through relay K1005 as an input to the range accelerometer integral amplifier. This output is fed to the motor-generator group MG1 in the memory integration unit A1029. Generator feedback is applied to the integral amplifier for stability. The motor drives a shaft, positioning the integral potentiometer R1 in the memory integration unit. The signal taken from R1 is proportional to acceleration and is applied to the servoamplifier. In conjunction with the proportional channel, the output from R1 continues positioning R102 at an increasing rate and turning the accelerometer case end-over-end until the float is nulled. After the float is nulled, the output of the signal microsyn will be zero. The integral channel continues turning the accelerometer case at a constant rate, keeping the float nulled as long as the acceleration remains constant. The output of R102 continues to increase at a rate proportional to acceleration. The output of R102, range velocity, is fed to the second integration loop.

During alignment and erection of the stable platform, the memory potentiometer R2 in the memory integration unit is positioned to store earth rate and gyro unbalances. During flight, R2 is held fixed by brake-clutch MP2. The output of R2 is sent to the heated resistor network, where it is combined with the range accelerometer compensations. The signal from the network is sent to the range accelerometer torque amplifier, where it is amplified and sent to the torque microsyn. The torque microsyn then turns the float at a rate to correct for gyro unbalances, earth rate, and Coriolis.

The track accelerometer servo loop operates the same way as the range loop except for the compensations. Both accelerometers receive memory, vertical, and horizontal Coriolis compensations. The track accelerometer receives, in addition, on- and off-course gravity compensations and sin MTA. The relays in both loops are used during checkout to reset the velocity potentiometers and the integral channels.

Computers and Programmers

Computers and programmers constitute the "brains" of the inertial guidance system. These units compute elapsed distance and guidance system corrections and decide when the missile should begin to assume a new course or altitude flight path. The system must contain computer circuits to resolve the navigation problems.

COMPUTER CIRCUITS. The term analog, used many times in reference to the inertial guidance (IG) system computers, means the system is a mechanism which operates to solve mathematical problems involving certain quantities in terms of other quantities. For example, the guidance system performs the job of computing missile course, including the heading, altitude, and distance information. These computations are not in terms of thousands of feet, or miles per hour, or degrees of magnetic heading, but rather in terms of electrical signals with different phases, polarities, or amplitudes.

Once appropriate data is translated into electromechanical terms, the system can perform a wide variety of operations. Addition, subtraction, multiplication, integration, and differentiation are representative of the computations that are necessary to generate accurate navigational information. The computations are completed almost instantaneously, even as the input quantities are changing.

The system receives its information in the form of shaft rotations and potentiometer and amplifier voltages. The individual circuits make up the computers in the IG system and are designed to perform one or more of three general functions: mixing, integration, and rate control.

Mixing. Often two or more sources of information must be considered in computing a desired output. Whether the output is to represent the sum or difference of the separate units, a common reference must be available to bring together the different points. The most common means of accomplishing mixing action is through the use of parallel resistor or amplifier networks.

Integration. One definition of the term integrate is "To indicate the whole of, to give the sum or total of." For this purpose, the definition may be expanded to read: The integral or sum total effect of a quantity is proportional to input amplitude multiplied by the time the input exists.

Applying this definition to missile quantities, the integral of missile acceleration at a rate of 5 miles per hour per second (assuming zero velocity at the start) for a duration of 5 seconds equals a velocity of 25 miles per hour. A velocity of 25 mph is the integral or total effect of that rate of acceleration lasting that period of time.

To put the above problem in computer terms, an electrical voltage representing the rate of acceleration is applied to a variable speed motor. A motor is inherently an integrating device. The higher the input signal, the faster its output shaft turns per unit of time. The longer a given input is applied, the more total turns its shaft will make before stopping.

Assume that an integrator is a motor which turns exactly 10 revolutions with an input signal of 1 volt lasting 1 second. An input of 5 volts for 5 seconds represents the missile acceleration conditions. The motor should make 250 revolutions before stopping. The output of an integrator is directly proportional to the product of the amount of the quantity to be integrated and the length of time that particular amount is present. If the input is of a small constant value lasting for an extremely long period of time, then the integral output will steadily increase at a constant rate during the entire period that the input is present.

Another basic characteristic of an integrator is linearity. This means that if a motor-integrator drives a total of 10 turns with an input of 1 volt, then it should drive 20 turns for 2 volts, 30 turns for 3 volts, and so on. This linearity of response can be accomplished by the addition of an amplifier and a tachometer-generator.

The amplifier increases the amplitude of the input, causing the motor to drive at a faster rate. The tach-generator produces an output voltage which is proportional in amplitude to the rate of drive of the motor shaft. The faster the motor drives, the larger is the tach-generator output. If the motor is not driving at all, then the tach does not produce an output.

A feedback path is provided for the tach-generator output back to the amplifier input. As the input signal increases, the motor drives faster. The tach-generator feedback also increases, opposing the increase. With carefully matched amplifiers, motors, and tach-generators, an extremely accurate linear integrator can be made (figure 9-22).

Rate control. The tachometer-generator and resistive feedback path to the input of the integrator amplifier serve to perform the task of controlling the rate of response of the integrator to a given input. By proper adjustment of the rate-controlling elements in the IG system, the linearity and accuracy of the computer circuits can be controlled. Rate controlling circuitry most often takes the form of attenuating resistances in amplifier feedback and signal input paths (figure 9-23).

PROGRAMMERS. Basically, a programmer is a form of timer, or more specifically, it can be described as a time-delay device used to control a

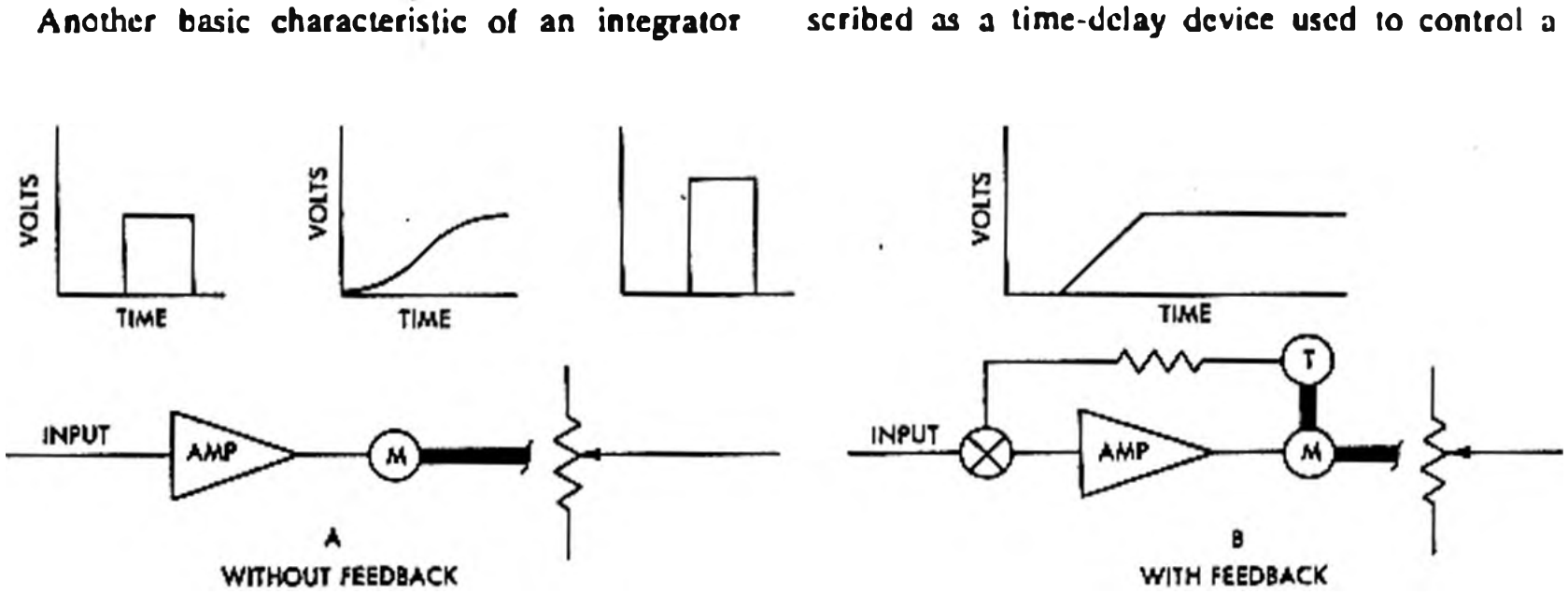


Figure 9-22. Linear Integrator

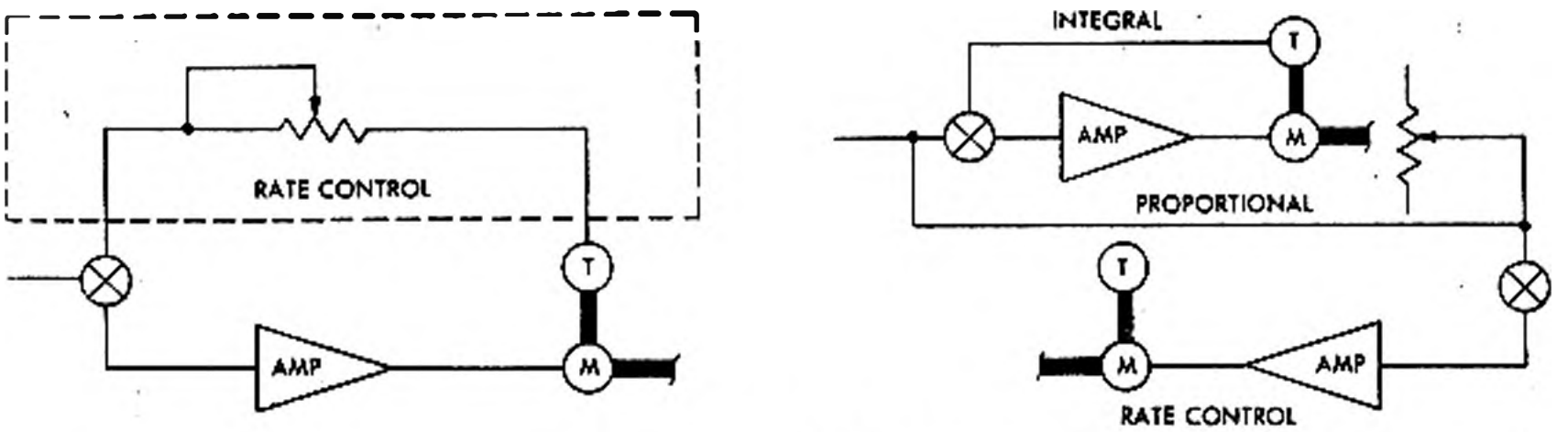


Figure 9-23. Rate Control

specific function during a predetermined time period or condition. The word condition is important because it describes the principle of operation of the programmers in the IG system. These components control the starting and stopping of certain inflight guidance functions—not in terms of clock time but in terms of distance the missile has flown since leaving the launch point.

Both altitude and track deviations, in addition to other functions, are initiated by the track and altitude programmers which must carry out their missions without benefit of any outside timing or triggering signals. They are surprisingly simple in construction and operation. The guidance system programmers provide a means of steering commands concerning variations in flight profile and certain other functions such as warhead arming, guidance monitor, and limited range-to-go signals.

A guidance programmer command could be as follows: "Fly 400 miles down the range plane; then go to a new course 20 miles to the left." Notice that this command contains two pieces of information: (1) when to change course and (2) where the new course is to be. In other words, the time at which the deviation is to begin is in terms of range distance.

A very simple mechanism can be constructed by mounting a movable cam on the output shaft of the range distance integrator. The cam being movable can be offset from its zero position to different values of shaft rotations or range distance.

The distance the missile should travel after it changes course can be set in by an electromechanical analog computer which uses electrical signals to represent quantities in the guidance problem. A voltage taken from a potentiometer can be used to represent the value of deviation the missile is to make. Through the use of a manually variable potentiometer, the desired value of deviation (figure 9-24) can be set in the system.

The arrangement for altitude deviations is identical to the one just described except that the output of the programmer is routed to the climb-dive section of flight controls rather than to the

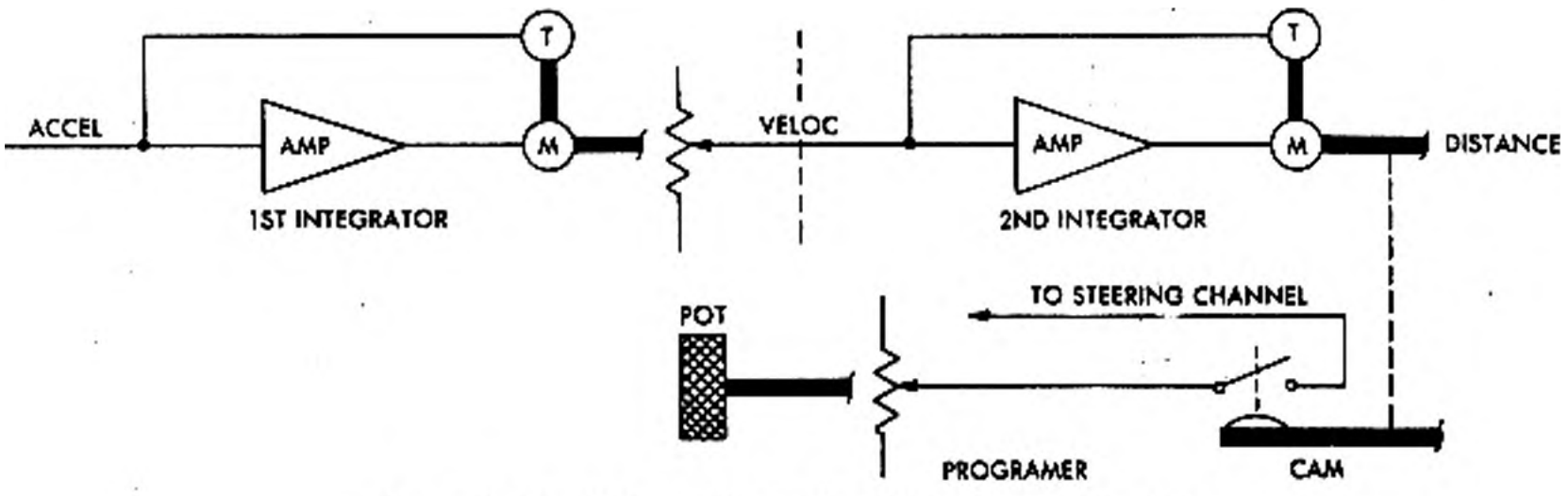


Figure 9-24. Distance Channel and Programmer Circuitry

steering section. The initiation of altitude deviations is also timed in terms of range distance.

The examples shown in figure 9-24 apply the same basic principles as those found in the track and altitude programmers of the guidance system. The actual circuits appear more complex because

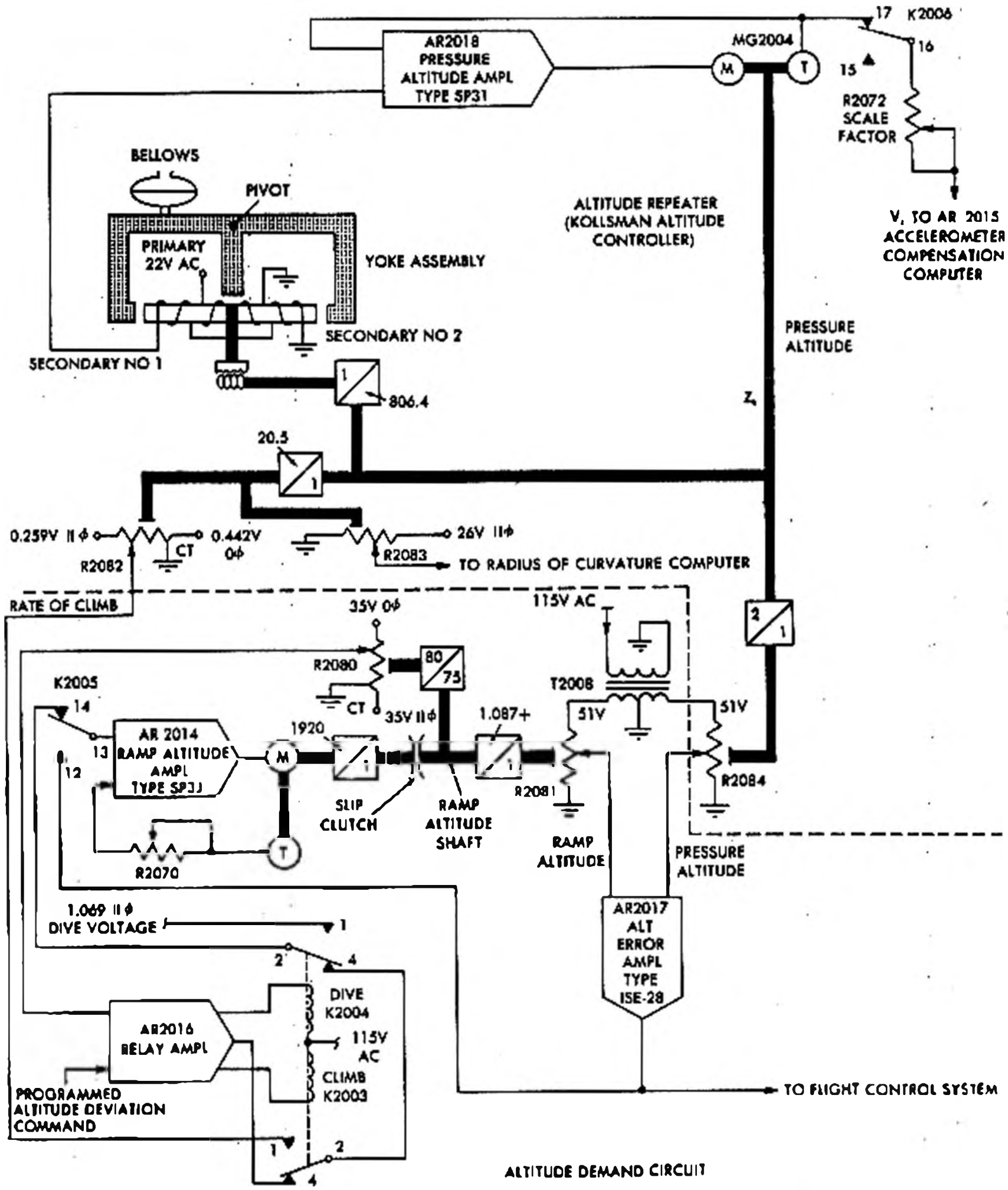


Figure 9-25. Altitude Computer

they must allow programming for more than one track or altitude deviation. Each is composed of a series of duplications of a very simple circuit.

ALTITUDE COMPUTER. The altitude computer shown in figure 9-25 consists of two sections: the altitude repeater circuit and the altitude demand circuit. The inputs to the computer are pressure altitude to the altimeter and programmed altitude deviation signals from the altitude programmer. The outputs are altitude demand signals to flight controls, pressure altitude (Zh) to the radius of curvature computer, and vertical velocity to the accelerometer compensation computer. The altitude computer operates as a difference computer. It compares the desired altitude with the actual altitude to provide error (demand) signals to the flight controls.

The altitude repeater circuit contains an altimeter which measures the rate of change of altitude and sends the signal to an integrator circuit. This gives Zh as an output from the motor and Vz as an output from the tachometer. The output from the motor portion of the integrator positions three potentiometers (R2082, R2083, and R2084) and also nulls the altimeter as a function of altitude so it can sense further changes in altitude.

If the missile is on altitude control, a difference between ramp altitude (R2081) and Zh (R2084) will be measured by the altitude error amplifier and sent to flight controls as an error (demand) signal. If the missile is on air-speed control, the output difference of the amplifier is fed back into the ramp altitude circuit to keep the ramp circuit matched with Zh. As the missile goes on altitude control, the ramp circuit is set up and ready to accept altitude deviation signals.

The altitude demand circuit operates as a summation network when the missile is on altitude control. Programmed altitude deviation commands from the altitude programmer are compared with the voltage from the ramp altitude feedback, R2080, in the relay amplifier. When there is a difference, the climb or dive relay is energized and the specific signal is sent to the ramp altitude circuit. This signal positions R2081, and R2084 is measured in the altitude error amplifier and sent to flight controls. When the missile reaches the desired altitude, Zh and ramp altitude are again the same. No error signal is sent to flight controls until a different command is received from the altitude programmer or the missile unintentionally changes altitude.

The sensing device in the altitude computer shown in figure 9-25 is the Kollsman altitude controller which has two secondary windings wound on a 2-section core separated by a liber insulator. Secondaries 1 and 2 are wound series opposing. The primary winding is wound over both sections of the core. In a nulled condition the core is centered between the ends of the yoke assembly so the couplings between the primary and each of the secondaries are equal, and the coupled signals are of opposite phase. The input to pressure altitude amplifier AR2018 is then zero. Assume that air pressure decreases (missile climbs). This causes a bellows in the altitude computer to expand. The expanded bellows makes the yoke assembly pivot, decreasing the air gap between the yoke and core No. 1 and increasing the air gap between the yoke and core No. 2. The differential in air gap results in a greater coupling of energy into secondary No. 1, causing a pi-phase signal to be applied to pressure altitude amplifier AR2018 through normally closed contacts 8 and 7 of relay K2006. The signal is amplified and applied to motor-tachometer MG2004, and the motor drives. Tachometer feedback, which is proportional to the rate of change in pressure, is applied to amplifier AR2018 for stability and also is applied through scale factor potentiometer R2072 to the accelerometer compensation computer as a vertical velocity signal. The motor drives the pressure altitude shaft which positions three potentiometers:

- R2083—supplies a pressure altitude signal for the radius-of-curvature computer.
- R2082—supplies a rate-of-climb signal for the altitude demand circuit.
- R2084—supplies a pressure altitude signal for the altitude demand circuit.

To put the system in a constant airspeed flight condition, relay K2005 is energized by the ground support equipment prior to flight. During the constant airspeed portion of the flight, the altitude demand circuit positions the wiper arm of potentiometer R2080 so the potentiometer output always represents missile altitude. The output of potentiometer R2084 of the altitude repeater is applied to altitude error amplifier AR2017. Any error signal developed by amplifier AR2017 is

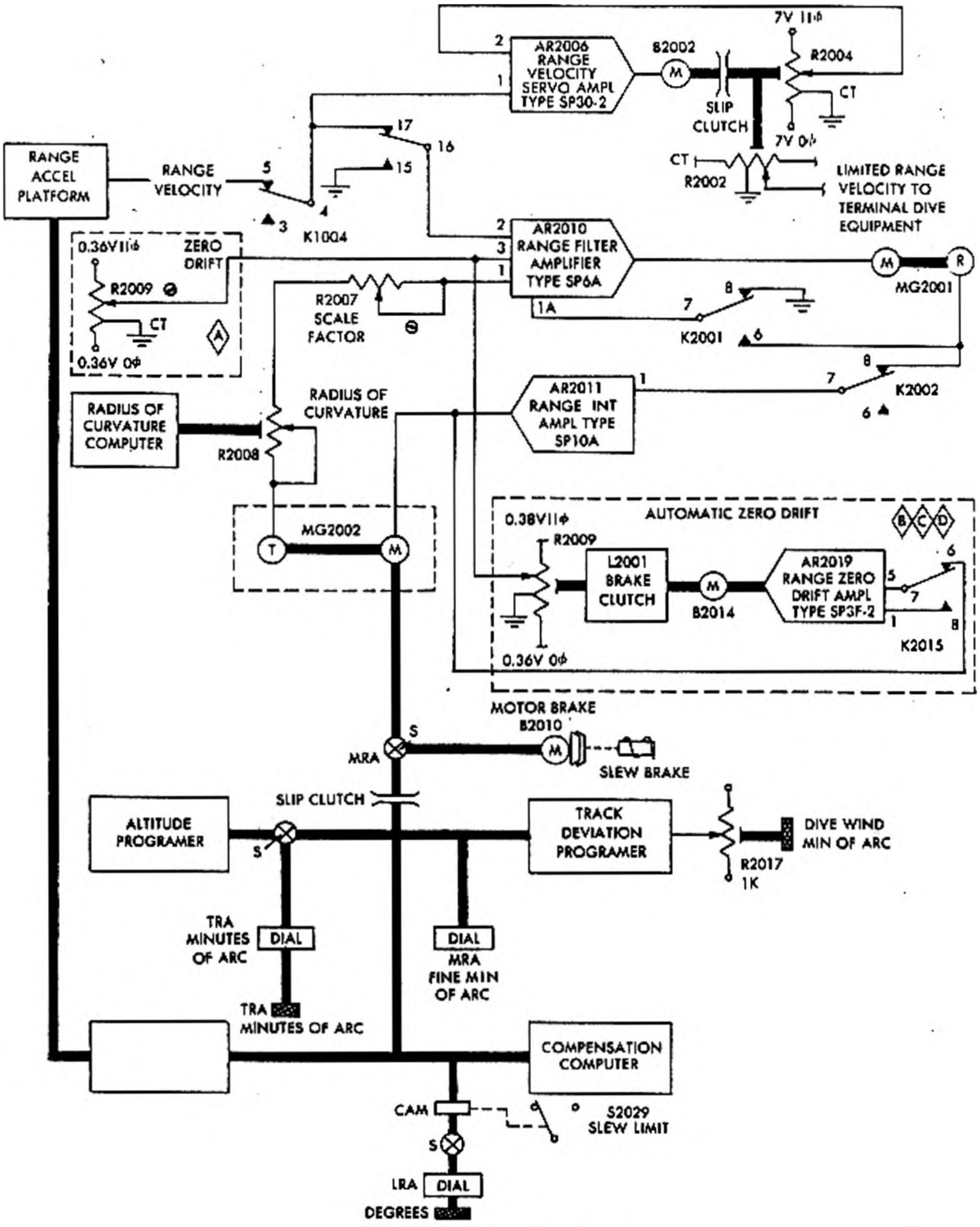


Figure 9-26. Range Computer

applied through energized relay K2005 to ramp altitude amplifier AR2014, where it is amplified and applied to motor-tachometer MG2003. Motor-tachometer MG2003 positions the wiper arms of potentiometers R2081 and 2080. The output of potentiometer R2081 is applied to amplifier AR2017. The wiper arm of potentiometer R2081 is driven until its output is equal to, but opposite in phase to, the signal from potentiometer R2084. With this condition, the output of amplifier AR2017 is at null. When potentiometer R2081 was positioned, potentiometer R2080 was also positioned. The signal from potentiometer R2080 is applied to relay amplifier AR2016. Since relay K2005 is energized in a constant airspeed flight, the output of amplifier AR2016 is not applied to amplifier AR2014.

Before a programmed altitude deviation command can be used, certain conditions must be met. The traveling cam nut must have reached the setting to initiate the altitude selector switch. When this position is reached, the holding path for relay K2005 is broken, allowing the relay to deenergize. When relay K2005 is deenergized, deviation command signals can be applied to amplifier AR2016 for comparison with altitude signals from potentiometer R2080, and resultant climb or dive signals are routed through relay K2005 to amplifier AR2014. Assume that the missile is flying at 40,000 feet when the first altitude deviation command is given. The command signal from the altitude programmer calls for an altitude of 20,000 feet, and this command is applied through relays K2006 and K2005 to amplifier AR2016. Amplifier AR2016 compares the command signal with the existing missile altitude signal from potentiometer R2080. The output of amplifier AR2016 energizes dive relay K2004, and a dive signal is applied through relay K2004 and relay K2005 to ramp altitude amplifier AR2014. The output of amplifier AR2014 drives motor-tachometer MG2003, driving a shaft which positions potentiometers R2080 and R2081. The output from potentiometer R2081, which is now different from the output of R2084, is applied to altitude error amplifier AR2017. As a result, a 0-phase error signal is developed in amplifier AR2017 and sent to the flight control system, commanding the missile to dive. The dive rate is controlled by a limiting circuit in amplifier AR2016. Since the rate of rotation of potentiometer R2081 is nearly equal to the rate of rotation of potentiometer R2084, the output of amplifier AR2017 is nearly constant. The missile continues to dive until the output from potentiometer R2080 is equal in magnitude, and opposite in phase, to the command signal from the altitude programmer. Potentiometer R2081 output becomes constant and R2084 output becomes equal in magnitude and opposite in phase to the R2081 output. At this time there is no output from amplifier AR2017, and the missile levels off at 20,000 feet.

RANGE COMPUTER. The range computer, known as the second integration loop, is used to integrate range velocity to determine the range distance the missile has traveled. The range computer operates as an electro-mechanical circuit to compute MRA (missile range angle).

The range computer (figure 9-26) is a standard integration circuit (with modifications) to insure the accurate measurement of MRA. The inputs to the range computer are Vx from the range first integration loop and a mechanical input from the radius of curvature computer. The MRA mechanical output is sent to the (a) range platform drive loop, (b) track programmer, (c) stabilization gyro compensation computer, (d) accelerometer compensation computer, and (e) altitude programmer. The altitude programmer is initially positioned by a data set-in of TRA through a differential. MRA into the programmer through the same differential effectively subtracts from TRA. During flight, the position of the programmer at any given time is RTG (range to go), or as pointed out previously: RTG = TRA – MRA.

The input Vx is applied to AR2010 and a mechanical filter MG2001 which will remove noise and quadrature components from Vx. This filtered Vx is integrated in AR2011 and MG2002. The output of the motor portion of the integrator is MRA, while the output of the tach-generator is velocity. The tach-generator generates a voltage proportional to the rate of change of the input, MRA. This is, by definition, velocity, since V = time rate of change of distance, or Vx. The amount of negative Vx feedback from the tach-generator is initially set up during ground testing to provide for the correct Schuler tuning rate.

Since the input Vx has not been compensated

either for altitude variations or for the changing curvature of the earth, this must be done or MRA will not be equivalent to the actual surface distance that has been traveled. This is accomplished by the radius of curvature computer which sees altitude and latitude as inputs. The outputs adjust the amount of feedback Vx, insuring that MRA will be correct. This correction of the feedback Vx amounts to integrating Vx with respect to time and dividing by radius of curvature to get surface MRA, or

$$MRA = \frac{\int_{t=0}^{t=n} Vx\,dt}{\text{radius of curvature}}$$

During ground testing, adjustment must be made not only in case of circuit drift but also for gyro drift. This is accomplished by the automatic zero drift circuit which uses the output of AR2011 after the input to the range computer has been grounded by energizing K2001. If there is no output of AR2011 with input grounded, no correction is necessary. Any output of AR2011 is sent through energized K2013 to range zero drift amplifier, AR2019. This causes motor, B2014, to position R2009 through the brake-clutch, L2001. The output of the zero drift circuit provides the necessary negative feedback to eliminate any circuit drift. When the drift has been eliminated, the brake-clutch operates as a brake and locks the zero drift potentiometer to provide continuous elimination of drift.

The range computer integrates range velocity with respect to time and divides the integral by the radius-of-curvature. The quotient is represented by rotation of the MRA shaft which positions various potentiometers, switches, and resolvers to provide signals proportional to MRA. These signals are used both as correction signals in themselves and to compute other correction or demand signals for the missile guidance system.

The range velocity signal from relay K1004

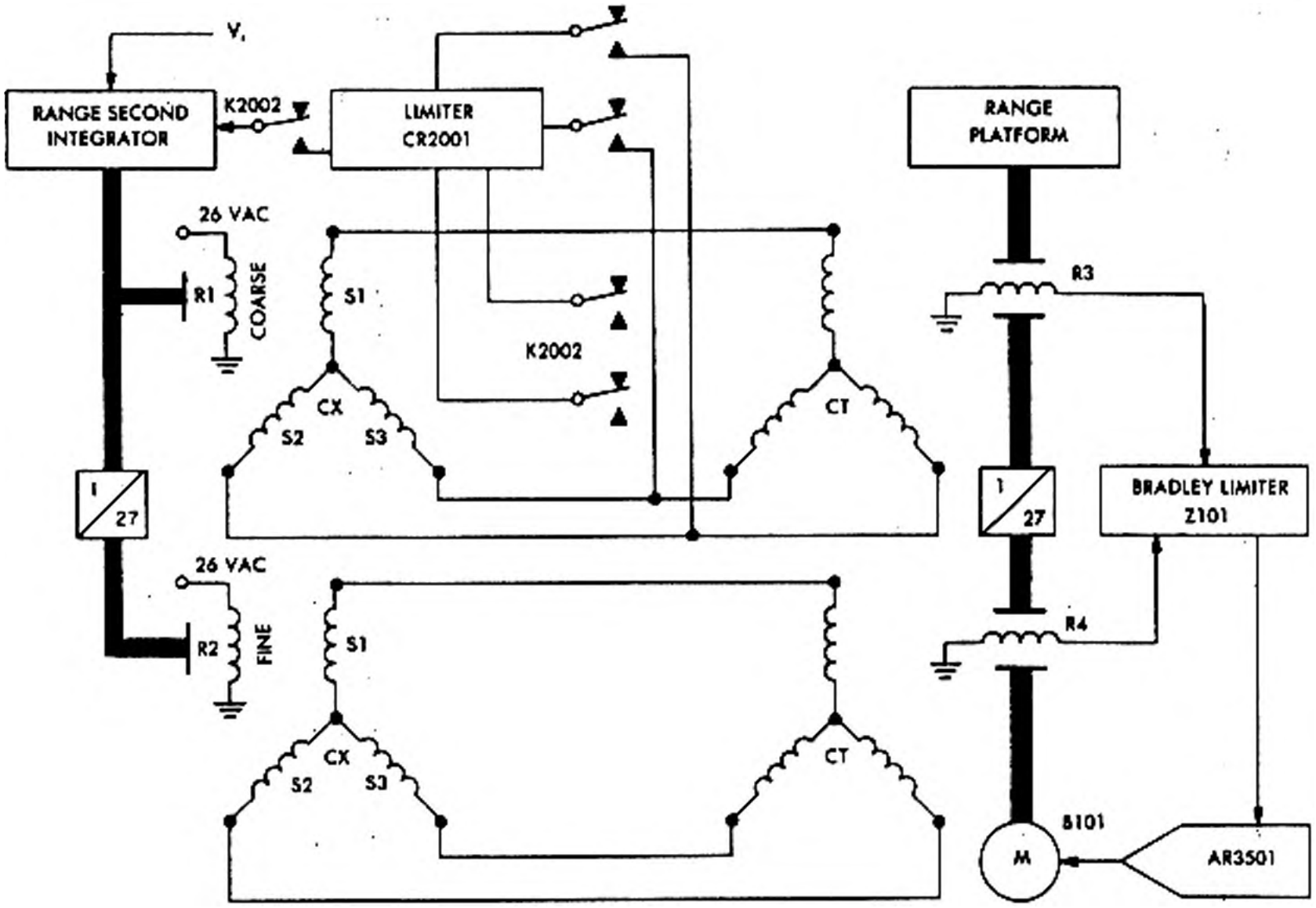


Figure 9-27. Range Platform Drive Loop

(figure 9-26) in the gyro-stabilized platform control is applied to amplifier AR2010 through relay K2001. The output of amplifier AR2010 drives the motor of range mechanical filter MG2001. The output of range mechanical filter MG2001 is supplied through relay K2002 to amplifier AR2011. The output of amplifier AR2011 drives motor-tachometer MG2002 which drives the MRA shaft at a rate that is proportional to distance traveled. The gain of the loop is controlled by adjusting the amount of negative feedback from motor-tachometer MG2002 to amplifier AR2010 through radius-of-curvature servo as changes in altitude and latitude occur, dividing the distance traveled by the radius-of-curvature of the flight path. The zero drift potentiometer R2009 is used to set the critical null condition of this high-gain loop. Reset of the MRA shaft is accomplished by synchro limiter CR2001 through the second integration loop. The outputs of coarse and fine synchro transmitters R2008 and R2009 are coupled through transformer T2002 to synchro limiter CR2001. The limiter selects either the coarse or fine signal and applies it to the second integration loop to reset the MRA shaft. During flight, the outputs of synchro transmitters R2008 and R2009 drive the range accelerometer through an angle equal to MRA.

RANGE PLATFORM DRIVE LOOP. The range platform drive loop (figure 9-27) causes the range platform to be driven through MRA by electrically connecting the MRA shaft to the range platform. Since the range platform serves as a mount for the range accelerometer, the input axis of the range accelerometer is maintained perpendicular

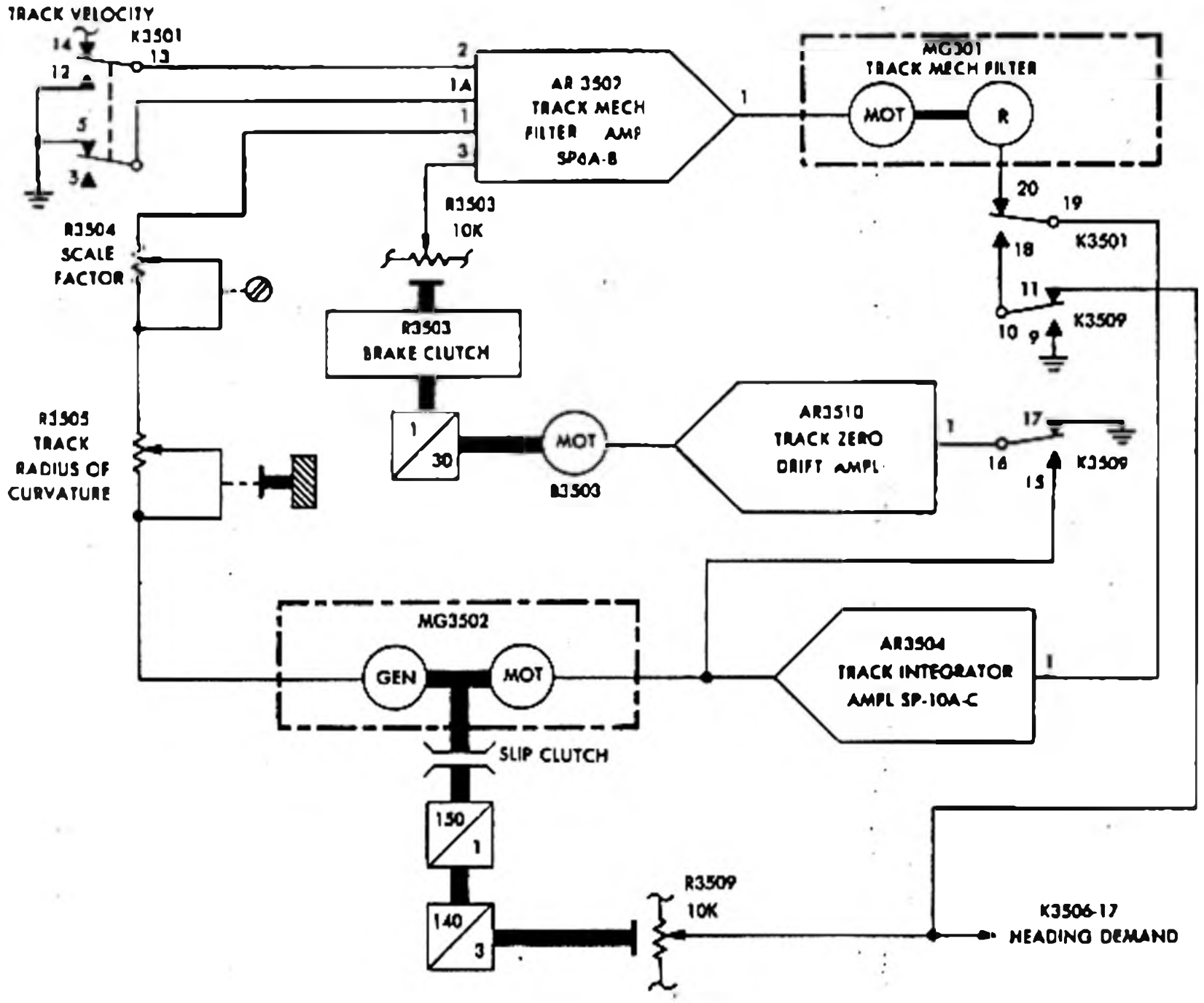


Figure 9-28. Track Computer

to local vertical. This is necessary to minimize the possibility of gravity having an effect upon the range accelerometer.

The range platform drive loop, a synchro coupling between the MRA shaft and the range platform, maintains the input axis of the range accelerometer perpendicular to local vertical by driving the range platform through the same angle that the MRA shaft has been driven. If the input axis of the range accelerometer is not maintained perpendicular to local vertical, the accelerometer may sense gravity as acceleration. This would result in an incorrect velocity and MRA with the result that the missile would miss the target.

The range platform drive loop (figure 9-27) is a 2-speed, 2-channel synchro chain, providing a system for both coarse and fine controls. When the MRA shaft begins driving, it causes the coarse and fine rotors, R1 and R2, to turn in relation to the stators, Cx. This movement causes a changing voltage to be induced to the stators of the control transformers mounted on the range platform. The control transformer rotors R3 and R4 detect a voltage equivalent to the angle turned by the MRA shaft and apply this voltage to the Bradley limiter. The limiter selects the larger of the two rotor outputs (which is normally the fine rotor R4 because of the 1 to 27 gear ratio between the synchro chains) and sends this output to motor B101 through amplifier AR3501. The motor turns both the range platform and rotors R3 and R4 through the same angle that the MRA shaft has turned.

The output stator logs S2 and S3 are used during ground testing to reset the range second integrator (or MRA). If R1 and R2 are not at zero MRA, there will be a differential voltage between S2 and S3. By energizing K2002, this difference is sent through a limiter and back to the second integrator. The second integrator then drives the MRA shaft until there is no difference in potential between S2 and S3. This is the standard method used to reset circuits; that is, feeding the output back to the input to force the circuit to drive to zero.

TRACK COMPUTER. The track computer (figure 9-28), known as the track second integration loop, is used to integrate track velocity to determine the track distance of MTA. The computer is also used to control the guidance-on signal.

Since the track computer is similar to the range computer, only the differences are presented. The inputs to the track computer are Vy from the track first integration loop and two data set-ins; track radius of curvature and guidance-on (preset MTA). The outputs of the track computer are MTA to the navigation computer, sin MTA to the accelerometer compensation computer and to the torque microsyn of the track accelerometer, and a signal to flight controls for guidance-on (figure 9-28).

TRACK PROGRAMMER. The purpose of the track programmer is to provide lateral deviations which are programmed to avoid obstacles or heavily defended enemy installations during flight. The track programmer provides track deviations and a return to the range plane signal.

In appearance and operation the track programmer resembles the odometer portion of an automobile speedometer. It consists of sets of wheels on a common shaft.

The outputs of the track programmer, dependent upon the data set in, are used to establish when a deviation is to occur and how much deviation is to be made.

NAVIGATION COMPUTER (HEADING DEMAND CIRCUIT). The heading demand circuit applies information to the flight controls system to maintain a missile on the desired heading. It also demands a programmed change of heading and/or develops a signal which enables the missile to assume a flight crab angle (when one is required) to counteract the effects of crosswinds.

The navigation computer (figure 9-29) contains two loops: (1) a proportional loop which is responsible for the major portion of the navigation computer's effect on the missile heading and (2) an integral loop (this has certain unique functions to be considered under various flight conditions).

To understand the operation of the navigation computer, you must review the various control signals that are applied to the flight controls system. Figure 9-30 shows in simplified block diagram form the relationship of the three signals involved.

The particular flight circumstances determine whether one, two, or all three of the control signals will be applied to the flight controls

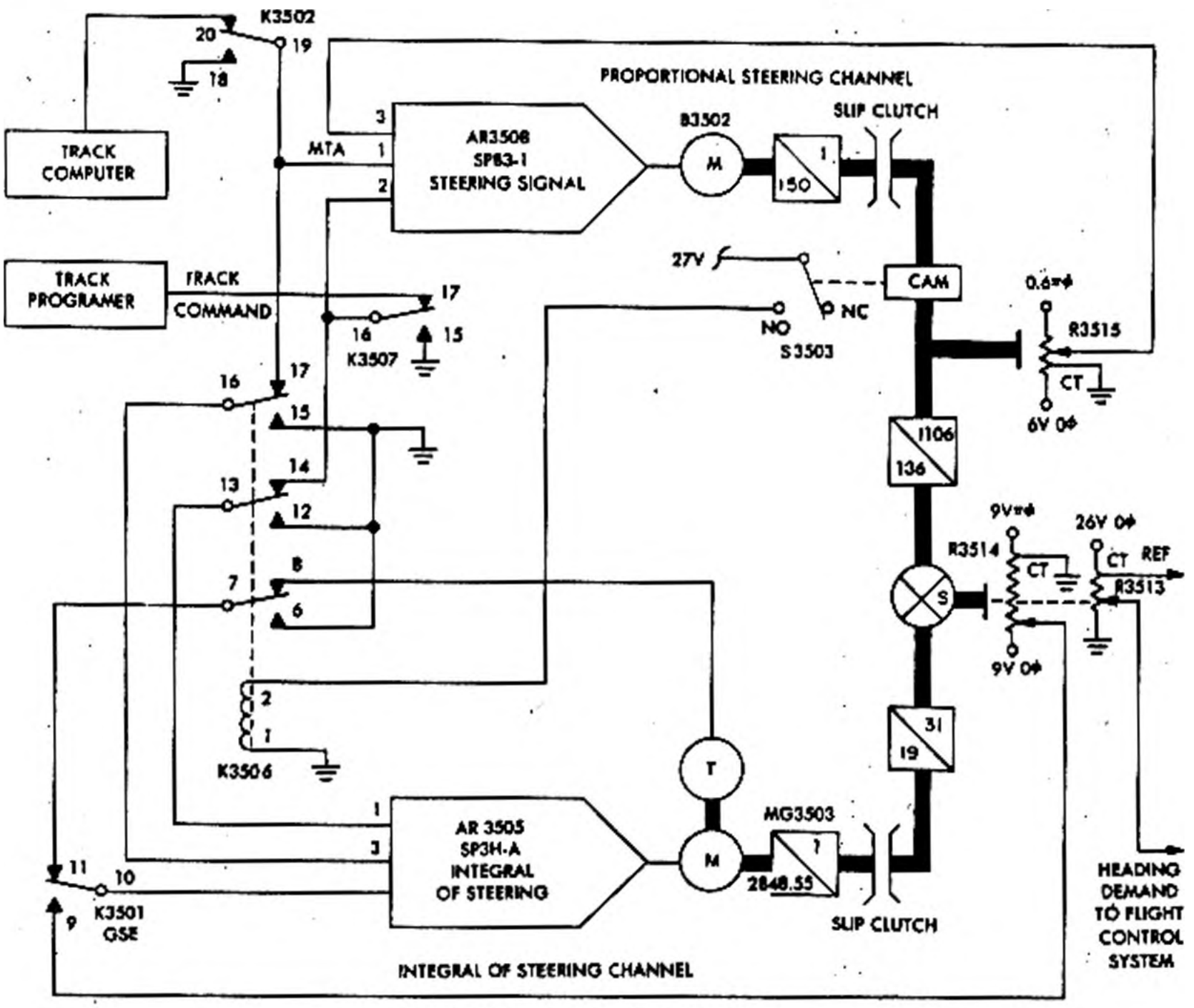


Figure 9-29. Navigation Computer

system. For example, when the missile deviates laterally from the flight course without a yaw or roll error, only a lateral correction signal is present. When the missile rolls without a yaw or MTA error, only a vertical gyro roll correction results. When the missile deviates from the flight course producing MTA, yaw, and roll errors, the flight control system is affected by a combination of all three signals. The yaw signal, which is produced by the yaw synchro, is combined with the lateral correction signal in the yaw lateral network in the flight control system. After the resultant signal is amplified and demodulated, it is combined with the demodulated roll signal from the vertical gyro. The output of the roll demand mixer causes actuation of either the right or left spoilers to correct the missile for roll or heading deviations.

To return the missile to the great circle course (assume there is no roll or yaw error to be considered), the amount of deviation, MTA, from the range plane, whether it results from programmed deviations or atmospheric conditions, is computed by the track computer. This signal is applied to the navigation computer through K3502 (figure 9-29). From here it is applied to AR3508 in the proportional loop and to AR3505 in the integral loop. The output of AR3508 causes motor B3502 to turn. The rate is proportional to the amount of error and the direction to the phase of the signal. B3502 positions the proportional shaft through a slip clutch arrangement. This shaft

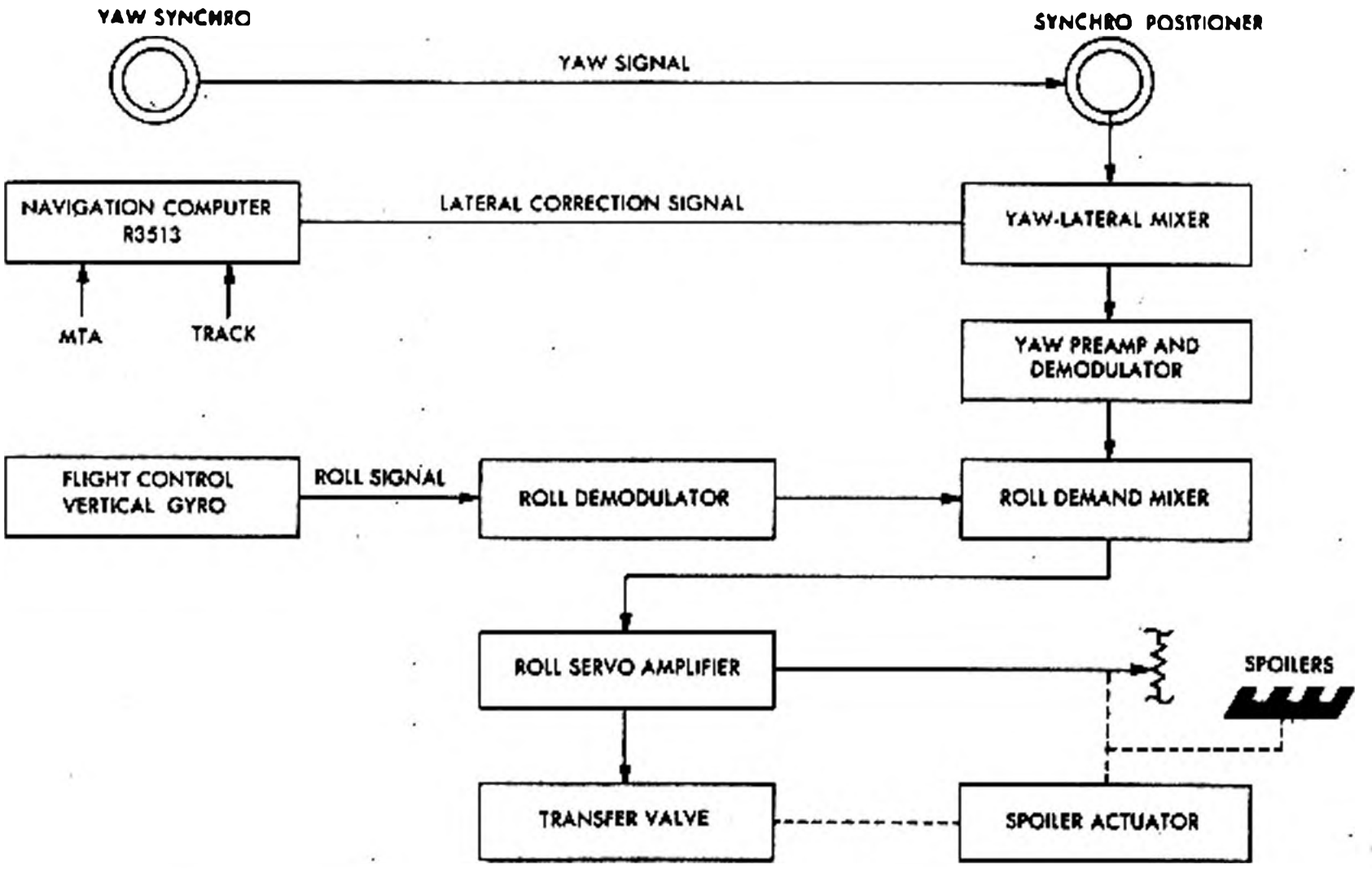


Figure 9-30. Heading Demand-Flight Controls

positions R3513 directly and R3514 through a mechanical differential. R3515 provides a proportional output by a degenerative feedback to AR3508. As the MTA signal increases so does the output of R3515. Conversely, when MTA begins to decrease, the output of R3515 becomes greater than MTA and reverses direction of the motor.

The output of the amplifier, in the integral loop, actuates MG3503 which positions the integral shaft through the slip clutch. This shaft also positions R3513 and R3514. The output of R3513, the heading demand signal, causes the flight control system to actuate the spoilers, causing missile roll and yaw. The missile then begins the return to the range plane, and the MTA signal amplitude begins to decrease.

The amount of spoiler action in the flight controls determines the angle at which the missile deviates from the line of flight. When the missile experiences a track error, it should eliminate this error as soon as possible. The greater the angle at which it returns to the range plane, the quicker the error is eliminated.

The missile is so designed to work best with positive control (the control surfaces constantly in action). For this reason the integral channel adds an additional setting on R3513 which causes, as the missile returns to the range plane, a certain amount of overshoot and actually oscillates or overshoots about 1 1/3 times. The integral pot also permits the function of establishing a crab angle or a compensation for crosswind. Consider the situation in which a crosswind is forcing the missile off course to the left and neglect roll error. An MTA signal is applied to both the proportional channel and the integral channel. Both signals produce heading demand signals to position R3513. The missile banks to the right trying to attain the desired flight course heading, causing a yaw error signal of equal magnitude to the heading demand signal and of opposite phase. This signal just cancels out the heading demand signal. As a result, the missile returns to the range plane at some small angle and the MTA and yaw errors diminish. The crosswind is still present and will again blow the missile off course to the left but not as far this time. This is be

cause the missile was never permitted to overshoot the range plane. As a result of this, the proportional channel was never permitted to wipe out the integral channel setting on R3513. Some slight amount of spoiler action is still present, causing the missile to head slightly into the wind. As the missile returns to the range plane again, the MTA and yaw errors reduce to zero although the signal on R3513 was not wiped out. The wind blows it off still less this time. The process continues until the setting on R3513 is the right magnitude to create the correct amount of spoiler action necessary to cause a crab angle. The missile remains on the range plane with no MTA error but some yaw error which is canceled out by the output of the navigation computer.

ALTITUDE PROGRAMMER. The altitude programmer provides altitude control signals to the altitude computer. It also provides such signals as dive wind, guidance monitor, warhead arming, and detonate signals. Altitude deviations can be programmed to occur at various decrements. The altitude programmer contains an initiate altitude switch, a track range angle (TRA) set-in, and a range-to-go assembly.

The altitude programmer, figure 9-31, is a pipe with threads on both the inside and outside. The inner thread distance is equal to 615 arc-minutes, and the outer thread distance is equal to 585 arc-minutes. A threaded screw is driven down the inside threads by RTG. When the ends of the threads are reached, a clutch plate is engaged which begins driving a cam nut (the traveling nut). As the traveling nut moves down the outside thread, it closes altitude deviation switches every 30 arc-minutes, and at the appropriate RTG the nut closes the termination switches.

ACCELEROMETER COMPENSATION COMPUTER. The accelerometer compensation computer, figure 9-32, generates various compensation signals for the range and track accelerometers and a latitude signal for the radius of curvature computer.

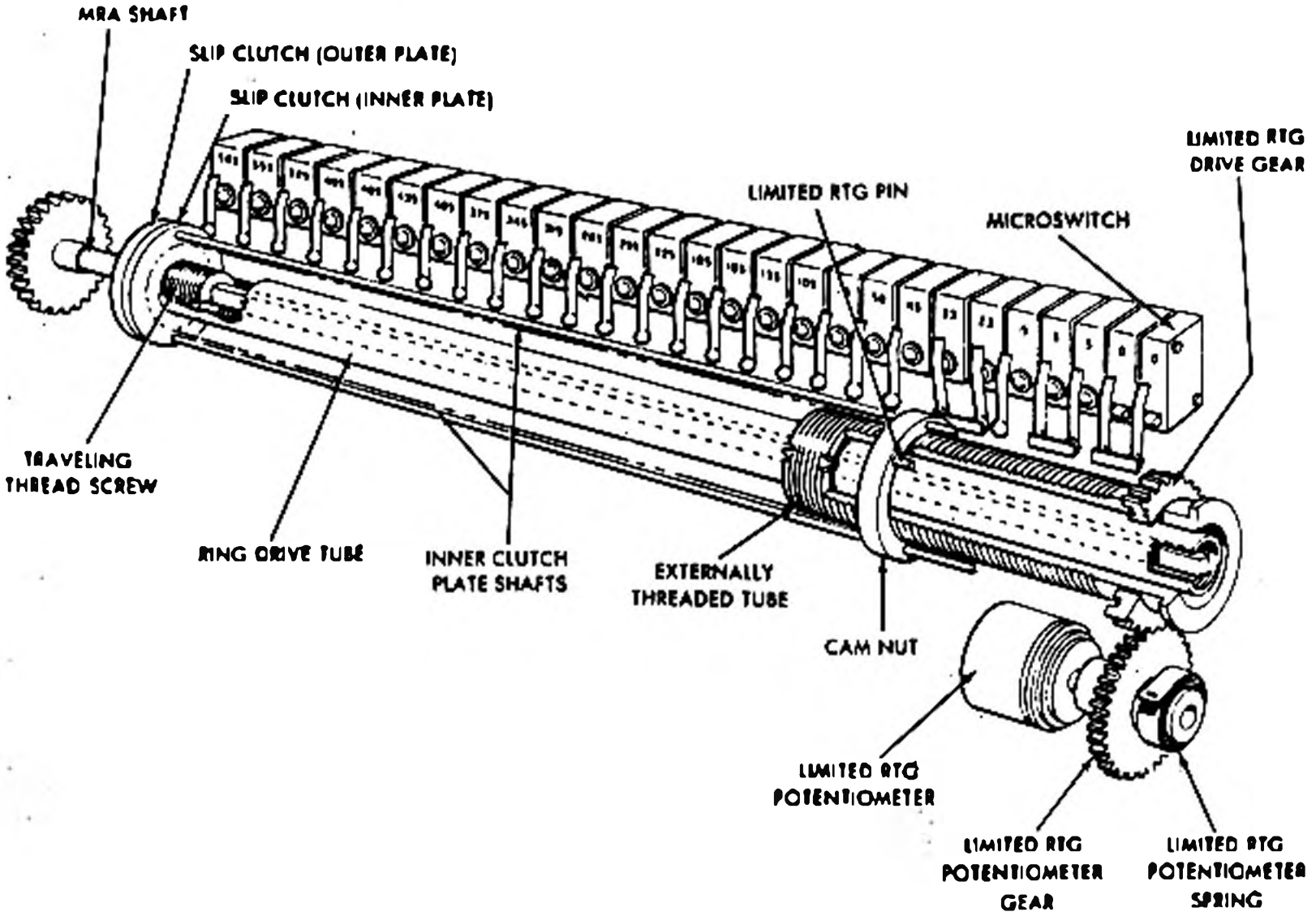


Figure 9-31. Altitude Programmer

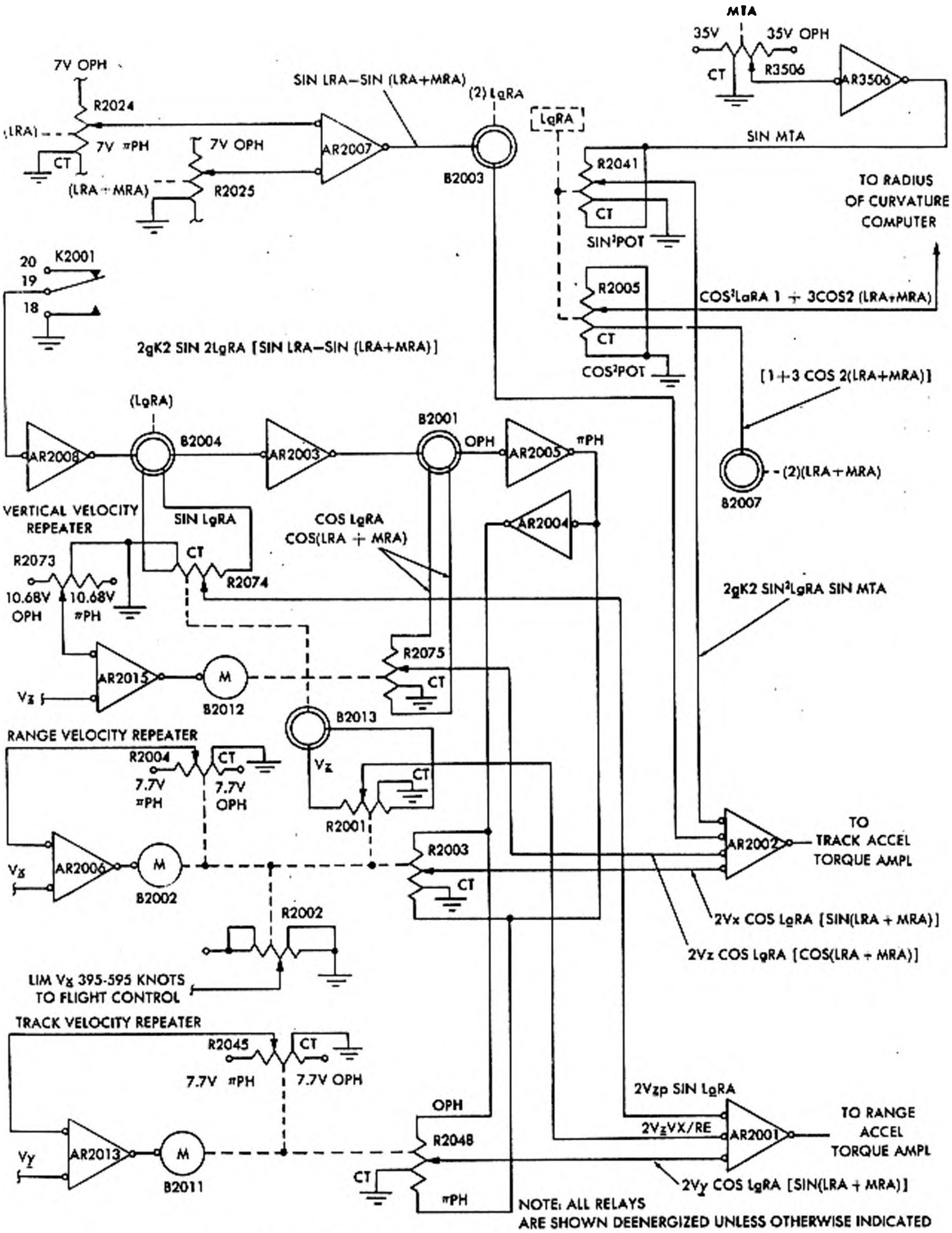


Figure 9-32. Accelerometer Compensation Computer

The inputs to the accelerometer compensation computer consist of three mechanical and four electrical signals as follows:

Mechanical Inputs	*Electrical Inputs*
MRA	Range velocity – Vx
LRA	Track velocity – Vy
LaRA	Vertical velocity – Vz
	Sin MTA

These inputs are combined by components in the computer, using the principles of addition, subtraction, multiplication, and/or trigonometric functions. The accelerometer compensation computer translates these input signals into the seven output correction signals described in the paragraphs that follow.

On-flight course gravity correction. When the launch range angle (LRA) set-in is made, potentiometers R2024 and R2025 and resolver B2003 are positioned. The output of sine potentiometer R2024 sin LRA is applied to amplifier AR2007. The output of sine potentiometer R2025, which is further positioned by MRA sin (LRA + MRA), is also supplied to amplifier AR2007. The difference in these inputs, sin LRA – sin (LRA + MRA), is applied as excitation to resolver B2003 which is positioned to an angle twice the LaRA set-in. (Latitude of the Range Axis. This is the angle present between the range axis and the co-equatorial plane.) The output of resolver B2003 is therefore sin 2LaRA [sin LRA – sin (LRA + MRA)], and it is applied to track accelerometer compensation amplifier AR2002. The relative channel gain of amplifier AR2002 effectively multiplies this input by $2gK_2$. The product, $2gK_2$ sin 2LaRA [sin (LRA + MRA) – sin LRA], is the correction signal for the track accelerometer.

Vertical Coriolis correction for track. Resolver B2001 is positioned by LRA and MRA and is excited by cos LaRA which is one of two outputs of resolver B2004 on the LaRA shaft. Resolver has two outputs. One output, cos LaRA [cos (LRA + MRA)], is applied to potentiometer R2075 on the vertical velocity repeater shaft. The output of potentiometer R2075, 2Vzw cos LaRA [cos (LRA + MRA)], is applied to amplifier AR2002 as the correction signal.

Horizontal Coriolis correction for track. The second output of resolver B2001, cos LaRA [sin (LRA + MRA)], is applied to isolation amplifiers AR2004 and AR2005. The outputs from isolation amplifiers AR2004 and AR2005 are applied to potentiometer R2003 on the range velocity repeater shaft and to potentiometer R2048 on the track velocity repeater shaft. The output of potentiometer R2003, 2Vzw cos LaRA [sin LRA + MRA)], is applied to amplifier AR2002 as the correction signal.

Horizontal Coriolis correction for range. The output of potentiometer R2048, 2Vy cos LaRA [sin (LRA + MRA)], is applied to amplifier AR2001 as the correction signal.

Vertical Coriolis correction for range-1. The second output of resolver B2004 on the LaRA shaft, sin LaRA, is applied to potentiometer R2074 on the vertical velocity repeater shaft. The output of potentiometer R2074, 2Vz sin LaRA, is applied to amplifier AR2001 as part of the correction signal. The output of R2001, 2VzVx/Re, is applied to AR2001 as well as other signals.

Off-flight course gravity correction. $\sin^2$ potentiometer R2041 on the LaRA shaft is positioned by setting in LaRA, and it is excited by the sin MTA signal from the power supply-computer. The output of potentiometer R2041, $2gK_2 \sin^2$ LaRA sin MTA, is applied to amplifier AR2002 and becomes the correction signal for the track accelerometer.

Accelerometer compensation computer outputs. Resolver B2013, on the vertical velocity shaft, is positioned by the Vz signal. The output of resolver B2013 is applied to potentiometer R2001 on the range velocity repeater shaft. The output of potentiometer R2001, 2VzVx/Re, is applied to amplifier AR2001 and becomes the second part of the correction signal. The three signals (horizontal Coriolis correction for range and two components of vertical Coriolis correction for range) are mixed in amplifier AR2001 and applied to the range accelerometer torque amplifier. The four signals (horizontal Coriolis correction for track, vertical correction for track, off-flight course gravity correction, and on-flight course gravity correction) applied to amplifier AR2002 are mixed and applied to the track accelerometer torque amplifier A1024. The output of potentiometer R2005 is applied to radius-of-curvature amplifier AR2009. Potentiometer R2002 on the range velocity repeater shaft is positioned by a Vx signal. The output of potentiometer R2002,

limited range velocity, is applied to the terminal dive computer in the flight control system.

RANGE RADIUS-OF-CURVATURE COMPUTER. The radius-of-curvature computer of figure 9-33 has two inputs: pressure altitude and latitude. The pressure altitude input comes from potentiometer R2083 in the altitude computer, and the latitude input comes from the $\cos^2$ potentiometer R2006 in the accelerometer compensation computer. The two input signals are mixed and amplified in radius-of-curvature amplifier AR2009 and applied to motor B2005. Motor B2005 drives a shaft which positions radius-of-curvature potentiometer R2008 in the range second integration loop and feedback potentiometer R2028. Potentiometer R2028 provides feedback to amplifier AR2009 so that the output of amplifier AR2009 will be proportional to the input rather than an integral of the input. Potentiometer R2008 is part of the feedback circuit of the range second integration loop. The gain of the loop is kept proportional to the reciprocal of the radius-of-curvature of the flight path by potentiometer R2008.

STABILIZATION GYRO COMPENSATION COMPUTER. The stabilization gyro compensation computer (figure 9-34) provides a compensation to offset the gravity unbalances along the spin reference axis of the tangent gyro (gUsra). Any unbalances contained in the gyros can normally be eliminated by the earth memory units of each gyro (provided the gyro does not change drift rate during flight). The mounting of the tangent gyro on the stable platform causes the pull of gravity to be felt as a force about the input axis.

As the missile leaves the launch point and flies down range, gravity is pulling from a different direction directly proportional to the change in MRA. To eliminate gUsra, the maximum compensation voltage is set in as the tangent gUsra signal on R1015. The output of R1015 is applied as excitation to R2020 which is driven by the MRA shaft. The output of R2020 will be 0 at launch and will increase proportionally to MRA. As a result, gravity does not change the drift rate of the tangent gyro. The other stabilization gyros do not require a gUsra compensation since the change in gravity is not felt about their input axes and no precession occurs.

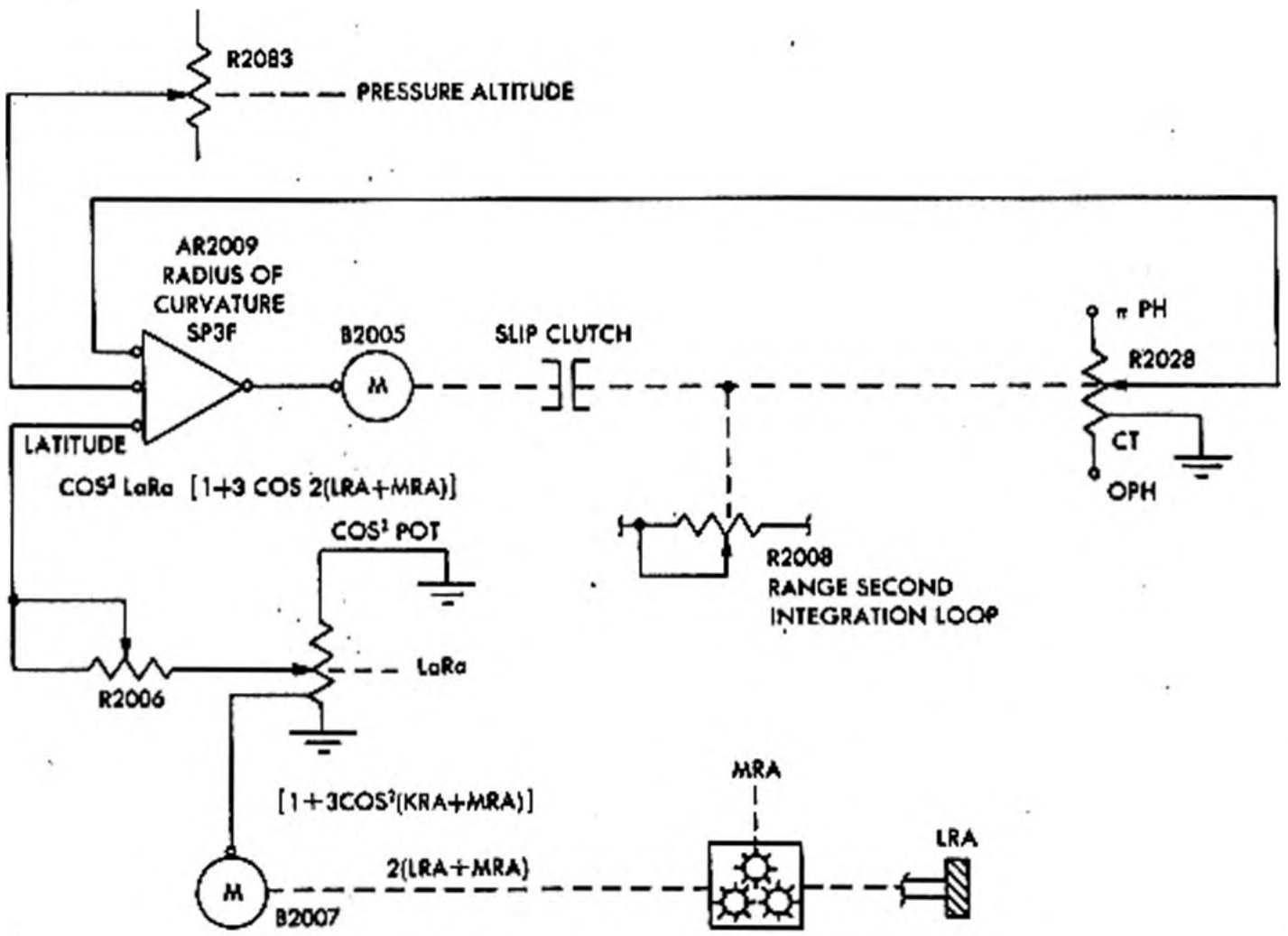


Figure 9-33. Range Radius-of-Curvature Computer

Each stabilization gyro has an earth memory unit which memorizes the drift rate of that specific gyro during checkout. The earth memory units with the test equipment measure the output of the gyro due to drift and position a memory potentiometer. The output of the potentiometer is applied to the torque microsyn of the specific gyro, causing it to torque in opposition to the drift. When there is no longer any output from the gyro, memorization has been accomplished and the memory pots are locked with a brake-clutch mechanism.

POWER SUPPLY COMPUTER. The power supply computer consists of a track computer, a navigation computer, and a power supply. The operating principles for the track computer are divided into

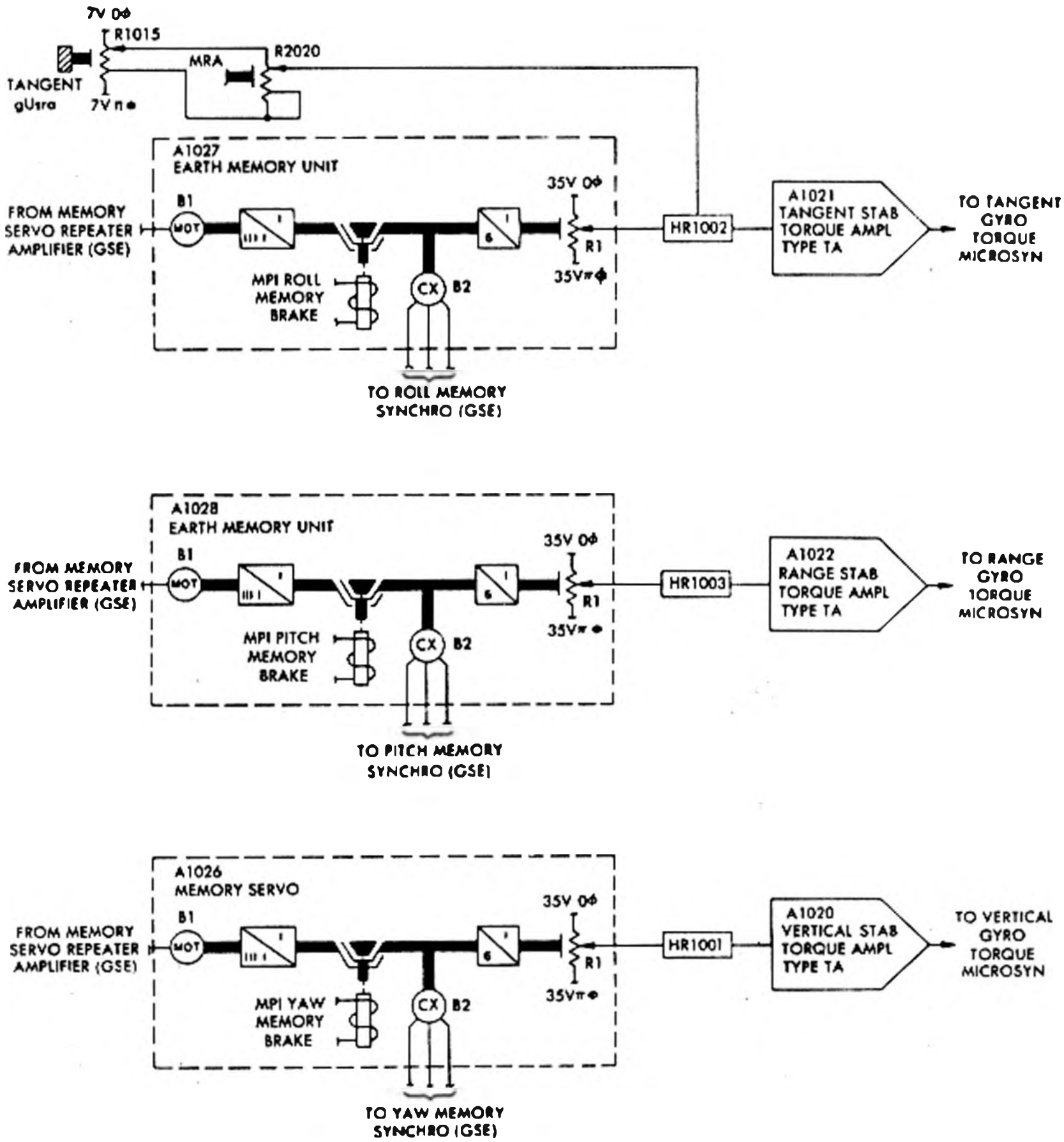


Figure 9-34. Stabilization Gyro Compensation Computer

two functional assemblies—the track displacement computer and the guidance-on circuitry for the navigation computer.

Track displacement computer. The track displacement computer contains a track mechanical filter and a track second integration loop. The track mechanical filter is a damping device that filters transient quadrature voltages to provide a signal proportional to track velocity for the track second integration loop. The track second integration loop constantly integrates this signal with respect to time over a radius-of-curvature value that is determined and set into the track computer prior to missile launch. The result of this calculation in the track second integration loop is a signal that represents the lateral angular distance of the missile from the range plane. This signal is the MTA signal. This MTA signal drives a motor-generator that mechanically rotates an MTA shaft at a rate proportional to MTA. The MTA shaft rotation positions the dials of a track displacement indicator, the wiper arm for each of three potentiometers, and guidance-on differentials.

A portion of the range platform drive loop is contained in the computer power supply. This portion consists of an amplifier and a transformer. The output from the amplifier provides excitation to the primary coil of the transformer. The induced signal from the secondary coil of the transformer is routed to the gyro-stabilized platform where it is used to drive the range platform drive servomotor.

Navigation computer. The navigation computer contains a proportional steering channel and an integral of steering channel. Each channel contains an amplifier that provides a signal to drive a motor, and each motor is mechanically connected through a differential and a slip-clutch to the heading demand shaft. Incoming signals to these channels rotate the heading demand shaft clockwise or counterclockwise through a portion of one revolution. As the heading demand shaft rotates, it positions the wiper arm of a potentiometer. If guidance-on has occurred, the signal from the wiper arm is routed to the flight control system as the heading demand signal.

Incoming signals to the proportional steering channel and integral of steering channel are provided by the track programmer in the missile guidance computer when a scheduled track deviation occurs and by the track displacement computer when the missile is not on the range plane or on a prescribed course parallel to the range plane. During large, rapid changes in MTA, the integral of steering channel is opened and only the proportional steering channel drives the heading demand shaft. As the missile returns to the range plane or a prescribed course parallel to the range plane, the integral of steering channel will close. If a small steady state error signal due to crosswind exists, the integral channel drives the heading demand shaft until a null condition is achieved. When the null condition is achieved, the wiper arm of the heading demand potentiometer comes to rest away from the center tap. The steady state output from the wiper arm to the flight control system causes the missile to maintain a sufficient angle to the range plane to compensate for the crosswind.

The abbreviated discussion of inertial guidance systems is very complex. Sufficient information has been presented to understand a signal path or block diagram analysis without getting too involved in the circuitry. The inertial guidance functional data (table 9-1) should be of some help as a form of review.

CELESTIAL NAVIGATION SYSTEMS

A simplified approach to a celestial navigation system assumes an inertial system supervised in a continuous series of fixes. Two systems used in missiles are known as stellar supervised inertial autonavigator (SSIA) and another known as automatic celestial navigation (ACN).

SSIA System of Missile Guidance

In the stellar supervised inertial autonavigator, periodic stellar sights are taken to check on gyro drift. This additional supervision is necessary because not all gyro drift is constant. Gyro drift that is not constant varies as to direction and magnitude. The inability of the slow loop correction to "predict" this random drift leaves an error that tends to increase with the passage of time. For a short flight of 45 minutes or so, random drift would introduce a probable error of about a half mile with modern gyro design. This error naturally increases as the time of flight increases.

Table 9-1. Inertial Guidance Functional Data

NAME	PURPOSE
Gyro-Stabilized Platform	Provides a fixed launch point coordinate reference for the range and track accelerometers.
Stabilization Gyros (Range, Tangent and Vertical)	Slave the gyro-stabilized platform to the launch point coordinates.
Gyro Accelerometers (Range and Track)	Sense accelerations in the range and track directions.
Range Platform	Maintains the input axis of the range accelerometer perpendicular to the local vertical (gUsre).
Stabilization Compensation	Generates a compensation for tangent Gyro Unbalance about the Spin Reference Axis (gUsre).
Meterization	Compensates the stabilization and accelerometer gyros for gyro unbalances and earth rate.
Vz Computations	Convert the sensed range accelerometer accelerations to range velocity.
Vy Computations	Convert the sensed track accelerations to track velocity.
MRA Computations	Convert range velocity to Missile Range Angle (MRA).
MTA Computations	Convert track velocity to Missile Track Angle (MTA).
Track Programming	Provides track deviation signals.
Altitude Programming	Provides altitude deviation signals.
Accelerometer Compensations	Provide correction signals to minimize the effects of horizontal and vertical Coriolis (range and track accelerometers) and correction signals for on-course and off-course gravity errors (track accelerometer).
Altitude Computations	Provide altitude command signals to flight control.
Navigation Computations	Provide heading demand signals to flight controls.
Radius-of-Curvature Computations	Compute radius of curvature for MRA computations.

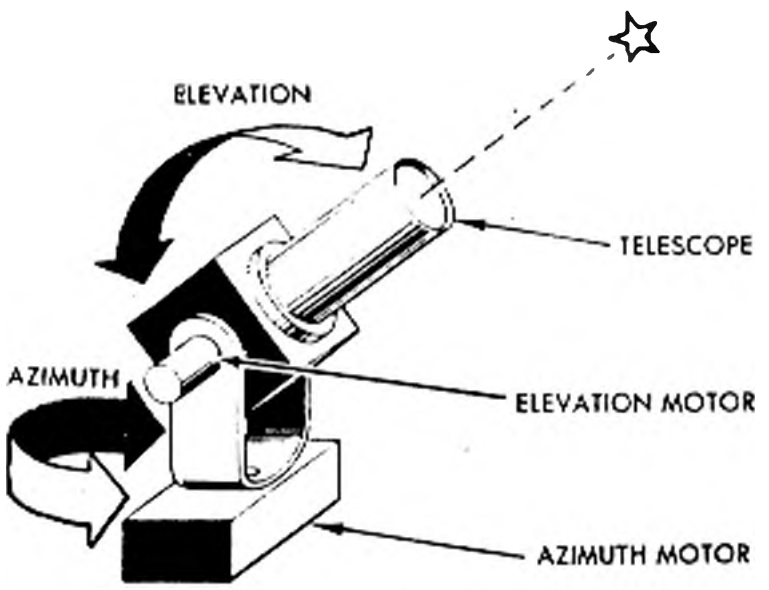


Figure 9-35. Automatic Sextant

One possible method of overcoming the random drift error is by the use of celestial sights. This is done much as a human navigator would check his dead-reckoning position with a sight on an object of known position, such as a star.

A physical addition to the stable platform consists of an automatic sextant, so mounted that it can be varied in the elevation and azimuth axes. A combination mounting that gives the variation is shown in figure 9-35.

The azimuth and elevation motors are precision torquers connected to a sextant positioning system. The system, as shown in figure 9-36, receives its signal from a tape playback which is prerecorded with the necessary position and rate data for the complete flight.

Telescope azimuth and elevation information must be read from the tape at the proper time. This timing of the tape playback is an important function because a star is at a particular angle with respect to a certain spot on earth only at a particular instant of time.

The telescope is aimed at the star by the information read from the tape and is then programmed through the tape information to follow the star.

A scanning system detects whether or not the star is centered in the telescope field. The error signal developed can be detected and processed to give an indication of the sextant error.

The outputs of the stellar error-detection circuit, shown in figure 9-37, are several voltages that are proportional to the missile error in pitch, roll, and yaw. After passing through the scanner, light from the star falls on the phototube causing a voltage output proportional to the light inten-

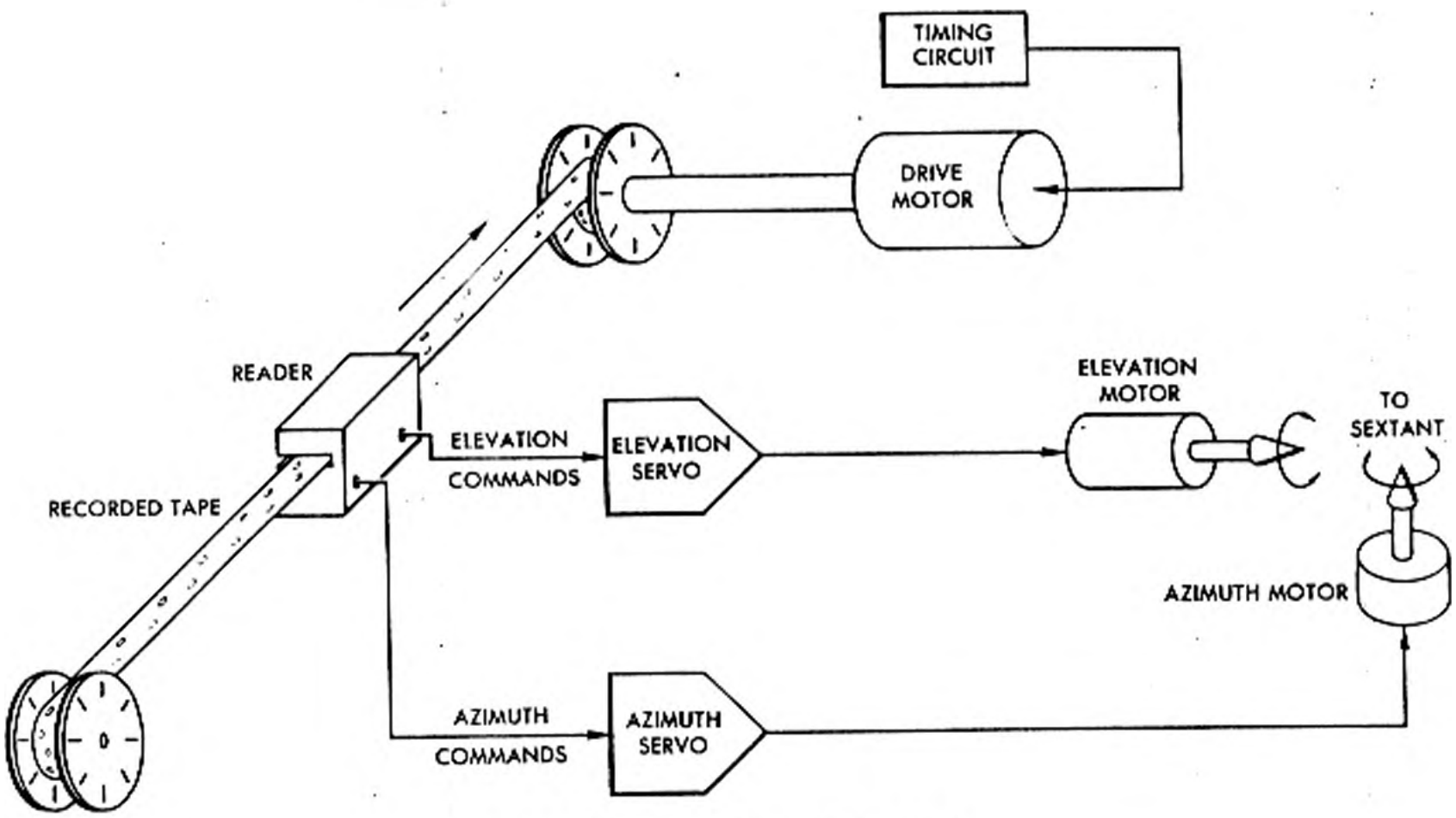


Figure 9-36. Sextant-Positioning System

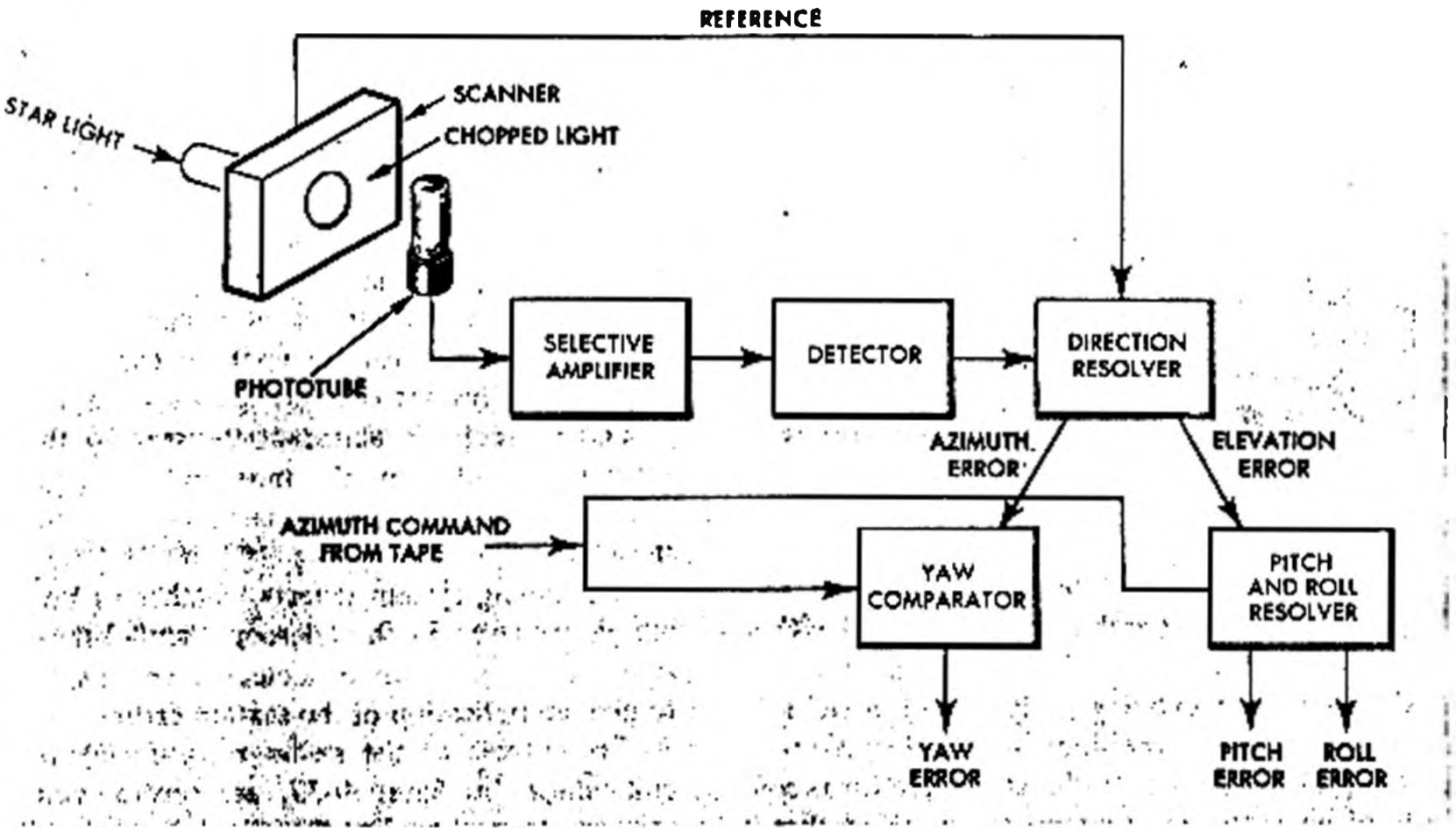


Figure 9-37. Stellar Error-Detection Circuit

sity. The phototube output passes through a selective amplifier. The selective amplifier separates the signal from the noise. The desired signal is then detected to obtain the information of the error. In the first resolver the detected signal is resolved into azimuth and elevation error signals by comparison with a reference voltage from the scanner. The direction resolver can be a bridge demodulator circuit or some other type of phase detector.

The errors are again resolved to make them usable in the missile system of coordinates after the azimuth and elevation errors in the sextant have been determined. The second resolvers are controlled from the tape reader. The same signal that positions the sextant sets a resolver for the elevation error output. If the elevation signal were not resolved in this manner, there would be no way of determining whether the error was in the pitch or the roll axis.

If the telescope were elevated and pointed directly forward along the missile heading, any elevation error signal out of it would be attributed to error about the pitch axis. Also, if the telescope were pointed out the side in the lateral direction, any elevation error would be only a function of roll. The resolver is needed to determine whether the elevation error signal is due to pitch or roll, or a combination of the two.

Consideration of the problem shows that there has been no great effort made to preserve the quantity (amplitude) of error signal. The direction and axis of error have been carefully retained but the magnitude has not. Keep in mind that the ideal output of the system is a zero voltage. It would only be an expensive and unnecessary refinement to retain the magnitude because, in a nulling system of guidance, the error voltage approaches zero and balances out when the missile is on course. Proportional control is not attained. Because of the delaying of the signals through the guidance circuitry and damping by rate function, the missile tends to return to the desired course as fast as possible but without overcontrol oscillations.

An ideal situation in a stellar sighting system would be to check first one star whose line of position (LOP) ran tangent to the course of the missile and then another star whose LOP would be normal to the missile course. In such a situation, the information from the first star would be applied to the computer direction channel. In the second

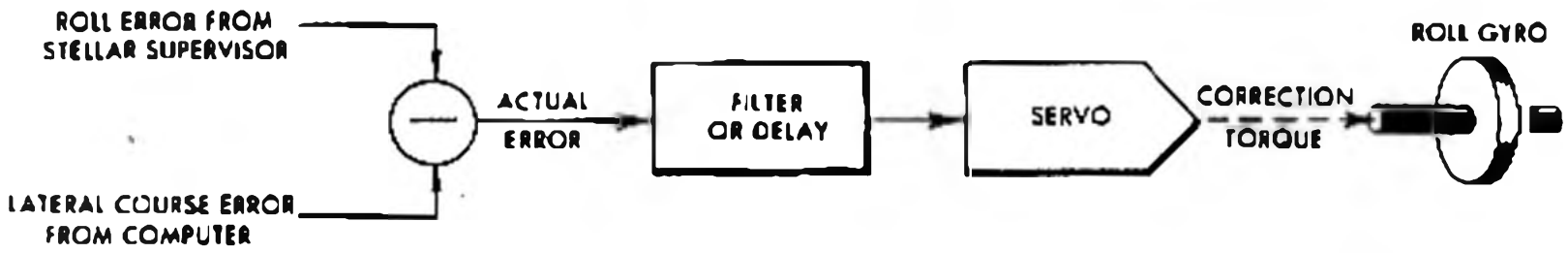


Figure 9-38. Correction Circuit for Determining True Error

case, the stellar error signal would go to the distance channel in the computer. These signals would then alter the autonavigator operation by correcting the gyros to a new position, compensating for any drift that has occurred. To be assured that the correction torque is accurate, the error signal would have to be averaged over a period of time.

Figure 9-38 shows the principle of deriving the difference between the roll error as detected in the stellar supervisor and the course error in the computer. Both errors must be in the same terms. The scale factor of the signals from each circuit must be alike. The feedback loop for such a system is via the new platform position obtained through the star loop and the accelerometer loop back to the computer. If the star is held sufficiently long, the error is damped out to an insignificant amount.

The pitch and yaw gyros are corrected in the same fashion. Some additional consideration for the operation of the yaw circuits is necessary. The azimuth detection of a star is not nearly as accurate as the elevation detection, so in some cases a complex crossmixing of pitch and roll stellar outputs is used to correct the yaw gyro.

With the stellar-supervised inertial autonavigator, the missile can fly for long periods under cloud cover with little effect on accuracy. The difficult part is the development and manufacture of gyros which can fulfill the requisites of such a system. The techniques needed to produce gyros with zero bearing friction and no unbalanced forces quite understandably tax the ingenuity of the industry. The gas-lubricated (gas-bearing) gyros come the closest to the ideal gyros for such a system. The gyros that are easier to produce, and therefore more plentiful, also are more inaccurate. The mass-production gyro is only suitable for use in a continuously (or nearly so) supervised system.

ACN System of Missile Guidance

Based on what has already been covered, the most logical step to overcome the principal problem of a SSIA system would be to use a stellar system having continuous supervision.

The idea of inertial supervision is usually overlooked in considering automatic celestial navigation (ACN) because the system is continually referenced by stellar fixes. But, in any system, the inertial principle still is present in the action of the autopilot between guidance commands.

The platform equipment for ACN requires the addition of one or more automatic sextants. With two sextants operating simultaneously, a series of fixes are obtained rather than just LOP. With simultaneous fix on two stars, there is less chance for inaccuracy and less necessity for averaging out errors. A spare or standby sextant might be incorporated into the equipment so that it can be zeroing in on the next star in the navigation sequence without detracting from the previous fix.

A disadvantage of the multiple telescope system is the need for a relatively large window to view the celestial sphere. The cost of such a perfect-plane window increases considerably with size. The large window also finds itself subject to more forces and flections caused by high speeds.

Effects of Refracted Light and Bias

Refraction of light takes place as it passes through conducting media of varying indices. Since shock waves which occur at or above the sonic speed are, in effect, a conductor of a different index than the ordinary conductor, the dispersion of light rays caused by shock waves may limit the use of celestial or stellar systems. Note this effect in figure 9-39.

In practical applications, some noise or bias exists in the output of the velocity-measuring component. The noise that exists as short peaks of energy is effectively filtered out by the time con-

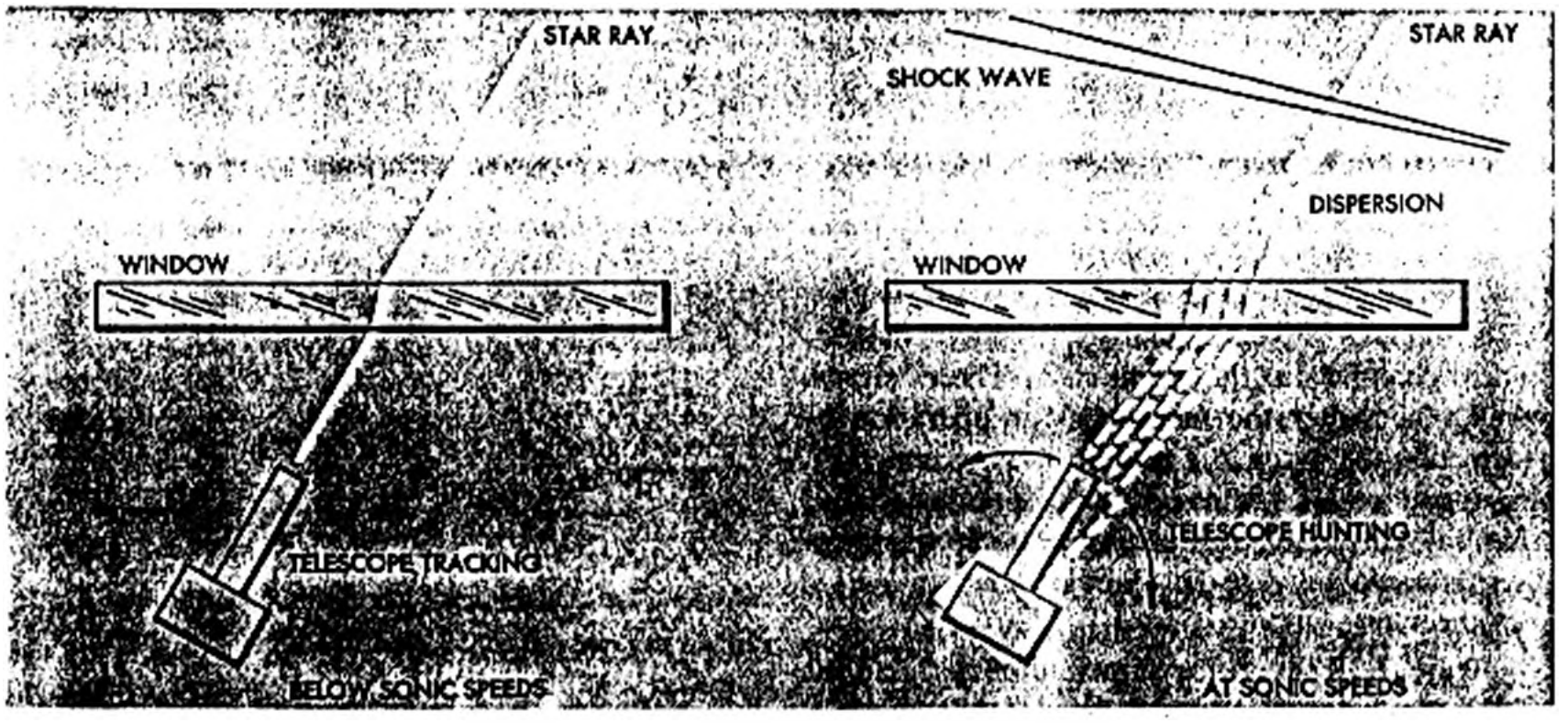


Figure 9-39. Effect of High Speeds on Stellar Systems

stants (delays) of the circuitry. Steady noises (errors) are not filtered by such action. If some steady error exists on the signal, indicating the measured velocity, then the whole computer output would be in error. An additional circuit is required that does not let the bias error through but admits other signals. Such a circuit has to have the characteristics of a high-pass filter which uniformly passes AC current of every frequency. A unique arrangement of amplifier elements gives the desired result, as shown in figure 9-40.

The desired high-pass filter characteristic can be obtained using a DC amplifier with integrator feedback. A constant positive voltage (E_i) at the

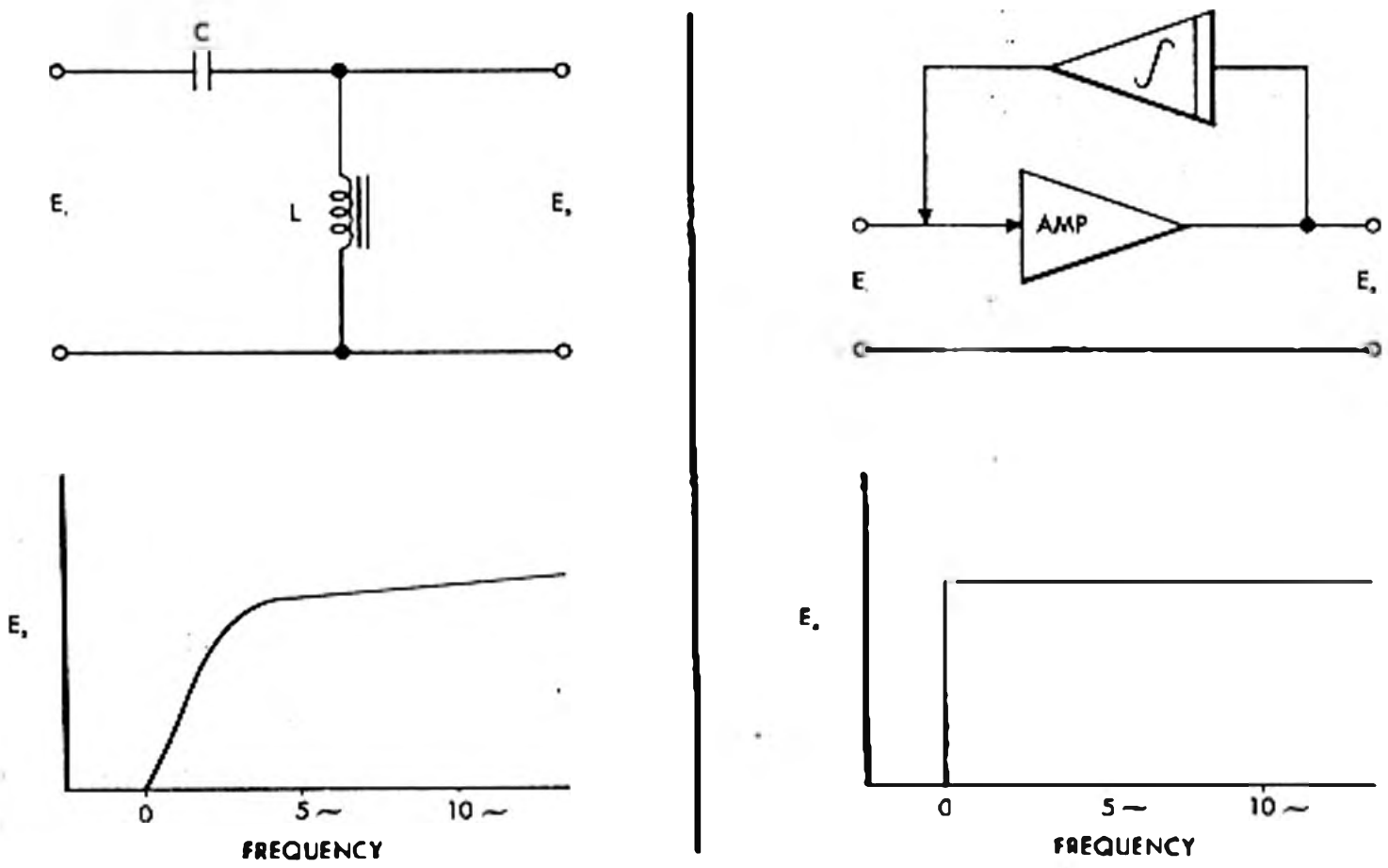


Figure 9-40. Characteristics of Two Types of High-Pass Filters

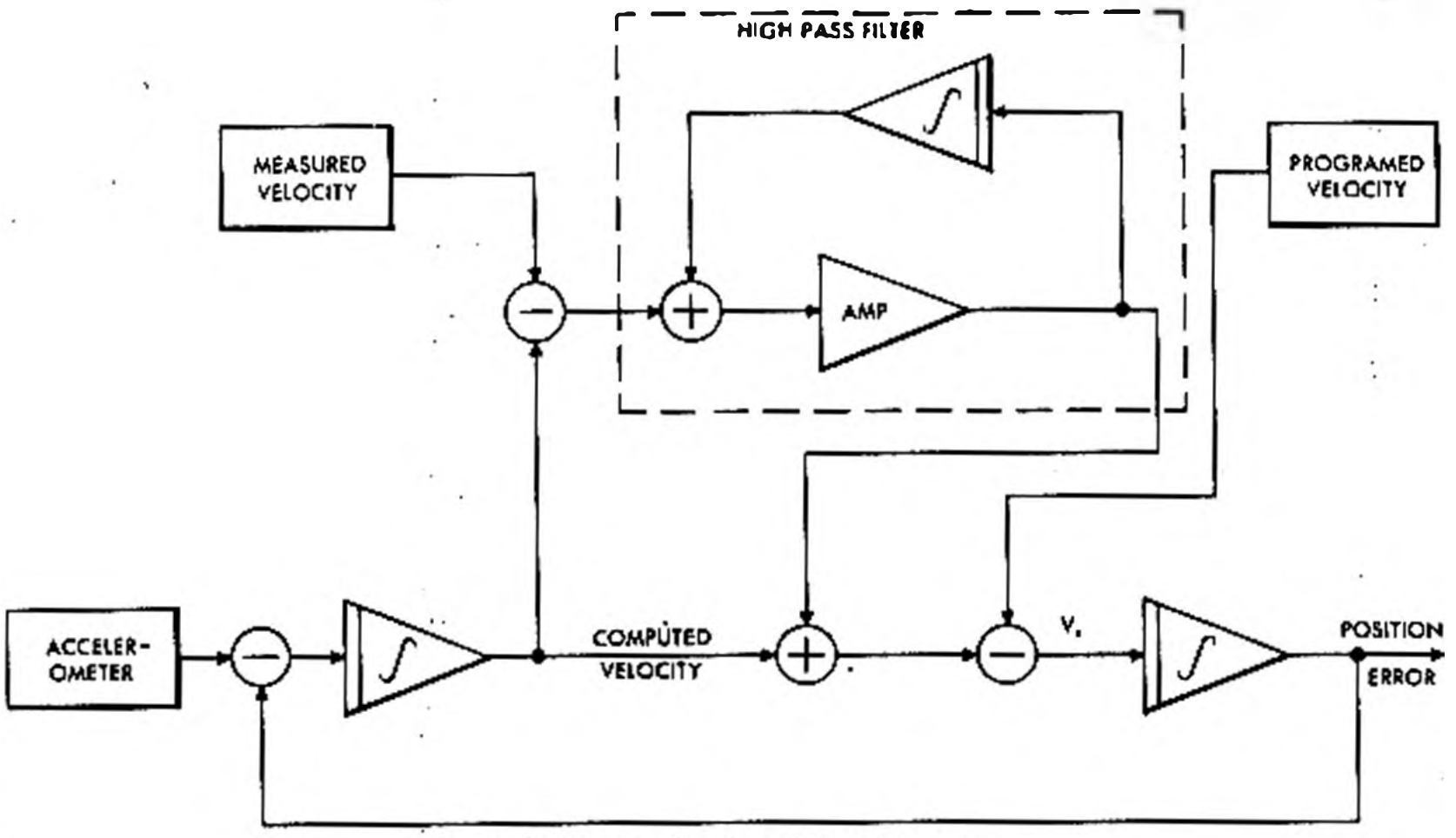


Figure 9-41. Application of Error Damping

input tends to give a like positive voltage at the output (E_o). The amplifier output becomes the input to the integrator. The output of the integrator with this constant positive input is a negative going voltage. The negative going voltage is added to the input of the unity gain amplifier, where it cancels the input, driving it to zero.

The integrator responds slowly to an input signal. It takes about 10 minutes for the integrator output to build up sufficiently to cancel a steady amplifier input. This means all voltages that varied faster than this would overcome the feedback and pass through the circuit before feedback takes effect. These voltages would continue to pass through the circuit as long as they varied at this rate.

In figure 9-41, a DC filter is applied to an inertial computing channel. Note that the computed velocity which is summed with the measured velocity is varying at the same rate as the computed velocity. In such a case, the filter input would be DC.

Since the varying difference between the two velocities is the required factor, the addition of the computed velocity is necessary and creates no error in the connection shown. By damping with externally measured velocity, the summing of errors in an inertial system is eliminated.

LONG-RANGE HYPERBOLIC GUIDANCE

The utility of hyperbolic navigation exists only for the distance at which the direct or groundwave can be distinguished. Long-range navigation by the Loran principle gives optimum operation on the low frequencies, in the 100 kHz to 200 kHz region.

Long-range missile guidance equipment, to make use of the Loran principle, must use an automatic receiver. There is no need for the regular indicator unit, since the outputs must be in the form of shaft rotations or voltages usable in the control systems.

Pairs of transmitting stations (master and slave) generate the pulses of RF energy separated by a delay interval. The delay exists after the master pulse. The delay exists while the pulse signal travels to the slave site and is used to key the slave transmitter. This delay is used to insure that the transmitted pulse of the master station is always received first at a receiver located in the area serviced by the system. Pulses of a particular time difference set up an LOP that is hyperbolic in na-

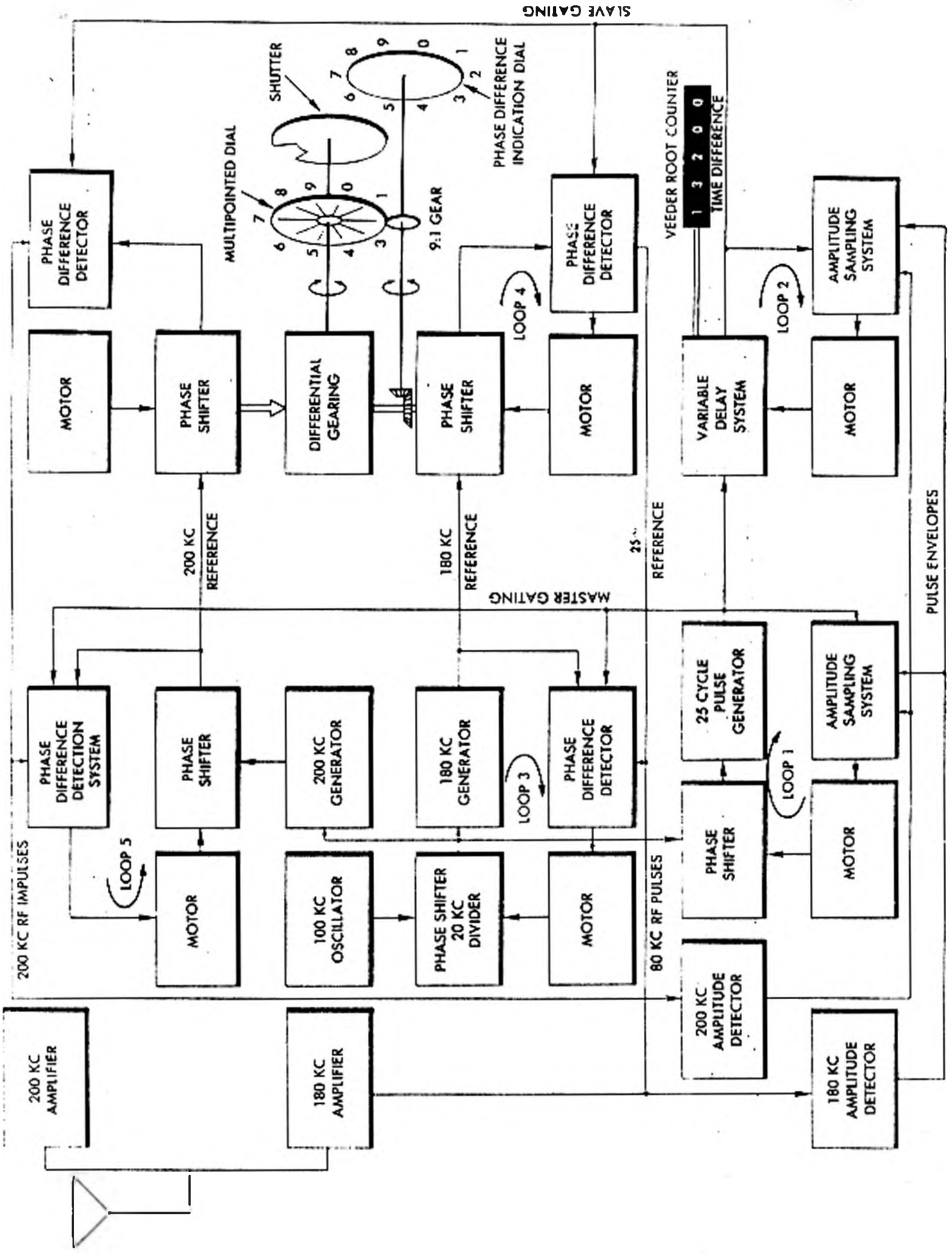


Figure 9-42. Indicator-Receiver for Automatic Phase-Matching System

ture. In the functioning of a Loran system, the pulse time differences are first measured for the coarse indication. The beginning RF cycles of the two transmitted pulses (master and slave) are compared for the fine range indication. In a manual-type Loran system, this matching is done by the operator using the presentation on a cathode-ray tube equipped with a precision sweep. To accomplish this automatically requires a more complicated receiver.

Some method of minimizing noise should be added to assure accurate operation under as wide a range of conditions as possible. Locally generated signals which are equal in time, frequency, and phase to the transmitted pulses are used. These local signals, which are noise-free, are compared to the transmitted slave signals, giving a better and more accurate comparison.

Automatic Long-Range Loran

An automatic phase-matching type of Loran system breaks up doubtful cycles of information, and it separates them by transmitting alternately on two different radio-frequencies. The pulse envelope timing and the phase of both radio-frequencies are controlled by the transmitting timer. By proper adjustment of phase and envelope timing, the envelope match and the RF match on both frequencies can be satisfied simultaneously in the receivers.

Shown in figure 9-42 is a simplified block diagram of the receiver for the automatic phase matching system. Servo loop 1 operates to bring the 25-Hertz pulse generator into time coincidence with the 25-Hertz modulation envelopes of the received master signals of 180 kHz and 200 kHz carriers. Servo loop 2 operates the variable delay system to bring its delayed output pulses for slave gating into coincidence with the received slave signals. The cycle matching on the master signal involves comparing the noisy master signal with the noise-free local reference signal. The same operation is carried out with the slave signal. This permits operation under worse signal-to-noise conditions than in a manual, visual, long-range Loran system. The amount of delay is indicated on the Veeder Root counter, and this is the envelope time difference reading between master and slave pulses, furnishing a coarse indication of the LOPs.

Servo loop 3 brings the 180 kHz CW reference wave into phase match with the received master 180 kHz signal by shifting the phase of the 100 kHz oscillator output. The 180 kHz CW reference and the received 180 kHz slave pulses are fed into the servo loop. The phase shifter is driven to bring these waves into phase match. The rotation of the phase-shifter shaft of loop 4 represents the phase difference between 180 kHz master and slave pulse signals.

In a similar fashion, servo loop 5 sets up a 200 kHz CW reference wave which is in correct match with the received master 200 kHz RF pulses. As with servo loop 4, servo loop 6 drives its phase shifter to bring the two inputs of the phase-difference detection system into matching phase. Again, the shaft rotation of the phase shifter of servo loop 6 represents the phase difference between the received master and slave 200 kHz signals.

Servo loops 3 and 5 are gated by the 25-Hertz pulse generator to operate only during the time of reception of master signals. Servo loops 4 and 6 are gated by the time-delayed slave gating 25-Hertz pulses so as to operate only during time of reception of the slave signals.

The mechanical outputs of the phase shifters of servo loops 4 and 6 are fed into the differential gearing. The output of the differential gearing is the difference between the two inputs or, electrically, the difference between the two phase-difference measurements.

The shaft of the differential output has a circular shutter with a 40° sector opening cut out. This opening represents the 20° of tolerance required for cycle-difference measurements.

The multipointed dial located behind the differential shutter has 9 equally spaced pointers indicated on a scale. It is driven through a 9:1 reduction gear from the 180 kHz phase-shifter shaft. Consider a fractional 9-cycle interval of the received time difference which is to be indicated by the cycle difference and the phase difference indicators. The phase difference or the fractional cycle difference is indicated on the phase-difference indicator. The number of whole cycles is on the multipointer dial. The shutter, and its 40° opening, indicates the presence of errors in the phase-difference measurements by the position of the arrows of the multipointer dial with respect to the shutter opening. No error caused by low signal-to-noise ratios exists if the arrows are present. If no

arrow is present, errors sufficient to prevent correct cycle-difference measurements are present in the phase-difference measurements.

The positions of the phase shifters that indicate the intervals can now be used in a computer. The information from this computer could be used to obtain distance and direction to target, to plot true position automatically, to control the automatic pilot, or to give whatever information of this order that may be required in the system.

The base line for a system accurate within a few hundred feet at 1500-mile range would need to be 600 to 800 miles long. The size antenna required for the receiver in a missile would also impose some problems. An adequate receiving antenna would need to be 30 to 50 feet long, or the equivalent. Using sections of the missile airframe to fulfill this purpose is possible.

FM Loran

An FM Loran system is similar in practice to the one just presented, but it uses a unique approach to eliminate ambiguities. In this system, the outputs of three transmitters are frequency-modulated by a sine wave.

These three transmitters, with different low-frequency carriers, are frequency-modulated by the same AF signal so as to obtain identical modulation in frequency and phase for the three transmitters. The time for one cycle of the modulating frequency must be sufficient to allow the RF signal to get to maximum range of the system. The receiver automatically tunes out the skywaves. Because of the longer path traveled by the skywaves, the signal produced by them has a different frequency than the groundwave at any instant of time.

Comparison between a pair of transmitted signals is accomplished by measuring the relative delay required to produce a phase match in their modulating functions. The phase match is indicated by a constant difference-frequency output of maximum amplitude from the mixer. The phase of this difference-frequency is also an indication of the RF phase difference of the two signals being compared. Phase changes in this difference signal can be detected by comparison with an equal frequency signal which is locally generated. In this manner, a method of fine control is provided by RF cycle matching. Time difference is measured by the phase shifts required to maintain the phase conditions for signal pairs.

A reference oscillator system, using only two base stations but requiring a precision reference oscillator in the missile, integrates the time difference to measure range absolutely. It establishes circular lines of position about the two base stations.

The term "precision reference oscillator" is a slight understatement. The requirement for this airborne oscillator is such that it starts its cycles at the regular period intervals throughout its operating life. Practical oscillators have a slight phase angle change with changes in load, power supplies, temperature, or physical conditions. Any slight phase shift in this system would create error that would be difficult to detect.

TERRESTRIAL REFERENCES FOR LONG-RANGE GUIDANCE

Various mapmatching guidance systems have been suggested and devised. The main idea of an electronic unit of this sort is the comparison of a photo or map contained in the equipment with an image of the area the missile is flying over.

If a negative and a positive transparency of a scene were to be matched and held up to the light, the combination would appear opaque. Now if either transparency were moved somewhat in respect to the other, light would show where images were out of register. If one transparency were contained in a frame activated by a servo, it would be possible to devise a detecting control mechanism that would automatically rematch the images. Instead of a positive transparency, the projected image from a lens or a radarscope could be used to show the actual picture of the area being traversed.

At this point, it would be best to rule out daylight systems; that is, the use of photographs of the actual course or target area. A daylight system would put restrictions on operations caused by visibility conditions and would limit operations to a time when interception by target defenses would be easier. A radar mapmatching system would have no limitations as to the effects of weather or darkness. It has some other drawbacks which are covered later.

Magnetic guidance is another type of mid-

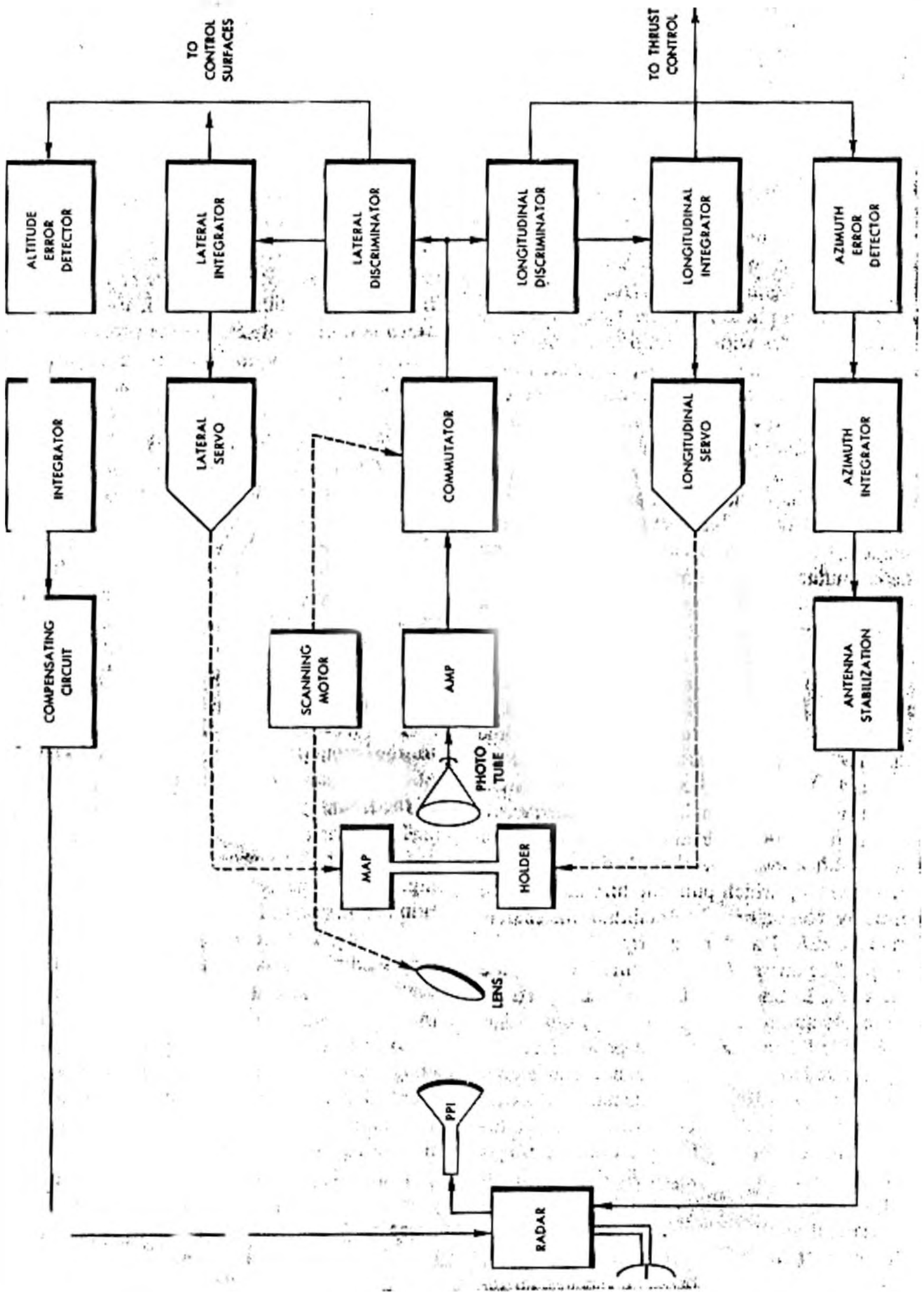


Figure 9-43. Radar Map-Matching System

course terrestrial guidance used in missiles. This type is discussed after the mapmatching system.

Radar Mapmatching Systems

The block diagram in figure 9-43 illustrates a workable guidance system which uses radar mapmatching. Although more adaptable as a homing system, radar mapmatching can be used for long-range guidance.

The radar map comparison is made by projecting the PPI image, by means of a rotating offset lens, through a negative transparency of the same region and onto a photomultiplier tube. When the PPI image coincides with the map image, the light transmitted is minimum. The map is printed on a strip of film and the lens rotated, causing the PPI image to be moved in a small circular pattern on the film. The effect produced in the photomultiplier tube output is similar to that obtained by an offset radar scan. When the output is properly commutated, left-right and fore-aft information is obtained.

The commutated pulses from the output of the photomultiplier tube are applied to DC discriminators and integrators; then, following through the block diagram, this information is supplied to two loops, lateral and longitudinal. The left-right information is applied to a servo-amplifier which drives the film carriage laterally to maintain the match. The position of this carriage is picked off as an error voltage for the control system. As the missile turns on its yaw axis, the film carriage is moved and the error canceled out.

The fore-aft information is applied to the longitudinal servo loop which pulls the film through the holder at the correct speed to maintain the match. The film is moved at a rate proportional to the groundspeed of the missile. The film may be keyed to indicate the location of a change of course or to initiate dive action.

Errors in indicated longitudinal position can result from a difference in altitude between reconnaissance and tracking runs, because of slant range distortion and altitude-return delay. The error is greatly decreased if the altitude error is compensated for in the longitudinal and lateral loops. It is known that altitude and azimuth errors produce lateral and longitudinal errors, which are sine and cosine functions of the antenna (vertical) scan angle. These false error signals cause an interaction between the loops. The interaction varies with different loop gains. Altitude and azimuth error signals are not derived conventionally from an altitude or azimuth discriminator. Altitude signals come from the lateral discriminator and are fed into the altitude error detector. The altitude integrator uses this output to furnish the altitude compensating circuit with a signal for expanding or contracting the video presentation. The video presentation is expanded or contracted to make it correspond to the scale of the transparency. The azimuth signal is derived in the same sort of functional circuitry, but its output is used to orient the antenna horizontal scan on the proper bearing.

It is necessary to have good angular match (within about 1°) before accurate left-right and fore-aft information can be obtained. The original angular match can be obtained by means of a magnetic auxiliary such as a compass. The match is maintained by the azimuth loop of the system.

Two types of holders for the film are possible. A frame-type film holder is the larger and more complicated, mechanically, of the two. It switches separate frames into the system and affords an easier initial lockon. The best method for investigating the filmstrip seems to be by scanning it through a mask with a semicircular opening, as shown in figure 9-44.

The filmstrip is pulled at a rate corresponding to the groundspeed of the missile. The length of a strip may need to be only about 1/20 of that of the frame-type map.

The radar maps used are the result of two procedures. Of these procedures, actual mapmaking flights over the terrain to be used is the best primary method. The actual reconnaissance of a target area, a procedure required by this method, may be difficult; so synthetic maps are often used.

Synthetic thiokol base compositions of relief maps of the area to be flown over are built up, using ordinary maps, aerial photos, and other intelligence information. The map is then photographed, using ultrasonic radar-trainer techniques. This map can be used with a success that is only slightly inferior to the actual maps.

The radar equipment used with this system would have the same requirements as one for a bombing navigation system. The CRT indicator would have to give minimum distortion to the picture. The same type system would be used in

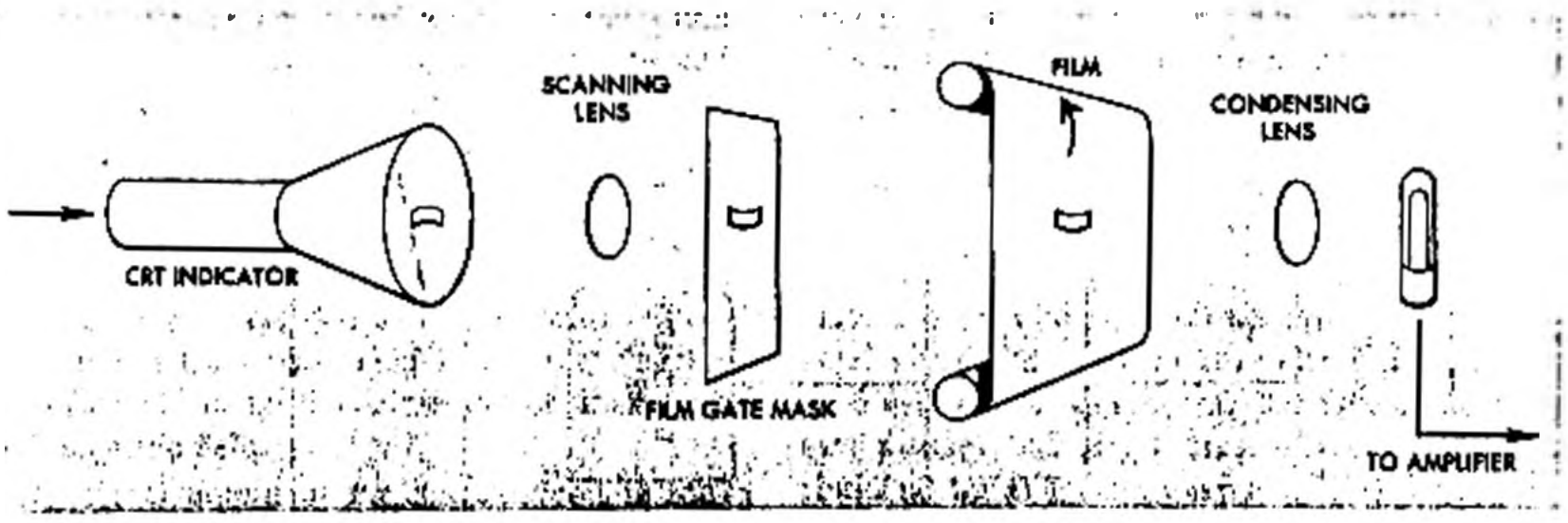


Figure 9-44. Operation of Film Holder for Strip Map

reconnaissance and actual matching flight. This indicator unit must have (1) precision ground range sweeps, (2) PPI presentation, (3) precision range and azimuth markers, (4) high-resolution flat-faced cathode-ray tubes, (5) provision for moving the indicator presentation proportionally to the groundspeed of the airborne vehicle, and (6) provisions for use as an accessory to a standard airborne radar set. The other radar-system components would be conventional with emphasis on reliability.

Radar mapmatching provides accuracy equal to homing systems, and this type of operation may be most useful. The system is better suited as a terminal guidance system than a midcourse system because it does not give control on overwater flights of any length or on flights over terrain lacking distinguishing features. Electronic countermeasures are a big drawback to the use of the system; a protection against countermeasures is the highly directional antenna.

As a result of the drawbacks, this system may be used in combination with a fairly accurate and nonradiating midcourse system. The terminal guidance would be used for a minimum of time prior to explosion so as to afford the greatest surprise since this would be the best way of circumventing possible countermeasures.

Magnetic Guidance Systems

A comparatively economical, although not too accurate, system primarily suited to a long-range midcourse guidance is a magnetic guidance system. The characteristics of the magnetic field around the earth are fairly well known and quite predictable. The magnetic field provides another method of determining a line of position. Three characteristics of the earth magnetic field that would be useful in guidance are (1) lines of equal magnetic deviation (isogonic), (2) lines of equal magnetic inclination or dip (isoclinic), and (3) lines of equal magnetic intensity.

A magnetic compass and its refinements, the flux gate compass and gyrosyn compass, furnish one method of using the earth magnetic field for navigational purposes. Through use of the compass alone, no means can be devised to recognize or compensate for drift that occurs in a missile. A line of equal magnetic intensity exists uniquely through a set of points and can be measured and charted. Equipment designed to measure the intensity of the earth magnetic field, an example of which is the flux valve, has been devised for various purposes and can be used in missile guidance systems.

With an external magnetic field present, the signal output at a flux valve is the second harmonic of the excitation frequency. Without the external magnetic field, no second harmonic signal would be evident. The strength of any second harmonic output is proportional to the intensity of the external magnetic field. The foregoing only holds true if the coil axis of the flux valve is properly aligned with the lines of force of the external field.

Consider the structure of a system which could guide a missile using these magnetic lines. The block diagram in figure 9-45 is an example of such a system. To measure the total intensity of the earth magnetic field, three flux valves are re-

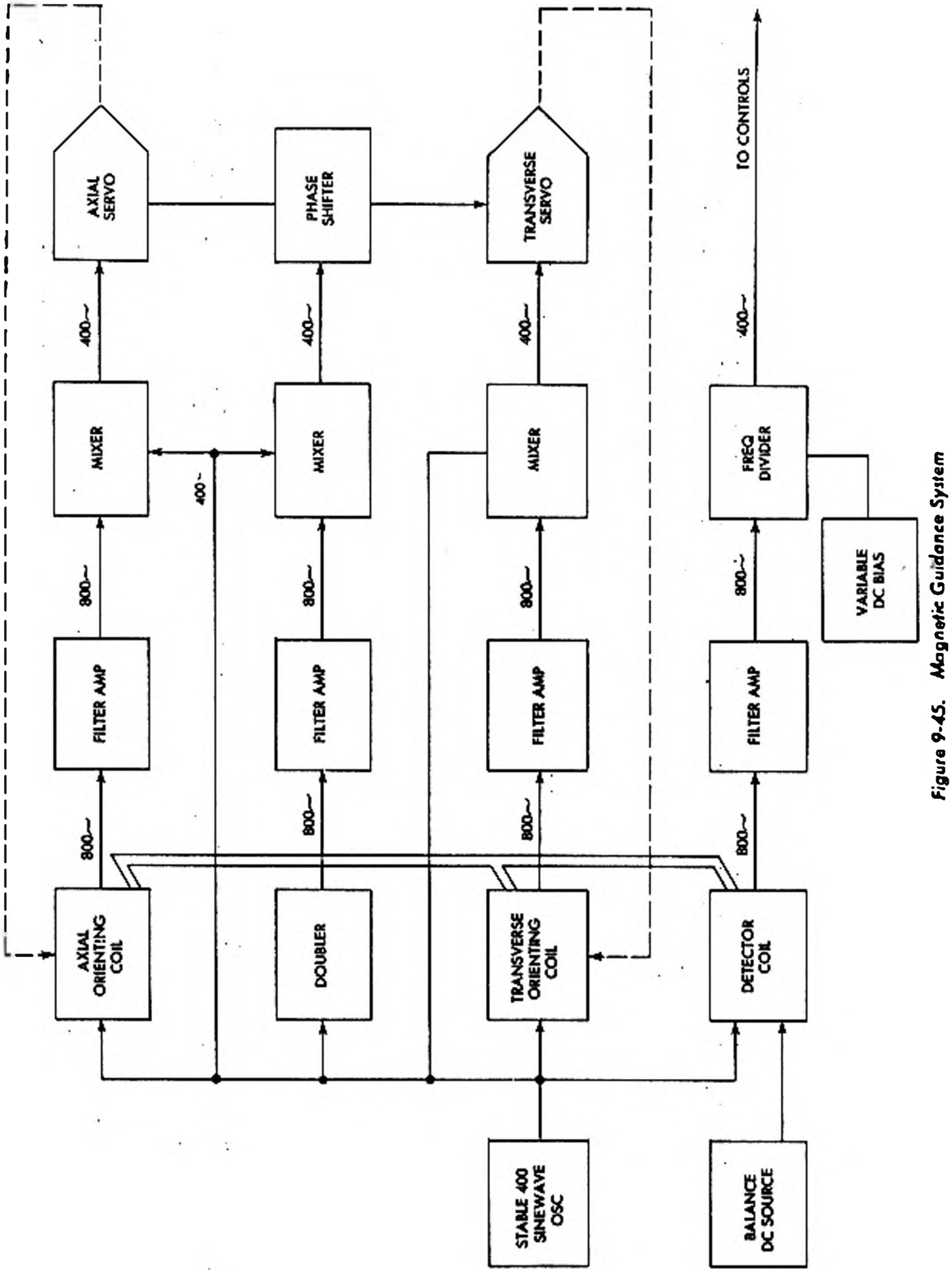


Figure 9-45. Magnetic Guidance System

quired. These elements are aligned along three mutually perpendicular axes and rigidly fastened to one another. These elements, shown in the diagram, are known as the transverse orienting, the axial orienting, and the detector coils. The transverse and axial coils are connected into servo loops in such a way as to null out. The nulling out of the transverse and axial coils with respect to an external magnetic field leaves the detector coil oriented along the direction of the earth total magnetic field.

The signal output of the detector coil is connected to an electronic unit which produces error voltage signals (shown as 400-Hertz signals to the controls). The output of this unit may be adjusted to zero volts for any given magnetic intensity of the adjustment of a variable DC bias. The output error signal is an AC voltage whose amplitude is a measure of the amount of deviation from the desired line of position and whose phase is a measure of the direction of error.

The electronic unit consists of the detector and axial orienting channels. Note that the oscillator supplies signals of the fundamental frequency to all three channels. Automatic volume control and filter elements which are used to insure signals of constant amplitude free from harmonics are incorporated in the stable oscillating unit.

The detector coil output, distorted by the effect of magnetic fields described previously, is connected to a filter which removes the fundamental frequency. The resultant second harmonic signal is connected through an amplifier to a frequency divider, which converts the signal frequency to that required by the control system.

A source of regulated DC, shown being fed into the detector coil, is provided to balance out part of the earth magnetic field and also to provide the zero adjustment for any field-strength measurements.

The axial and transverse orienting channels are identical in structure to the detector channel. The second harmonic outputs of these channels are heterodyned with a proper signal to feed one phase of the positioning servos.

A magnetic system consists of two components, a magnetometer and an electronic unit. The magnetometer would be so located as to isolate it from stray fields arising in the missile equipment. Tactically, magnetic guidance increases system security. For midcourse flight it permits a silent missile which makes detection difficult and, for all practical purposes, eliminates the possibility of jamming. It has the added advantage of unlimited traffic and target-handling ability.

Accuracy can be good up to about seven miles from the target but is limited to the course line only. No method of determining range is available. The missile would have to be launched near, or flown to the vicinity of, the line of intensity that crosses the target area. No evasive action could be taken by the missile.

While the magnetic field at the surface of the earth is accurately charted, the field at higher altitudes is not so well known. The local magnetic irregularities would have less effect at high altitude, so the field of the earth would be more regular and predictable at these altitudes. Magnetic storms and the area near the magnetic poles create unpredictable conditions that prohibit the use of magnetic guidance under such conditions.

SECTION C
Terminal Guidance

In the final phase of its flight, the missile may be guided all the way to the target. These systems are referred to as terminal guidance systems. There are several guided systems that fall into this category. The most common ones are the short-range homing systems or some type of inertial system. These terminal systems may also be the only guidance systems used in short-range missiles.

HOMING GUIDANCE SYSTEMS

A homing system is a specialized form of guidance—selecting and identifying a target through some distinguishing characteristic of the target. Such identifying characteristics as heat or sound from a factory, light from a city, or reflections of radar waves from a ship or aircraft are used as the intelligence to direct the missile to the target.

Homing systems may be classified in three groups: passive, semiactive, and active. A *passive homing system* is one which is designed to select and identify the target by means of natural emanations or radiations from the target itself. Such radiations as heat waves, light waves, and sound waves have been used in passive homing systems.

A *semiactive homing system* is one which selects a target by means of energy from an external source, such as a tracking radar, reflecting from the target. This radar may be ground based or airborne. Equipment used in semiactive homing systems is more complex and bulky than that used in passive systems. It provides homing guidance over much greater ranges and with fewer external limitations in its application.

In an *active homing system*, the target is illuminated by equipment in the missile itself. In an active radar homing system, for example, both the transmitter and the receiver are contained within the missile.

Another way to divide homing systems is by the frequency spectrum to which the system is sensitive (seeks out). A brief description follows of the various types of seekers.

Moving through the spectrum from low to high frequency, *sound* has had some use in seeker systems. Naval torpedoes have been developed as passive sound seekers, but such seekers have certain drawbacks. The sound-seeking missile is limited in range and utility because it must be shielded or built so that its own motor noises and sound from the launching point will not affect the seeker head.

Electromagnetic radiations are the most popular of the media of the homing systems. *Radio* has use in the passive homing system. The seeker acts as an automatic direction finder on a frequency being transmitted from the target area, and it homes on that frequency. There are no weather or visibility restrictions, but it is unlikely that there would be a radio transmitter conveniently operating in a target area. Radio jamming can do a thorough job of "confusing" such a unit.

Radar can be used in any of the three classes of systems but is best suited for semiactive and active. At the present this use of electromagnetic radiation in a target seeker is foremost in effectiveness. Radar is little restricted by weather or visibility; but it is susceptible to enemy jamming.

Heat is best used with a passive type seeker. It is difficult to jam or decoy heat-seeking systems when used against aerial targets because the heat developed by engines and rockets of the aerial targets is difficult to shield. With a sufficiently sensitive sensor, this system is very effective.

Light is useful in a passive seeker system. However, its use is restricted by both weather and visibility. Such a system is quite susceptible to countermeasure techniques.

Various flight paths or trajectories may be determined with respect to fixed targets, but for moving targets special requirements must be met. In homing systems, the sensing elements must be sharply directional to perceive small angular displacements between a missile and a target.

Conical Scanning

To make the sensor directional, it is rotated about its central axis in such a way as to produce conical scanning. This type of scanning is called "conical" because the area viewed or scanned by the rotating sensing element is in the form of a cone with the sensor at its apex. If the target is slightly off the axis of the scanner, the signal emitted by the target (or reflected from it) is stronger at one position of the scanner than at any other position. The amount and direction of its deviation from the axis can be measured and converted into electrical error signals or error voltages.

In this manner conical scanning can be used to provide the control voltage to make the antenna or sensing device follow a target automatically. As in the case of a beam rider, either a rotating parabolic reflector or a rotating sensor (dipole) can be used to develop the conical scan.

When conical scanning is employed in homing systems designed to respond to waves of heat, light, and sound, relationships similar to those found in radar exist between the energy received and pointing errors. Conical scanning provides means for generating error voltages proportional to position errors between the missile and the target and may be used with many different types of sensing devices. The accuracy obtained from conical scanning systems, in addition to their adaptability and ruggedness, has led to their common use in missile homing systems.

Two or more sensing devices may be fixed in position with respect to one reference axis, such as the longitudinal axis of the missile, to provide up-down or right-left position data. Such a homing system would function only in a missile that was perfectly roll stabilized. If the missile should roll about its longitudinal axis, the sensing element

designed to view a sector around the elevation or pitch axis might be displaced so that it would view a sector around the azimuth or yaw axis and, therefore, cause actuation of the wrong set of controls. In other instances, especially in tracking a moving target, the fixed scanner would require a constant reference plane to develop accurate position signals.

Applications of Homing Systems

The infrared homing devices are suitable for use against such targets as mills, factories, bridges, railroad yards, jet aircraft, troop concentrations, ships, or any targets which present large temperature differentials with their surroundings.

The degree of temperature of the target is not important, but the *difference in temperature* be-

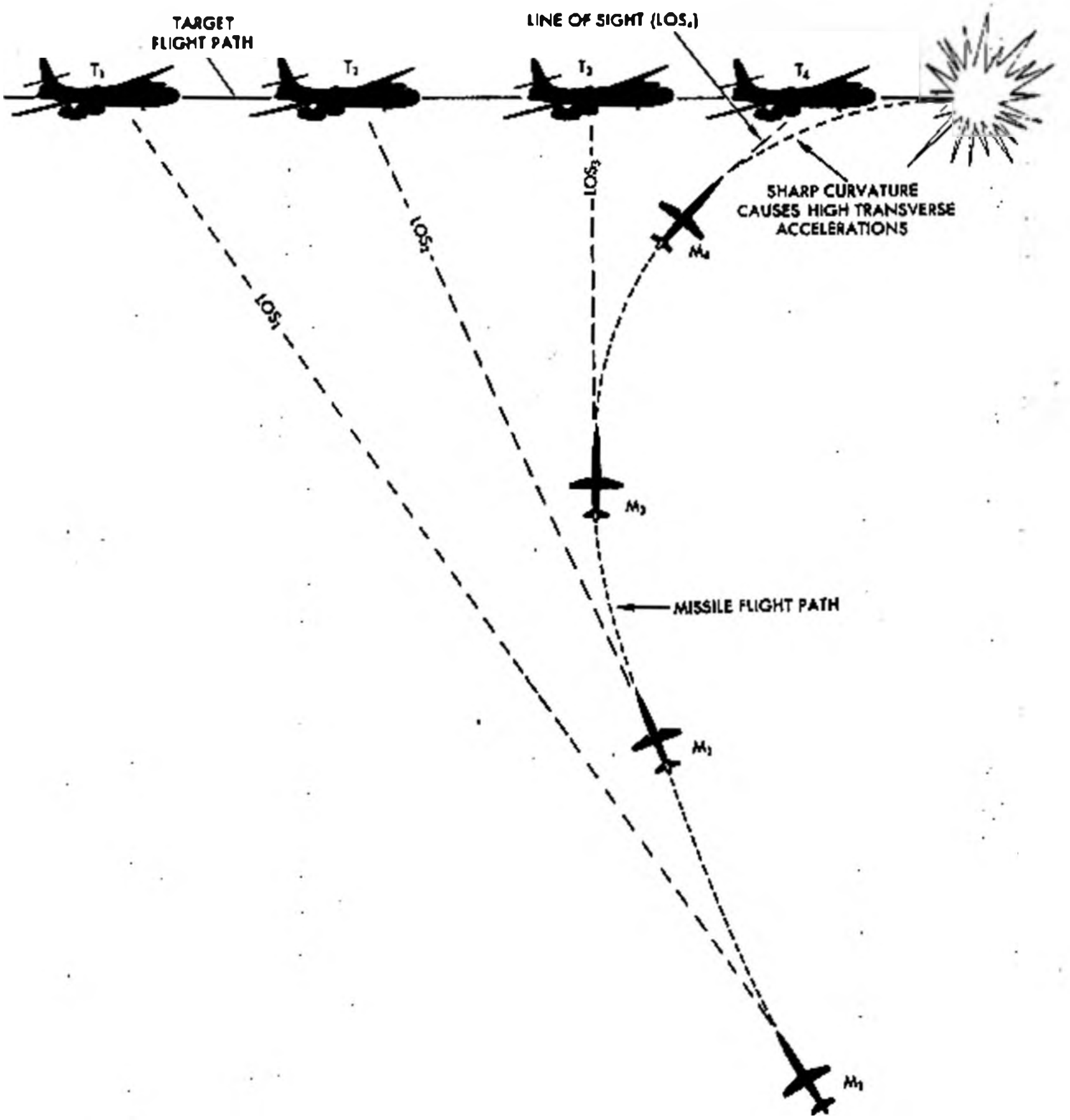


Figure 9-46. Pursuit Homing

tween the target and its surroundings is the factor which enables the heat seeker to identify the target.

Compare this principle to light seekers, which are used in missiles designed for use against targets which present a *contrast in illumination* against their background light. Again, it is the contrast of the light and not its intensity which enables the light seeker to home on its target.

Sound trackers are suitable for use against such targets as ships, wherein the sound tracker homes on the sound of the propellers. Such a device would be used on a torpedo. Although underwater attack missiles are included in the designation of missiles, the use of sound trackers by USAF is unlikely.

All homing systems are subject to limitations in use. For example, the heat seeker requires a clear, moisture-free atmosphere and could be led astray by countermeaures such as fires set to guide it away from its target. The light seeker may be decoyed away from its objective by bright flares. This system requires an unobstructed view of a light-emitting target for accurate homing. Sound seekers are subject to diversion unless they are designed to home on sound of a specific frequency, such as that produced by a ship's propellers turning at a constant rate.

Because of the need for an accurate homing procedure during the terminal phase against an air target, several methods of guidance have been tried. Two methods which have proved satisfactory so far are pursuit homing and lead homing. Either may use infrared, light, or radar homing heads.

PURSUIT HOMING. Pursuit homing is that type of guidance which requires that a missile travel directly toward the target at all times, as shown in figure 9-46. A fundamental requirement of any homing system is that the sensing scanner be accurately alined with the longitudinal axis of the missile in which it is installed. The controls are actuated so that its longitudinal axis is always in line with the target. If the lateral or sideways accelerations required on such a course are computed, they are found to become increasingly larger on the last part of the flight path. It is this portion of the flight path which is most critical in obtaining a hit. Since the antiaircraft application is one of the most important of the fire control problems, the use of pursuit homing against fast-moving targets presents certain difficulties. Pursuit

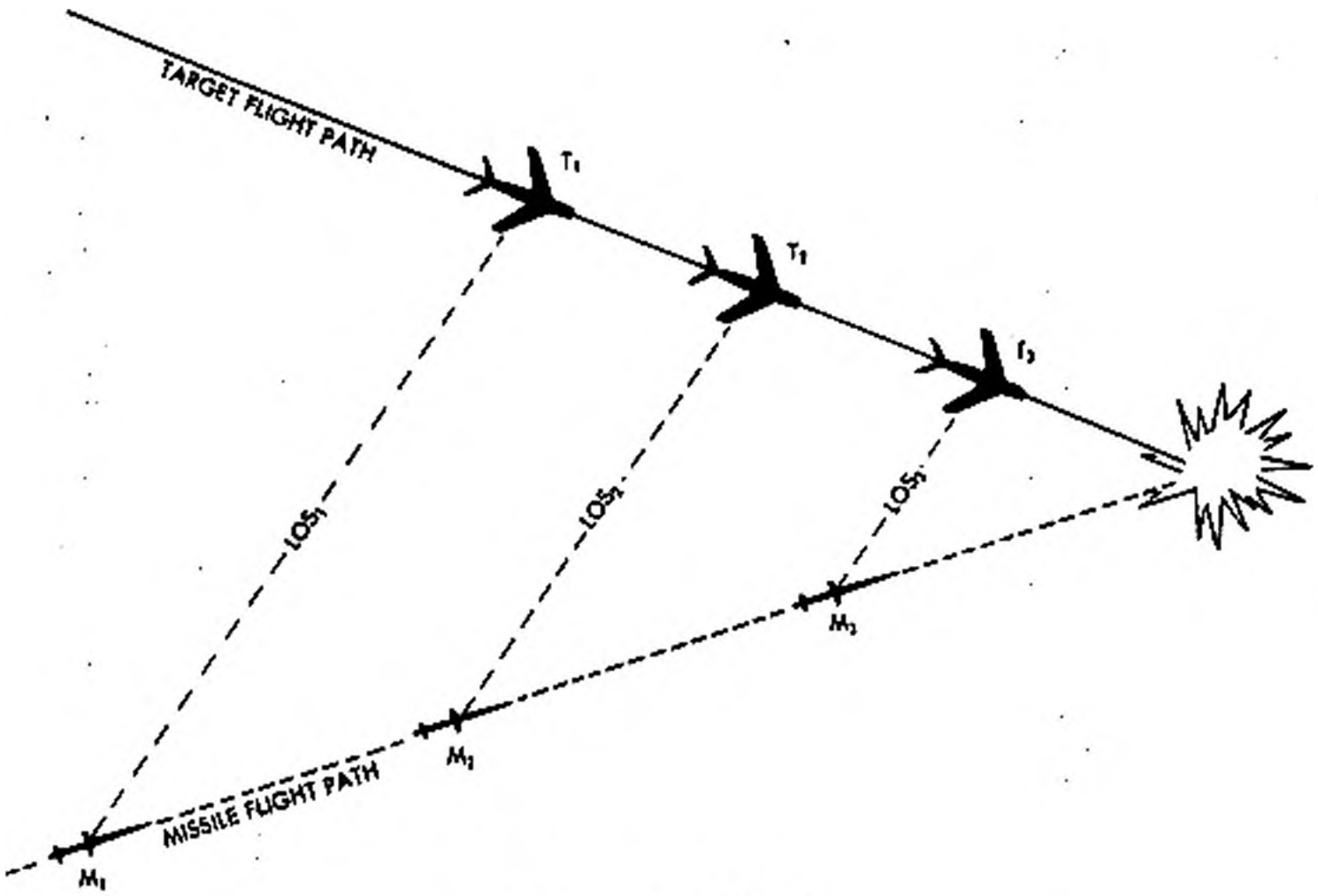


Figure 9-47. Lead Homing

homing at the present time appears limited in application.

LEAD HOMING. Compare this type of homing to pursuit homing. Assume that the homing head or scanner is able to determine the angular rate of change of its course direction as the missile pursues the target. This must be in addition to determining the direction of the target. If the missile can be set on a course which keeps its angular rate of change at zero, the condition under which the missile travels directly toward a collision point is established.

See figure 9-47. Notice that the missile completes the final and critical portion of its trajectory in a straight line. Both vertical and lateral accelerations are small and the most efficient flight path is attained. To obtain this increased intelligence, the homing head requires more complex equipment.

Most missiles equipped with active or semiactive homing systems of the radar type use lead homing guidance and follow a trajectory known as a constant true bearing (CTB) course. On a missile, this type of guidance seldom calls for lateral accelerations which are greater than those of the target. CTB is defined as guidance along a path such that the absolute bearing angle between the missile and the target remains constant.

It is not possible to construct a control system which will guide a missile along a pure CTB path. If the absolute bearing angle is to be kept constant, all movements of the target would have to be instantly sensed and counteracted without any lag whatsoever. The missile cannot detect changes in velocity of the target but only changes in its target path. To effectively use CTB, the missile must detect a small change in the absolute bearing angle before it can begin to follow the movement of the target.

The resulting trajectory is only an approximation of a pure CTB path and is known as proportional guidance. When there is an angular displacement in the line of sight, the missile angular velocity or lateral acceleration will be made proportional to this error angle. The error angle is the difference between the new angle of sight and the original absolute bearing angle. This system requires a certain amount of lead if it is to operate with stability.

The "tracking loop" in the equipment acts to track the target and then give an output voltage to the homing computer which is proportional to the rate of change of the bearing angle. Once the target-bearing information has been acquired by the sensor unit, the computer converts the information into electrical signals (voltages) suitable for controlling the flight path to the target.

PASSIVE HOMING SYSTEMS

The basic components of a passive homing system includes such items as a sensing element, filters, a scanning system, a reference generator, amplifiers, phase inverters, a square wave gen-

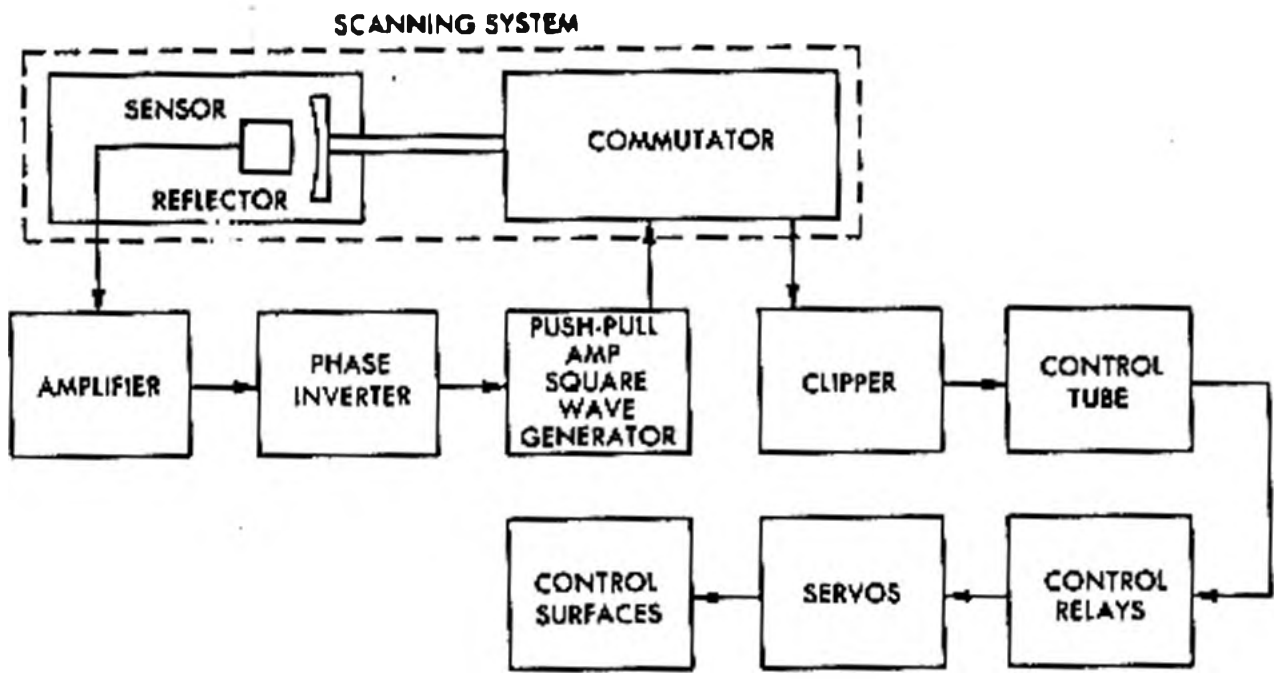


Figure 9-48. Passive Homing System

erator, error signal computers, control circuits, and relays.

The principal differences within passive homing systems lie in the type of sensing elements and scanning devices employed. The block diagram shown in figure 9-48 is representative of passive homing systems in general. The sensor to be used in such a system is determined by the characteristic radiation from the target. Heat sensitive or infrared sensors are sometimes of the lead sulphide type, or more often of the nickel oxide type commonly known as bolometers.

Scanners, Reference Units, and Computers

The parabolic reflector surface or mirror used in scanners must be of some material which will not absorb the radiations received from the target but will reflect them to the sensing element. Refer to figure 9-49. Notice how the sensor is located on the axis of rotation of the reflector. The entire scanning assembly should be well balanced and free from mechanical and optical defects. It should be mounted in a chamber to protect it from external pressure, moisture, and any other factors which might interfere with its operation. Other scanning methods applicable to homing systems include the fixed sector scanner, wherein several sensors are used, with each sensor mounted or focused to view a specific portion of the total area scanned.

The reference unit or reference generator used in most homing systems is a two-phase generator with output voltages 90° apart. One output may be used as a reference for elevation and the other for azimuth. Some systems use regulated voltages from a power supply for reference, while other systems may use batteries which are carefully calibrated to maintain fixed DC reference levels. Regardless of type, the reference unit supplies the standard with which the received signal is compared to determine the variations in amplitude and phase which result from position error. Passive homing systems, such as a heat seeker employed in small missiles, generally use battery power supplies from which the reference voltages are obtained. In such systems mechanical switching devices or commutators may be used to provide phase comparison with the DC reference.

The course computer or comparator unit contains the circuitry in which the signals from the target are compared with the reference voltage. This comparison produces the error voltages which are applied to the missile control actuators. The error voltages should be well filtered to provide a steady DC level for activation of the relays. These components may be classified as mixers, integrators, and rate measuring devices.

Power Supplies

Power for the electronic circuits and the servo

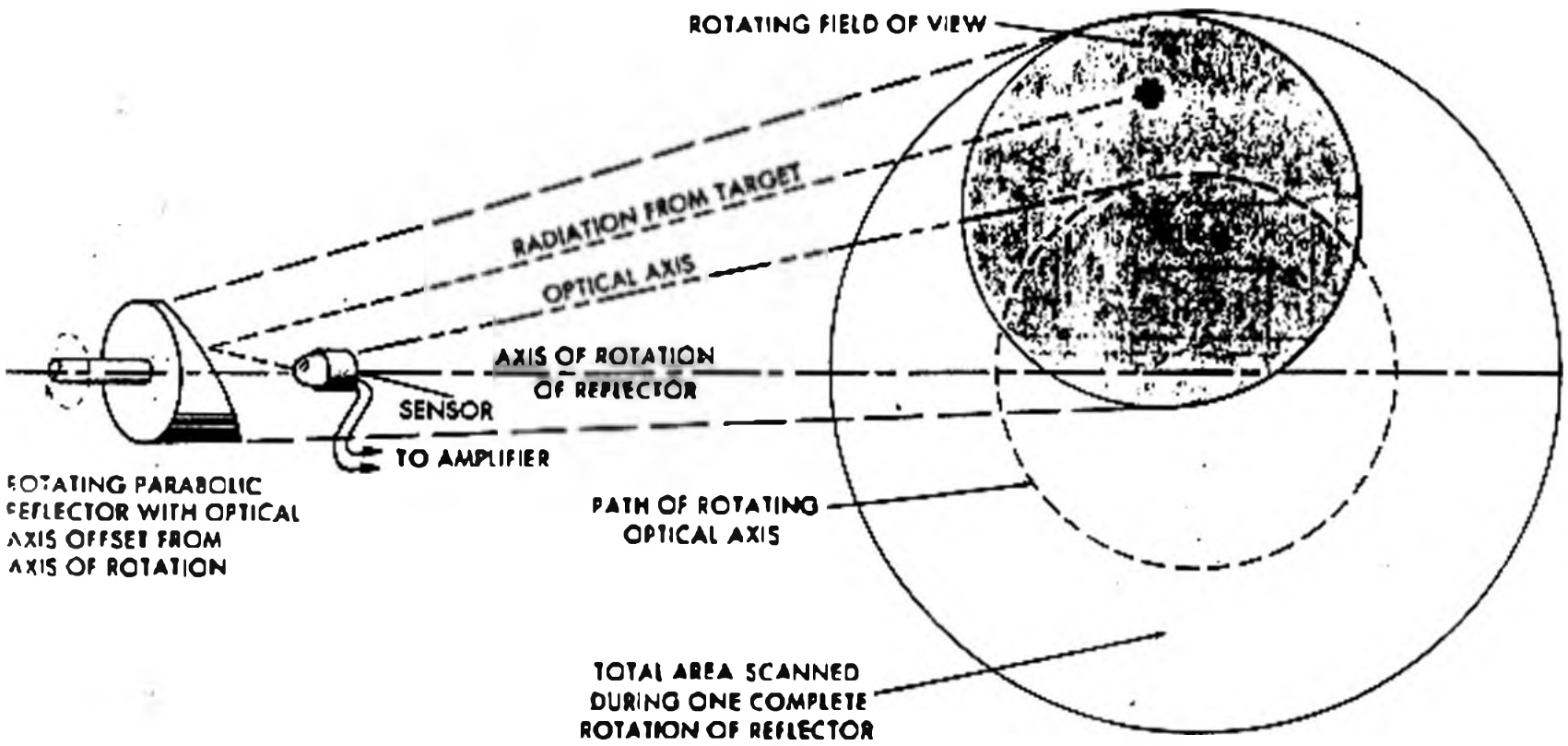


Figure 9-49. Conical Scanning System

controls associated with passive homing systems is primarily supplied from lead-acid storage cells. For circuits requiring alternating currents, the AC may be supplied by small inverters or by vibrator-type power units similar to those used in automobile radios. Low-voltage pulsating DC may be obtained from vibrators raised to higher potentials by means of transformers. The brief period of operation of a passive homing system makes possible the use of relatively small, high-amperage storage cells.

AIM-4D Falcon Missile

The heat-seeker (infrared) homing system is used in the AIM-9B *Sidewinder* and the AIM-4D *Falcon* missiles. Both of these missiles have proven their capability in combat. Figure 9-50 gives a block diagram of the guidance system used in the AIM-4D.

GUIDANCE FUNCTION. The guidance function contains the elements which track the target and steer the missile. Infrared energy emitted by the target is processed into the two signals required for missile operation: the precession error signals for tracking the target and the steering error signals for steering the missile. An error in the missile's collision course, or a maneuver by the target, will cause rotation of the line-of-sight as seen by the missile. This line-of-sight rotation develops signals which cause the receiver to track and produce proportional deflections of the missile control surfaces. The missile then attempts to counteract any change in the target's bearing by flying a new collision course.

The guidance function consists of five sections: receiving section, error signal section, receiver gyro control section, steering signal section, and steering servo section. The receiving section converts the target energy into an electrical signal, the modulated carrier signal, and develops AGC voltages which maintain the carrier signal at constant strength. In the error signal section, the carrier is demodulated to separate the modulation envelope, which contains the useful tracking and steering information, from the carrier. In the yaw and pitch error signal channels the modulation information is compared with the reference pips from the phase reference generator channel and resolved.

RECEIVING SECTION. The receiving section detects infrared energy emitted by the target and produces the carrier signal with modulation phased according to the target position. The modulated carrier is sent to an error signal section for phase detection into its yaw and pitch components.

ERROR SIGNAL SECTION. Refer to figure 9-50. The received signal is demodulated, amplified, and resolved into yaw and pitch error signals by comparison with signals developed by a phase reference generator. These error signals are applied to the receiver gyro control section to correct optical system tracking, and to the steering section to effect guidance of the missile. The error signal consists of the tracking error signal channel, yaw and pitch error signal channels, and the phase reference channel.

Tracking error signal channel. The tracking error signal channel detects the error signal, amplifies it, and applies it to the yaw and pitch error signal channels.

The tracking error detector, CR 505, demodulates the carrier signal and produces a tracking error signal that is applied to K$_t$GAIN adjust R401 (figure 9-50). R401 sets the total gain of the tracking error signal channel and contributes to the overall tracking time constant.

The X TALK R402, in conjunction with K$_t$GAIN R401 and C402, forms a phase shifting network which corrects the phase of the tracking error signal so it can be properly compared with the phase reference pips in the error signal channels. The phase reference pips come from the phase reference generator in the phase reference channel. These pips control the sampling of the error signal by the error signal channel and represent the nutational position of the telescope optical axis relative to the missile fuselage, thus providing tracking orientation. The tracking error signal, in passing through the receiving section, is delayed because of the inherent phase shifting characteristics of each stage. Without further compensation, the tracking error signal would arrive at the error signal channels in a phase that would give incorrect coordination with the phase reference pips. This would result in cross-talk between the yaw error signal channel and the pitch error signal channel.

The tracking error signal is amplified by amplifier V401. The signal is then fed to cathode

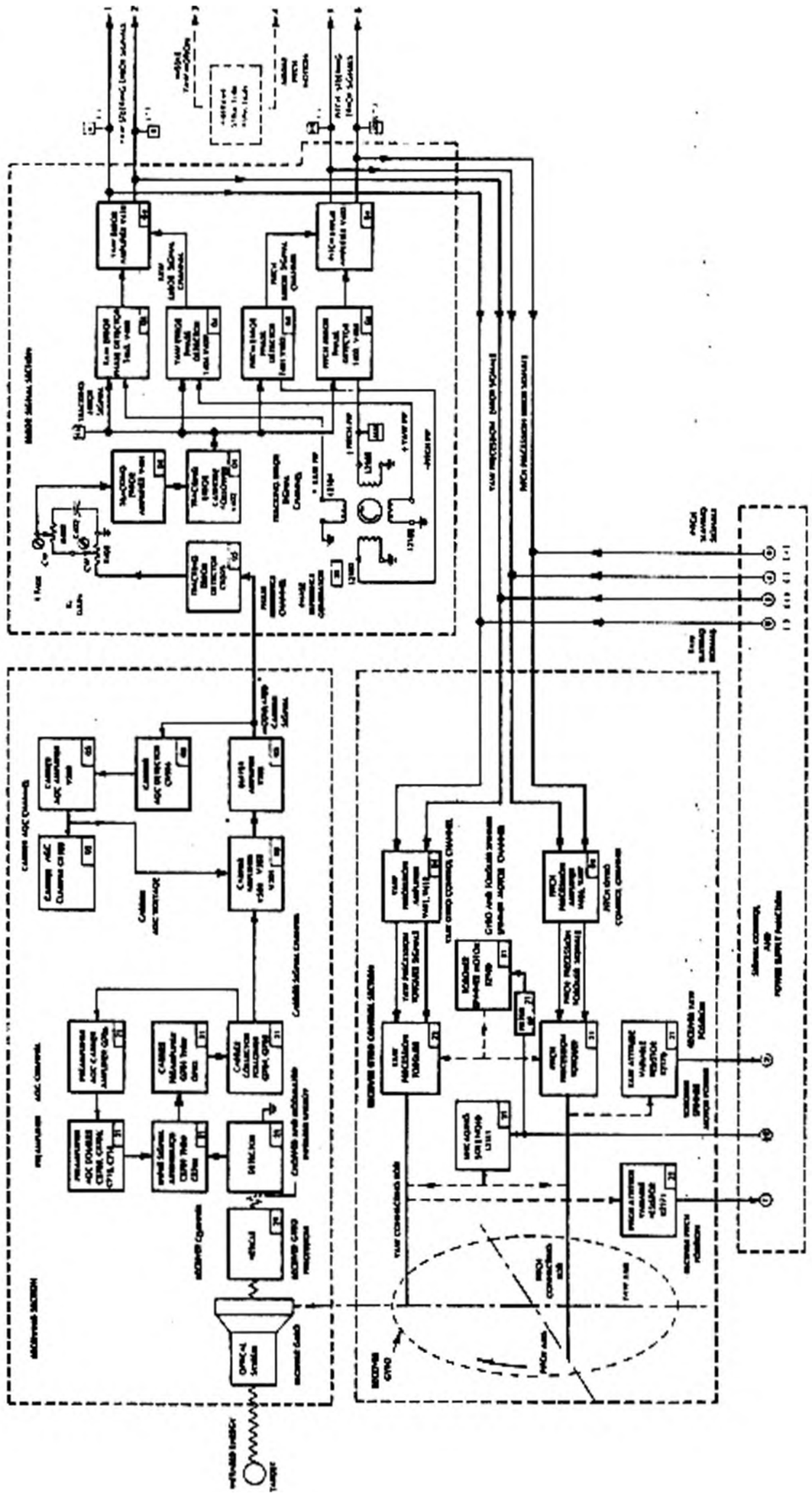


Figure 9-30. Guidance Function Detailed Block Diagram

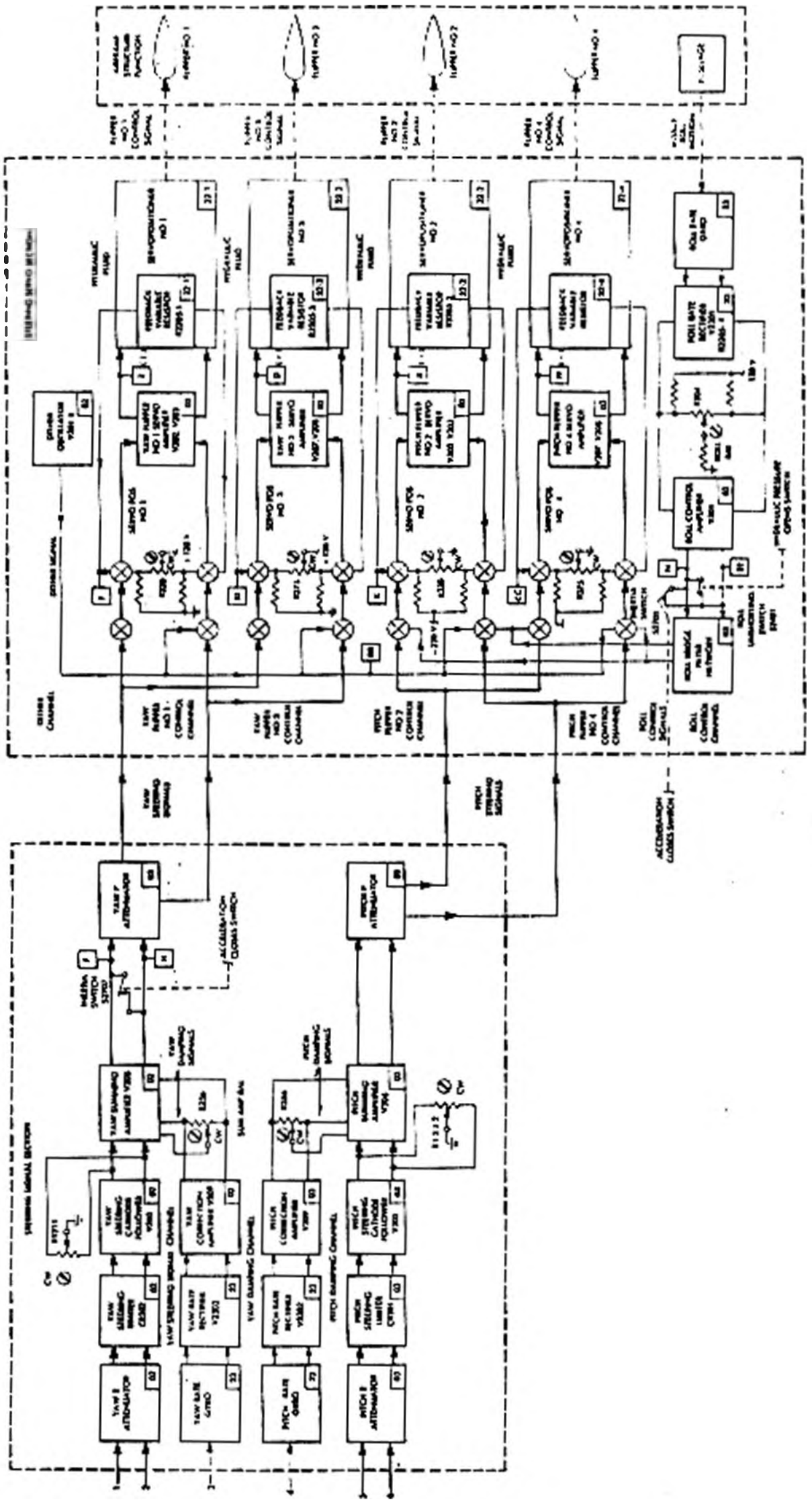
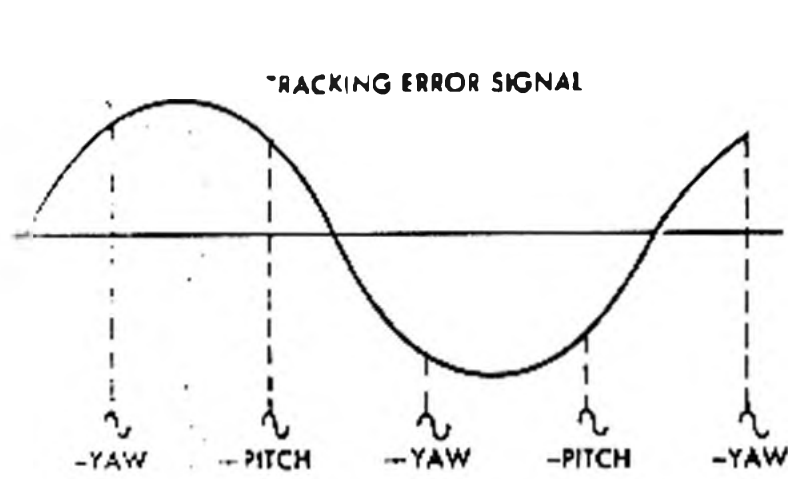
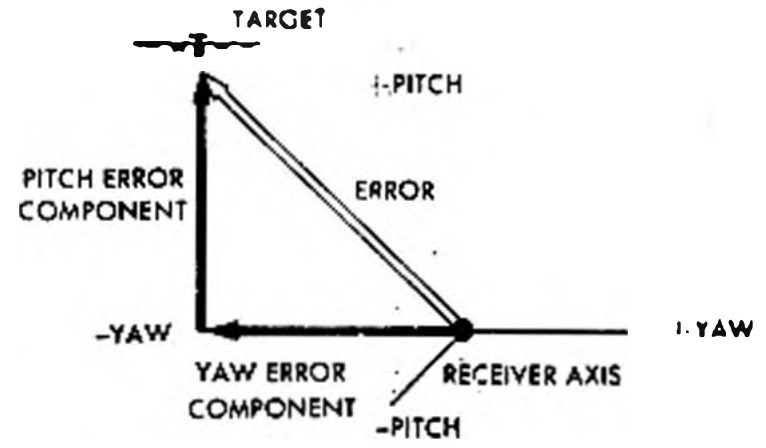
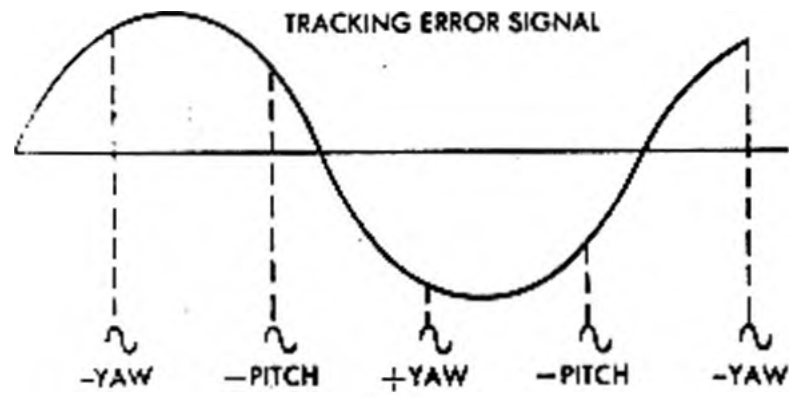
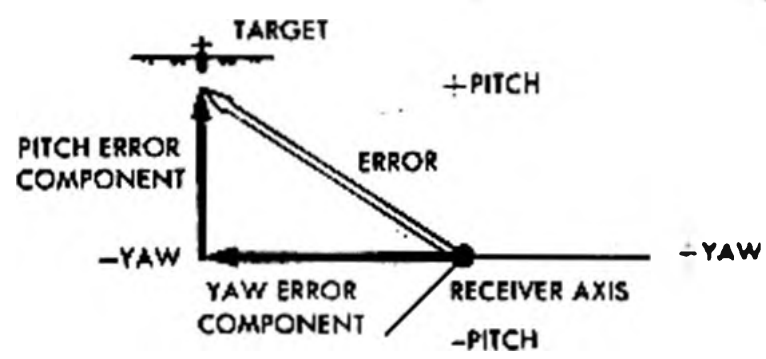


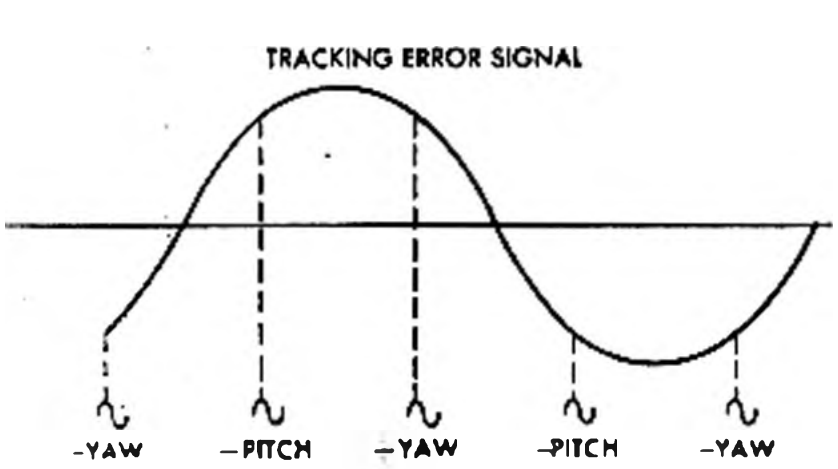
Figure 9-50. (continued)

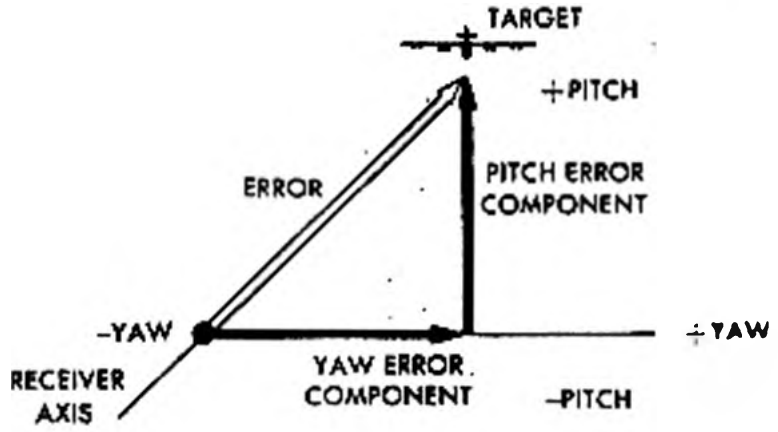


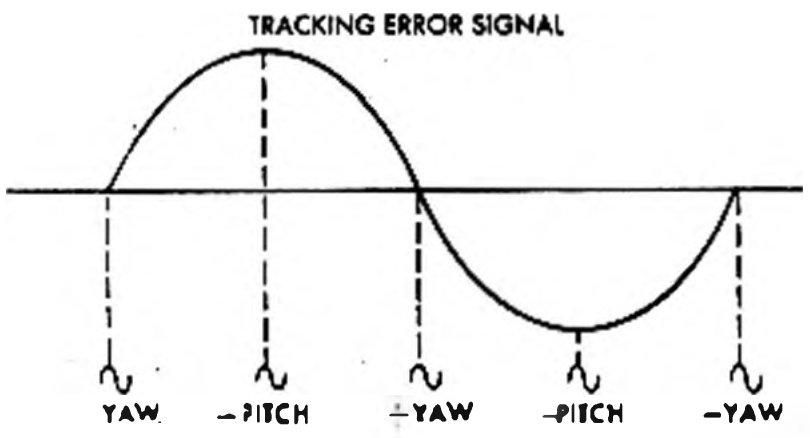












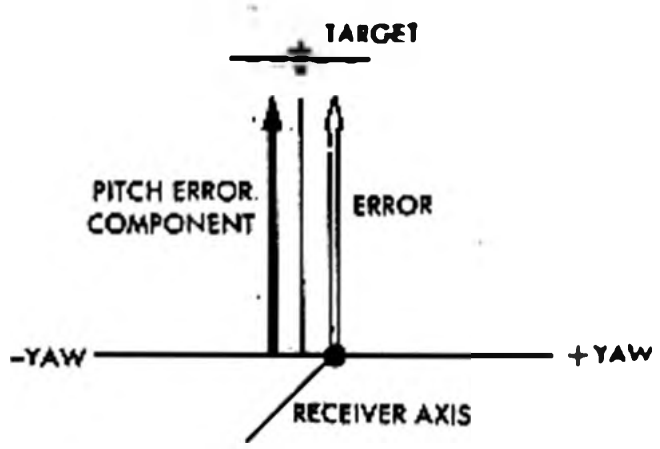


Figure 9-51. Resolution of Tracking Error

follower V402 which provides a low-impedance source for driving the error phase detectors. The voltage developed across the cathode resistor is coupled to the yaw and pitch error phase detectors.

Yaw error signal channel. This channel extracts the yaw information from the tracking error signal and develops signals which are used to steer the missile and reposition the receiver. The yaw error phase detector, V408, is gated on by the +yaw pips and conducts for a short time, either charging or discharging a storage capacitor. The direction of conduction depends upon the relative time of occurrence of the pip and the phase of the tracking error signal at that instant. The magnitude of conduction depends upon the amount of pointing error of the receiver gyro. The pip voltages indicate the instant at which the optical system is pointing in either the +yaw, +pitch, -yaw, or -pitch direction. If the missile sees the target at a point above it and to the left, figure 9-51, the error may be represented by a vector which can be resolved into a pitch error component and a yaw error component. In this case, the -yaw pip would occur while the tracking error signal is positive. Because of the position of the target, the pitch error component and the yaw error component are nearly equal. The -yaw pip and the + pitch pip occur at such a time that yaw error phase detector V408 and pitch error phase dector V404 conduct equally and put equal charges on their respective storage capacitors. The +yaw pip and -pitch pip occur 180° later and cause yaw error phase detector V409 and pitch error phase detector V403 to conduct equally but in the opposite direction to V408 and V404. The output of the yaw error detectors is a push-pull DC voltage. If the target appeared closer to the -yaw axis, the yaw error component would be larger than the pitch error component. The -yaw pip would then occur closer to the maximum of the tracking error signal and the +pitch pip would be closer to the crossover point. This would be manifest by greater conduction in the yaw error phase detectors than in the pitch error phase detectors. Because the target would still appear between the +pitch and -yaw axis, the +pitch pip and -yaw pip would still occur during the positive portion of the tracking error signal.

If the target appeared between the +yaw and the +pitch axes, phase of the tracking error signal would change and the signal would be positive during the +yaw and +pitch pips. If the error was in pitch only, the +pitch and -pitch pips would occur during the positive and negative maximums, respectively, of the tracking error signal and the yaw pips would occur at the crossover points since there is no yaw component.

Pitch error signal channel. The pitch error signal channel operates in the same manner as the yaw error signal channel. The channel receives the same tracking error that is applied to the yaw channel, but the reference pips are the + and - pitch pips from the phase reference generator.

Phase reference channel. This channel consists of a phase reference generator which develops the pips used to determine the times at which the tracking error signal is sampled. The phase reference generator is physically an integral part of the receiver gyro. The pips are related to the rotational position of the telescope, and provide spare reference for the steering error and precession error signals.

The generator (figure 9-52) consists of four coils mounted on the stator of the receiver gyro and a small permanent magnet mounted on the rotor. The four coils are equally spaced at 90° intervals around the stator. As the rotor spins, the magnet induces a voltage in each coil. A positive voltage is induced while the magnet is approaching the coil. When the magnet is directly opposite the coil, the voltage decreases to zero. As the magnet recedes, it induces a negative voltage. Because of the speed of the motor these voltages are very short positive or negative pips. The pips are applied to the phase detectors where they serve as reference voltages to key the sampling of the tracking error signals.

RECEIVER GYRO CONTROL SECTION. This section (refer to figure 9-50) positions the receiver gyro (and thus the optical system which is mounted on it) in response to slaving signals or precessing error signals. The torquer spinner motor provides the force that positions the gyro by precession. Torque is split into two directional forces to cause yaw and pitch gyro precession by torquers in the yaw and pitch gyro control channels.

GYRO TRACKING. See figure 9-53. The receiver

gyro is mounted on bearings which allows it to rotate about the pitch and yaw of the missile. Since the optical system is mounted on the gyro, any movement of the gyro is in fact a movement of the optical system. The gyro stabilizes the optical system in space; it maintains its spin axis in a fixed direction in space regardless of the turning motions of the gyro mount (missile fuselage). The gyro is positioned by using its properties of precession. Any force which attempts to turn the gyro is resisted. The force does have an effect which is felt 90° away in the direction in which

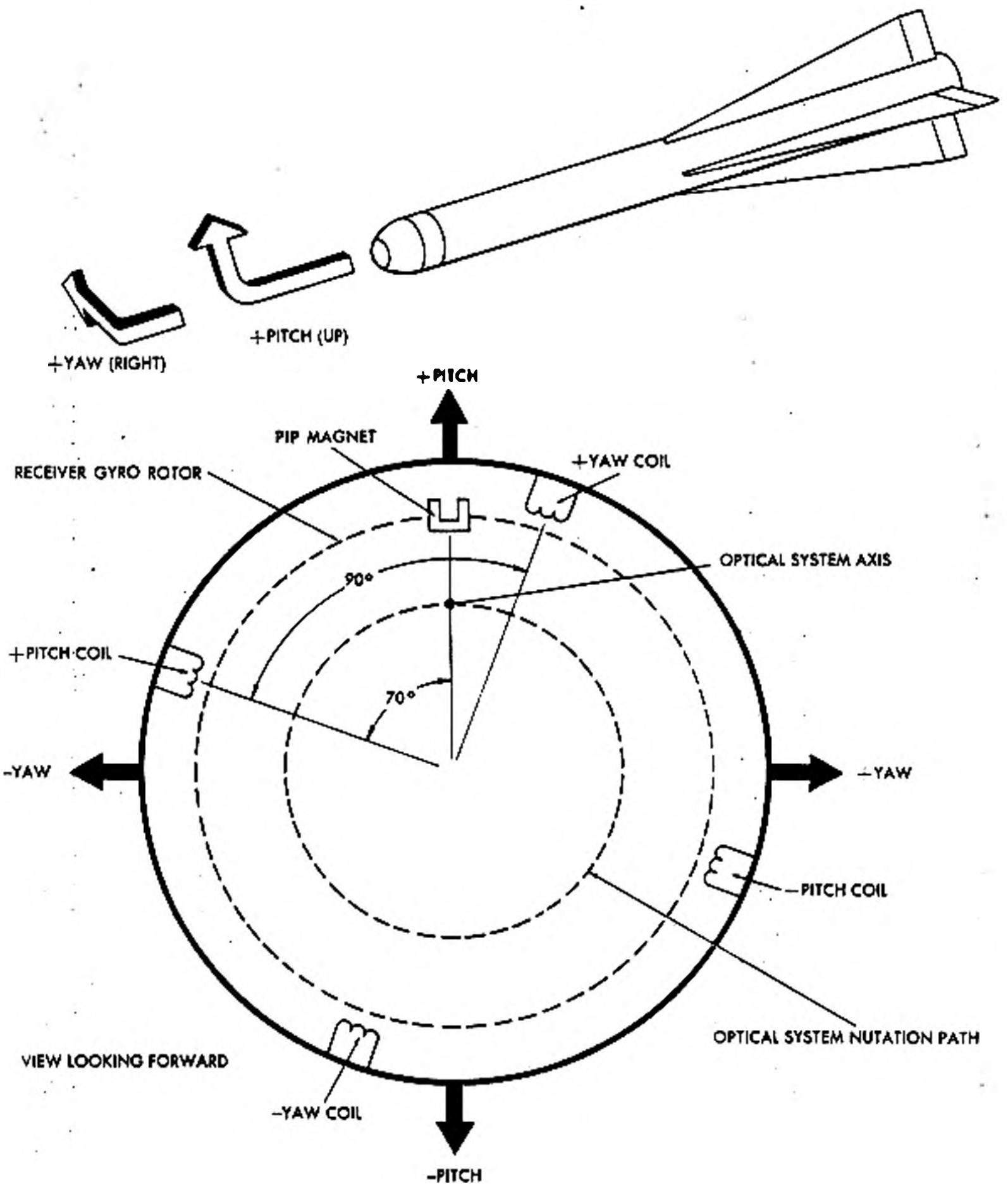


Figure 9-52. Generation of Phase Reference Pips

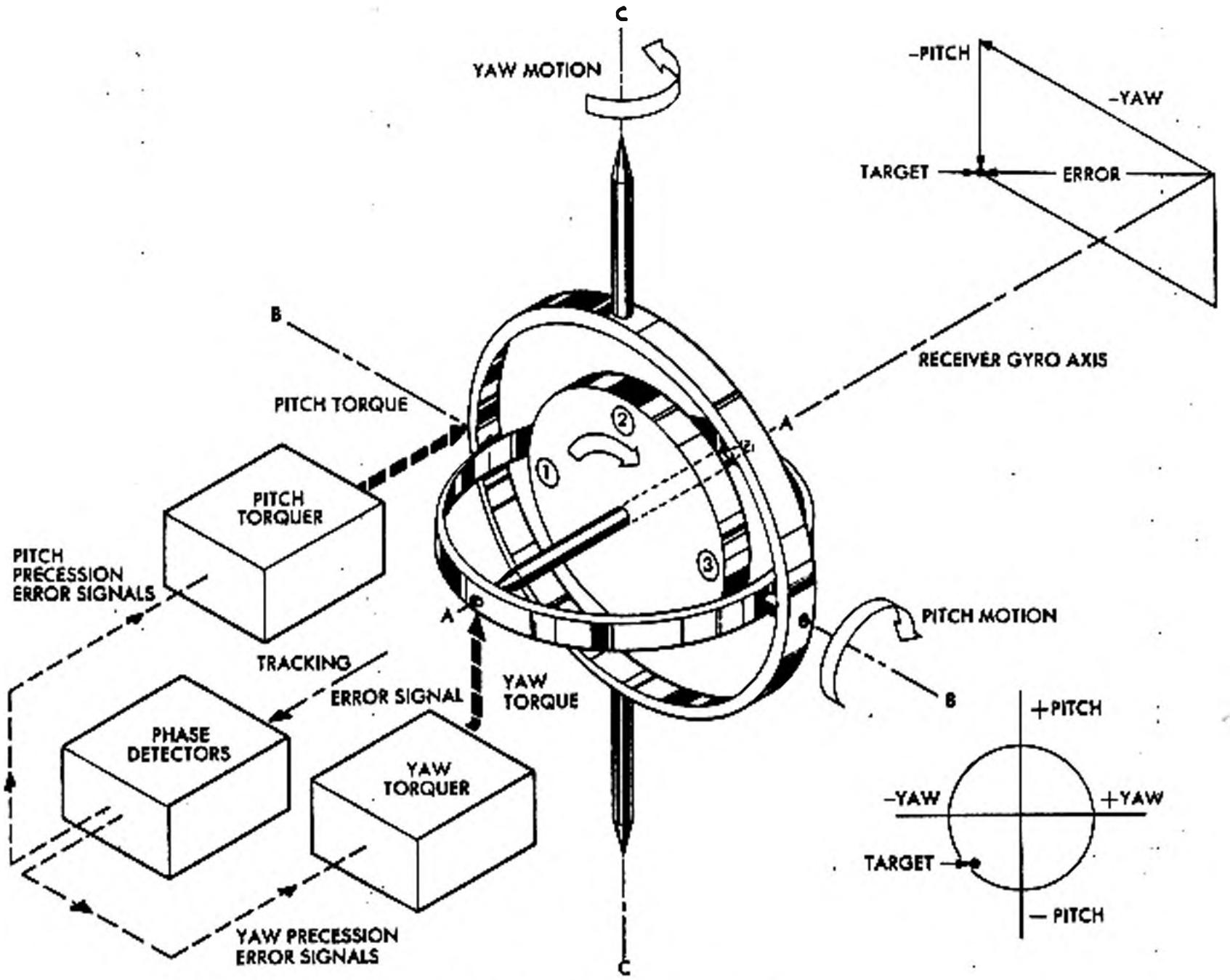


Figure 9-53. Gyro Tracking

the rotor is spinning. For example, if a move in the yaw direction is desired, a torque is applied about the pitch axis. This torque is felt 90° away, resulting in yaw precession.

In figure 9-53 the gimbal mountings are equivalent to the yaw and pitch bearings. The pitch torquer is energized and is applying torque about axis C-C. This torque is attempting to rotate the gyro assembly to the right and is equivalent to applying a force at point (1) in the forward direction. The effect of this force is felt at point (2) and causes the gyro to rotate about axis B-B in the direction shown by the pitch motion arrow. Similarly, yaw torque from the yaw torquer is equivalent to a force in the forward direction at point (2).

The effect of this force is felt at point (3) and causes yaw motion about the C-C axis. The combination of these motions causes the gyro to point down and toward the left, which is the correct direction to reduce the tracking error.

STEERING SIGNAL SECTION. The steering signal section processes the push-pull steering error signals developed in the error signal section and applies them to the steering servo section. (Refer to figure 9-50 continued.) The steering signal section consists of the yaw and pitch steering signal channels, and the damping channels.

The steering servo section feeds the yaw and pitch error signals to the respective "flipper controls" to correctly steer the missile on its flight path.

The automatic tracking and steering channels are similar in all passive homing systems. The

circuitry and methods to accomplish the steering functions will differ as to the choice of different manufacturers and Air Force requirements.

SEMIACTIVE HOMING SYSTEMS

Semiactive homing systems are generally of the radar type. They are employed in missiles designed for longer ranges than those using passive homing systems. The basic components include tracking radar, sensing units, scanner, reference generator, receiver, range circuits, gating and sweep circuits, control circuits, servo system, and computers. Except for the sensing element, most of the circuit functions are similar to those presented under the passive homing system.

Tracking Radar

The tracking radar used to illuminate the target for semiactive homing may be any conventional radar of the antiaircraft or airborne fire control type designed to track a target automatically. Its radio-frequency and pulse-repetition rate must be acceptable to the receiving equipment used in the missile. Radar sets of this type radiate a high-power RF beam to provide strong echo signals and usually employ conical scanning to track the target.

Sensing Units

In homing systems using radar, the antenna is the sensing element. Antennas used in semiactive homing systems may vary widely in type and configuration. The type of antenna will depend on the electrical characteristics of the radar system as well as on the configuration of the missile.

DIRECT ANTENNA. Direct pulse antennas must provide for unobstructed reception of pulses from the tracking radar throughout a wide range of angular deviations of the missile with respect to the ground radar position or a radar-equipped fighter. These antennas are sometimes mounted in an aft section of the airframe and are usually slotted waveguides.

ECHO ANTENNA. The echo or forward antenna is located in the nose of a missile and usually is a fixed dipole with a rotating offset parabolic reflector to provide conical scanning. In homing systems designed to guide a missile by proportional guidance, the echo antenna and scanner are gimbal-mounted to provide a wide range of angular movement. This enables the target-seeking antenna to lock on the target at an initial bearing angle and thereafter provide the angular changes in the missile trajectory to bring it into collision with the target.

Scanner

Conical scanning is used in semiactive radar-homing systems because it can be attained with a minimum of moving parts, and the components are readily adaptable to most airframe configurations. The high frequencies employed in radar permit the use of very small antennas without impairing the efficiency of the system. This minimizes the size of the parabolic reflector, spinner motor, and reference generator required for conical scanning. The power requirements of such a system are comparatively small.

The functions of other basic components to include reference generators, receivers, range circuits, gating and sweep circuits, control circuits, servos, and computers are explained in a typical semiactive homing system used by the AIM-4A Falcon missile. This guidance system is described in the following paragraphs.

The guidance function steers the missile into collision with the target. The guidance function is composed of five sections: the receiving section, range tracking section, error processing and resolving section, flight control section, and the antenna positioning section. Figure 9-54 shows an overall functional diagram of the system. The receiving section converts the reflected target energy into target error signals. The range tracking section operates in conjunction with the receiving section and provides a means of selecting the reflected target signal and rejecting all others. The error signal from the receiving section is resolved into its pitch and yaw components and converted to DC signals by the error processing and resolving section. These DC signals are applied to the antenna positioning section, causing the missile antenna to follow that target and are also applied to the flight control section to maintain correct steering.

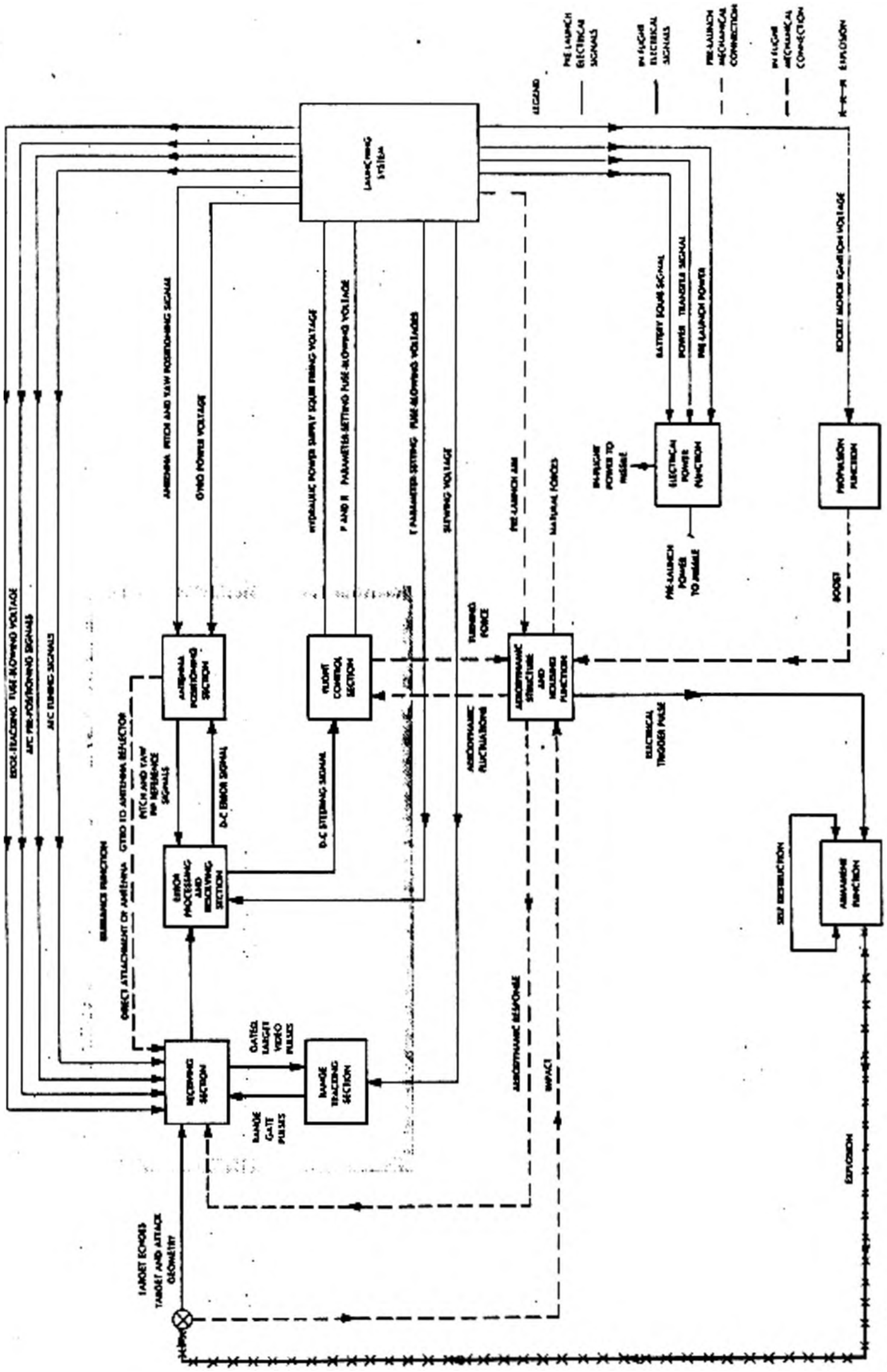


Figure 9-54. Missile Model AIM-4A Functional Block Diagram

Receiving Section

The receiving section receives the echo signals reflected from the target and prepares them for use by the range tracking and error processing and resolving sections. The range tracking section uses the output of the receiver section as a reference in generating range gate pulses to gate the receiver, and the error processing and resolving section uses the output to develop steering signals for the flight control section and error signals for the antenna positioning section. The receiving section consists of the antenna channel, mixer circuit, IF channel, video channel, and AGC channel.

ANTENNA CHANNEL. The antenna channel receives the target echoes, error modulates and feeds them to the mixer circuit. The antenna is a crossed-dipole to give a circular receiving pattern, necessary to make the reception independent of the roll of the missile during flight.

MIXER CIRCUIT. Two microwave signals of different frequencies, when applied to a crystal mixer, produce an IF whose value is the difference of the two incoming frequencies. See figure 9-55. The new signal is detected by the crystal mixer and applied to a coupling network at the output of the crystal which is tuned to the frequency to be passed. This frequency is then fed through the IF stages and several video stages to a gated video circuit, shown in figure 9-56. The video cathode follower V609-A isolates the gated video amplifier from the stages that follow and separates the gated target video pulse from the range gate pedestal. The pulse stretching diode CR602 (see figure 9-57), third video amplifier V401-A, and cathode follower V407-A stretch and amplify the video pulse.

Range Tracking Section

The range-tracking section (refer to figure 9-54) generates receiver gating pulses, time coordinated to select only the desired target video. This section consists of a range-tracking video amplifier circuit, time discriminator circuit, integrator circuit, range gate timing circuit, and a range gate generating circuit. The range-tracking video circuit amplifies the selected target video pulses and applies them to the time discriminator circuit. The time discriminator generates a DC error signal if the range gate timing pulses applied to it do not occur in proper synchronization with selected target video pulses. If each target video pulse from the range-tracking video circuit is evenly bracketed by an early gate pulse and a

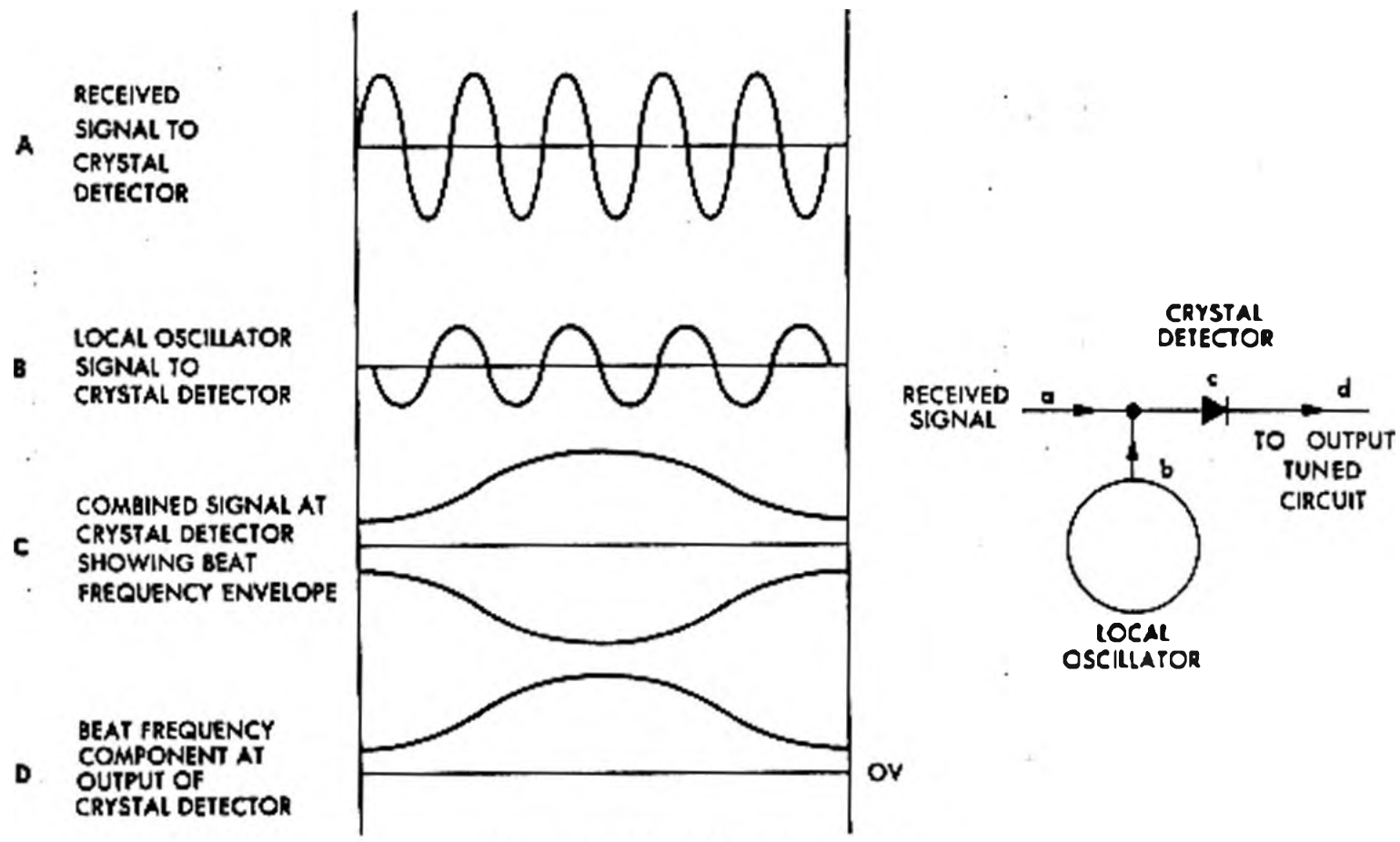


Figure 9-55. Mixer Operation

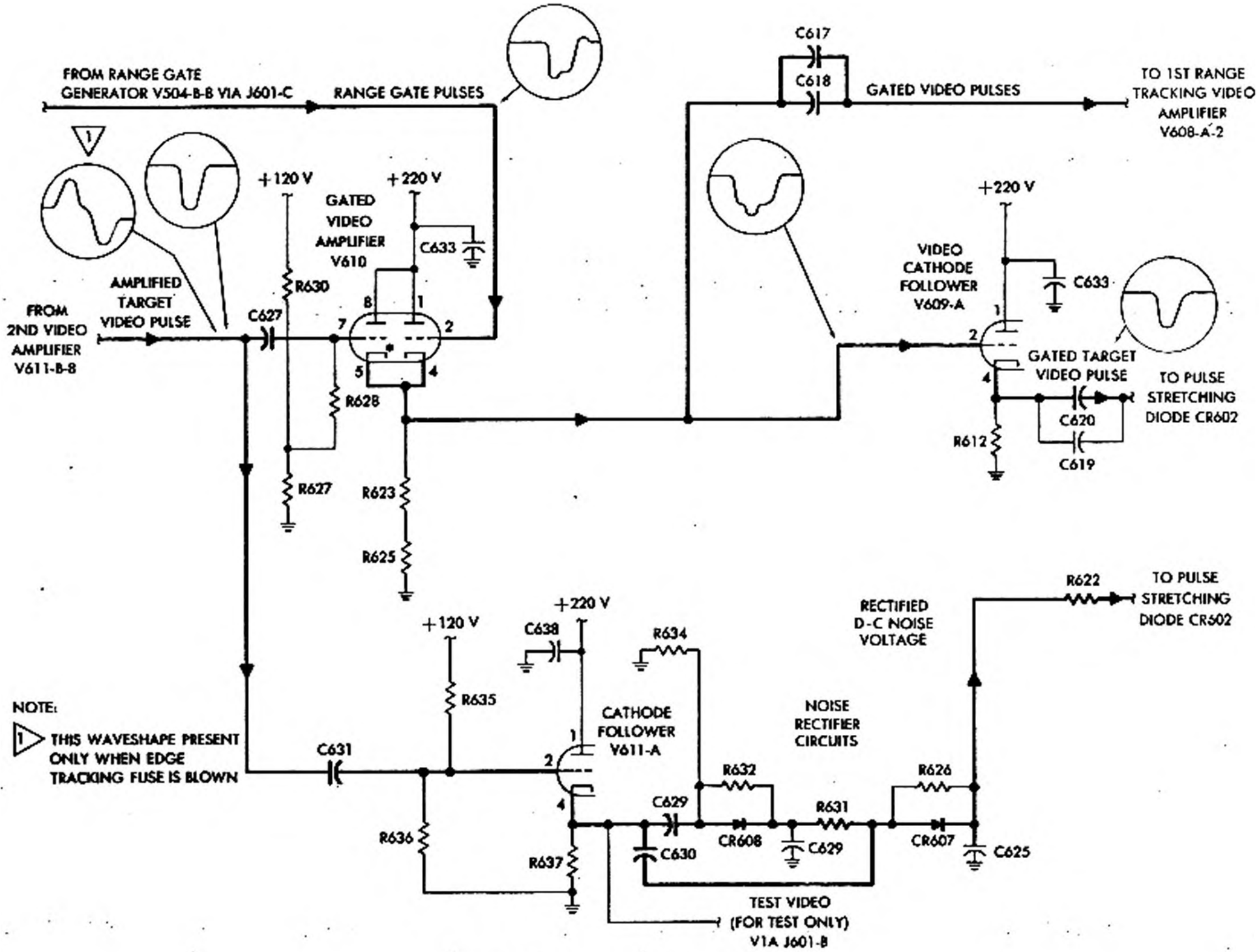


Figure 9-56. Video Circuit Partial Schematic

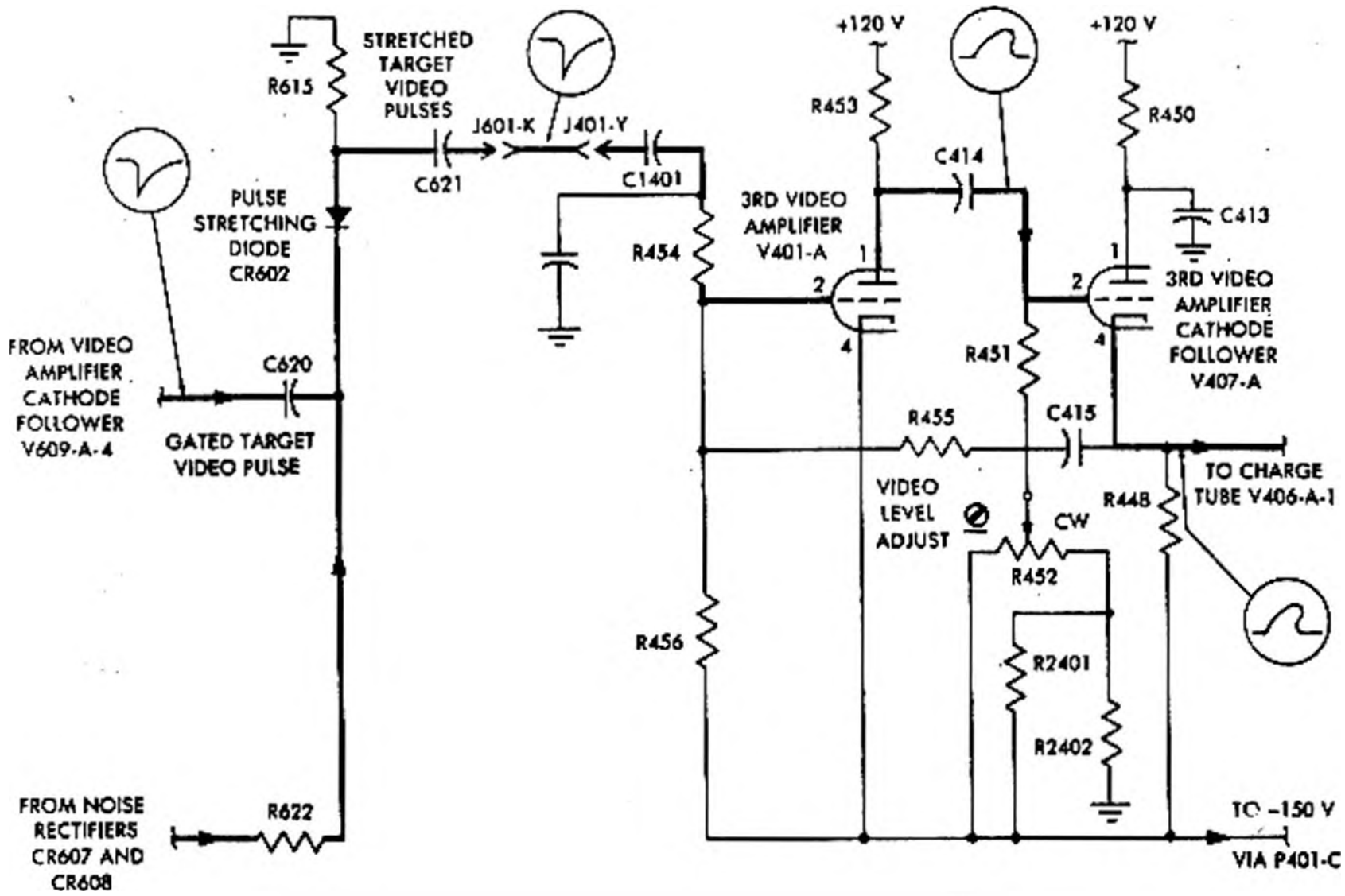


Figure 9-57. Video Amplifier and Shaping Circuit

late gate pulse, the discriminator circuit produces no output error signal. If the pulses are not evenly bracketed, a DC error signal is fed to the integrator circuit.

The integrator circuit smooths and amplifies the error signal and also provides lock-on memory. The output of the integrator circuit is a DC control current, which is applied to the range gate timing circuit, where it is used to control the frequency of a crystal oscillator. The integrator circuit controls the output frequency of the range gate timing circuit.

Error Processing and Resolving Section

This section receives a nominal error signal from the receiving section and converts it to a DC error signal which is applied to the antenna positioning section and flight control section (refer to figure 9-54). The section consists of the error processing channel, and the pitch and yaw error resolving channels. The pitch and yaw error channels compare the error signal with a reference signal and generate the DC error signal used to position the antenna and control the flight of the missile.

Flight Control Section

The flight control section (refer to figure 9-54) provides the turning force for steering the missile to the target. The section receives steering signals from the error processing and resolving section. It uses them to position the control surfaces which control the flight of the missile. The flight control section consists of the pitch and yaw steering signal channels, pitch and yaw positioning channel, aerodynamics stabilizing channel, and the hydraulic circuit.

ACTIVE HOMING SYSTEMS

Basic components of active radar homing systems are essentially the same as those employed in the semiactive systems. In the active homing system, the illuminating radar and all other system

components are located within the missile. The range and power of the system are limited by the capabilities of the missile electrical power supply and the type of components for which space may be available.

Some space is saved because no antennas or direct pulse receiver channels are required. The target-seeking forward antenna is used for both transmission of the target-illuminating pulse and reception of the echo signals. This is made possible by the use of a TR box (duplexer) to switch the antenna from transmitter to receiver automatically. The TR box protects the receiver against the high-power pulse from the transmitter and prevents absorption of the echo signal by the transmitter circuits.

In an active homing system, range data are determined by the interval of time elapsing between the transmission of the pulse and the reception of the echo from the target. The instant that the pulse leaves the antenna is considered as time zero, because it serves as the reference of base time for the sequence of operations of the homing system. Target position error is determined by conical scanning as in the semiactive homing system.

A modified active radar homing system which employs some of the principles of television is the map-matching system. Echoes from the target area are presented on a PPI cathode-ray oscilloscope. Figure 9-58 shows how the PPI presentation, or radar map as it is sometimes called, is compared with a previously prepared photographic map of the target area. This is indicated by the film map which is unrolled or played back between the PPI scope and the camera tube. If the radar map and the photographic map coincide, no light passes through them from the sensor (camera tube). If they do not match, light passes through the photofilm at the points of "mismatch" and falls upon a photoelectric cell or other light-sensitive pickup. This produces error signals which are used to reposition the missile until the maps coincide, indicating an on-course condition. If this system is used in a missile for homing, a map of the immediate target area is required. Use of the system is limited to targets which have been previously photographed or mapped, such as cities, canals, islands, and so forth. An important consideration is to use only areas where no significant changes in the surface of the mapped area have occurred subsequent to the time of mapping. Surface changes, such as flooding in large areas of the target zone, might produce large errors in the homing guidance. Such areas would produce different echo intensities on the PPI scope presentation than would normal reflections. A PPI presentation, when compared with a photographic

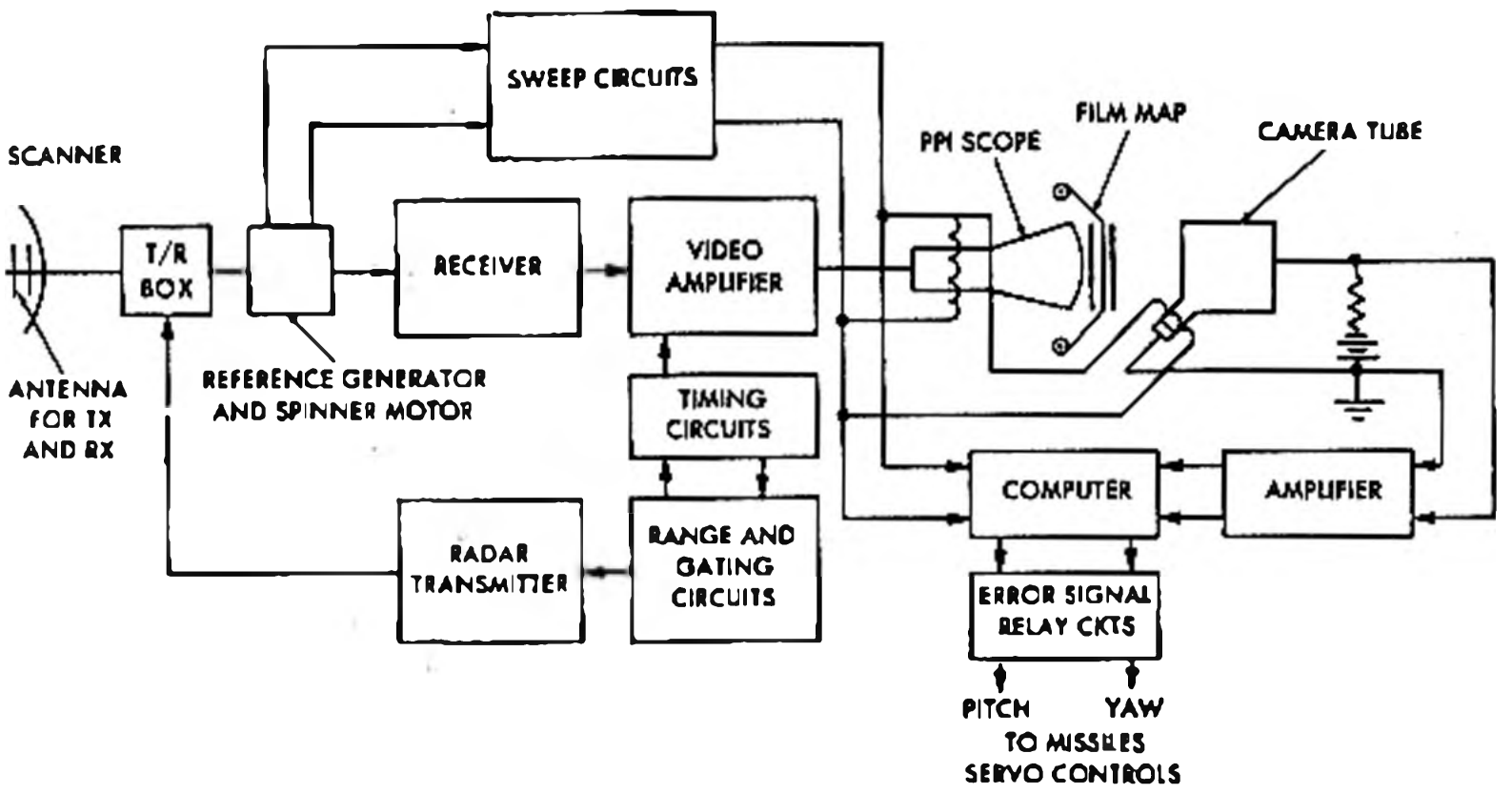


Figure 9-58. Map-Matching Homing System

film map of normal surface conditions, would indicate a mismatch when none actually existed.

TERMINAL INERTIAL SYSTEMS

Two specific terminal inertial guidance systems are known as the constant dive-angle system and the zero-lift (ballistic) inertial system.

Constant Dive-Angle System

Figure 9-59 is a general block diagram of a constant dive-angle system. As in any terminal flight system, midcourse guidance equipment directs the missile to some point in space known as the release point. At this point the midcourse guidance equipment is disabled, and the terminal guidance system takes over control of the missile.

As in other inertial systems, a stabilized platform is the reference plane containing the accelerometer sensors for a constant dive-angle system. Missile circuitry is then able to compute the missile position during the dive with respect to the release point.

At release point, the outputs of the accelerometers are fed to their respective computing channels which are somewhat similar to those in the simple inertial system. One accelerometer measures lateral (direction) acceleration, and the other accelerometer measures longitudinal (distance) acceleration.

These acceleration signals can be changed by the integrators within the computer into velocity signals and then into signals representing positions. The distance computer channel generally does not need an output representing the position error. The computer channel for the distance channel of the constant dive angle has only one integrator, and the velocity signal is sent to the pitch servo.

When the velocity signal is of the correct constant value, there is no output from the computer to the pitch servo. The method of accomplishing this error-only output is explained in the preceding section in the discussion of simple inertial systems. If the computer output should differ from the desired value, there is an error output at the pitch servo. The error output causes a control action that corrects the dive angle to the proper amount.

VERTICAL-DIVE SYSTEM. A variation of the constant dive-angle system is a vertical-dive system. This system makes use of the distance com-

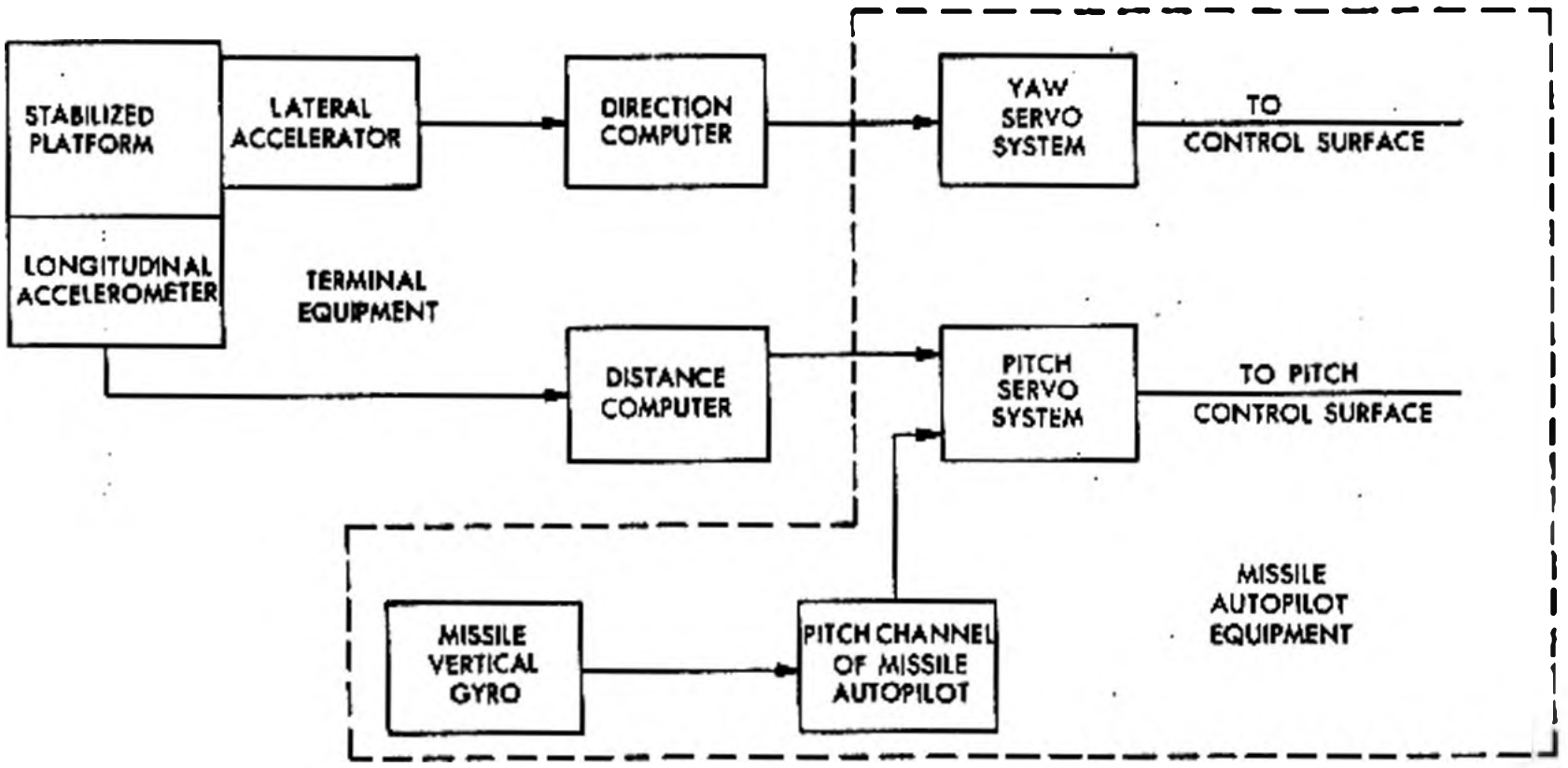


Figure 9-59. Constant Dive-Angle System

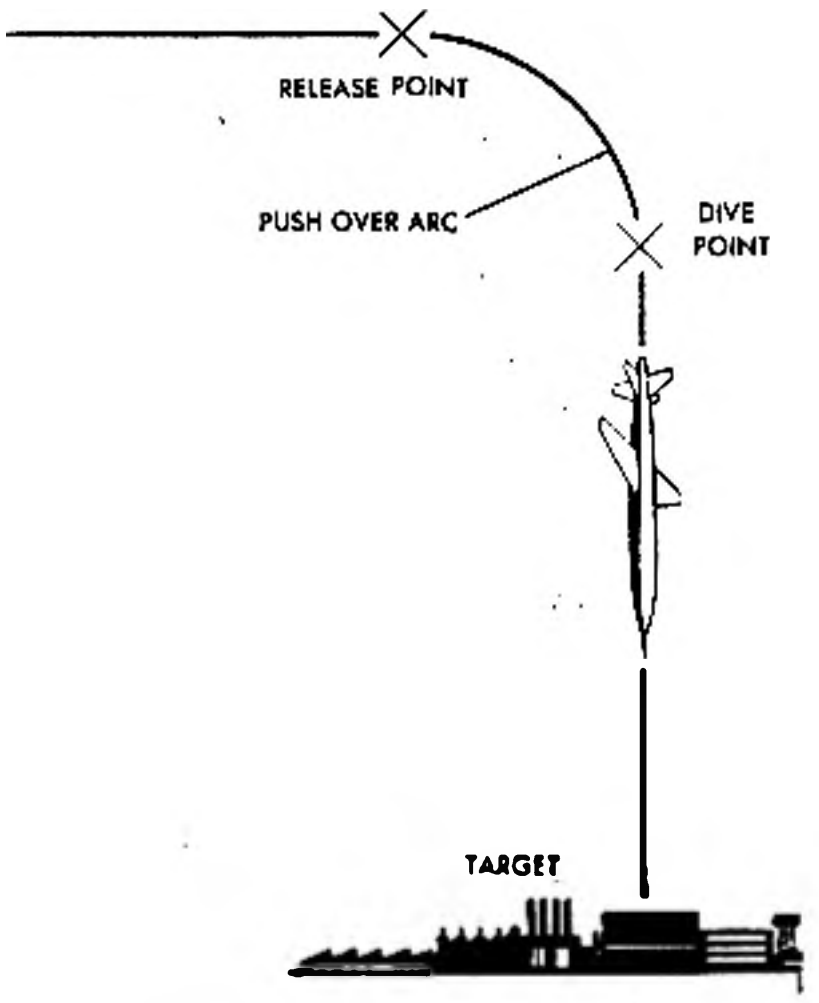


Figure 9-60. Flight Path using Vertical Dive Inertial System

puter which has two integrating blocks so that position error is the signal output. The desired signal output of the computer channels is a zero error signal signifying that the missile is on the vertical path.

Figures 9-60 and 9-61 show the flight paths of vertical-dive and constant dive-angle systems. They make noticeable the fact that the vertical dive is a special condition of the constant dive-angle system. After the missile has completed its push-over arc and is diving straight toward the earth, any acceleration in a horizontal plane causes the missile control surfaces to react and bring the missile back to its original vertical path.

PUSH-OVER ARC. The missile performs its movement through an arc before taking a straight-line course to the ground. This movement, known as push-over arc, is accomplished by precessing the vertical gyro of the automatic pilot about the pitch axis only.

There are many factors that determine the number of degrees of arc and the rate of precession of the vertical gyro. The dive angle that the missile is to take is the primary factor in determining the number of degrees of vertical precession. Another factor is the angle of incidence of the wings.

Notice the upper missile in figure 9-62. If a missile longitudinal axis is pointing straight down, the wings have some lift because of the angle of incidence of the wings; thus the missile does not dive vertically. To correct for this, the missile

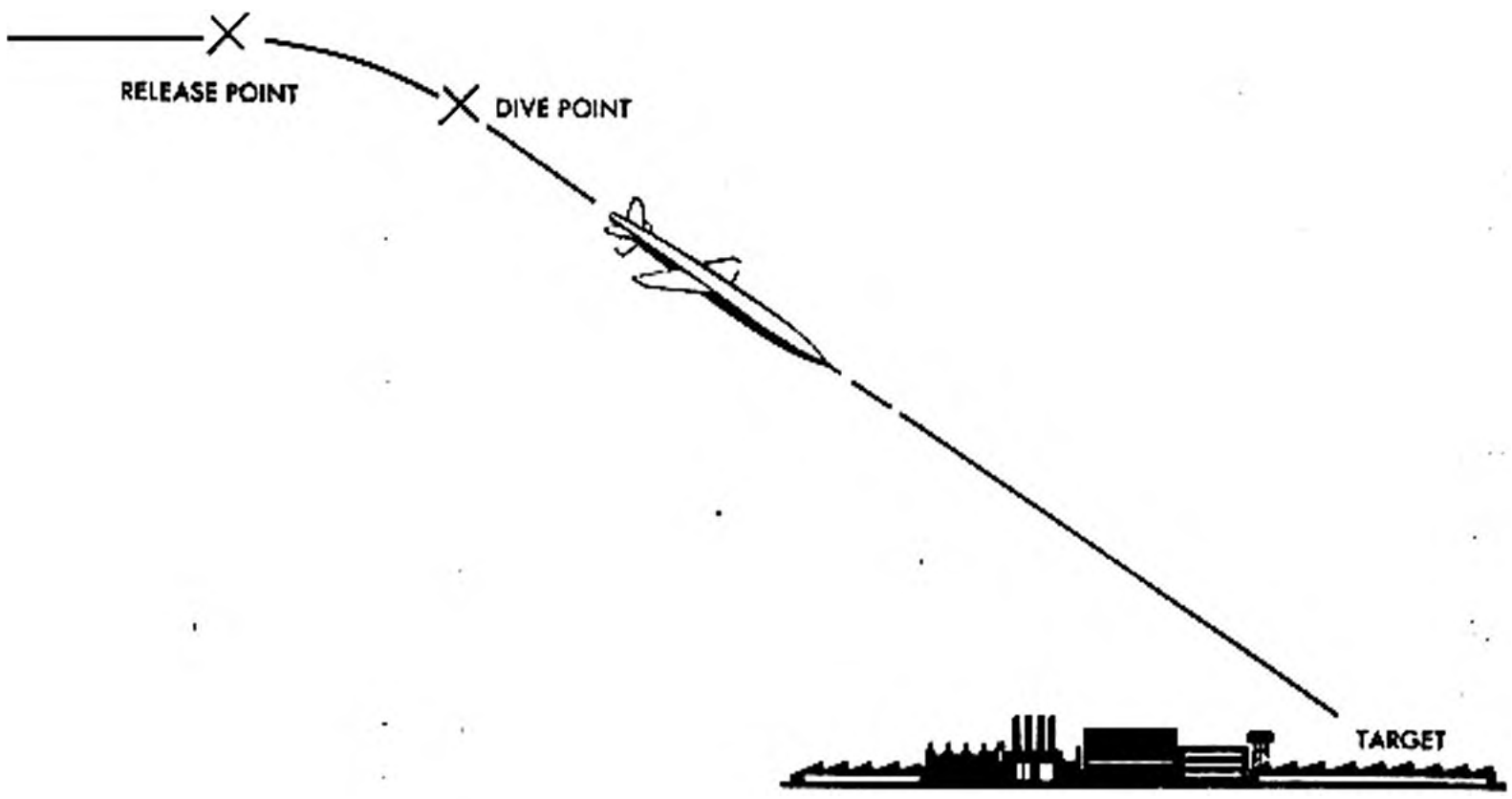


Figure 9-61. Flight Path using Constant Dive-Angle System

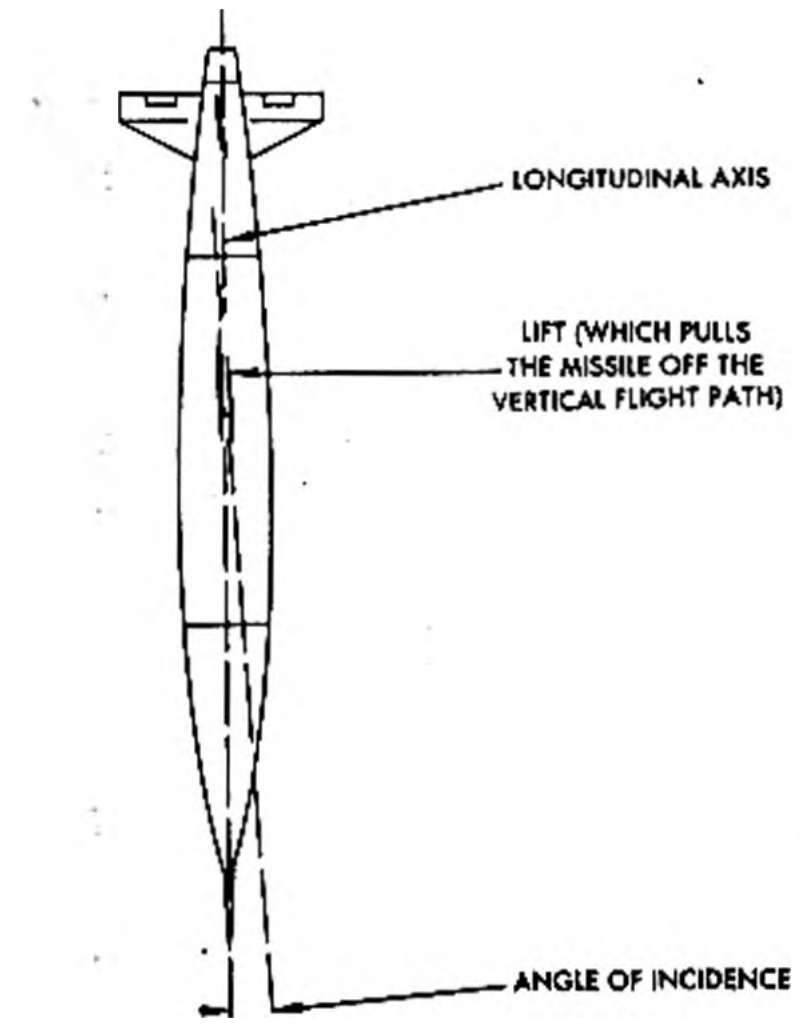


Figure 9-62. Dive Attitude

actually has to nose over farther than the angle of dive as shown by the lower missile in the illustration. This is true of any constant angle dive. This compensating effect also is calculated to do away with the added lift effected by the wings resulting from the faster dive speed.

The rate of precession, or how fast the missile is to nose over, is an engineering problem involving, among other things, the aerodynamic characteristics of the missile.

When this arc through which the missile moves is completed, a point is reached called dive point. At dive point, the automatic pilot is cut off from the yaw and pitch servos and has no further effect on the control surfaces.

Zero-Lift Inertial System

A typical zero-lift inertial system is shown in figure 9-63. The diagram also shows the relationship between the terminal system and the missile control system (the autopilot).

The terminal equipment can be broken down into two separate functions. The first function is to establish the flight path, or programmed path, by means of a constant-speed motor whose rotor moves the wiper arm of a potentiometer. The second function is to keep the missile on this programmed path by means of an accelerometer.

If the desired path is to be the theoretical ballistic path, which is a parabolic curve, the wiper arm of the potentiometer has to be moved from the ground end of the potentiometer at a constant rate of speed. If the voltage taken from the wiper arm (the programmed signal) is plotted on a graph with respect to time, the result is a straight line, as shown in figure 9-64. When this voltage is fed into a motor, the resultant displacement of the motor's rotor is an integration of the input voltage. As you will remember, the integral of a constant-slope line is a parabolic curve. The graph of this curve is plotted in figure 9-65.

The motor to which this programmed signal is fed is a part of the vertical gyro called a torquer motor (reference figure 9-63). The displacement of this motor rotor is used to precess the vertical gyro of the missile control system in the pitch axis; therefore, the gyro presents a parabolic path as a reference to the pitch channel of the missile control system.

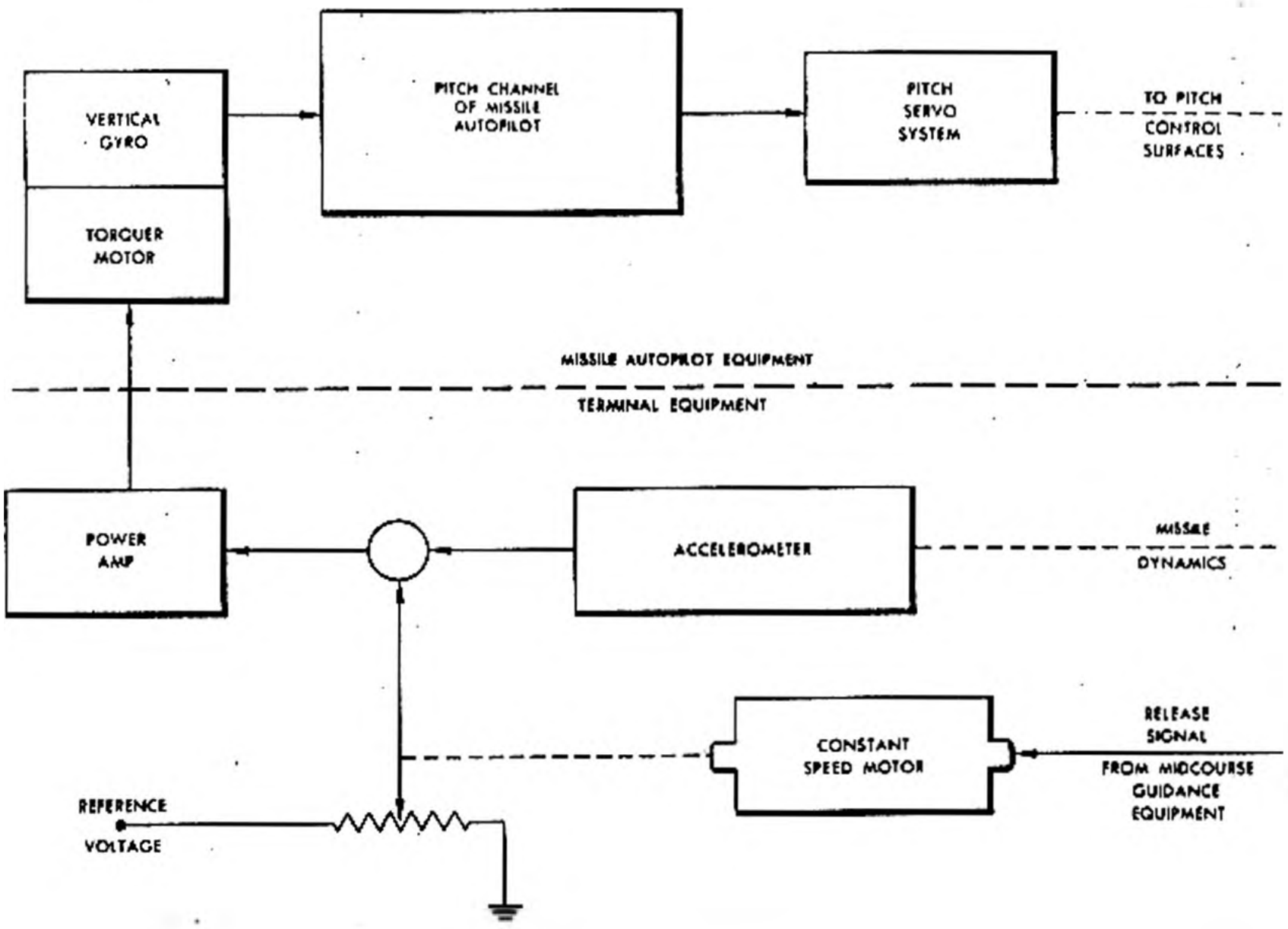


Figure 9-63. Zero-Lift Inertial System

With a parabolic path as reference in the pitch axis, the missile tends to fly that path. Because of the angle of attack of the wings and the thrust of the engine, the missile actually flies some other path, unless some compensation is made for these factors.

This compensation is accomplished by means of an accelerometer which is mounted so that it is sensitive to accelerations in the vertical. If the wings have any lift, the accelerometer senses this lift and originates a signal which corrects the precession of the vertical gyro. If there is some lift present, the signal from the accelerometer adds to the programmed signal in the mixer and precesses the gyro at a faster rate. If the missile noses over too far and negative lift is present, the accelerometer originates a signal which sub-

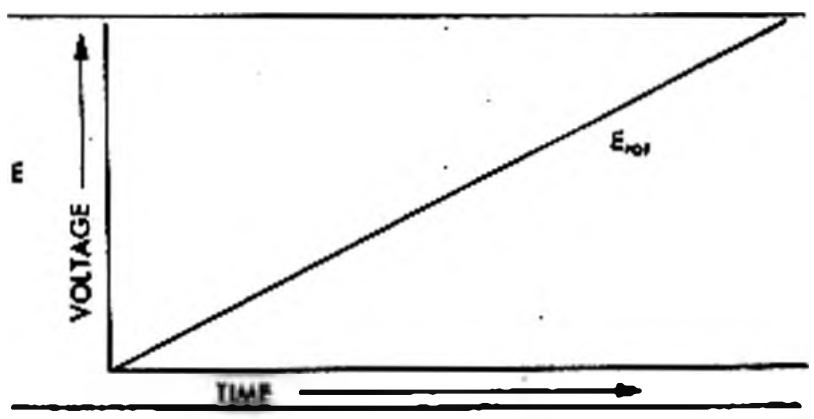


Figure 9-64. Voltage from Potentiometer

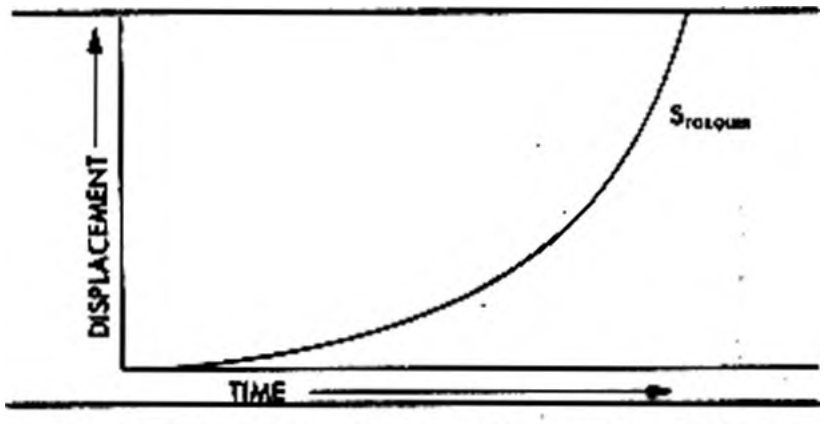


Figure 9-65. Displacement of Torque-Motor Rotor

tracts from the programmed signal in the mixer and slows up the precession rate of the gyro. The missile therefore flies a predetermined (theoretical ballistic) path to the ground as shown in figure 9-66. The term zero-lift is applied to this system because the accelerometer signal compensates for any lift along the vertical axis.

Midcourse guidance brings a missile to the desired release point, and it also sets up the circuitry for the terminal guidance equipment, including the release signal to drive the constant-speed motor. The purpose of the power amplifier in figure 9-63 is, as its name implies, to take the input signal and amplify its power so that it can drive the torquer motor.

In the case of either inertial system, the accuracy of the terminal guidance equipment cannot exceed the accuracy of the midcourse guidance in determining release point. Of course, the terminal inertial systems themselves have inaccuracies which require compensation.

SECTION D
Composite Guidance Systems

As indicated previously, more than one guidance system is used in long-range missiles. These are sometimes referred to as composite guidance systems. Composite systems are considered as several individual guidance systems working together or in sequence to guide the missile through all phases of flight. A combination of several systems is often necessary because of the wide difference in requirements that have to be met. During the *immediate launch or boost period*, extreme acceleration sometimes prevents the use of normal guidance components. Such accelerations act to close relays, precess gyros, and saturate accelerometers which have the required sensitivities for ordinary flight conditions.

Usually, regular guidance system components that contain movable parts are locked in position or, at least, the circuits into which they work are neutralized to withstand the strain of launching operations. The missile airframes are designed sufficiently stable so that control settings made before launch will enable the missile to make a stable flight throughout the boost period. Even in a system which is "dead" during the boost period, any component with a movable element is carefully balanced and positioned so that the longitudinal acceleration will not have a damaging effect.

After the boost period, the guidance equipment starts to function, either because of timer operations or the loss of the boost acceleration. Midcourse guidance is not always put into operation at this time. In ground launched missiles, the *climb out period* is necessary to get up to the proper altitude and position for midcourse guidance to take over. The climb out period can be supervised either by a preset guidance system or by a form

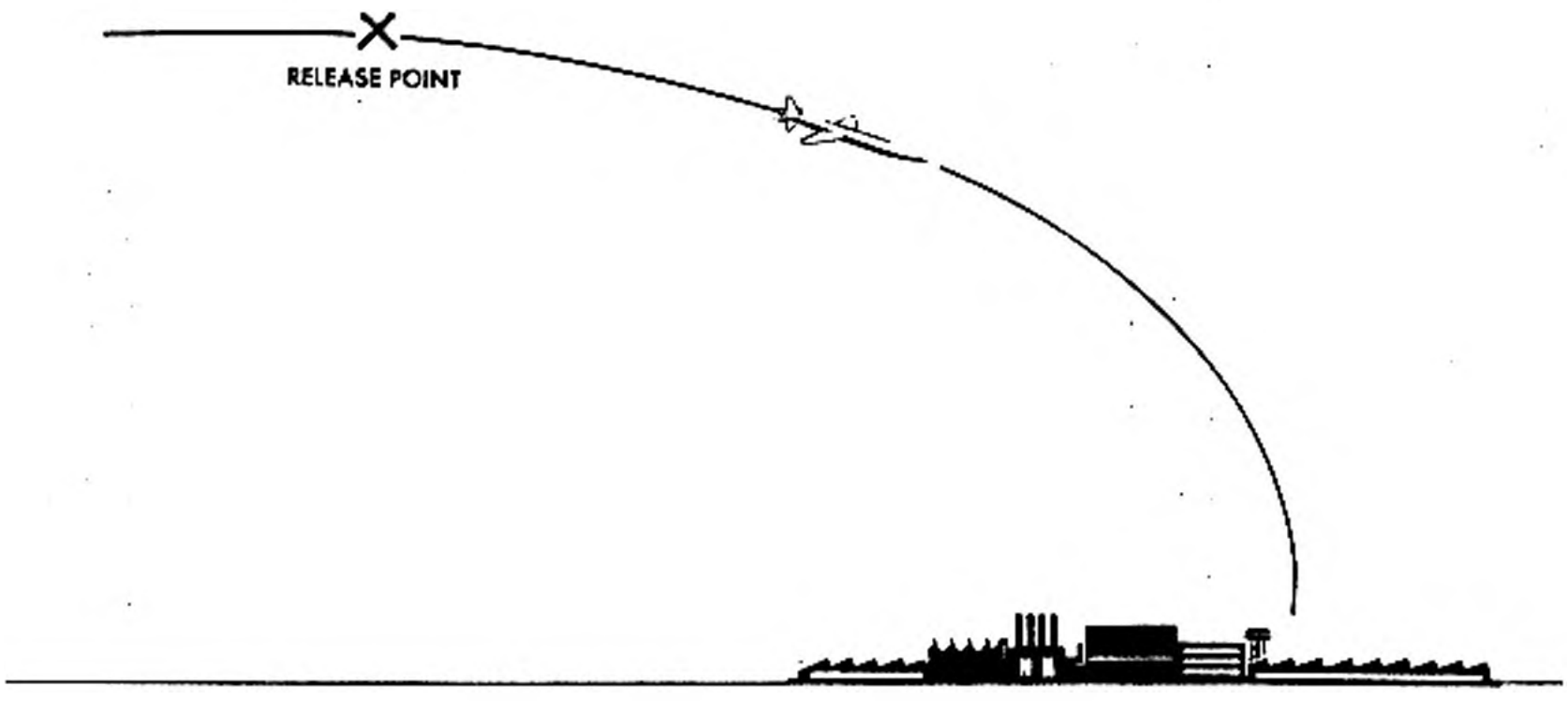


Figure 9-66. Flight Path using Zero Lift Inertial System

of remote control. In some missiles there is no climb out period as a separate part of guidance. In these missiles the midcourse system takes over immediately following launch.

The *midcourse trajectory* usually is calculated to start at some point remote from the launching area. There are several reasons for this. Since midcourse trajectory follows a great circle path or some similar defined characteristic, a launching on the trajectory would destroy the security of the launching location. By plotting the missile course and searching along its path, an enemy could locate the launching site. There also can be technical reasons for delaying the start of the midcourse trajectory. A midcourse system using some terrestrial feature, such as magnetism or map-matching, requires a particular, outstanding characteristic which might not be present in the vicinity of the launcher.

To align the missile with the selected course and to calibrate the *midcourse guidance equipment* are other guidance functions. These operations could take place during climb-out, or they could be continuous, starting at the instant of launching. The midcourse guidance equipment can be as self-calibrating or correcting as any homing equipment. In homing, the course is also constantly changing, but the missile can see its target directly and act to intercept it. In other guidance problems where the missile cannot see the target, the guidance equipment can be aligned and corrected to the actual missile position or to the missile position compared with the desired position at a particular instant of time.

In one system, correction is made after reference to a source of information external to the missile. The whole correction system may be external such as a radar, radio, or stellar tracking device. The device computes the existing error in terms of missile guidance parameters. Then it codes them and transmits them to the missile on a command link. In another system, the correction is contained in the missile. Such a system, for example, uses Doppler radar to measure ground velocities in transverse and longitudinal directions, and it also uses airspeed indicators or fixed beacon plotting radar.

In many missiles, the midcourse guidance system is considered the dominant one, with other guidance systems relegated to the position of auxiliaries. The midcourse system operates over the longest range and time. In short-range missiles, the midcourse is often the only actual guidance equipment installed in the missile. In a composite guidance system, however, the accuracy of the flight depends on each of the auxiliaries to do its full job.

A midcourse system must, in itself, be accurate because of the length of time it operates. An error-producing drift would produce an error which is proportional to the length of time it is accumulated. For example, a given error at the target corresponds to a very small drift in the midcourse system and relatively large drifts in the other systems which only operate a small fraction of time of the midcourse system.

The possibility of failure of the midcourse (main) guidance system also exists, so some means of providing a guidance signal in spite of the primary system failure is sometimes incorporated in a missile. The system that serves this purpose is a true auxiliary. It may be used during the correction and alignment period, or it may exist for the sole purpose of a *standby guidance system*. This auxiliary might simply be an arrangement set to take over if primary signals are lost, or it may provide for inertial continuation of the last correct signal conditions. As stated previously, it could be a completely separate system. Standby guidance is so arranged as to return the guidance function to the primary system if the primary system begins to function properly again.

CONTROL MATRIX

The selection of the guidance system which can control the path of the missile is a function of a special system. Guidance-system selection circuits could be scattered around within the electronic equipment of the missile, with each system-enabling circuitry located in close proximity to the particular system it causes to function. Or all the selection circuits could be organized into one block, in which case the circuitry is called the control matrix. Separated or unified, the functions of the matrix components are the same. Their functions are to select a guidance system operation in the proper sequence for the flight.

Control of a missile during flight can be provided for by more than one source. Some signal

is set up to designate the phase of flight the missile is going through. This signal may come from playback equipment which uses a recorded tape to determine phases, or it may come from a combination of signals from various units such as radio control or primary guidance. A control matrix acts upon the signal, whatever its source, to provide the command which switches in the correct guidance function for the particular phase or period of flight.

A control matrix can be likened to an automatic telephone switchboard, automatically transferring the correct signal to the control system regardless of conditions. By being able to provide a guidance function in spite of any eventuality, the control matrix furnishes the missile with the most correct guidance function.

If the midcourse guidance should become inoperative or unusable, the matrix switches in an auxiliary guidance system to hold the missile on course with an accuracy only secondary to that of the midcourse system. If the primary midcourse system again becomes operative, the control matrix returns the guidance function to it.

The matrix can be considered as the automatic guidance switchboard or guidance sequencing computer. Inasmuch as the signals are of the on-off variety, they can be considered as binary digits. A succession of these binary digits results in a signal, the nature of which can be considered to be dyadic (consisting of two elements). The control matrix operation is analogous to a digital computer, and for a combination of digits fed in, the proper digit comes out.

VELOCITY-DAMPING DOPPLER RADAR

A Doppler radar used for velocity damping of an inertial system is somewhat different from the homing type of Doppler. The velocity-damping Doppler requires antennas to measure the forward and lateral components of velocity. These antennas have to be direction-stabilized so that the velocities along, and normal to, the flight path can be measured to nullify errors that a drift angle creates.

The antenna mounting has two antennas looking down and forward at a slight angle away from the roll axis of the missile. A third antenna looks to the rear. A comparison between the signals from the two front antennas is used to align the direction of the missile. A comparison between the forward and rear antenna signals gives the forward velocity. Drift is computed from the angle between the antennas fore and aft axis and aircraft heading.

A Doppler radar similar to the equipment used in bomb navigation equipment is used with a special marriage unit employed to couple the Doppler output into the missile computer.

Doppler would not be used continually on a flight, for such use would violate security. On long, overwater flights, Doppler is of little use because insufficient return is received from a relatively smooth, flat surface such as an ocean.

This section ends the treatment of missile guidance systems. A new aspect of missile guidance and control components has been introduced by the study of lasers with their seemingly unlimited potential of use and accuracy. The principles of laser operation are presented in chapter 11.

CHAPTER 10

Guided Missiles Operations

With the advent of the guided missile, two popular concepts came into existance relative to the nature of future wars. Probably the more common was the push-button concept. This concept envisioned future wars being won or lost by use of intercontinental, supersonic, long range missiles carrying A-bomb or H-bomb warheads. There would be very little or no ground fighting involving the use of foot troops, small arms, tanks, and the other common types of armament. It was even envisioned that missiles could completely replace manned aircraft.

The other popular concept was that nuclear and thermonuclear bombs were an improvement over conventional bombs of the past and that guided missiles were merely more effective carriers. Advocates of this concept maintained that the general mode of warfare would not change. This concept reasoned that even if decisive military action shifted to missile weapons, ground troops would be required to repel enemy invasion attempts, seize bases needed for launching missiles, and physically occupy critical territories to prevent resumption of hostilities.

Both of these concepts were subject to considerable discussion. The latter concept, the more popular, eventually evolved into what is now the "mixed force" concept. Many missiles have been developed to augment the manned aircraft and the ground forces, while large intercontinental ballistic missiles have been strategically located as deterrent weapons. Missiles have come to play an important role in the isolation of the battlefield and neutralization of important targets.

SURFACE TARGET SELECTION

Longe-range missile attacks are designed to bring about, through systematic application of force to selected vital targets, the progressive destruction and disintegration of the enemy's war-making capabilities to a point at which he no longer has the ability or will to wage war. The part played by missiles in attaining this goal consists of surface-to-surface and air-to-surface missiles employed against selected surface targets.

Considerations given to selection of surface targets for missiles differ little from considerations given proposed targets of conventional aircraft. Guided missiles are needed not only for attacking new ultra-modern ground targets, but also those existing strategic targets which are highly vulnerable from the air.

Surface-Target Classification

Generally, surface targets are classified as either tactical or strategic. Tactical targets are those which have a direct influence on the course of a battle. For instance, military personnel, tanks, and weapon emplacements in the actual battle zone are tactical targets. These targets normally are located near the point of enemy contact. They are generally temporary and transient, requiring attack by weapons which can reach them in a minimum of time. Some tactical targets are scattered over a front and extend deep into enemy territory. An enemy missile launching site, for example, may require immediate attack even though it is hundreds of miles from the battlelines.

Strategic targets are normally located many miles from the battle area, and their destruction

does not directly or immediately influence the course of a particular battle. These targets are usually located deep in an enemy country. They are generally large installations of importance to the productive power of the nation. Industrial factories, mines, and oil fields are examples of strategic targets.

With the development of guided missiles, the tactical area has moved farther from the battle lines. If troops in a battle zone are classified as a tactical target, then surely troops loading aircraft and manning missile launch sites with missiles that may be immediately thrown into the battle must also be tactical targets even though they may be located hundreds of miles behind the actual battle front. It becomes ever more difficult to differentiate between tactical and strategic targets on the basis of distance from the battle line. This chapter presents the basic considerations given to selection of missile surface targets with no attempt to differentiate between tactical and strategic targets.

Specific Target Systems

The ultimate goal of attacks on surface targets is to weaken and immobilize enemy troops so that they are more vulnerable to defeat. Major targets for consideration are industrial centers located in the heart of the enemy country. Successful attacks on such targets not only drastically curtail the flow of supplies and war implements to the enemy troops but also affect the morale of the enemy. Striking major industry centers destroys factories, communication systems, food stockpiles, and fuel and arms supplies.

Generally, the productive capacity of a nation depends on its strength in four broad categories:

- Raw materials.
- Processing and end-products manufacture.
- Supporting services and production of basic equipments.
- Imports.

Each of these categories is a source of targets. The relative importance of these targets depends on the specific objectives to be attained and the war situation at the specific time.

The first group is comprised of mines, forests, oil fields and farms from which are extracted a nation's raw materials.

The processing and end-products group pertains to the factories which convert the raw materials into finished products for civilian and military use. Aircraft plants, ship building yards, armament plants, and refineries are a few of the military targets in this group.

The supporting services and basic equipment group consists of such facilities as powerplants, internal transportation systems, and research and development activities. This group is not directly engaged in the production of weapons, but the activities support the producing units.

The nation's imports are also of major importance during time of war. Targets in this category consist mainly of ships, rail cars, and pipelines carrying critical materials into the enemy nation. These targets are most vulnerable during transit.

Target Considerations

Some of the criteria considered in the selection of surface targets are presented below. Generally a surface target selected for a missile mission is beyond the capability of a more conventional and less expensive weapon.

IMPORTANCE OF TARGET TO THE ENEMY. A target should be of such importance that its destruction will deprive the enemy of the means of sustaining a war or a particular campaign. The importance of a target may vary with the progress of the war or the theatre of operation. At one stage of the war, cutting off the nation's imports may be of major importance. At some later date, destruction of oil refineries may be the prime mission. In any case, the importance of the target should justify the use of missiles. Also, the nature of the target should not be such that the enemy can easily substitute another item or another source of supply, or can merely eliminate the use of the facility with no strategic effect on the war effort.

RESERVE CUSHION MAINTAINED BY THE ENEMY. The target selected should be one for which the enemy has no ready large reserve production capacity nor large reserve supply. Among other things reserve production capacity may exist in idle or near idle facilities that may be put into full operation on short notice. A certain amount of reserves exist where facilities used for low-priority materials may be easily converted to high-priority production. Large stockpiling of materials can make it possible for the enemy to

repair damaged facilities or convert other facilities with little effect on the flow of materials and equipment to the battle lines. The amount of alternate facilities available must be considered to determine how much of the overall industry must be destroyed to effectively retard the flow of materials from that particular industry into the enemy's war machine.

IMPACT OF TARGET DESTRUCTION. Impact here refers to how soon and to what extent an attack on a particular target will impede the enemy's war effort. An attack on a target engaged in the manufacture of the components of a weapon should result in the critical reduction of the output of that weapon. The components should not be available from a stockpile large enough to permit uninterrupted production of the finished weapon even though the output of components from the target is delayed.

VULNERABILITY OF THE TARGET. A target under consideration should have physical characteristics which will allow damage to be achieved by the missile. A highly vulnerable target can be damaged effectively with a minimum expenditure of missiles. Consider, for instance, an electric powerplant. This target is highly vulnerable since a single missile hit might easily render it unusable. On the other hand, a target located well underground and housed in reinforced concrete would be much less vulnerable and would require the use of a larger number of missiles of greater destructive power. Even then damage may not be sufficient to completely halt operations.

RECUPERATIVE POWERS. Recuperative powers pertain to how quickly and easily the damage to a target may be repaired and the facility returned to normal or near normal operation. In the case of an air strip, for example, filling of craters and repairing the strip would take a relatively short time and the facility could be back in operation. The electric powerplant cited previously would have a low recuperative capability. It could be put out of commission for a long period of time with a single hit. Also, factories and other agencies relying on the powerplant would be inoperative during the recuperation period.

There are cases in which planned attacks on an enemy target having a high recuperative potential would be desirable. This situation would exist if friendly forces expect to occupy and use the facilities as soon as possible after the attack.

DISPERSAL OF TARGET SYSTEM COMPONENTS. A target system's components should be concentrated in few enough places that a reasonable number of attacks can destroy it. The percentage distribution of production among plants producing a major weapon when the plants are distantly separated is an important factor. The plants producing the greater percentage of the nation's total output would be a top priority target. On the other hand, if component parts of some weapon are produced separately by a number of widely dispersed plants, then the plants producing the most important parts would probably be the first priority.

IDENTIFICATION OF THE TARGET. The ability to identify the target is of utmost importance. In the past, physical characteristics of the target have been used as identifying characteristics. For example, targets emitting heat, light, or some other form of energy were ideal targets for missiles using homing guidance systems sensitive to these forms of energy. Nearly all surface-to-surface and air-to-surface missiles now use inertial guidance and the target is "identified" by reference coordinates or carefully plotted geographic location.

ENEMY PROTECTION OF TARGETS. The enemy's defense of a target is an important consideration in determining its suitability for attack. The concentration of defenses often indicates the importance of a target to the enemy. Ground and air defenses of a target must be evaluated in planning an attack. Enemy countermeasures may consist of interceptor aircraft, antiaircraft batteries, interceptor missiles, decoying systems, and electronic jamming systems. The type and number of defense installations protecting a potential target must be analyzed when planning an attack to determine the most effective method of penetrating or neutralizing the defenses. The use of decoy weapons may be effective in drawing defensive aircraft or missiles from the primary attack missile. The use of air-to-surface missiles may be used to destroy the defense installations prior to the employment of a missile or missiles to strike the primary target. The elements of surprise, speed, and jam-proof guidance systems greatly decrease the possibility of effective enemy opposition.

SIZE AND LOCATION OF THE TARGET. A target must be of a size commensurate with the accuracy of the missile used, and the principal elements of the target must be within the range of the missile. Suppose a surface-to-surface missile with a 150-mile range has a deviation of 0.1 percent of the distance from launch site to target. At a target range of 100 miles, the missile will hit within 528 feet of the calculated center of the target area. If the target is large enough, the center of dispersion of the missile's warhead will still be on target. If the target is small relative to the accuracy of the missile, a complete miss may result.

If the accuracy by which a target can be located is not comparable with a missile's guidance system, the missile is of no value to plan an attack on that particular target. The Germans encountered this problem during World War II when attempting to locate London with respect to their missile launch sites. Target location may be a greater limiting factor than accuracy of a missile.

Within a war zone, targets are usually detected and located by direct observation or map data. In the case of long-range targets within an enemy nation, such means are more difficult and unreliable. Chances of survival of observers is questionable, and existing map data may be inadequate. The only accurate representation of the earth's surface is a globe which is impractical for aeronautical planning. The spherical surface must be represented on a flat surface and projections of this type introduce distortion in geographical coordinates. Various methods of projection are used to keep distortion to a minimum.

With the development of reconnaissance type aircraft capable of extremely high altitude flight and extended flight time capability, the direct observation method has been made more usable. Also, the era of space flight has resulted in the capability of a country to orbit observation satellites that transmit direct television pictures of selected areas. These pictures provide information regarding installation of new missile launching sites, concentration of munitions and supplies, as well as up-to-date weather conditions. Photographs taken by flying observers and up-to-the-minute data supplied by orbiting satellites, have improved the planning of strategic and tactical missions.

The role of military intelligence is to detect and accurately locate targets. The success of a mission depends upon the accuracy of the intelligence data furnished to the mission planners. All the sources available must be used to determine *where*, *how*, and *how hard* a target must be hit. Successful interpretation of intelligence data such as photographs, reports, and charts depends on the skill and experience of the individual interpreter. The interpreter must extract pertinent information required for mission planning from many photographs and other data furnished to him. These interpreters can obtain amazing amounts of accurate information both from what they see on the photographs and what they piece together from indications and discernible features associated with certain activities which may actually be hidden. Photo-interpretation provides a very accurate and extensive source of information relative to the strength, disposition, and activities of the enemy in areas not readily accessible to ground observers.

A detailed knowledge of the enemy's customs and military techniques also aids the interpreter in obtaining information from photographic studies. He must know what to look for and be able to identify what he finds.

Priority Assignment to Surface Targets

To obtain a desired effect during hostilities, some targets are attacked while others are ignored. Then because of a changing situation, a previously ignored target may become a target of great importance. Usually there are many more potential targets available than there are missiles to fire at them. In any given situation, targets must be considered in view of the factors previously presented and assigned a priority rating. Each important industry of the enemy is a potential target system based on the considerations mentioned earlier. Each subsystem of the target systems receiving the highest priorities is, in turn, rated by the same screening process. Top priority for missile attack is then assigned to one or more of the systems in attempting to cripple or destroy the output of the finished product of that industry. Figure 10-1 depicts representative target systems. Notice that higher priorities have been assigned to rubber, guided missiles, and oil industries. Next, note the breakdown of the guided missile industry into

TARGET CODE	PRIORITY
AIRFRAME ASSEMBLY	1 1ST PRIORITY
OIL	2 2ND PRIORITY
RUBBER	3 3RD PRIORITY
GUIDED MISSILE LAUNCH SITES	4 4TH PRIORITY
SHIPBUILDING	5 5TH PRIORITY
ELECTRONIC EQUIPMENT	6 OTHERS

Figure 10-1. Target Systems Analysis and Target Selection

various subsystems. For simplicity, only one component breakdown is illustrated. The various subsystems of the missile industry are analyzed and assigned priority. In this case the highest priority has been assigned to airframe factories.

Postattack Target Analysis

Reconnaissance of the target must be accomplished after an attack to determine the success of the mission and to obtain information of value to subsequent preattack target analysis. The best means of obtaining the desired data are from reconnaissance photos at various intervals following attacks, captured documents and signal communications.

Knowledge of physical damage to the various structures in a target area gives information relative to particular warhead and fuze characteristics, method of attack on targets, and methods of target location. This information is not all that is desired. It is also desirable to know the production loss and the effect on military personnel. The progress of repair and reconstructions must be followed to determine the enemy's repair priorities and to establish the time for additional attacks.

This postattack analysis also provides the basis for future priority assignment to targets. For example, should the attack continue against aircraft factories, or should the attack be shifted to oil refineries or some other industry.

AERIAL TARGET CONSIDERATIONS

It has often been said "that the best defense is a good offense." The only completely successful defense against enemy missiles would be vigorous offensive action to prevent all enemy missile launchings. This can only be accomplished by the destruction of all launching sites, unfired missiles, supply lines supporting missile units, and missile production facilities. The success of such an offensive action during a war is improbable. Consideration must be given to the use of surface-to-air and air-to-air missiles as defensive weapons against attacks by enemy missiles.

The high speeds, high altitudes, and evasive maneuverability of missiles have surpassed the capabilities of ordinary antiaircraft artillery as effective defensive weapons. Many surface-to-air and air-to-air missiles have been developed to counteract the surface-to-surface and air-to-surface missiles in other nations' weapons inventories.

Areas or installations to be defended must be assigned defense priorities. Such a priority for a particular installation is based on estimates of enemy tactics to be used against the installation, the importance attached to it by the enemy, its vulnerability to enemy weapons, and recuperability factors. Plans are then established as to maximum allowable enemy effort and defensive measures necessary for effective protection.

The initial requirement to restablishing an air-defense unit is the determination of the volume of firepower the unit must be capable of delivering on the expected enemy effort. The volume of firepower is determined by the characteristics of the proposed defensive missile and its kill probability relative to certain targets. Consideration of estimated enemy tactics, area to be defended, and missile characteristics are aids in estimating the damage that can be expected on an aggressor target. These factors make possible a mathematical calculation of the number of defensive missiles necessary to provide the desired target-destruction probability.

Various estimates and mathematical probabilities relative to anticipated enemy attacks on installations serve as a basis or guide in establishing guided missile defenses against attacking aircraft and missiles.

Aerial Target Detection and Evaluation

The problems of utmost importance in an operational situation are detection and evaluation of attacking missiles. Detection must be achieved at

ranges great enough to permit necessary defensive action. Before such action can be taken, the attacking missile must be identified, and it must be evaluated as to speed, altitude, and course. This can be done by radar. A vast network of early warning radar systems is operated by North American Air Defense Command to provide detection, identification, and interception of any unknown aircraft or missile approaching our nation. In the event of a missile flying on a course toward our country the data collected by this warning network must be coordinated with launching and guidance equipment of defensive missiles so that their launching time, launching attitude, flight path, and other factors are properly related for a successful interception resulting in the timely destruction of the attacking missile.

In connection with aerial targets, the phrase "range of accomplishment" is often used. Range of accomplishment of a mission refers to how near an attacking target must approach a defense installation in order to succeed in its attack. Of two targets otherwise equally difficult to destroy, it may be highly advantageous to destroy one target at a greater range than the other. For example, an aircraft carrying guided missiles to be launched at a relatively long range has a longer range of accomplishment than a bomber carrying conventional bombs. The craft carrying the guided missiles should therefore be destroyed at longer range before it releases the missiles. After the missiles are released, each one becomes a new target itself —faster, smaller, and more difficult to destroy.

Postattack Target Analysis

Postattack analysis is also an important consideration relevant to aerial targets. If three launch crews defending an installation are launching surface-to-air missiles at a maximum rate of fire, those missiles launched after the target is destroyed are useless, expensive expenditures. Fast evaluation of the mission results and communication between air defense units are essential to prevent excessive expenditure of missiles. In the case of air-to-air missiles the range is usually shorter, permitting visual observation of the result of the attack.

MISSILE SELECTION FOR A SPECIFIC TARGET

The selection of a particular missile for attack on some surface or aerial target actually originates with the consideration of the anticipated targets. Target characteristics must play an important part in the design and construction of various missiles if effective performance is to be expected. This is especially true in the type of guidance system selected. Some of the criteria evaluated as aids to missile selection are presented here in a general sense rather than discussing specific missiles and the targets for which they are designed.

Target Consideration for Missile Selection

Some of the target characteristics on which missile selection depends are speed, maneuverability, range, and vulnerability.

Speed and Maneuverability. Speed and maneuverability of any target are two factors which very likely will be known. Targets can be grouped in accordance with speed and maneuverability as follows:

Group I targets. This group consists of fixed area-land targets, such as large cities which are densely populated and highly industrial. Because of its large size and the fact that the target is fixed, guidance information can be preset into a missile before launching with a reasonable certainty of hitting within a certain area. Such targets make the least demands on a guidance system for accuracy. The higher the speed of a missile selected for use against area targets, the less vulnerable is the missile to enemy countermeasures.

Group II targets. Point-land targets, ships, submarines, trains, and other moving targets are placed in this category of targets. The speed range of targets in this group vary from zero in the case of point-land targets to upwards of 100 miles per hour in the case of moving targets. Even though some of these targets may be highly maneuverable, their comparative low speed will still allow a good possibility of a direct hit with the right type of missile. Guidance systems for missiles used against these targets must be much more accurate than missiles used against group I targets. Missiles used against group II targets may be air launched or ground launched with a predetermined trajectory. They must be capable of guidance in flight by radio, radar, or homing devices to compensate for errors or for movement of the target.

Group III targets. Aircraft and certain missiles which are limited to subsonic speeds make up

group III targets. A missile selected for destruction of such targets must have greater speed and must possess more sensitive and accurate guidance. The guidance system may be basically one of the types listed under group II but must compensate for added speed and maneuverability of the target.

Group IV targets. Generally, jet aircraft and supersonic missiles are placed in group IV. Missiles selected to attack group IV targets must be capable of supersonic speeds and high altitudes. Guidance systems must be highly sensitive and accurate.

RANGE. Range is another consideration in selection of a missile. Range is basically the distance from the launch site to the target. In the case of surface targets and large missiles this range is fixed. Smaller tactical missiles may be carried on mobile launchers in which case, if the missile meets all the other requirements for attacking a surface target except range, the launcher may be relocated at a point closer to the target. Air-launch techniques provide a means of extending the range of certain missiles as well as providing selectivity of targets during an actual mission. In any case, adequate range of the selected missile is a necessity.

VULNERABILITY. Considerations as to the vulnerability of targets are covered earlier in this chapter. In selecting a missile for attacking various targets, these considerations must be weighed against the effectiveness of the warhead and the arming and fuzing system of the missile. The number of missiles of a certain type necessary to insure destruction of the target must also be considered. A particular target might be highly vulnerable to a blast-type warhead, while another target may be more vulnerable to a fragmentation warhead. Still another target may require the warhead to penetrate through a structure prior to detonation for effective destructive force to be applied. Vulnerability can now be defined as the probability of effective damage being caused to a target by a given missile at a given burst location with respect to the target.

IDENTIFYING CHARACTERISTICS OF TARGET COMPONENTS. Certain characteristics of the components that make up a target are valuable aids in the selection of a missile from the standpoint of guidance. For example, if a target component emits one form of radiant energy, a missile using a homing guidance system sensitive to the particular form of energy may be most effective to strike that target.

Other Factors Which Influence Missile Selection

Other factors which must be considered in missile selection are the rate of fire based on the availability of a particular missile, and handling and launching procedures. Still another factor is the required dependability of the missile under varying temperature, weather, and field conditions.

After consideration of all influencing factors, a missile is designed or selected to most nearly meet the qualifications necessary to achieve the desired result against a target.

MISSILE WARHEADS

The objective of any missile is to deliver a warhead to a designated target. The missile is actually just a transporting vehicle. The most sophisticated guidance and propulsion systems are of limited value if the warhead cannot produce sufficient lethal effect at the right time to destroy or incapacitate the target.

During the early development of guided missiles in World War II, standard bombs of the era were adapted as warheads. These bombs, installed in suitable airframes, often equipped with some type of guidance system, constituted the first missiles. Figure 10-2 is an illustration of such an adaptation. It shows a conventional bomb used as the warhead, an airframe for transportation, and a radio or radar for guidance.

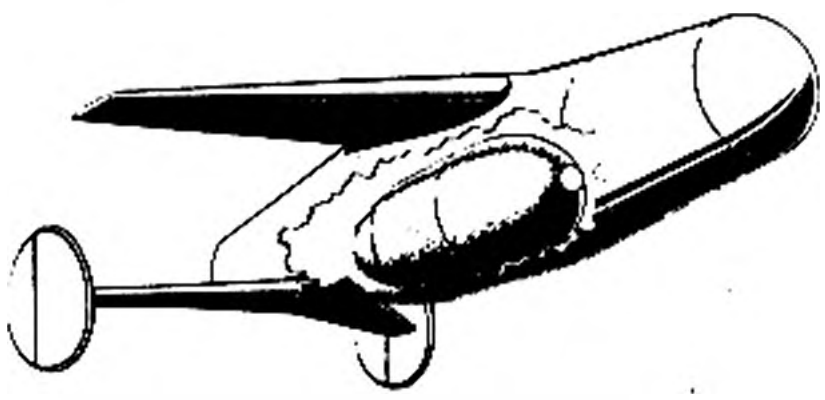


Figure 10-2. Employment of Conventional Bomb as Weapon in Early Missile

Constant research and experimentation in warhead design make it possible to ascertain optimum effect of warheads on targets of various nature. The warhead selected for a given target depends on the characteristics of the target. Types of armor found on the target, penetration desired, speed of the target, type of destruction desired, and other factors must be considered in warhead design. Special design considerations apply to each type of target. Surface targets present the need for fewest differences from conventional weapons. Aerial targets impose more difficulties upon warhead design and missile components.

Desirable Characteristics for Warheads

Certain characteristics are desired in a warhead depending upon its mission requirements. All warheads do not necessarily possess all the desired properties listed below:

- Ideal design would permit interchangeability of warhead and rapid installation onto the missile in the field.
- The warhead must be able to withstand accelerations of the missile, and the filling agent should not deteriorate during storage and flight.
- The warhead should be designed to produce maximum distribution of destructive elements on the target area.

Types of Warheads

The word *warhead* generally implies a quantity of high explosives encased in a metallic casing. This is not always true. A warhead may contain high explosives, incendiary, or chemical agents. In the high explosive warhead category, the warhead can be designed to produce varied results. It may be designed to produce destruction caused by blast or concussion, or it may be designed to spread fragments of the warhead. The following presentation covers the various types of warheads from the standpoint of basic physical construction and the effects produced on various targets.

BLAST-EFFECT WARHEADS. A blast-effect warhead consists of a quantity of high-explosive material encased in a metallic case. The force of the explosion of this type of warhead produces a pressure wave in the air or other surrounding medium, which causes damage to the target by concussion. This action is somewhat analogous to the waves that result from dropping a large stone in a pool of water. Blast-effect warheads are especially effective against underwater targets because of the greater density of water as compared to the density of air. It is also effective against many ground targets where the structures may be susceptible to high-pressure or overpressure as this condition is called. Figure 10-3 is an illustration of a representative blast-effect warhead. The metallic case contains a high-explosive substance and a suitable fuze unit. (Various types of fuzes are presented later in this chapter.) On impact with the target the plunger strikes and detonates the percussion composition which in turn fires the igniting charge. The igniting charge detonates the main explosive to produce the destructive pressure waves through the surrounding medium.

FRAGMENTATION WARHEAD. A fragmentation warhead depends on the force of an explosive charge to disperse metallic fragments at high velocity. The fragments may be contained within the warhead casing, consist of the casing itself, or be a combination of both. The size of the fragments, velocity of the fragments, and the dispersal pattern can be controlled by variations in design and construction of the warhead. If the casing is used for fragmentation, the size of the fragments can

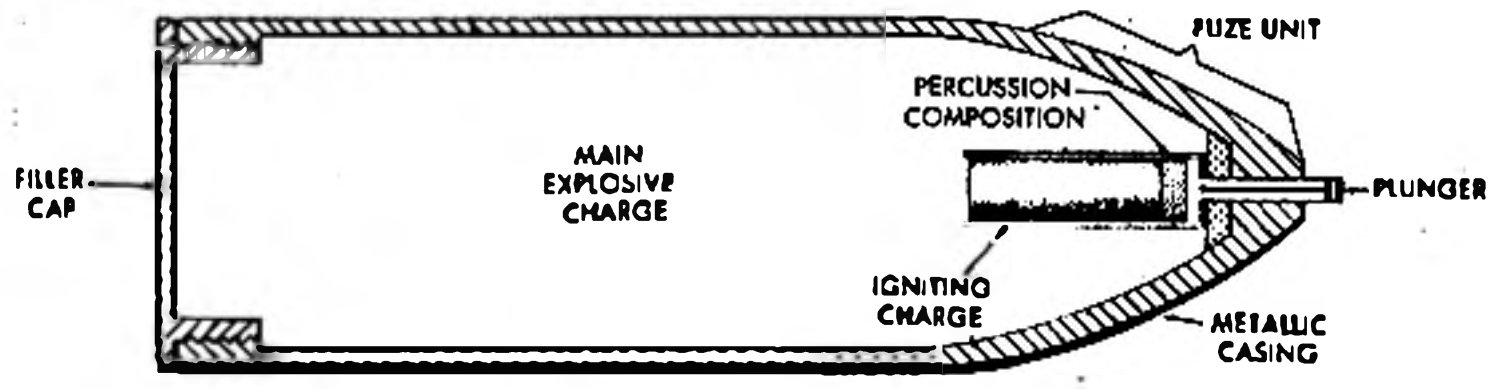


Figure 10-3. Basic Construction of Blast-Effect Warhead

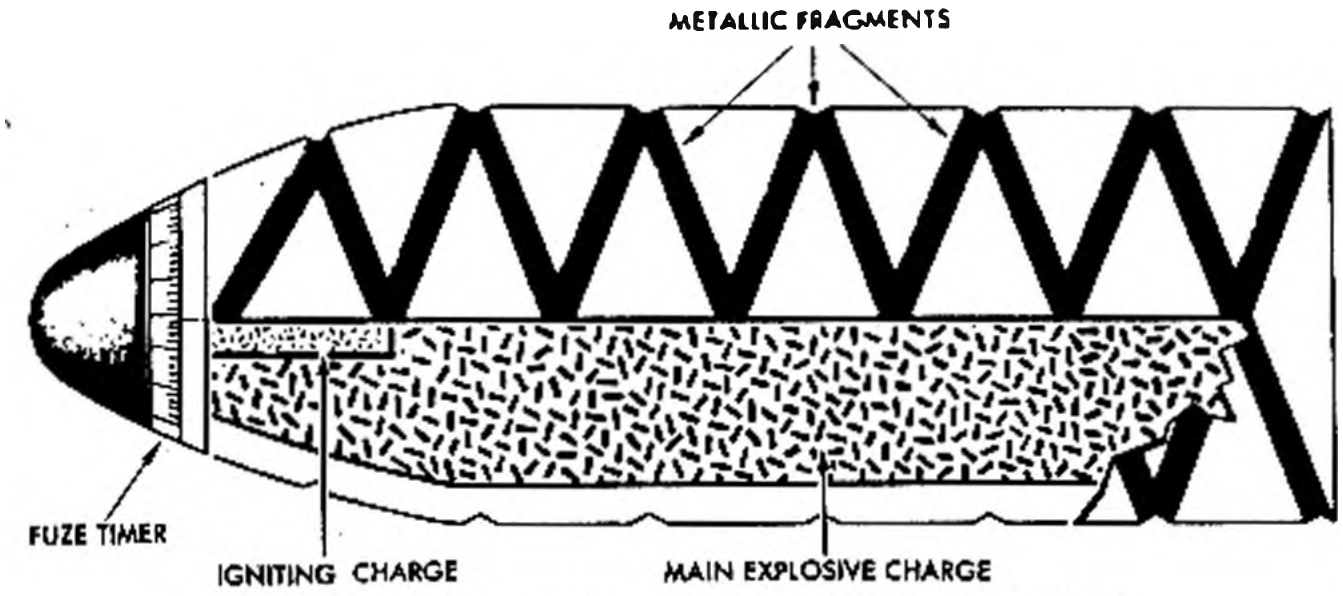


Figure 10-4. Basic Construction of Fragmentation Warhead

be controlled by weakening the case at selected points to cause fractures to occur at these points when the warhead explodes. The velocity of the fragments is controlled by the type of explosive and the explosive-to-metal ratio. The dispersal pattern is determined by the point of detonation and the shape of the warhead.

Figure 10-4 is a sketch of a basic fragmentation warhead showing the metallic shell designed to break into fragments of a desired size. It is loaded with an explosive material and contains a fuze unit that can be set to activate at a predetermined time.

Performance of the fragmentation warhead is very effective against personnel and light vehicles and weapons. It is most effective when detonated above the surface of the earth. Under this condition, the fragments are dispersed over an area in a downward and outward pattern.

Some types of fragmentation warheads consist of a casing containing many small bomblets. Each of these bomblets is a small fragmentation type unit in itself. The warhead is normally activated by explosive or other means above the ground and the bomblets are dispersed in a wide pattern. These bomblets then are triggered and spread their fragments over a large area.

SHAPED-CHARGE WARHEAD. A shaped-charge consists of a casing containing highly explosive material and is so designed that the force of the explosion is concentrated in one direction. Damage is produced by the compression force of the blast and/or fragments ejected in the direction of the blast. The shaped charge is widely used against armored surface targets and heavily reinforced structures. Figure 10-5 illustrates the effect that the shape of the charge has on the distribution of the explosive force. Only the configuration of the explosive material is shown, not the entire warhead. Assuming that the amount of explosive material is the same, the arrows indicate the direction and magnitude of the force.

Note that in the case of the *flat charge*, the forces are distributed evenly over the area of the target resulting in very little penetration. The shallow cone in the middle illustration concentrates the force in one direction and produces deeper penetration of the target. The *deep cone* to the right of the picture illustrates adequate concentration of the explosive force to completely penetrate the armor of the target. Metallic fragments of the warhead as well as fragments of the armor can now reach the inside of the target to do additional damage.

CHEMICAL WARHEADS. A chemical warhead is designed to produce personnel casualties by inhalation of or physical contact with toxic materials. This warhead may also be designed to destroy combustible targets by use of incendiary materials. More specifically, chemical agents are referred to as casualty agents, smokes, and incendiaries. The type and amount of agent placed in a warhead depends on the nature of the target and the results desired.

Incendiary warheads contain chemicals that combust violently, produce extremely high temperatures, cover a large area when released, and are difficult to extinguish. Incendiary warheads

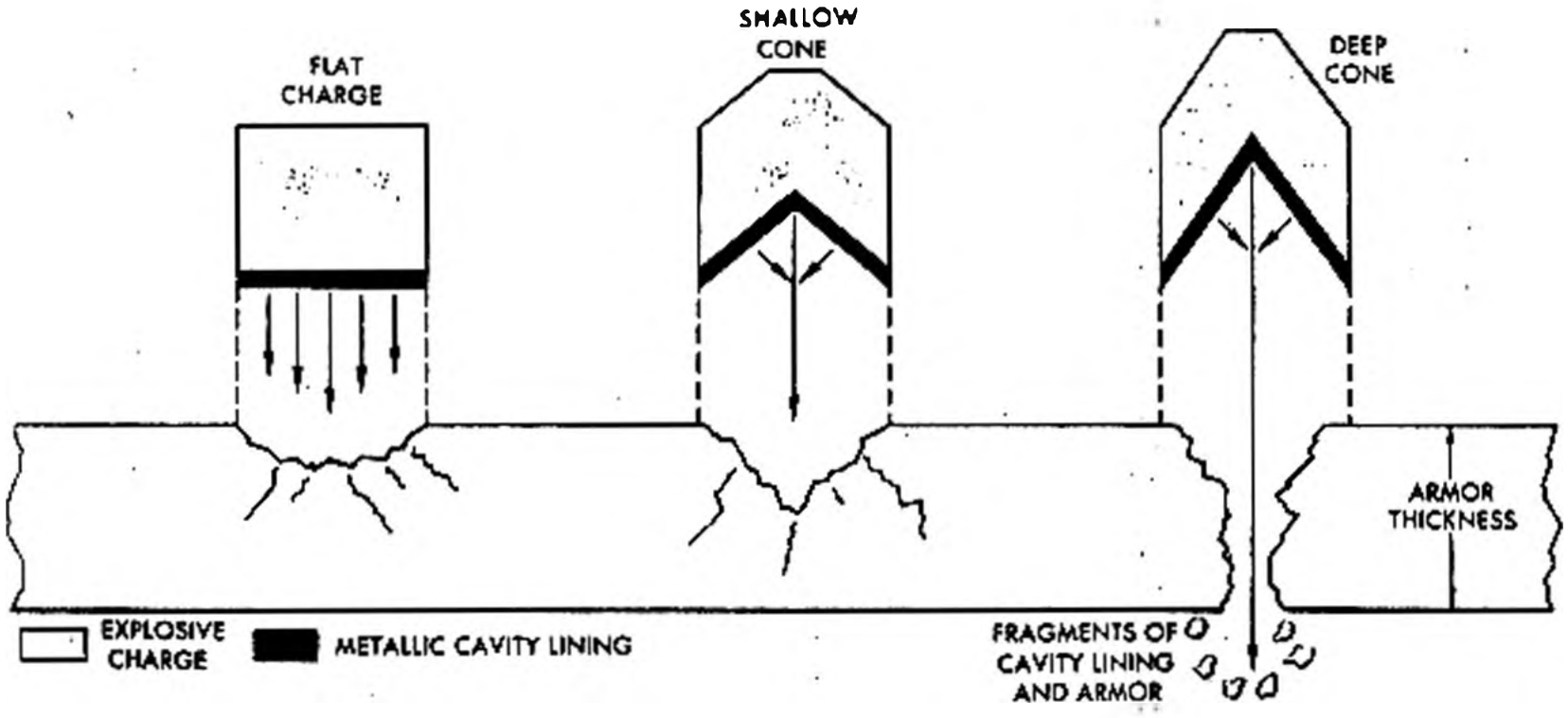


Figure 10-5. Effects of Various Shaped Charges

are used mainly against surface targets. Magnesium, which burns at a temperature of approximately 2000°C, is a common incendiary material. Other substances that are used in incendiary warheads are napalm (jellied gasoline) and phosphorus. The type of incendiary material used and the point of detonation of the warhead are determined by the characteristics of the intended target.

The basic construction of a magnesium incendiary warhead is illustrated in figure 10-6. It consists of a magnesium casing packed with thermite (aluminum and iron oxide) and a fuze with a suitable igniting mixture (powdered magnesium or black powder). The fuze ignites the thermite which burns violently. The burning thermite ignites the magnesium casing. An explosive charge may be used in the warhead to produce greater dispersion of the incendiary materials over the target area.

NUCLEAR WARHEADS. Because of national security, the amount of material available for publication relative to nuclear warheads is limited. No attempt is made here to present information pertaining to design characteristics or assembly of the components of a nuclear warhead. Instead, the general characteristics and effects of a nuclear explosion are explained.

When a nuclear detonation occurs, a great amount of radiation energy is released. Exposure to nuclear radiation can cause severe radiation sickness and death. The effect of radiation varies with the dosage. A low concentration for a long period of time can be as fatal as a high concentration for a short time.

A great quantity of heat is developed at the instant of a nuclear explosion. When the explosion occurs a large "ball of fire" develops. The tem-

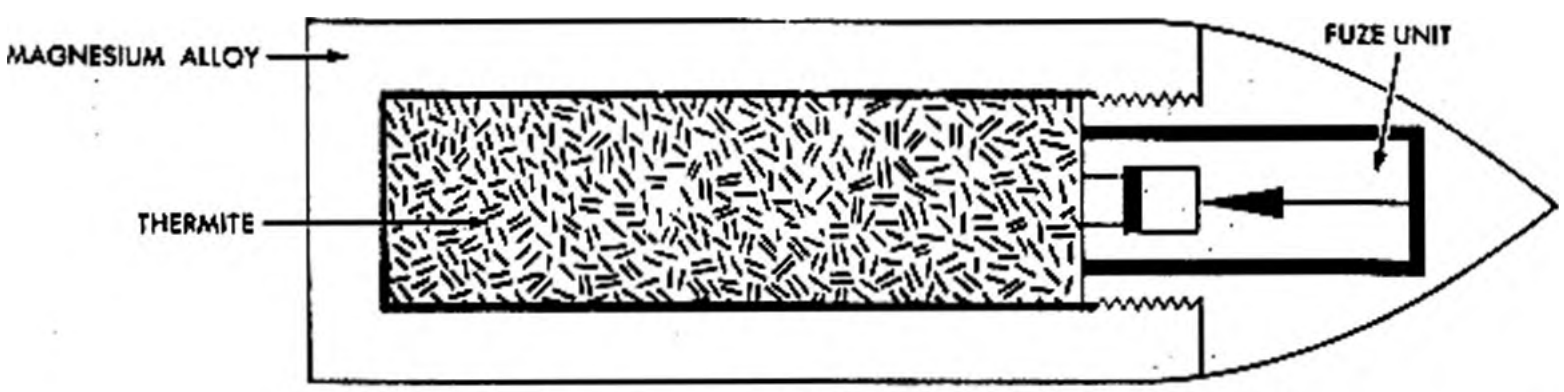


Figure 10-6. Basic Construction of Magnesium Incendiary Warhead

perature in the center of these hot gases reaches millions of degrees. The fireball rises vertically and radiates its energy in the form of visible, ultraviolet, and infrared waves. These intense heat waves cause blindness and skin burns. At the same time, the heat causes combustion of flammable materials up to several miles from the blast.

The temperature and pressure of the gases generated by a nuclear explosion develop a destructive blast wave. This wave consists of two pressure regions, one higher than normal atmospheric pressure and one lower. The blast wave moves outward from the center of the explosion with velocity and pressure differentials that vary directly with the intensity of the explosion. Figure 10-7 illustrates the propagation of the blast waves just described.

The explosion illustrated in figure 10-8 is initiated when a neutron (electrically neutral particle of matter) is caused to enter the nucleus of an atom of a suitable material such as uranium or hydrogen. The atom "blows up," ejecting two lighter atoms and a number of particles of matter including additional neutrons. This atom-splitting process is known as nuclear fission. Millions of electron volts of energy are given up for each atom which undergoes fission. As the fission reaction is produced by a neutron and since the reaction throws out more neutrons, other atoms of the original mass are split, producing two or three more neutrons each. This reaction grows to explosive proportions within a very small fraction of a second.



Figure 10-7. Blast Waves Formed by Nuclear Explosion

The effects of a nuclear detonation are divided into two phases, as shown in figure 10-8.

Phase 1 consists of those phenomena which accompany the actual explosion. Pressure effects have already been explained. Alpha and beta particles are emitted at the instant of the explosion. The range of these rays are relatively short and, while harmful, do not present a serious threat. Some of the electromagnetic radiations such as infrared, visible light, and ultraviolet radiations have been explained previously. The danger of X-rays is well known by all of us. Gamma radiation has the same general characteristics as X-ray, but has more energy and will penetrate greater thicknesses of materials.

Phase 2 consists of radiation that persists after the explosion. Unfissioned radioactive material is reduced to a powder and scattered by the explosion. Alpha particles emitted by this material prove to be a great hazard if taken into the body through the mouth or through open cuts.

The portion of the original matter that undergoes fission produces radioactive particles that are scattered by the explosion. These particles emit beta and gamma radiation. As stated previously, the beta rays are not particularly dangerous, but the gamma radiation may be hazardous over a large area after the explosion.

A hazard resulting from radioactivity induced in the components of the target area also develops from a nuclear explosion. This radioactivity is induced by the neutrons which are released at the instant of the detonation. These neutrons enter the nuclei of some atoms found in some target components, causing a physical reaction which leaves a radioactive atom as a product. High levels of gamma radiation may result from this action.

Energy is also available from *combining fr particles* (protons and neutrons) to form a nucleus. Vast amounts of energy are released by combining isotopes of the lighter element (hydrogen). This process is called nuclear fusion.

These reactions do not take place as easily as the fission process because of the repulsive forces of the nuclei. The neutron, being uncharged, easily enters a uranium nucleus, but positively charged nuclei repel each other. To overcome this repelling force, temperature must be increased to excite

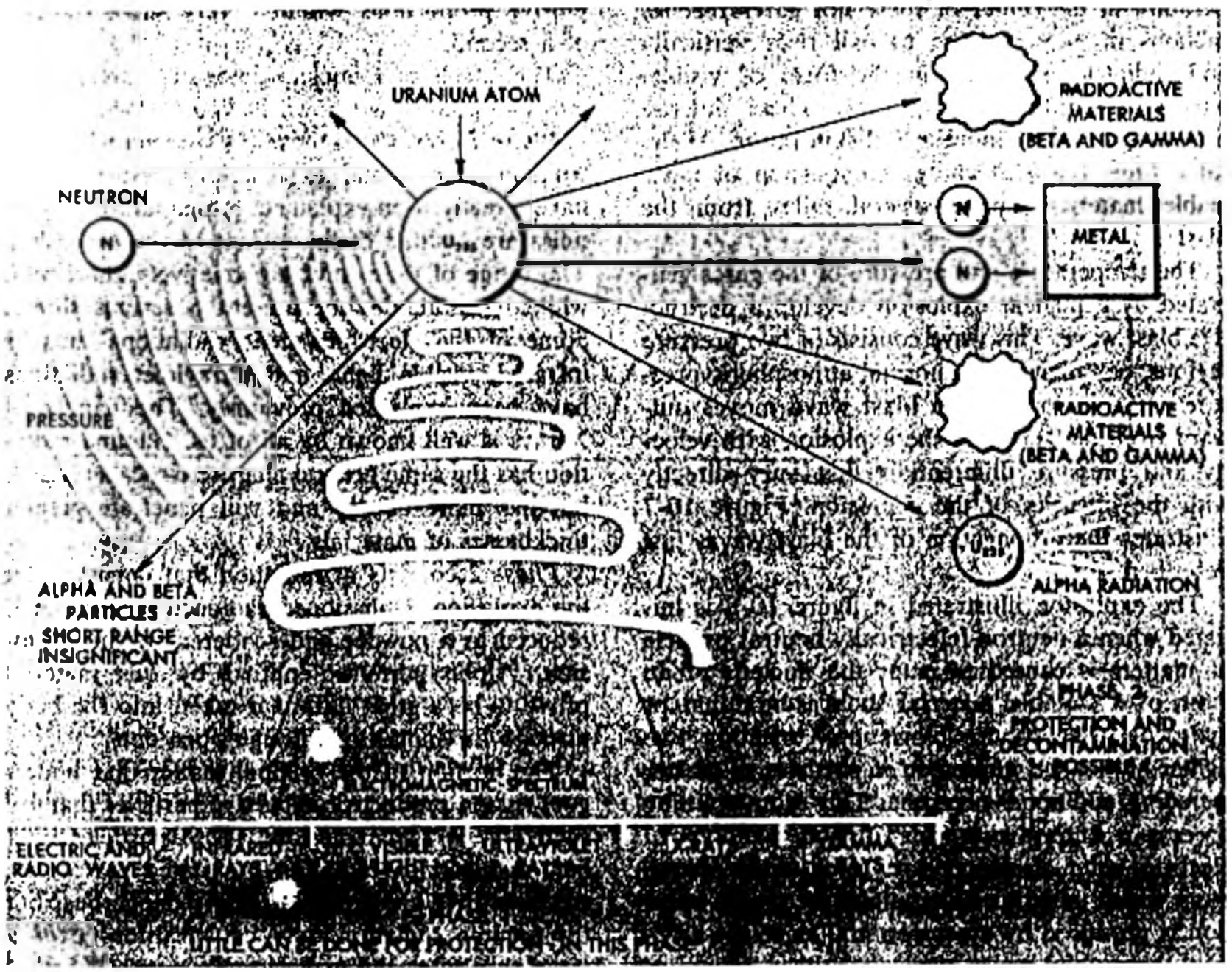


Figure 10-8. Activity Stemming from Fission Reaction

the atom to sufficient energy to fuse. On a weight comparison basis, much more energy is released by fusion than by fission. Temperature induced (thermonuclear) reactions therefore are much more efficient.

Extremely high temperatures are required for fusion, and only a fission reaction can provide these temperatures. Therefore, a fission weapon is used in conjunction with a thermonuclear weapon to achieve large yields.

FUZES FOR MISSILE WARHEADS

A fuze is a device designed to activate the bursting charge in a warhead at the proper time for maximum effect against a particular target. Special design characteristics apply to each type of target.

Description of Various Fuzes

Fuzing of a warhead depends on the overall characteristics of the target, attacking missile, and warhead. Target considerations are location, vulnerability, speed, and physical structure, along with the capabilities of the missile and warhead, in order to make the probability of damage to the target as high as possible. Various types of fuzes are explained below.

IMPACT FUZE. An impact fuze is actuated by inertia when the warhead strikes the target. For example, consider a cylindrical tube located in the warhead with a shock-sensitive explosive-percussion charge fixed in the front end of the tube and a movable metallic plunger at the rear end of the tube. While the missile is in flight, the metallic plunger is retained at the rear end of the tube.

When the missile strikes the target, inertia forces the plunger forward where it strikes and detonates the shock-sensitive fuze charge. The fuze charge, in turn, detonates the bursting charge of the warhead.

A time delay element may be used with an impact fuze to allow the warhead to penetrate the target before detonation.

TIME DELAY FUZE. A time delay fuze is designed to detonate the warhead at some predetermined time after the missile flight has begun. There are various types of time delay elements in use with missile warheads. One type of time delay element consists of a burning powder train. Time delay is determined by the length and burning rate of the train. Another type of delay element resembles a clock movement. The time delay interval must be set before the missile is launched and cannot be changed after the missile is in flight. The characteristics of the target are an important consideration in determining the requirements of a time delay fuze.

PROXIMITY FUZES. Proximity fuzes are actuated by some characteristic feature of the target or target area. Listed below are some basic proximity fuzes:

- Photoelectric proximity fuze.
- Accoustic proximity fuze.
- Pressure proximity fuze.
- Radio proximity fuze.
- Electrostatic proximity fuze.

Each of these fuze types is preset to function when the intensity of the target characteristic to which the fuze is sensitive reaches a certain magnitude. Proximity fuzes are designed so that the warhead burst pattern will occur at the most effective time and location relative to the target.

Classification of Fuzes by Position in Warhead

When classified according to assembled position in the warhead, fuzes are generally classified as *point fuzes* (assembled in the nose of the warhead), or *base fuzes* (assembled in the rear of the warhead). The fuze or combination of fuzes used in a warhead and the fuze location depend on the mission of the missile and the effect desired.

Warhead Detonation Points Relative to Fuze Type

Now that the various types of fuzes have been described from the standpoint of structure, operation and location, let's look at the influence of various fuzes on the point of warhead detonation.

Figure 10-9 illustrates the influence of the fuze on the point of warhead detonation. In the case of an impact-fuzed warhead (middle drawing), the explosion occurs when the missile strikes the target.

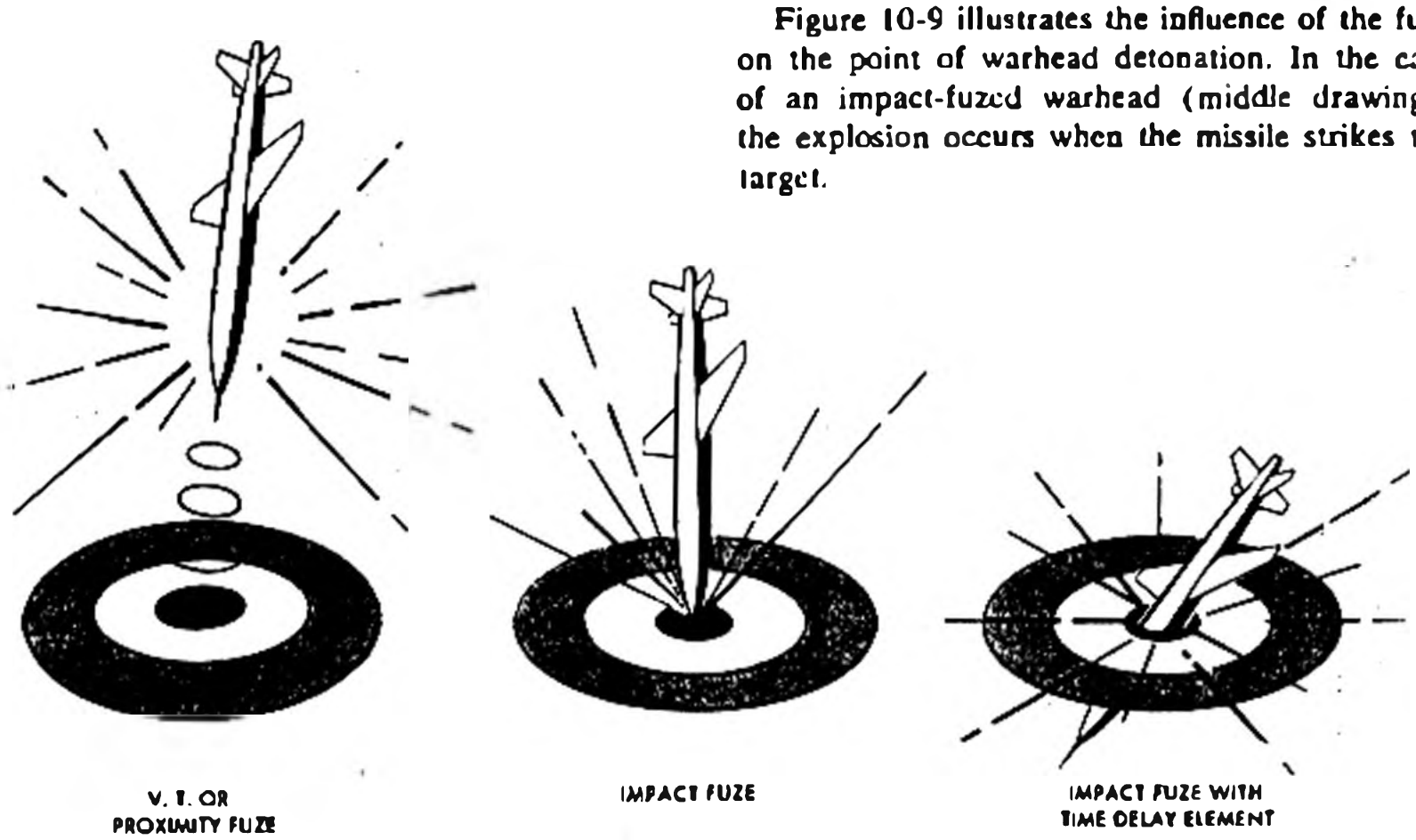


Figure 10-9. Influence of Fuze on Point of Warhead Detonation

In the right drawing, a time delay element is used in conjunction with an impact fuze. The delay element allows the warhead to penetrate the target before detonation. In this case the fuze would probably be the *base* type fuze located in the rear of the warhead. This location precludes damage to the fuze by impact with the target. In the left drawing, the proximity fuze is actuated by some characteristic of the target such as radio-frequency energy. The warhead in this instance is detonated at a predetermined distance from the target.

Many air-to-air missiles are designed so that the missile penetrates the target before the warhead detonates. These designs generally use a fuze actuator that is located somewhere on the aft-body of the missile. When the missile penetrates the fuselage of a target aircraft this actuator is contacted and in turn detonates the warhead. Proximity fuzes are also used on some air-to-air missiles. In the event that an air-to-air missile misses its target, it is possible that it could strike a friendly aircraft or friendly forces on the ground. To prevent this, these missiles are equipped with a self-destruct circuit in the fuzing system. When the missile is fired a timer actuates. If the missile fails to hit the target within a set number of seconds the warhead is detonated.

AEROSPACE GROUND EQUIPMENT

Present missile weapon systems are a complex organization of electrical, electronic, pneumatic, hydraulic, and propulsion systems. Equally complex are the ground support equipment (GSE) systems and facilities needed to prepare the missile for launch. The following discussion presents the general requirements and layout of typical missile launch environments.

GSE for Air-Launched Missiles

Air-launched missiles are launched from aircraft. The target may be another aircraft, a ground facility, or the missile may be a decoy used to confuse the enemy and allow a bomber to increase its capability to penetrate the enemy defenses. The missile must be matched to its mission and also to the particular carrier aircraft. In general, the following paragraphs cover the areas of checkout and maintenance, and launching the missile.

CHECKOUT AND MAINTENANCE FACILITY. The missile is little different from any other weapon in its requirement for maintenance. A missile needs basically the same checks as an aircraft. These checks include complete operation of the guidance system, control system, propulsion system, and warhead system.

If the missile uses a solid fuel for propellant, it cannot be actually fired during checkout, but the system can be checked to determine if it is ready for operation when needed.

Figure 10-10 shows a typical air-to-air missile checkout facility. Within this one building the missile can be completely checked out except for an engine run and the detonation. Most of these missiles use solid fuel engines, so there is no requirement for the engine check.

Notice in the figure that there is room provided for removal of the missile from shipping containers upon initial receipt, as well as handling those missiles returned from the flight line.

From this point the missile follows a preset path for its functional checkout prior to return to an aircraft or to the storage area. The missile may be sidetracked as needed for repairs or overhaul.

Figure 10-11 shows the missile checkout console. The console is designed to check out a Falcon missile, which uses an infrared guidance system. Within this one console, all systems of the missile can be checked out. Similar equipment is used to check out other air-launched missiles.

Notice that the missile can be placed on the roller conveyor and cabled to the console, and complete systems checks performed. Malfunctions or discrepancies can be corrected by repair or replacement, and a thoroughly checked and operational missile is removed from the console.

Since the Falcon comes in several circuit configurations, an identification circuit is built into the missile. This enables the console to identify the circuit configuration that must be used to check out the missile. The checkout then becomes a semiautomatic procedure with the technician checking the console display devices (meters, lights, or digital readout) for the proper indication. Discrepancies are noted and corrected as they occur.

LAUNCHER. One launcher used for air-launched missiles is similar to that shown in figure 10-12. It is an external pylon installed on the aircraft

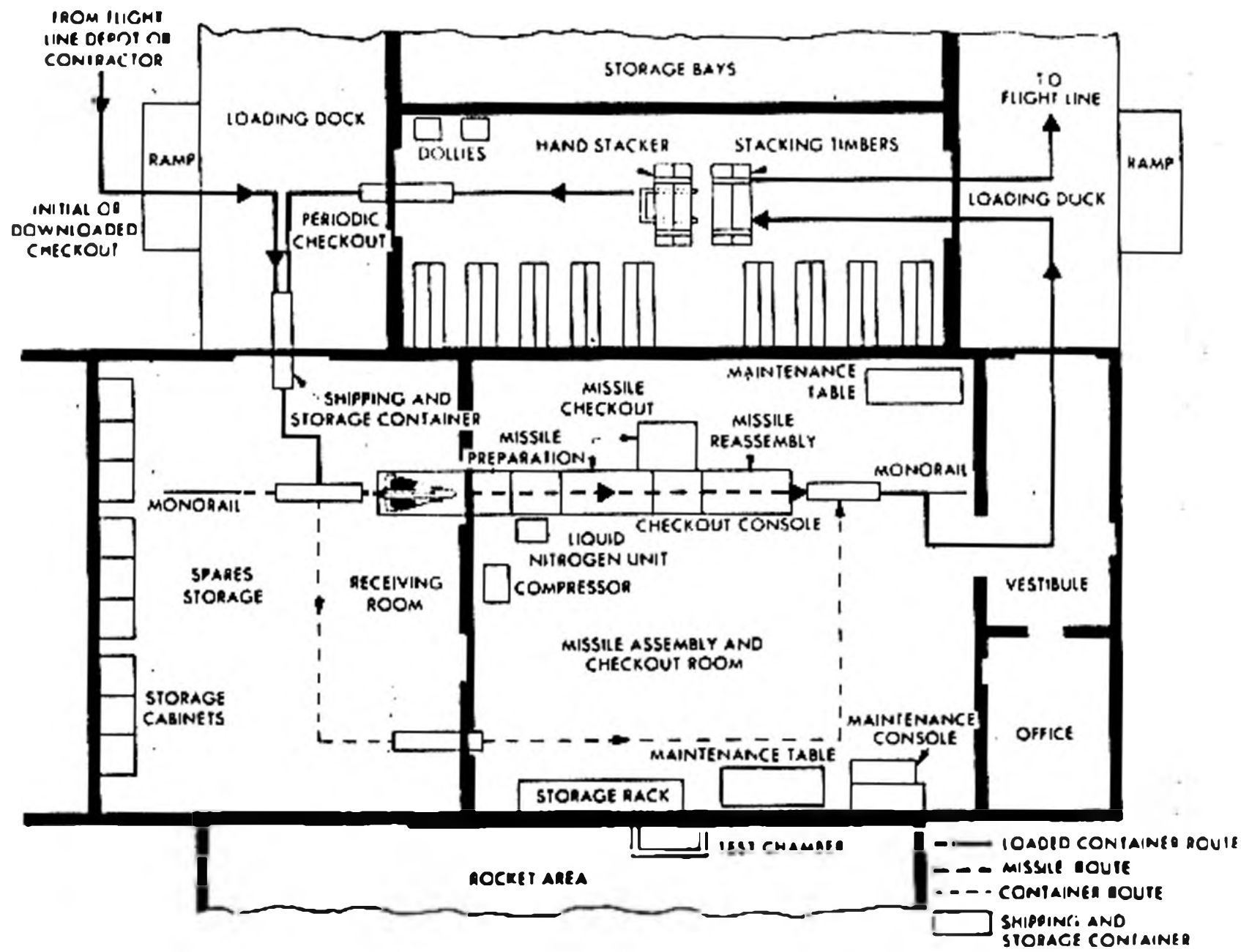


Figure 10-10. AIM Checkout Facility

wing. The missile is mechanically and electrically connected through the pylon to the aircraft. Firing circuits actuated by the pilot release the missile and ignite the propellant.

As with the installation of any munitions on or in an aircraft, care must be exercised by the loading crew to make sure that the missile remains capable of being fired.

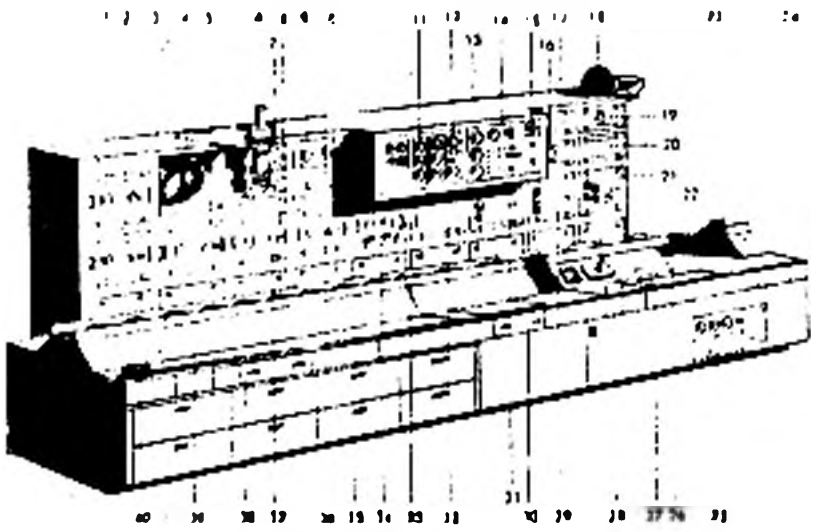
The appropriate technical orders give the proper steps and safety precautions to be followed to accomplish the loading. Since these are constantly changing, they are not discussed here except to point out two items. These are the mechanical and electrical connections.

The mechanical connection must be correctly made to allow the missile to make a clean release from the aircraft. Since the release may occur either as a drop with motor ignition occurring after clearance from the aircraft or as a propelled release with motor ignition occurring prior to release, you can see that any hindrance could cause damage to the missile or the aircraft.

The electrical connection is as important as the mechanical connection. Electrical connectors should be clean, pins straight, and proper fit assured at the time the cables are connected. Improper electrical connection can lead to a misfire, expending the missile prior to release, or inoperative electronic circuit function resulting in missile failure.

GSE For Ground-Launched Missiles

The headline-gathering missiles are those launched from the ground. These missiles, developed as weapon systems, have played an important role in our space research program. As



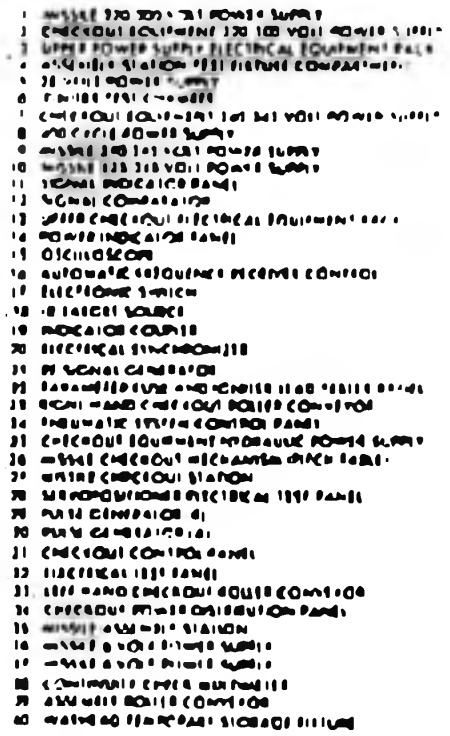


Figure 10-11. AIM Checkout Console

weapon systems, the type of launcher used for these missiles have been changed from above-ground structures to big holes in the ground. The days of having a missile weapon system installed above ground seem to have passed.

The facilities used to perform the missile checkout are similar to those discussed in the previous section. Because of the increased size of the missile, the buildings must be larger to allow the entire missile to be brought inside and disassembled as needed for the system checkout. System components may then be routed to their respective areas for checkout and maintenance.

The launch facility is where the greatest differences are evident. There are also considerable differences among the various weapon systems. The facilities that follow are for the CIM-10A, LGM-25, and LGM-30.

CIM-10A GROUND-TO-AIR MISSILE. The CIM-10A is popularly called the Bomarc. The launching facility is a coffin whose roof opens to allow missile erection and launch. Figure 10-13 shows a typical CIM-10A maintenance and launch area. Remember that this is an intercept missile and the launch is remotely controlled from some central control point far removed from the missile site. Because of this there is no launch control section designated in the figure.

Within each missile shelter are the needed facilities for completely monitoring the missile status by the remote missile control facility. Any abnormal status checks are reported back to the site so maintenance people can be immediately dispatched to correct the malfunction and return the missile to a ready status or replace the missile with a spare.

Without the launch crew facilities, the area is essentially only a combined ready-to-fire storage and maintenance facility. The maintenance people place the missile into the coffin and return to it only when necessary. The missile is fully fueled

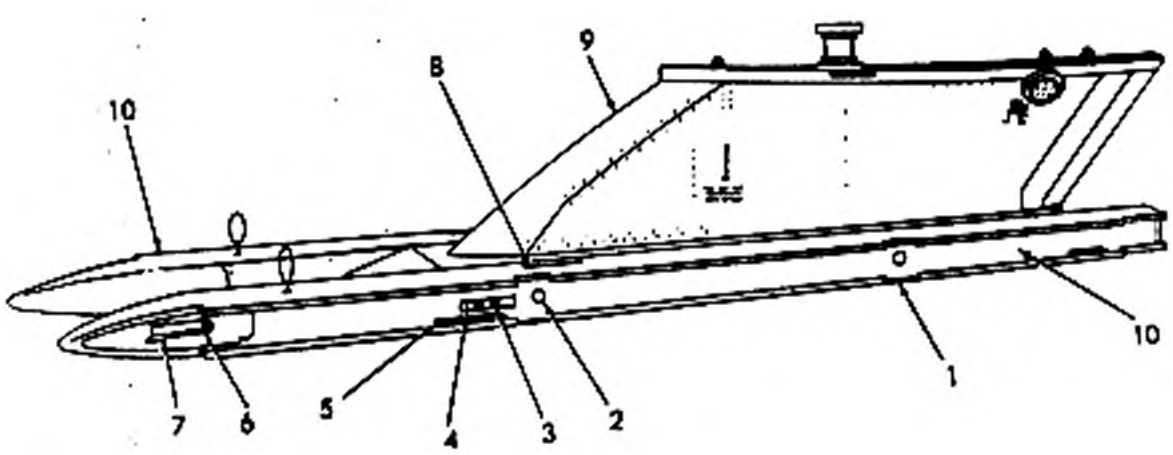


Figure 10-12. AIM Launcher

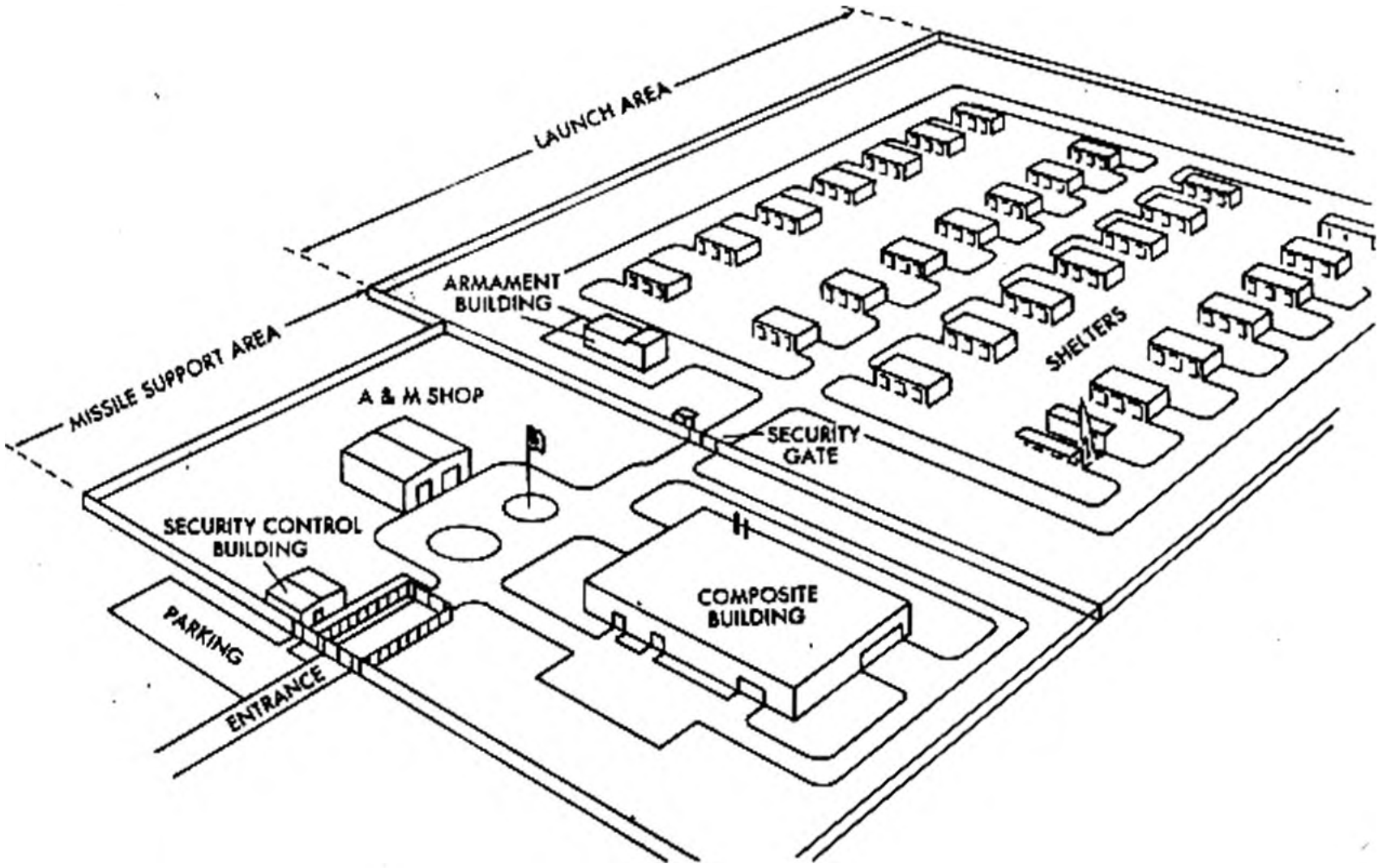


Figure 10-13. CIM-10A Launch and Support Facility

when put into place, and fuel storage and transfer facilities are not needed in the launch area.

Since the missile status is monitored by a remote organization, the monitoring equipment is not needed at the launch site. This again simplifies the launch facility equipment.

LGM-25 SURFACE-TO-SURFACE MISSILE. The second generation of the Titan is the LGM-25, TITAN II. A hardened site, called a silo, is used to launch this missile. Figure 10-14 shows a typical launch facility. Detail A relates the launch control center to the silo. Each silo has its own control center where missile status is monitored and periodic operational checks may be performed, as well as final commitment for launch.

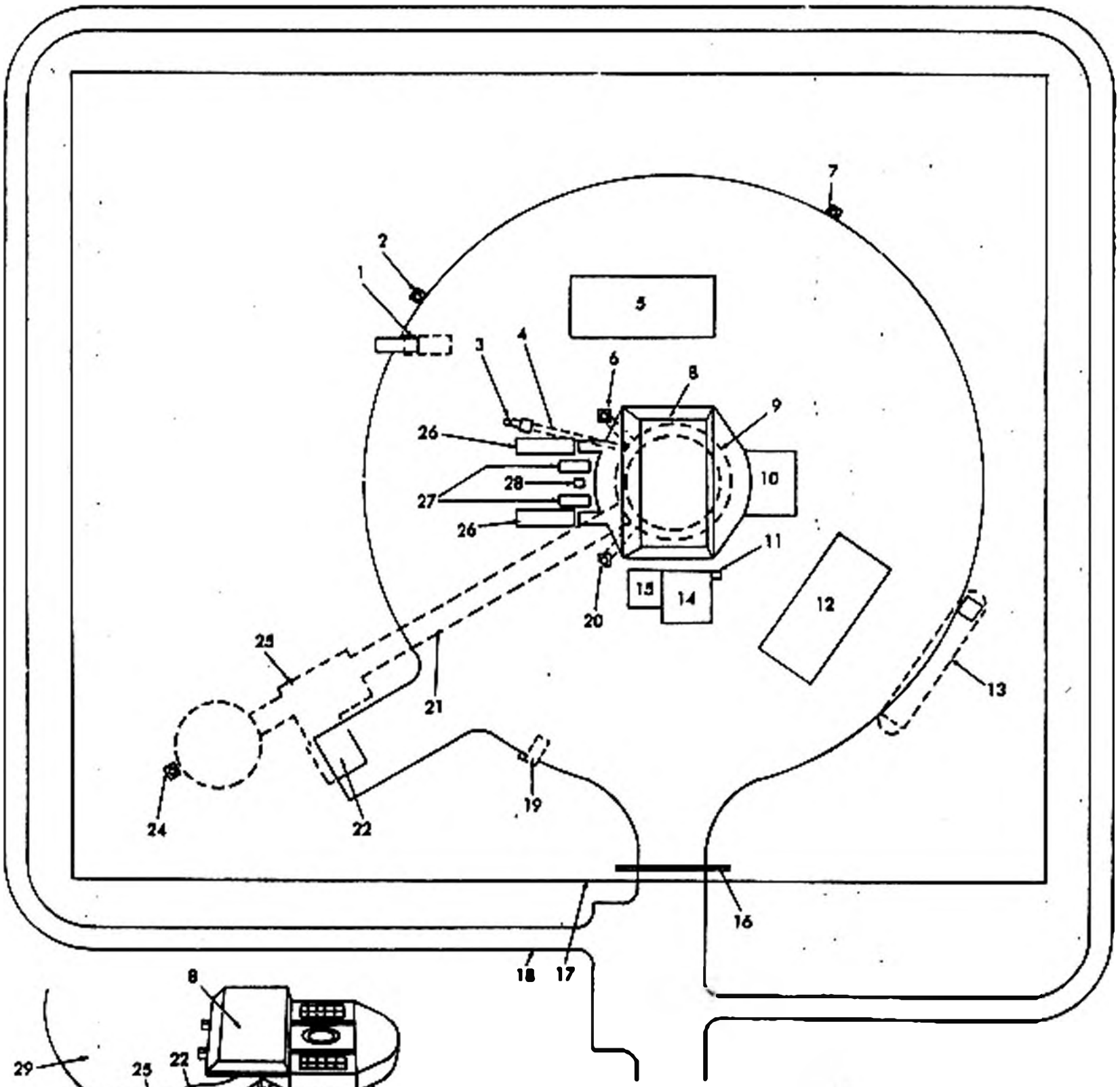
The control center is the heart of the complex. Here are the launch crew facilities for eating and sleeping as well as the complex electronic equipment for monitoring the missile. Also provided is a closed circuit television system monitoring the immediate terrain around the site.

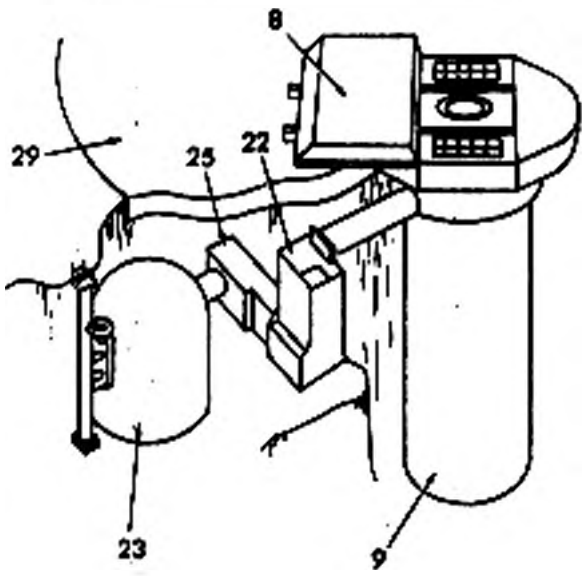
During the period of time that each launch crew is on duty, constant monitoring of the missile status is maintained. Operational readiness checks that include operation of all systems except engine start are also performed. Malfunctions that occur are isolated as much as possible, and maintenance personnel are called upon to repair or replace parts. These people are dispatched from the maintenance activity area to return the missile to a ready-to-launch status.

Fuel and oxidizer storage and transfer facilities are provided at each launch site. The missile is fueled completely after installation in the silo and remains in that status until removed.

The internal configuration of the missile silo is shown in figure 10-15. The work platforms at the various levels enable the maintenance personnel to reach the system components of the missile in the event a malfunction occurs.

The missile is held suspended above the bottom with a shock system. This is to prevent missile damage in case of a natural catastrophe such as an earthquake, as well as in case of attack. During the time the missile is installed in the silo, the launch crew subjects the missile to continuous





1. OXIDIZER DUMP TANK
2. OXIDIZER FLARE STACK
3. SURVEY TRANSFER STATION
4. SIGHT TUBE
5. OXIDIZER HARDSTAND
6. AIR EXHAUST
7. FUEL FLARE STACK
8. SILO CLOSURE DOOR
9. MISSILE SILO
10. CRANE HARDSTAND
11. VALVE PIT
12. FUEL HARDSTAND
13. FUEL DUMP TANK
14. COOLING TOWER PIT
15. TRANSFORMER PIT
16. SURFACE GATE
17. SECURITY FENCE
18. SERVICE ROAD
19. DIESEL STORAGE TANK
20. AIR INTAKE
21. CABLEWAY
22. ACCESS PORTAL
23. CONTROL CENTER
24. AIR INTAKE AND ESCAPE HATCH
25. BLAST LOCK
26. BUMPER PAD (2 PLACES)
27. DOOR ACCESS PIT (2 PLACES)
28. CATCH BASIN
29. HARDSTAND

Figure 10-14. LGM-25 Launch Complex

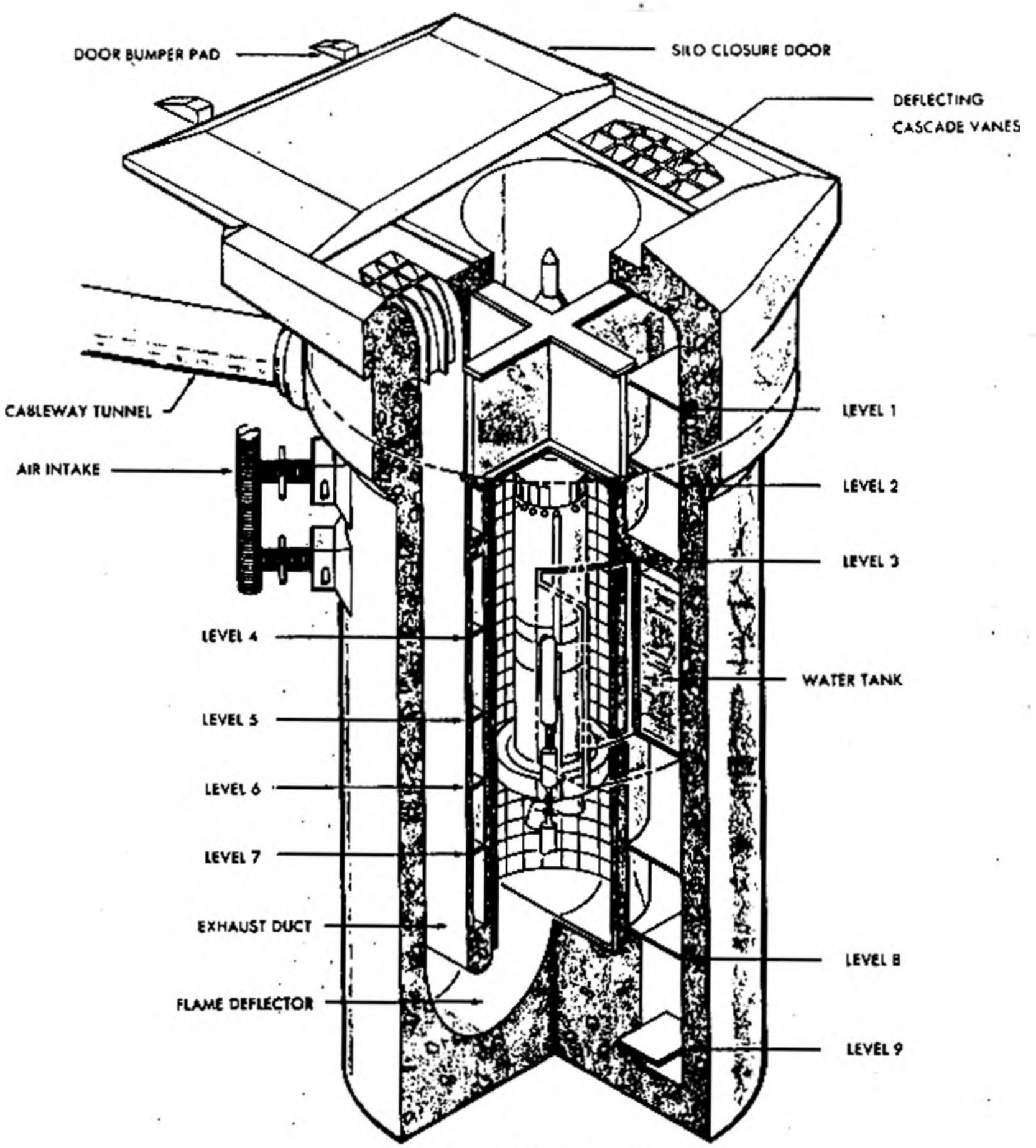
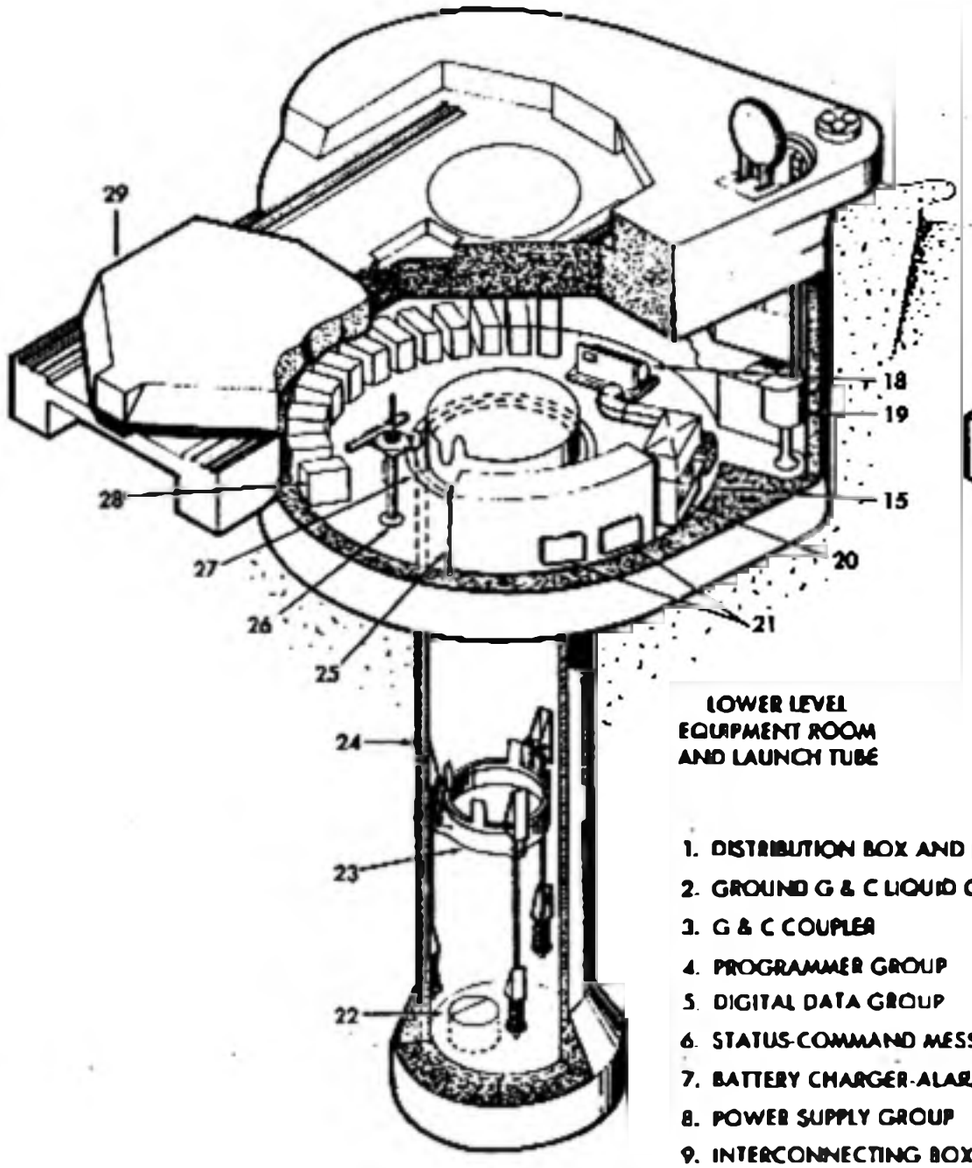


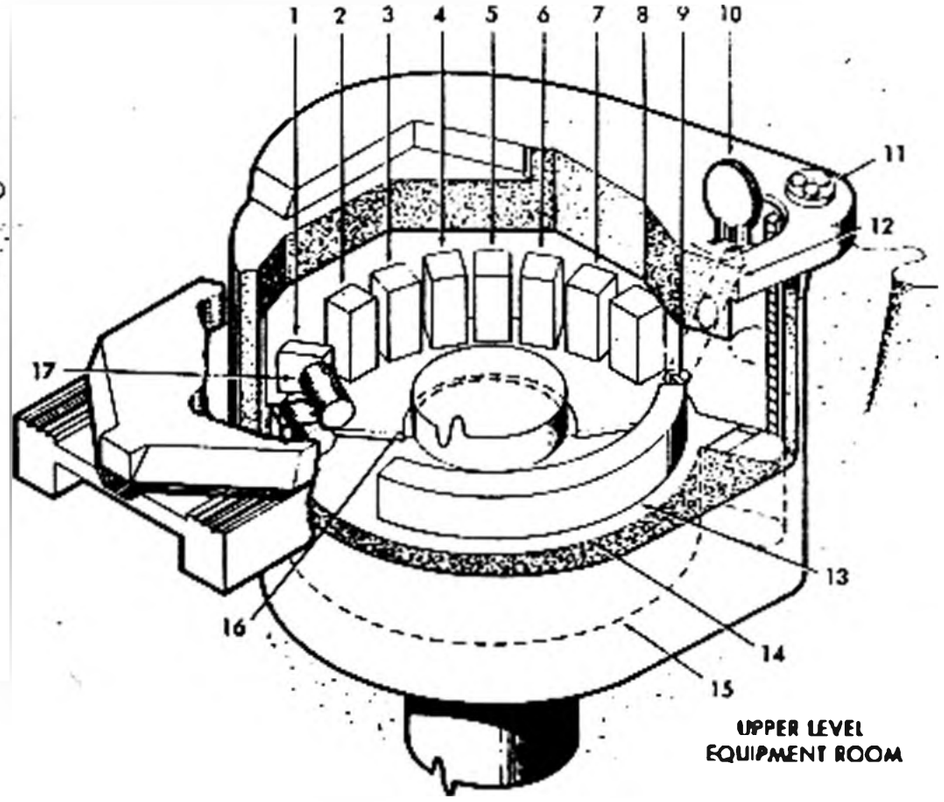
Figure 10-15. LGM-25 Launch Silo

monitoring. They may also perform periodic exercises in which the missile is made ready to go except for the final order.

LGM-30 MINUTEMAN MISSILE. This missile is a solid-fueled surface-to-surface missile encased in a concrete lined hole much the same as the Titan II. Since it is a solid-fueled missile loaded and ready to go at all times, there is no need for the fuel and oxidizer storage tanks. The other systems require as much monitoring and operating equipment as the preceding missiles.

A typical Minuteman silo is shown in figure





1. DISTRIBUTION BOX AND (VRSA CONTROL PANEL)
2. GROUND G & C LIQUID COOLING EQUIPMENT
3. G & C COUPLER
4. PROGRAMMER GROUP
5. DIGITAL DATA GROUP
6. STATUS-COMMAND MESSAGE PROCESSING GROUP
7. BATTERY CHARGER-ALARM SET GROUP
8. POWER SUPPLY GROUP
9. INTERCONNECTING BOX
10. PRIMARY DOOR
11. VAULT DOOR
12. LAUNCHER SIGHTING TUBE
13. COLLIMATOR BENCH
14. 1ST LEVEL FLOOR
15. 2ND LEVEL FLOOR
16. SHOCK MOUNTED 1ST LEVEL FLOOR
17. LIQUID STORAGE METAL TANK SET
18. MOTOR GENERATOR
19. SECONDARY DOOR
20. AIR CONDITIONING UNIT
21. ELECTRICAL SURGE ARRESTER SET
22. SUMP PUMP
23. MISSILE SUSPENSION SYSTEM
24. LAUNCH TUBE
25. CABLE PROTECTOR FRAME ENCLOSURE
26. CLOSURE ACTUATOR
27. BATTLE SPACE
28. STORAGE BATTERY SET
29. LAUNCHER CLOSURE

Figure 10-16. LGM-30 Launch Silo

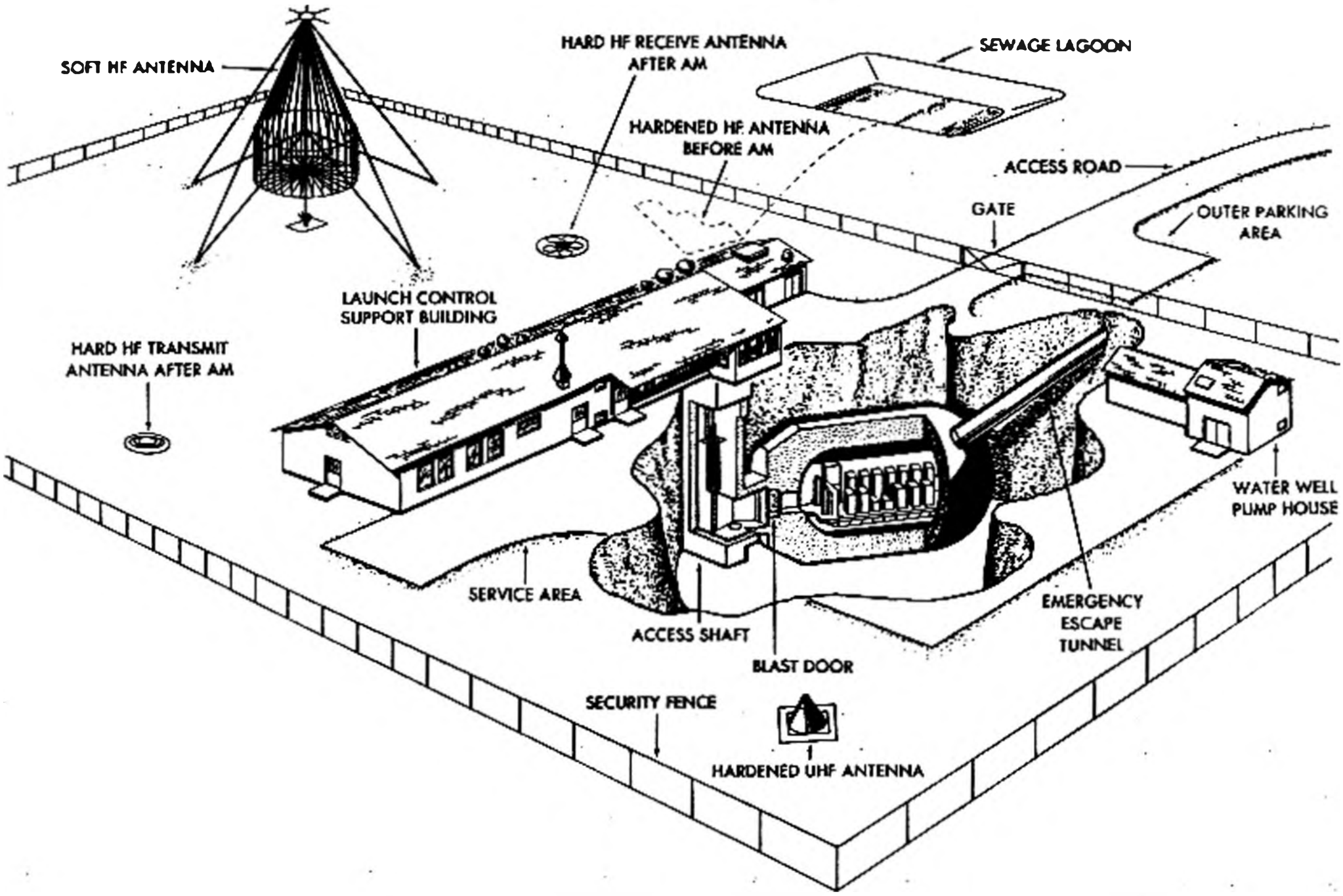


Figure 10-17. Launch Control Facility

10-16. The monitoring and operating equipment is contained on two levels at the top of the silo. Unlike the Titan II sites, the launch crew is located at some site other than the immediate silo location. All monitoring and control is done from this remote location.

The launch crew is able to perform certain operational readiness checks on each missile under its control. Maintenance area personnel are dispatched to the site as malfunctions are detected. Repair and/or replacement of the malfunctioning components is accomplished to return the missile to an operational status.

The launch control facility, shown in figure 10-17, is also a below-ground-level installation. From this point the launch crew monitors all missiles under its control.

With the launch facilities as described in the preceding paragraphs, extensive communication networks are needed. These networks must be available for instantaneous use and usually consist of telephones as well as radios. In all cases, safeguards are built in so that a launch crew cannot fire the missiles under its control without proper orders from higher authority.

CHAPTER 11

Principles of Lasers

INTRODUCTION

The summer of 1960 saw the world's first demonstration of an entirely new source of light, so concentrated and powerful that it could produce power densities millions of times as intense as those on the surface of the sun, and yet be controlled so precisely that surgeons could use it to perform delicate operations on the human eye. The beam from this device could burn holes in steel plates and set carbon on fire. It spreads out so little that if sent from the earth to the moon, it would illuminate an area of the moon's surface only ten miles in diameter. This new device is the laser. *Laser means light amplification by stimulated emission of radiation*, the process that takes place inside the device.

Simply stated, the laser is a beam of light. The light from a house light spreads in all directions in a number of various frequencies and waves. The very fact that light does travel in waves led to the development of the laser, which concentrates the waves of light into light beams of tremendous energy.

Originally developed by the use of ruby crystals, lasers are now produced by many solid materials, liquids, gases, and semiconductor devices.

The laser is one of the most important developments of modern science and is used extensively in medicine, by the armed forces, and in every facet of industry where precise measurements are needed. Hundreds of new applications are being proposed and developed by engineers and scientists and new uses appear unlimited in scope.

The armed forces are the largest buyers of the laser at the present time. The more important military applications include proximity fuzes, night-time vision and tracking, target illumination, and range-finding units that provide accurate measurements.

BRIEF HISTORY OF LASERS

The basic theory of lasers can be traced as far back as 1917 to atomic theories by Einstein, who pointed out that controlled radiation could be obtained from an atom under certain conditions.

Until the 1950s, lasers were strictly theory, but many scientists were working for reality. In June 1954, Dr. Charles H. Townes, then of Bell Laboratories, demonstrated an ammonia gas maser with adequate stimulation to sustain a maser beam of a signal equal to 24,000 megaHertz. The real significance of this demonstration was that Dr. Townes generated a signal without the use of a resonant cavity. This achievement opened the door to many experiments leading to the amplification and generation of electromagnetic radiation within the visible light spectrum wavelength.

Laser history actually begins with the *maser (microwave amplification by stimulated emission of radiation)*. Early masers were gaseous ammonia or rubidium. A maser was built in 1955, and a solid-state version was proposed in 1956. One of the first solid-state masers employed sapphire crystals (synthetic) doped with chromium oxide. It was later discovered, in 1959, that masers could amplify light as well as microwaves, although the extension of maser principles to permit operation in the light spectrum was suggested as early as 1953. This development was referred to as optical maser with the name changed to laser in 1965.

The light generating capability became more important than the amplifying ability. The pulsed

ruby laser, built in 1960, was the first to demonstrate the monochromatic and coherent light principle. This laser was of solid-state design and employed a synthetic sapphire crystal doped with chromium (ruby), Cr_2O_3 in Al_2O_3. The crystal was machined to an optical tolerance with both ends of the resulting rod silvered to enhance reflection. Bathed in white light from an electronic flash lamp, the green content provided energy to the crystal. This stimulated emission of coherent and monochromatic light was at 6943 Angstroms. The monochromaticity was about five times narrower than that of the ruby's natural fluorescence. Later developments brought this figure to 40 times, and then to 10,000 times the natural ruby fluorescence. Still greater monochromaticity is realized with the helium-neon gas laser.

The continuously operating gas laser was developed in 1961. It employed a mixture of helium and neon gas excited by an RF field, and it obtained an output in the infrared range. Laser action was produced with trivalent uranium in calcium fluoride in 1960. This was followed by other solid-state experiments. During 1961, amplification of light was performed using a ruby laser driven by a ruby laser oscillator. The generation of optical harmonics was achieved in 1961 using a laser source directed into crystalline quartz. Ruby laser light at 6943 Angstroms was raised to a shorter wavelength value of 3472 Angstroms.

The function of junction electro-luminescence in science and technology was altered rapidly back in the early 1960s. In 1961 gas lasers and optically pumped solid lasers stimulated research into the possibility of a semiconductor junction laser. Then, the discovery in 1961 and 1962 that the efficiency of recombination radiation was very high in gallium arsenide focused attention on this compound as a potential laser material. The first injection lasers were operated in late 1962. In the following years, great advancements were made in the development of basic materials, and in the application of these materials in workable laser devices for a wide variety of uses. Today, the principal types of lasers include solid-state lasers (for both pulsed and CW operation), injection lasers (for both pulsed and CW operation), gas lasers (for both pulsed and CW operation), and liquid, plastic, and Raman lasers. Each type uses a large variety of different materials. Some materials have been standardized in working laser devices, while other materials are used only for research effects. Development is still progressing at a very fast rate with new materials appearing constantly.

Atomic Theory

A brief review of atomic theory is essential to understand the principles of operation of the laser. You probably have studied the structure of matter in high school or college courses in physics, electronics, or other sciences. The treatment given here is only the essential knowledge.

THE ATOM. The atom is defined as the smallest particle of an element that retains all of the properties of the element. The atom is the smallest part of an element that enters into a chemical change, but it does so in the form of a charged particle. These charged particles are called *ions*, and they are of two types—*positive* and *negative*. A positive ion may be defined as an atom that has become positively charged. A negative ion may be defined as an atom that has become negatively charged. Ions of like charges tend to repel each other while ions of unlike charges attract, a fact well recognized from the study of magnetism and electricity.

Atoms have been found to be divisible into more fundamental particles called *electrons*, *protons*, and *neutrons*. The electron was first discovered as the basic unit of electricity. It is a small, negatively charged particle much lighter than an atom. The proton is a positively charged particle with the same magnitude of charge as the electron but is much larger than the electron. The neutron is formed when a proton and electron are combined, and is neutral in charge.

The *nucleus* of the atom consists of a group of positive and neutral particles (protons and neutrons), surrounded by one or more negative orbital electrons. Figure 11-1 shows the arrangement of these particles for an atom of the element boron. This concept of the atom can be likened to the solar system in which the sun is the central body around which the planets revolve in orbits at various distances from it. The electrons whirl about the nucleus of the atom much as the planets whirl about the sun.

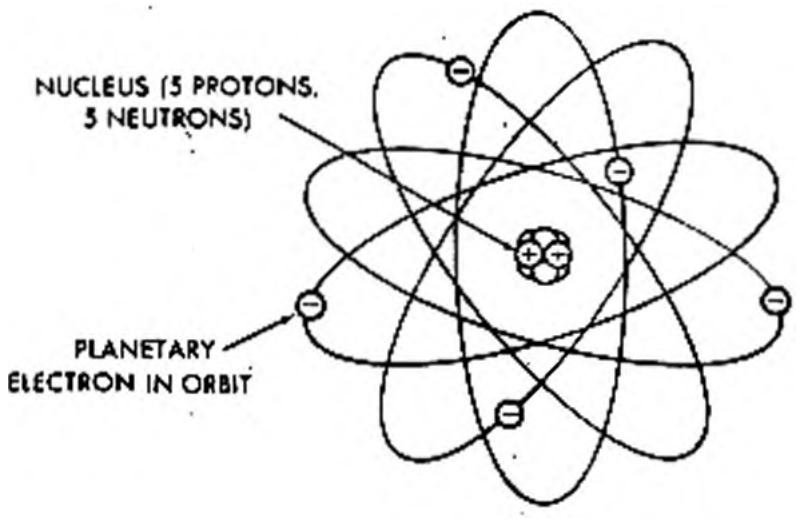


Figure 11-1. Boron Atom

In the lighter elements, the nucleus contains approximately one neutron for each proton while in heavier elements the neutrons will outnumber the protons. The nucleus of the helium atom consists of two neutrons and two protons. The mercury atom, a heavy element, has 80 protons and 120 neutrons in its nucleus.

PLANETARY ELECTRONS. Surrounding the positive nucleus of the atom is a negative-charged cloud made up of planetary electrons. Each of these electrons contains one unit of negative electricity equal in charge to the unit of positive electricity contained in the proton. In a normal atom, the number of planetary electrons is exactly equal to the number of protons in the nucleus. Therefore, the net charge of an atom is zero, since the equal and opposite effects of the positive and negative charges balance one another.

If an external force is applied to an atom, one or more of the outermost electrons may be removed. This is possible because the farther the electrons are from the nucleus, the less attraction they have to the nucleus. When atoms combine to form an elemental substance, the outer electrons of one atom will interact with the outer electrons of neighboring atoms to form bonds between the atoms. When bonding occurs in some substances, each atom retains its full complement of electrons. In other substances, one or more outer electrons will be gained or lost as a result of bonding. The electron configuration of the atom is of great importance as the chemical and electrical properties of a material depend almost wholly upon the electron arrangement within its atoms.

The nucleus of the atom is well shielded by the electron cloud and does not enter into chemical or electrical processes. To dissipate the nucleus of an atom requires a vast amount of energy such as is released by each atom in the explosion of an atomic bomb.

THE HYDROGEN ATOM. The simplest of all atoms is that of the element hydrogen. The hydrogen atom is composed of a nucleus containing one proton and a single planetary electron. According to a concept developed by Niels Bohr (figure 11-2), the electron travels about the nucleus in a circular orbit having a fixed radius. As the electron revolves around the nucleus it is held in this orbit by two counteracting forces. One of these forces is called *centrifugal force*, and is the force which tends to cause the electron to fly outward as it travels around its circular orbit. This is the same force which causes a car to roll off a highway when rounding a curve at too high a speed. The second force acting on the electron is *centripetal force*. This force tends to pull the electron in towards the nucleus and is provided by the mutual attraction between the positive nucleus and negative electron. At some given radius "r," the two forces will exactly balance each other, providing a stable path for the electron. For the hydrogen atom, the average radius is approximately 5.3×10^{-11} meter.

Bohr stated that since electrons favored certain energy levels, only certain orbits are possible. He maintained that an electron neither radiates nor absorbs energy as long as it stays in a particular orbit. When an electron goes from one orbit to another, however, it radiates or absorbs energy in discrete amounts called *quanta*. In 1922 Bohr was awarded the Nobel prize in physics for his

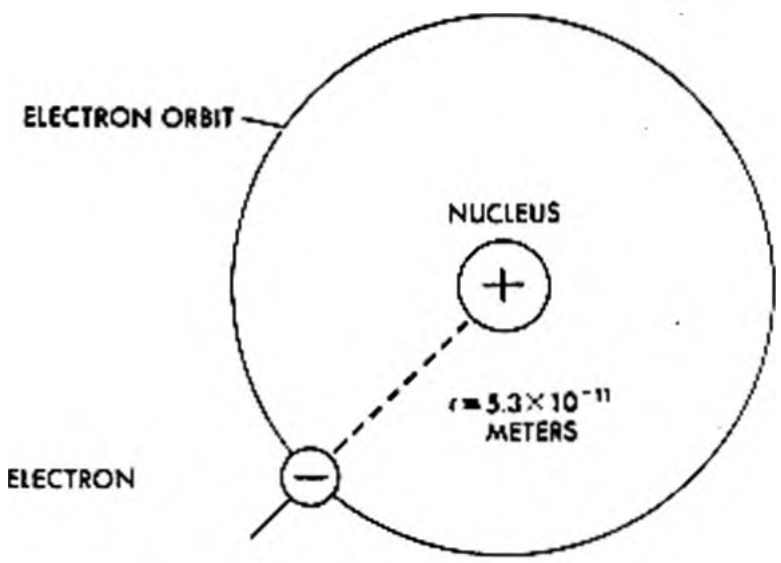


Figure 11-2. Hydrogen Atom

work. The model of the Bohr atom, figure 11-3, soon became the basis for the scientific explanation of matter.

QUANTUM PHYSICS CONCEPTS. The laser is often referred to as a *quantum electronic* device, thus necessitating the introduction of some important concepts of modern or quantum physics which directly apply to the laser. Certain pertinent facts relating to laser operation are presented here in expanded form as a foundation for the advanced topics treated in succeeding paragraphs. Max Planck was an obscure professor when he proposed the theory that if molecules of a material are excited, they will vibrate or oscillate and emit multiples of a unit amount of energy. This unit amount he designated as a quantum and defined it by a simple formula: $E = hf$, where E is the amount of energy (the quantum), h is a constant (Planck's constant) equal to 6.62×10^{-27} erg-sec, and f is the frequency of vibration or oscillation produced by the molecule. In effect, Planck described each molecule as a tiny oscillator that can generate E, 2E, 3E, 4E, etc. amounts of energy, but never a fractional amount of a quantum.

Planck's hypothesis that the energies of the electron oscillators responsible for radiation were quantized (restricted in certain integral multiples of a constant which now bears his name) were consistent with experimental findings. Planck's quantum hypothesis was successfully employed by Albert Einstein in explaining, among other phenomena, the photoelectric effect. In explaining the photoelectric effect, Einstein extended the quantum concept to the radiation itself. He assumed that light interacts with electrons in a metal as if the light were itself composed of discrete bundles of energy. Consequently, a light beam can be thought of as a stream of massless particles called *photons* which travel at the speed of light. Each particle contains the energy of hf joules.

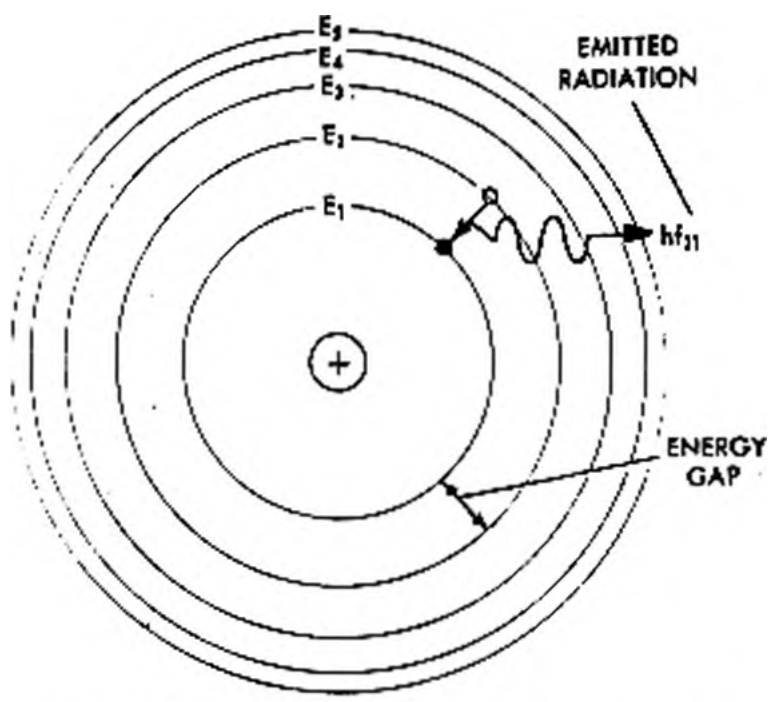


Figure 11-3. Bohr Model of Hydrogen Atom

ENERGY LEVELS. Since the electron in the hydrogen atom has both mass and motion, it contains two types of energy. By virtue of its motion the electron contains *kinetic* energy. Because of its position, it also contains *potential* energy. The total energy contained by the electron (kinetic plus potential) is the factor which determines the radius of the electron orbit. The orbit shown in figure 11-2 is the smallest possible orbit the hydrogen electron can have. For the electron to remain in this orbit, it must neither gain nor lose energy.

Light energy exists in tiny packets or bundles of energy called *photons*. Each photon contains a definite amount of energy depending on the color (wavelength) of light it represents. Should a photon of sufficient energy collide with the orbital hydrogen electron, the electron will absorb the photon's energy. The electron, which now has a greater than normal amount of energy, will jump to a new orbit farther from the nucleus. The first new orbit to which the electron can jump has a radius four times as large as the radius of the original orbit. Had the electron received a greater amount of energy, the next possible orbit to which it could jump would have a radius nine times the original. Each orbit may be considered to represent one of a large number of energy levels that the electron may attain. It must be emphasized that the electron cannot just jump to *any* orbit. The electron will remain in its lowest orbit until a sufficient amount of energy is available, at which time the electron will accept the energy and jump to one of a series of *permissible* orbits. An electron cannot exist in the space between permissible orbits or energy levels. This indicates that the electron will not accept a photon of energy unless it contains enough energy to elevate the electron to one of the allowable energy levels. Heat energy and collisions with other

particles can also cause the electron to jump orbits.

Once the electron has been elevated to any energy level higher than the lowest possible energy level, the atom is said to be in an *excited* state. The electron will not remain in this excited condition for more than a fraction of a second before it will radiate the excess energy and return to a lower energy orbit. To illustrate this principle, assume that a normal electron has just received a photon of energy sufficient to raise it from the first to the third energy level. In a short period of time, the electron may jump back to the first level emitting a new photon identical to the one it received.

A second alternative would be for the electron to return to the lower level in two jumps—from the third to the second, and then from the second to the first. In this case the electron would emit two photons, one for each jump. Each of these photons would have less energy than the original photon which excited the electron and would represent a longer wavelength of light.

This principle is used in the fluorescent light where ultraviolet light photons, which are not visible to the human eye, bombard a phosphor coating on the inside of a glass tube. The phosphor electrons, in returning to their normal orbits, emit photons of light that are of a visible wavelength (longer wavelength). By using the proper chemicals for the phosphor coating, any color of light may be obtained. The coloring of the screen of a television picture tube is an example.

In 1913, Niels Bohr gave the quantum theory new stimulus by proposing a quantized model for the hydrogen atom. In Bohr's model, the hydrogen atom is pictured as a small, positively charged nucleus, orbited by an electron. Bohr postulated that the rotational energy of the orbiting electron can have only certain discrete values. These values define a set of stable electron orbits; that is, while an electron is rotating in a stable orbit, it does not emit radiation. A region which separates allowed energy levels is called an *energy gap*. The electron can change energy only in a jump in which it either absorbs or emits a photon. The conservation of energy for an electron jump may be written $E_m - E_n = hf_{m,n}$ where E_m and E_n are two allowed energy states for the orbiting electron, and $f_{m,n}$ is the frequency of the emitted radiation. Figure 11-3 shows this for energy levels corresponding to m = 2 and n = 1. Using Bohr's model, the frequencies of the emitted radiation could be calculated. These calculated frequencies were found to agree almost perfectly with the observed characteristic spectrum of hydrogen.

In the years following the introduction of Bohr's model, the quantum theory grew in significance and scope. Some of the features of the Bohr atom provide a useful introduction to the important concepts necessary to the description of laser operation.

A concrete picture of the Bohr model in terms of energy is provided by figure 11-4, where energy is plotted on the vertical scale. The horizontal lines represent the allowed energy levels, and the vertical connecting lines represent examples of electron jumps, or transitions. The wavelength of the emitted radiation is given along the transition lines in Angstrom units (1 Angstrom = 10^{-10} meter). Energy level E denotes the lowest energy level, or *ground state*, for the atom. The other levels (E_2, E_3, etc) represent *excited states*. Such a representation is called an *energy-level diagram*. The usefulness of these diagrams in atomic physics can be appreciated when one realizes that the energy-level diagram for a given atom is peculiar to that type of atom. The energy-level diagram for

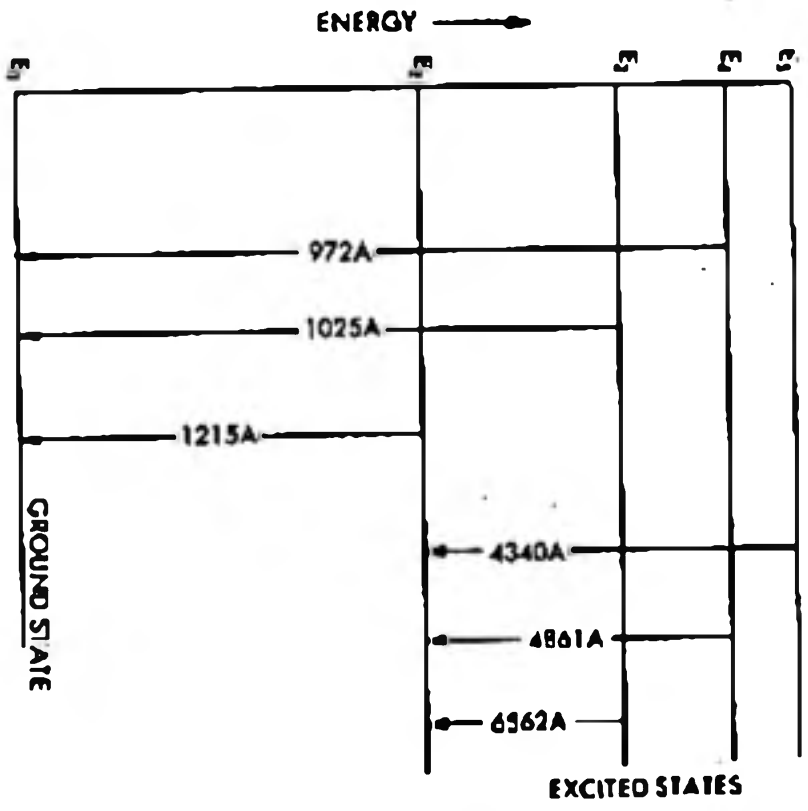


Figure 11-4. Energy Level Diagram for Bohr Model of Hydrogen Atom

an atom is to the atomic physicist what the schematic is to the electronics technician, because such a diagram provides important information concerning atomic behavior.

One feature of an atom's behavior not included in an energy diagram is the lifetime and population of levels. If the atom at some given time is in an excited state, its electrons are at any level except the ground state. The atom does not remain at this level if disturbed, but spontaneously jumps to some other level, and emits radiation. There is always a tendency, following radiation, for an atom to return to its lowest energy, or ground state. Consequently, it radiates and then spontaneously makes transitions downward, radiating at each level, until the ground state is reached. In a gas with unexcited atoms, almost all atoms are in the ground state.

The situation is complicated by the fact that the average time an atom remains in a state, before decaying to some lower level, depends on what state it is in to begin with. To complete the picture, a set of numbers must be made available which represents the mean lifetime of the electron in each of its possible states before it decays in spontaneous steps to the ground state. In general, these energy state lifetimes are quite short, less than 10^{-8} second. However, levels exist for which the lifetime is considerably longer. These levels are known as *metastable states*. Remember that each of the spontaneous transitions must conserve energy, and, therefore, each is accompanied by the emission of a photon. Photons emitted by spontaneous transitions are called *spontaneous emissions*. Since one atom does not "know" what another atom is emitting, there is a lack of cooperation among atoms, and the resulting emission produced is not coherent.

COMPLEX ATOMS. Although hydrogen has the simplest of all atoms, the basic principles just developed apply equally well to the atoms of more complex elements. The manner in which the orbits are established in an atom containing more than one electron is somewhat complicated and is part of the science known as quantum mechanics. In an atom containing two or more electrons, the electrons interact with each other and the exact path of any one electron is very difficult to predict. Each electron will lie in a specific energy band and the orbits may be considered as an average of the electron's position.

SHELLS AND SUBSHELLS. The difference between the atoms, insofar as their chemical activity and stability is concerned, depends upon the number and position of the particles included within the atom. Atoms range from the simplest, the hydrogen atom containing one proton and one electron, to the very complex atomic structures such as silver, containing 47 protons and 47 electrons. In general, the electrons within the atom reside in groups of orbits called *shells*. These shells are elliptically shaped and are assumed to be located at fixed intervals as predicted by the Bohr concept. Thus, the shells are arranged in steps that correspond to fixed energy levels. The shells, and the number of electrons required to fill them, may be predicted using *Pauli's exclusion principle*. Simply stated, this principle specifies that each shell will contain a maximum of $2n^2$ electrons, where (n) corresponds to the shell number starting with the one closest to the nucleus. By this principle the second shell, for example, would contain $2(2)^2$ or 8 electrons when full. In addition to being numbered, the shells are also given letter designations as pictured in figure 11-5. Starting with the shell closest to the nucleus and progressing outward, the shells are labeled K, L, M, N, O, P, and Q, respectively. The shells are considered to be full or complete when they contain the following quantities of electrons: 2 in the K shell, 8 in the L shell, 18 in the M shell, and so on, in accordance with the exclusion principle. Each of these shells is a major shell and can be divided into subshells

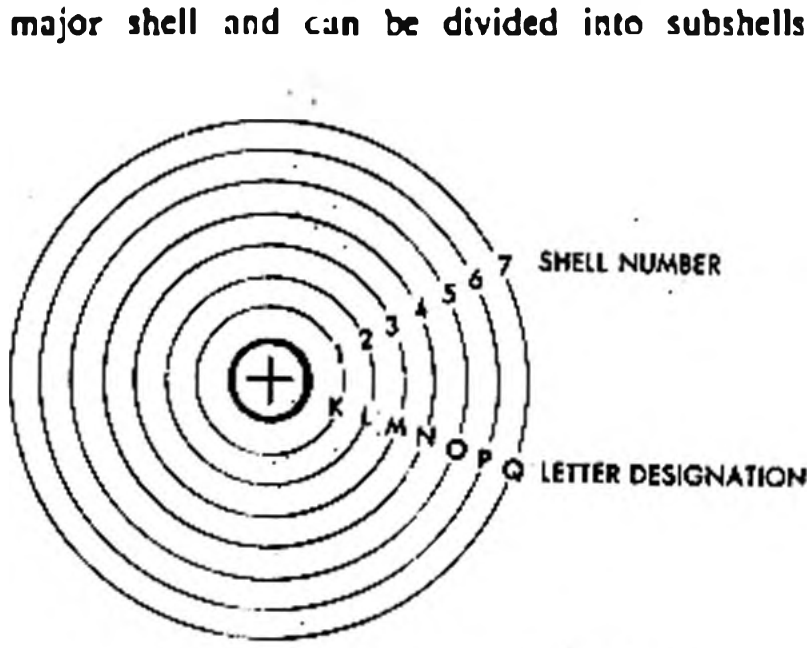


Figure 11-5. Shell Designations

of which there are four, labeled 1, 2, 3, and 4, respectively. Like the major shells, the subshells are also limited as to the number of electrons which they can contain. The "1" subshell is complete when it contains 2 electrons, the "2" subshell when it contains 6, the "3" subshell when it contains 10, and the last or "4" subshell when it contains 14 electrons.

Inasmuch as the K shell can contain no more than two electrons, it must have only one subshell. The M shell is composed of three subshells: 1, 2, and 3. If the electrons in the 1, 2, and 3 subshells are added, their total is 18, the exact number required to fill the M shell. This relationship exists between the shells and subshells up to and including the N shell. Beyond the N shell, the actual number of electrons required to fill a shell has not been experimentally determined.

To drive electrons out of the shells of an atom requires the internal energy of the atom to be raised. This raised energy may be obtained through bombardment by photons or by subjecting the atom to electronic fields. The amount of energy required to free electrons from an individual atom is called the *ionization potential*.

The ionization potential necessary to free an electron from an inner shell is much greater than that required to free an electron from an outer shell. Also, more energy is required to remove an electron from a complete shell than from an unfilled shell.

The dicussion of the atom, atomic structures, and energy levels is included because of the important role it has played in laser development. Laser operation requires (1) an active material that produces stimulated emission of radiation, (2) an excitation source that will pump power into the active material, and (3) a resonant structure.

Stimulated Emission

Ordinary light sources do not rely solely on spontaneous emissions for their output. Spontaneous transitions are not the only means by which a particular atom returns to its ground state. Consider the hypothetical case of a coherent light (photon) beam traversing atoms of a gas. When the frequency of such a source coincides with one of the frequencies of spontaneous emission, atoms are induced to make transitions between two particular energy levels whose difference is ΔE, and thus satisfy the relationship $\Delta E = hf$. An important result is that transitions from the upper to the lower energy states are induced, in addition to those in the opposite direction. The probability that the induced transition will be in one direction rather than the other depends only on which level the majority of atoms are located. When the transition is from a lower to an upper energy level (accompanied by the loss of one photon from the beam), it is called *absorption*. When the transition is from upper to lower energy levels a photon is emitted, and it is called *induced* or *stimulated emission*. Note that in this case the emission was induced by the presence of the binding energy of the atom.

If equal numbers of atoms are in two levels, the beam intensity remains constant in traversing the gas. With more atoms in the upper than in the lower state, the beam "sees" a net gain of photons, or is amplified by stimulated emissions in traversing the gas. Since photon emissions are induced by interaction with the photons in the beam, which are assumed to be coherent, they are all in phase, and the amplified beam is coherent, as shown in figure 11-6. In this way, a single spontaneous emission is amplified into an intense coherent beam. The remaining problem is to obtain the higher concentration of atoms in the upper energy level necessary for amplification. This is known as *population inversion*.

Stimulated emission is the basis of laser operation. When the photon is absorbed by the atom, the energy of the photon is converted to internal energy of the atom. The atom's electrons are raised to a "high energy" or "excited" state and later radiate this energy spontaneously, emitting a photon and returning to the ground state. Phosphorus illustrates this principle; it glows when hit by light and continues to glow for a period of time after the light is removed. Using ultraviolet light makes it glow as long as the ultraviolet light is present. However, as the energy source is increased, the glowing state ceases. The light energy is released in phase (polarized), but it is not lasting action even though this light is monochromatic (one frequency). This phenomenon is the beginning of the laser principle.

Coherence

You may recall from your study of electronics

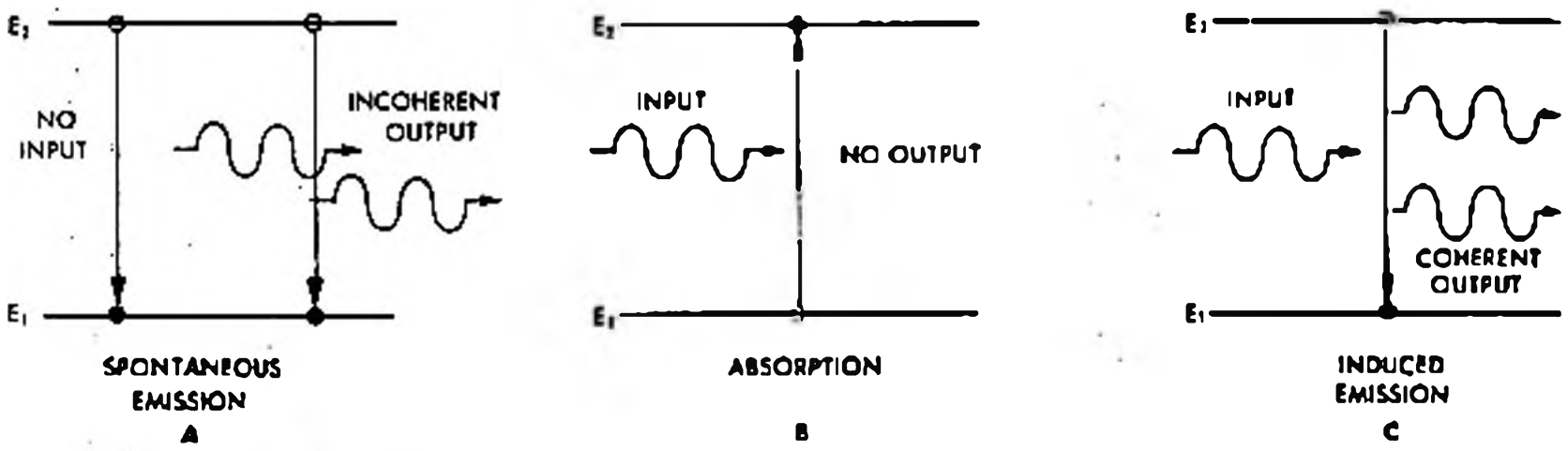


Figure 11-6. Atomic Transitions Producing Emission and Absorption

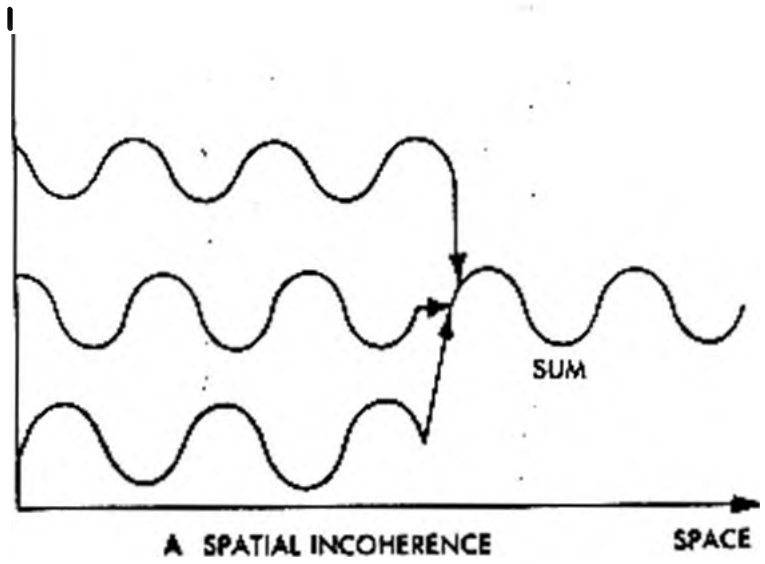
that an effective transmitter is a generator of electromagnetic waves which radiates a significant amount of coherent power in a narrow band of frequencies, including the one desired. The need for coherence in an efficient generator is sometimes overlooked. An example is the design of the antenna for a receiver. The assumption is sometimes made that increasing the antenna area increases the signal-to-noise ratio at the receiver detector. This is true only if the phase of the incoming signal is constant or varies in a predictable manner over all points of the antenna. If the phase of the signal changes in a random manner from point to point over the antenna, then the detector can only sum the absolute value of the incoming power. The sum of the phase angles could approximate zero and all modulation would be lost. This correlation of phase in the signal is precisely what is meant by *coherence*. An instantaneous correlation of phase from point to point in space is called *spatial coherence*, and a consistent correlation in phase at two neighboring points over a length of time is called *temporal coherence*. Refer to figure 11-7.

The generators of radio, television, and radar signals exhibit both temporal and spatial coherence in the emitted signal. Until the advent of the laser no sources of signal power operating in what is known as the *optical spectrum* were coherent in any but a statistical sense.

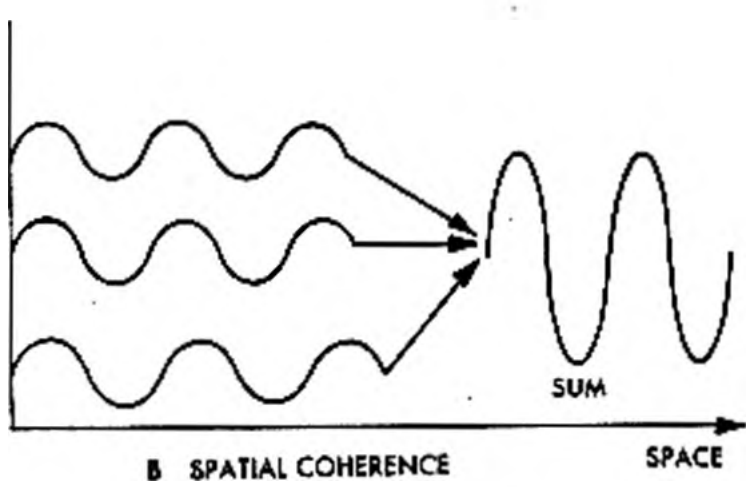
The reason for the lack of coherence in optical sources other than the laser is related to the lack of correlation among the motions of the electrons, since each electron behaves as a tiny oscillator which emits light. For familiar light sources, such as tungsten filaments and gas-discharge tubes, electrical energy is supplied to create conditions favorable to the emission of light; for example, by heating the filament or exciting the atoms in the gas. The actual emission process is not controlled. Each oscillator radiates independently of its neighbors. The emitted light, which is the sum of all the individual radiations, lacks both spatial and temporal coherence. The idea was first proposed in 1958 to maintain a constant phase relation over the oscillators by stimulating their emission with a wave of the frequency to be radiated. Two years later, using stimulated emission, pulses of coherent optical radiation from a single ruby crystal were achieved.

Resonant Cavities

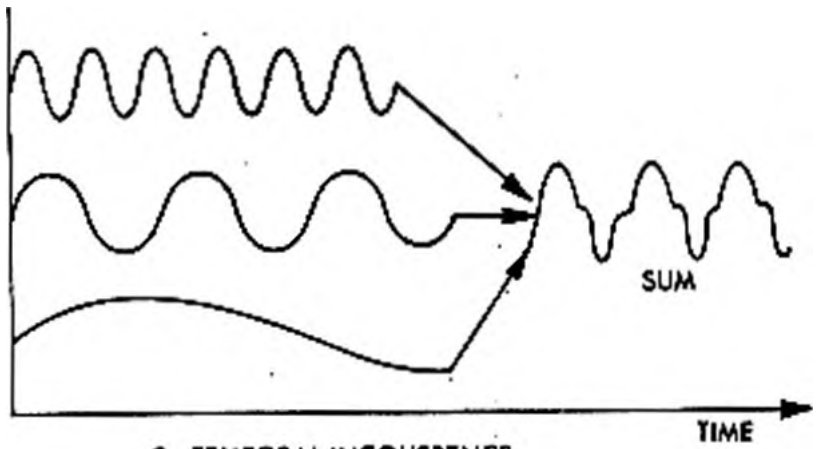
You may again draw on your knowledge of electronics to understand the use of microwave cavities as determining the frequency of oscillations, phase relationships, etc. Resonating structures (cavities) are also applied to lasers. Most lasers have parallel reflecting surfaces at the ends of the cavity. There must be a means of preventing the stream of electrons from hitting the cavity wall. To change the energy level of the electron, a method of stimulation or pumping is used. The term *pumping* is defined as a process whereby matter is raised from a lower to a higher energy state. Another major factor in laser design is a method to dissipate excess heat generated by the pumping action. The higher the temperature, the more activity within the atomic system. However, too much activity may cause the system to fail. This is especially true of ruby lasers. Tests have shown that the ruby laser does not function when it is overheated. The cavity must have some heat but the excess must be removed. To remove the heat, some lasers are bathed in a cryogenic



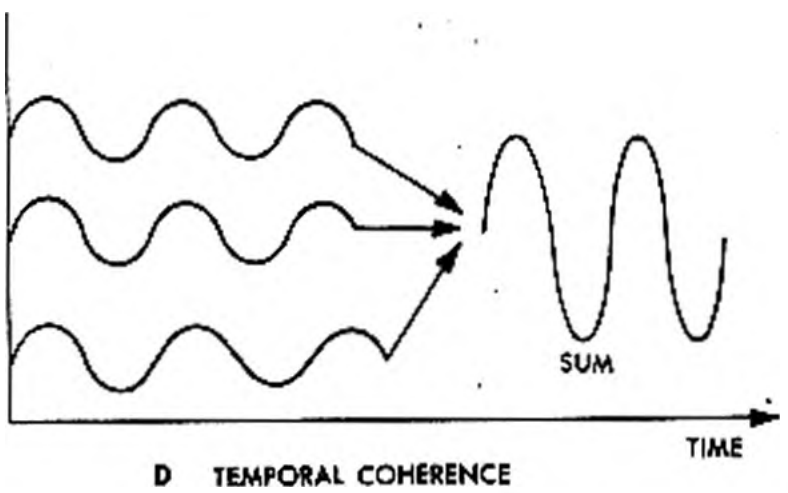
A SPATIAL INCOHERENCE



B SPATIAL COHERENCE



C TEMPORAL INCOHERENCE



D TEMPORAL COHERENCE

Figure 11-7. Spatial and Temporal Coherence Effects and Waveforms

medium, while others are cooled by heat exchangers.

At optical wavelengths, a single wavelength resonator would have dimensions inconveniently small. To overcome this problem, a special resonator can be built with dimensions thousands of times greater than the single wavelength. In the laser cavity, two mirrors are installed facing each other. The position of the mirrors is critical; they must be parallel, and a specific number of wavelengths apart. One of the mirrors is completely silvered while the other is only partially silvered and remains somewhat transparent. As the energy from the medium is reflected back into itself it increases. With repeated pulses of light applied to the medium, the energy continues to increase until "breakthrough" is accomplished. This breakthrough of light energy is coherent, nearly monochromatic, and polarized. These terms—coherent, monochromatic, and polarized—mean that there is a fixed phase relationship between the portion of the wave emitted at one instant and the wave emitted after a fixed time interval.

Obtaining population inversion involves the addition of energy to the gas. The process of populating an upper energy level at the expense of a lower one is called *pumping*. One possible method of pumping between two levels whose difference in energy is ΔE is by supplying electromagnetic energy of the frequency satisfying the relation $hf = \Delta E$, and thus raising the energy by absorption. This method is efficient only at the start, when the population of the lower level exceeds that of the upper. As equal population is obtained, the number of upward transitions becomes equal to the number of downward transitions, regardless of the pumping energy. To obtain inversion, a more sophisticated technique is necessary, involving at least one intermediate energy level in which the pumped atoms may be stored.

THE RUBY LASER

In its simplest form, the ruby laser is shown in figure 11-8. Figure 11-8A is the conventional drawing while figure 11-8B is more familiar to you as a schematic representation. The ruby (aluminum oxide) plus a few chromium atoms (.05%) sparsely located throughout the aluminum oxide is most commonly used. The heart of the

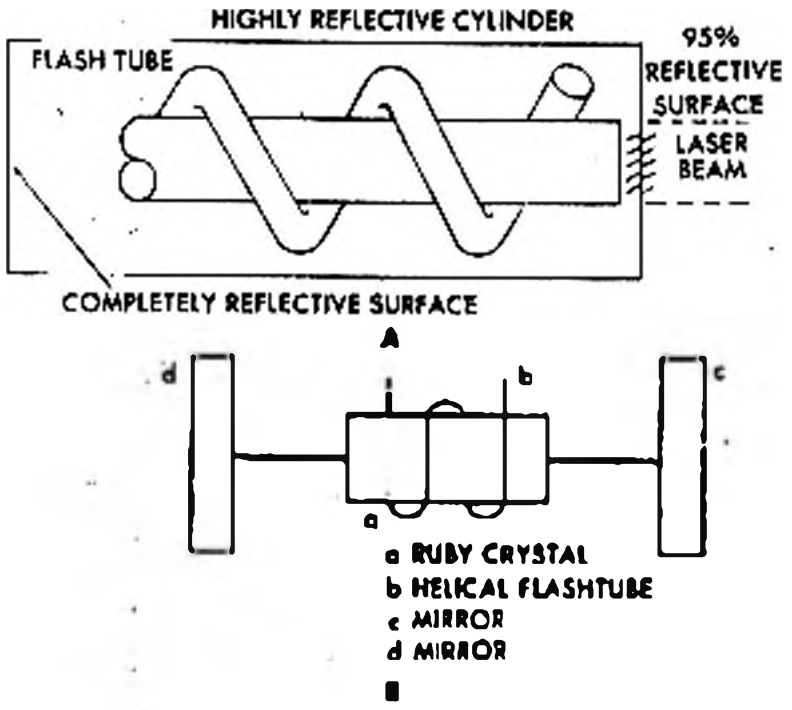


Figure 11-8. Simple Ruby Laser

device is the cylindrical ruby crystal (a), around which is the helical flash tube (b). Mirrors at each end (c and d) reflect the light back and forth through the crystal, and the laser beam emerges through one of the mirrors which is only partly silvered.

The helical flash tube is a device for producing very intense light. The ruby absorbs energy from the flash tube and in a very short time (thousandths of a second) re-emits it, some energy in the form of light and the rest as heat.

A small part of the light energy produced by the ruby consists of the red beams traveling parallel to its axis. This energy is reflected back and forth by the mirrors so that it passes through the crystal many times. As it passes, it is amplified; that is, it picks up more energy from the ruby. This energy travels in the form of red light along the beam as the intensity continues to build.

Light waves consist of all electromagnetic waves between 4000 and 7000 Angstrom units in wavelength. An Angstrom unit, denoted by the symbol Å, is 10^{-8} cm. The shorter 4000 Å waves produce the violet color, the 7000 Å waves are those of red light, and the wavelengths in between correspond to other colors of the visual spectrum. The ruby crystal presented here operates in the 6943 Angstrom range, which accounts for the characteristic red fluorescence.

Since the essential property of a gas is that the constituents do not interact with one another, a lightly doped (much less than one percent) crystal is essentially a gas of dopant atoms in a rather special container. Their energy levels are modified by the presence of the host material. A simplified energy level diagram of the chromium ion in a ruby crystal is shown in figure 11-9. The intermediate level (3) is metastable with a lifetime of approximately 10^{-3} second.

By examination of figure 11-9, you can see that a pumping light of frequency f_{13} causes a transition between levels 1 and 2. The atoms in the excited (2) state may return to the ground state spontaneously, either directly or by first stopping at the metastable state. Because the lifetime of the metastable state is 100,000 times longer than that of state 2, the atoms which fall there may be considered almost stationary. The rate at which atoms find themselves in state 3 is proportional to the rate at which they arrive in state 2, which, in turn, is proportional to the pumping power, and independent of the population. If sufficient pumping power is supplied, the population of the 3 state grows at the expense of the 1 state and population inversion is obtained.

As long as the population is inverted, the ruby can be an amplifier for radiation of frequency f_{31}, and, as with any amplifier, adding a positive feedback loop can cause sustained oscillation. In this case, the positive feedback is the return of some of the output light (f_{31} radiation) into the

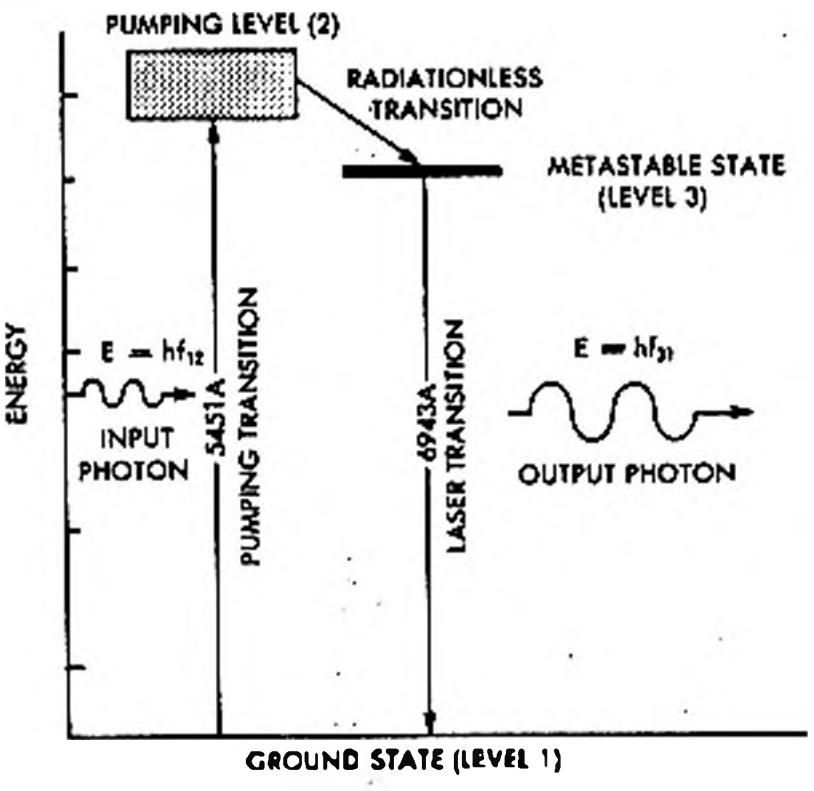


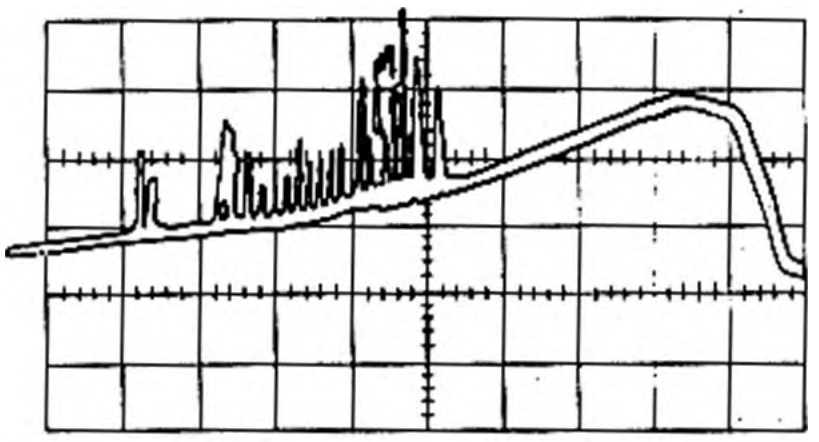
Figure 11-9. Three Levels of Chromium Ion in Ruby Rod (associated with pumping and laser action)

ruby. This can easily be accomplished with mirrors. By making use of the geometry of the mirrors so the feedback is directional, a resonant cavity is formed. The amplifier radiation, referred to as photon amplification, builds up in a standing-wave pattern that is familiar to you from your study of microwaves.

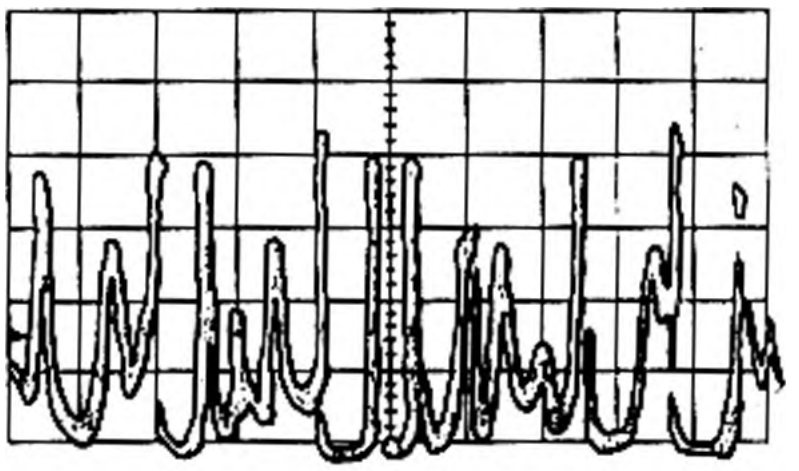
The resonant cavity formed in the crystal itself is made possible by carefully grinding and polishing the ruby, and silvering its ends. Because of the shortness of optical wavelengths, an essential difference exists between the crystal cavity and the more familiar microwave cavity. Calculating the wavelength from the energy-level diagram, the wavelength corresponding to f_{31} radiation is found to be 6943 Angstroms in a vacuum. A ruby ground to form a cavity 7.3-cm long (a typical size) has 100,000 nodes in the standing wave, and is resonant for every frequency that satisfies the standing wave condition $(n/2)\ \lambda = f_r$, where λ is the wavelength and n is an integer. For example, taking n as 10^5, the difference between resonant wavelengths, $\Delta\lambda$, is given by $(\Delta\lambda/\Delta) = (\Delta f/f) = 1/n = 10^{-5}$. The cavity is resonant for a large number of frequencies immediately around f_{31}, instead of being resonant for only one particular frequency, as in the microwave case.

The pumping power required to obtain population inversion for a reasonable-size crystal is considerable. This population inversion may be accomplished only in brief bursts of light from a flash lamp. First, the operating time of the ruby laser is limited to a couple of milliseconds. Second, while the ruby is operating as a laser, the metastable stage is being depopulated by stimulated emission, and very quickly (in 10^{-6} second) outruns the pump. This causes laser action to stop until the pump can again create a population inversion. The output of a ruby laser is composed of a series of irregularly spaced spikes about 10^{-6} second in duration, in an envelope defined by the pump lamp duration. Figures 11-10A and B show an oscilloscope trace of the output of a photo detector receiving light energy from a ruby laser.

Finally, consider the over-all efficiency of the ruby laser by forming a percentage from the ratio of total output of laser light energy to the electrical energy supplied to the pump. This efficiency is less than one percent with most of the lost energy dissipated in heating the ruby crystal. This makes cooling the crystal an important practical consideration.



A
OUTPUT OF RUBY LASER
TIME SCALE: 0.1 MILLISECOND/DIVISION.
ENVELOPE IS THE PUMP-LAMP FLASH.



B
EXPANDED PORTION OF TRACE
TIME SCALE: 1.0μ SEC/DIVISION

Figure 11-10. Output of a Ruby Laser

Additional information on the ruby laser is given later in this chapter. At this time, however, compare the theory of the ruby laser thus far to the theory of the gas laser.

THE GAS LASER

Gas lasers offer more avenues for exploration than solid-state lasers because the atoms are more accessible for stimulation by a variety of means.

To understand the operation of the helium-neon gas laser, a different means of obtaining population inversion, as well as of pumping, must be considered. The modification of the energy-level scheme is shown in figure 11-11. Note that a fourth or terminal level has been added above the ground state. The population inversion is now ob-

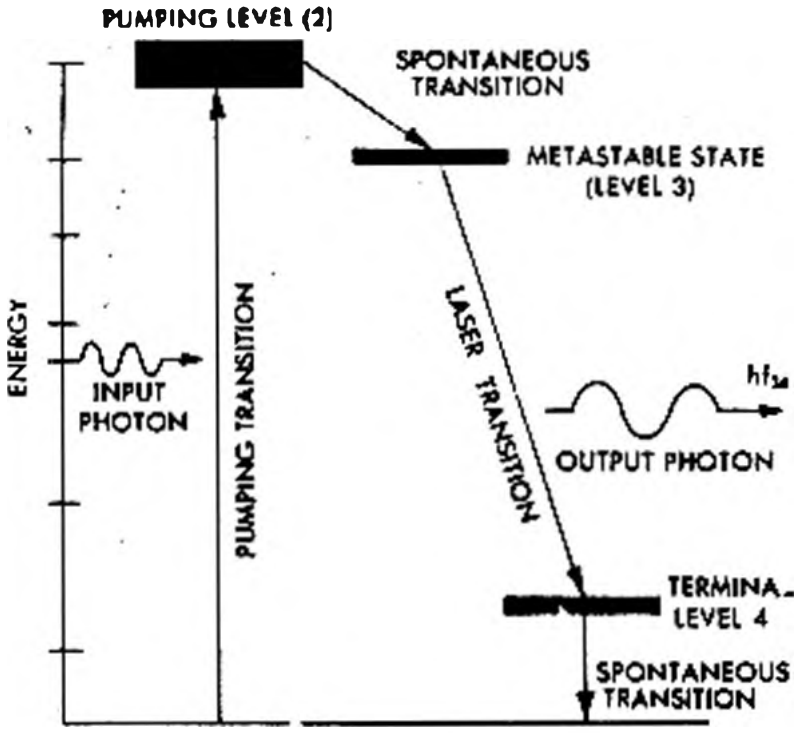


Figure 11-11. Energy Level Diagram of Four-Level Gas Laser

tained between the 3 and 4 levels. The advantage of the four-level scheme is that the initial population of the terminal level is negligible as compared with the ground state; therefore, inversion is more easily obtained. That is, fewer atoms in the 3 state are necessary for its population to exceed that of the 4 state than would be required at the ground level. This reduces the pumping power required, and opens up the possibility of pumping by a different method, called electron-collision pumping.

In quantum theory, an analogy is found for an electron beam traversing atoms of gas and that of a photon beam. Only when the kinetic energy of the electrons coincides with the differences in energy between any two levels are atoms induced by collisions with the electrons to make transitions between these levels. As in the electro-magnetic case, the most probable direction of the transition depends only on the relative populations of the states. Rather than the energy from a light beam, the energy of accelerated electrons can be used to pump atoms of gas in a discharge tube. These electrons may be the result of a glow discharge when RF energy is applied to the discharge tube. The advantage of this method is that energy can be maintained constant over extended periods of time to obtain a continuous laser output.

The first successful operation of a laser by collision pumping required the presence of two gases, such as helium and neon, in the discharge tube to realize the proper energy scheme. A slight digression is necessary to consider the transfer of energy between atoms of different gases. A sort of resonance phenomenon is encountered where energy transfer proceeds only when an energy gap is shared. That is, an atom of one type of gas (A) in a given energy level (2A) may transfer its energy to an atom of another type (B) in a stage (1B) via a collision if, and only if, there exists energy levels 1A and 2B of such value that $E_{2A} - E_{1A} = E_{1B} - E_{2B}$ (see figure 11-12). As before, the probability of the transition direction is determined by the population of the levels.

In the gas laser, the higher and lower energy levels are in different gases, and their population can be modified by changing the relative concentration of the different gases in the discharge tube. This additional control was fundamental in achieving the first observed laser action in a gaseous mixture of helium and neon.

Figure 11-13 shows the pertinent sections of the energy-level diagrams for helium and neon with the transitions indicated. Note that when the 3 level is well populated by pumping energy, there is amplification for two different frequencies, f_{34} and $f_{3'4'}$. Note also that the alternate route 1, 2', 3', 4' should result in amplification at the frequency $f_{3'4'}$. Laser action has been observed at all of these frequencies in helium-neon mixtures. Selection of oscillation between these frequencies is

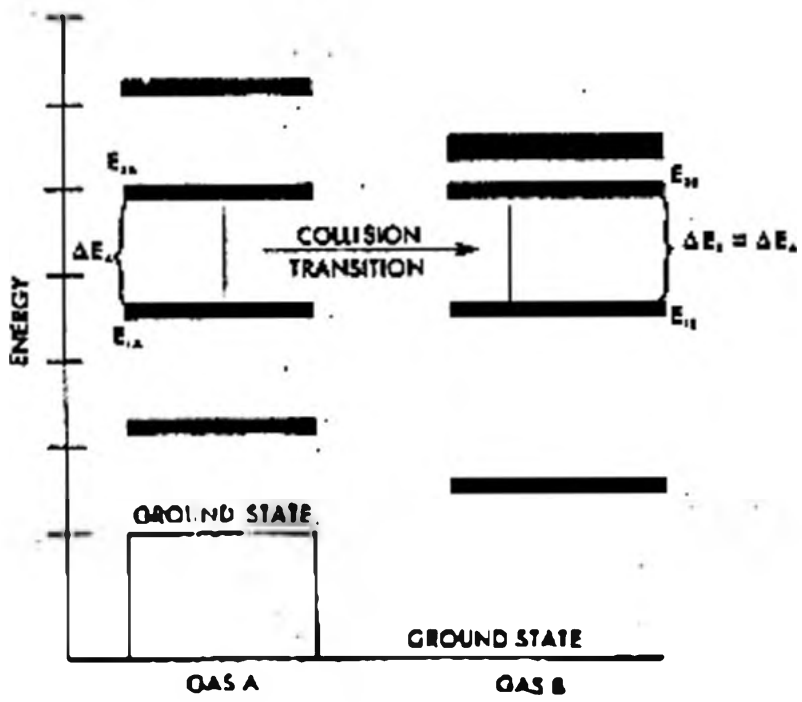


Figure 11-12. Resonant Transfer by Collision of Gas Atoms

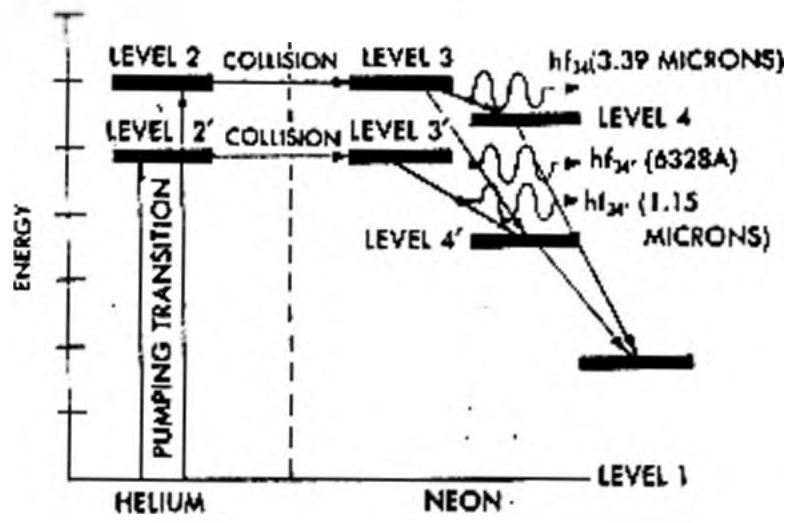


Figure 11-13. Energy Level Diagram of Helium and Neon during Laser Action

accomplished by using feedback mirrors with reflectivities at different frequencies.

The most distinctive difference between the helium-neon laser and the ruby laser is in their outputs. The ruby laser output is typified by irregular spiking, while the helium-neon laser is capable of a continuous wave of extremely narrow bandwidth. This distinction of emissions has determined the field of application of the particular laser. Equipments that require pulsed-type emissions of high instantaneous power for short durations have adopted the ruby laser. On the other hand, continuous communications that require sophisticated demodulation techniques, such as the practical use of light interferometry, find the gaseous type of lasers well suited.

The above discussion by no means covers all known lasers. In fact, the list of laser materials grows almost weekly. Laser action has been observed in other solids doped with small quantities of rare-earth ions, in almost all of the noble gases, and even in some special liquids. In addition, the observation of coherent light generated by injection currents in semiconductor diodes, such as gallium arsenide and gallium phosphide, has added another important class of laser, termed injection laser, to this fast-growing field.

LASER ENERGY AND EMISSION CHARACTERISTICS

Laser operation requires an active material that produces emission of radiation, an excitation source that pumps power into the active material, and a resonant structure. Both solid-state and gas lasers have these characteristics. The functioning of the solid-state lasers developed up to now is essentially the same as that of the original ruby laser. Figure 11-14 shows a block diagram of a typical ruby laser. The active material is ruby, the excitation source is a xenon flash tube, and the resonant structure is formed by a ruby rod whose ends are reflecting mirrors. One end of the rod has a heavy silver coat that forms an opaque mirror; the other end has a light silver coat that forms a 92-percent reflecting mirror.

Although the complete physical and mathematical description of laser action is complicated, it is possible to get a simplified picture of laser action by relating this action to the energy-level diagram of the laser material. Figure 11-15A shows a simplified energy-level diagram for $Cr+++$ (chromium ion) doped pink-ruby crystal ($Cr_2O_3:Al_2O_3$), the material used in the ruby laser. The heights of black bars 1 and 2 and of the area shown in cross hatching 3 indicate the possible energies that a $Cr+++$ ion can have. Energy is in units of 10^4 wavelengths per cm; 1 wavelength per cm ($1cm^{-1}$) is equivalent to 1.9858×10^{-16} erg. In their normal condition (the ground state), the $Cr+++$ ions have zero energy. This condition is indicated by level 1.

If light photons having a wavelength of 5600 Angstroms irradiate the ruby crystal, they will

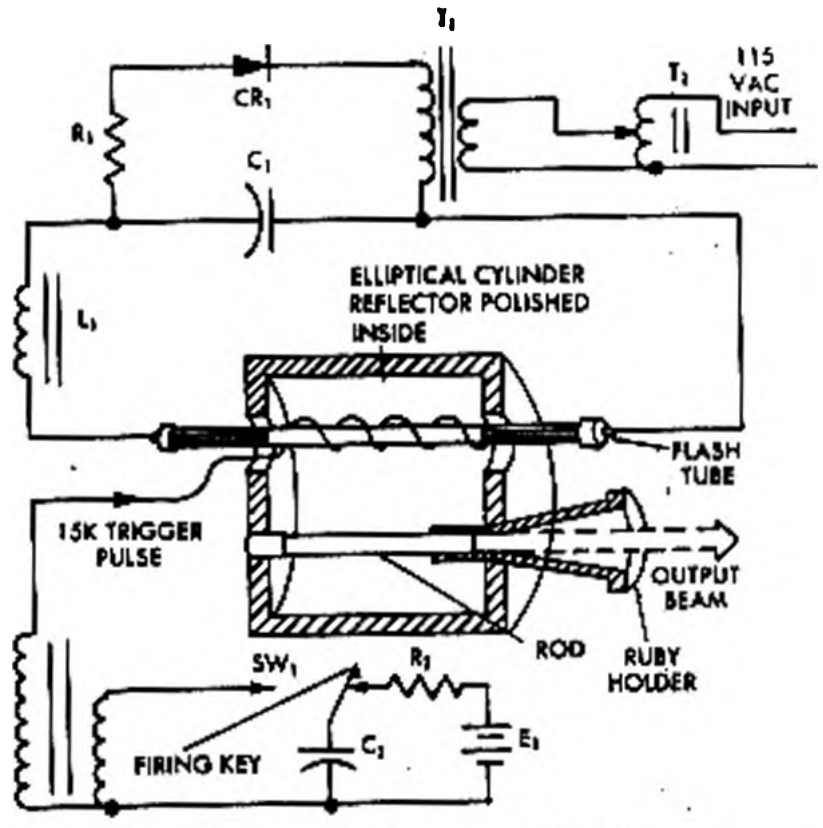


Figure 11-14. Typical Ruby Laser Block Diagram

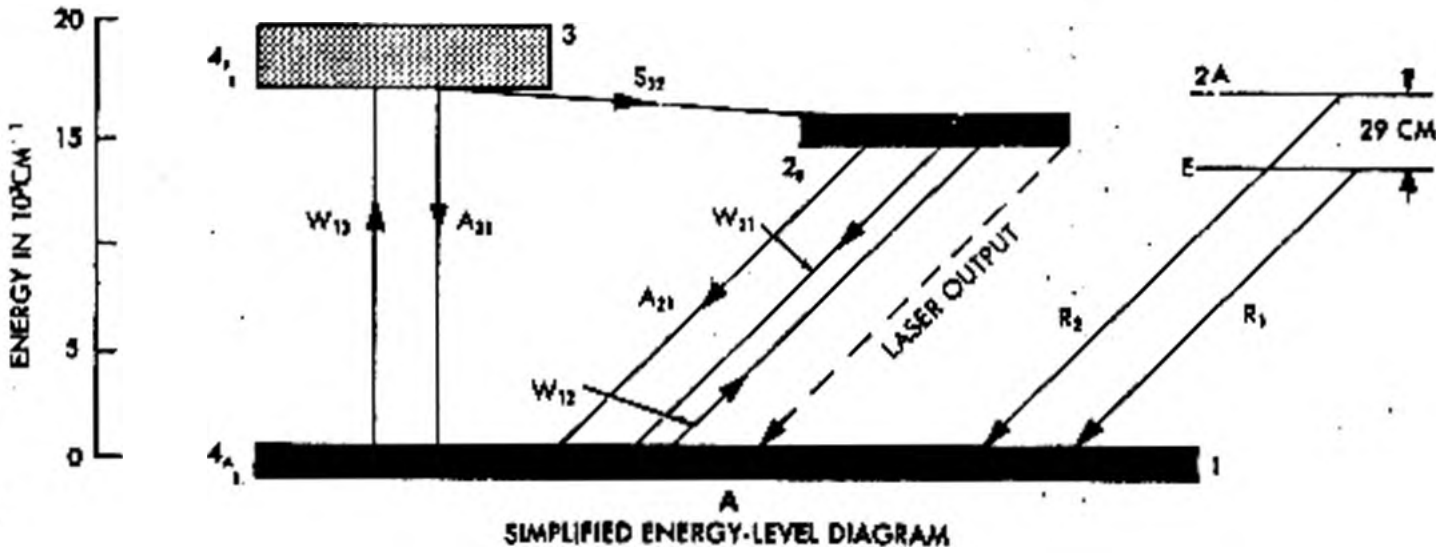


Figure 11-15. Energy Level Diagram for Ruby Laser

raise the energies of some $Cr+++$ ions to various energy levels lying in the absorption band of energies indicated by 3. Flash tubes supply this irradiating light, along with light composed of many other wavelengths. The left-hand arrow, W_{13}, going from level 1 to band 3 indicates the increase in energy acquired by a $Cr+++$ ion when it absorbs a 5600-Angstrom light photon. The use of light excitation to raise the energy level of atoms to higher levels is called optical pump-

ing. After short but finite times elapse (relaxation times), some of the Cr+++ ions in band 3 drop back to level 1 (shown by A_{31}), and some drop to level 2 (shown by S_{32}). The rate at which Cr+++ ions drop to level 2 is greater than the rate at which they drop to level 1. The Cr+++ ions in energy level 2 hold their energy for a short time before they drop to level 1. The rate that ions go from level 2 to level 1 (A_{21}) is less than the rate at which Cr+++ ions go from level 1 to level 3. Optical pumping builds up the number of ions having level 2 energies to a greater number than the number of ions having level 1 energies. In other words, the Cr+++ populations of levels 1 and 2 are inverted from their normal relationship. This population inversion is essential for producing stimulated emission.

In dropping from level 2 to level 1, Cr+++ ions radiate light. Level 2 (2_E) is actually composed of two levels, levels E and 2A, which emit radiation lines R_1 and R_2, respectively. If conditions were not completely correct for achieving laser action (for example, if an insufficient amount of excitation were applied), R_1 radiation would be spontaneous radiation rather than stimulated radiation, and would include a much broader band of wavelengths than the stimulated emission (parts B and C of figure 11-15). In both cases, the center frequency of the R_1 and R_2 energy radiated when ions drop from level 2 to level 1 is calculated from $\nu = (E_2 - E_1)/h$, where ν is the frequency, E_2 is the energy at the center of level E (for R_1) or level 2A (for R_2), E_1 is a constant representing the first energy level, and h is Planck's constant. The resonating character of laser action enhances radiation at the central wavelength of R_1, and diminishes other radiation. Arrow A_{21} in figure 11-15A indicates the spontaneous radiation of R_1 and R_2 that will be emitted if laser operating conditions are not correct. The broken arrow indicates the laser output, which is composed of coherent laser radiation at R_1 and spontaneous (incoherent) radiation at both R_1 and R_2.

The simplified sketches shown in figure 11-16 illustrate sequences of laser action. At the instant that pumping light is applied, all Cr+++ ions are in the ground state; the unshaded circles indicate this state. Optical pumping raises some Cr+++ ions to level E (figure 11-15A). The black circles in figure 11-16B indicate ions that

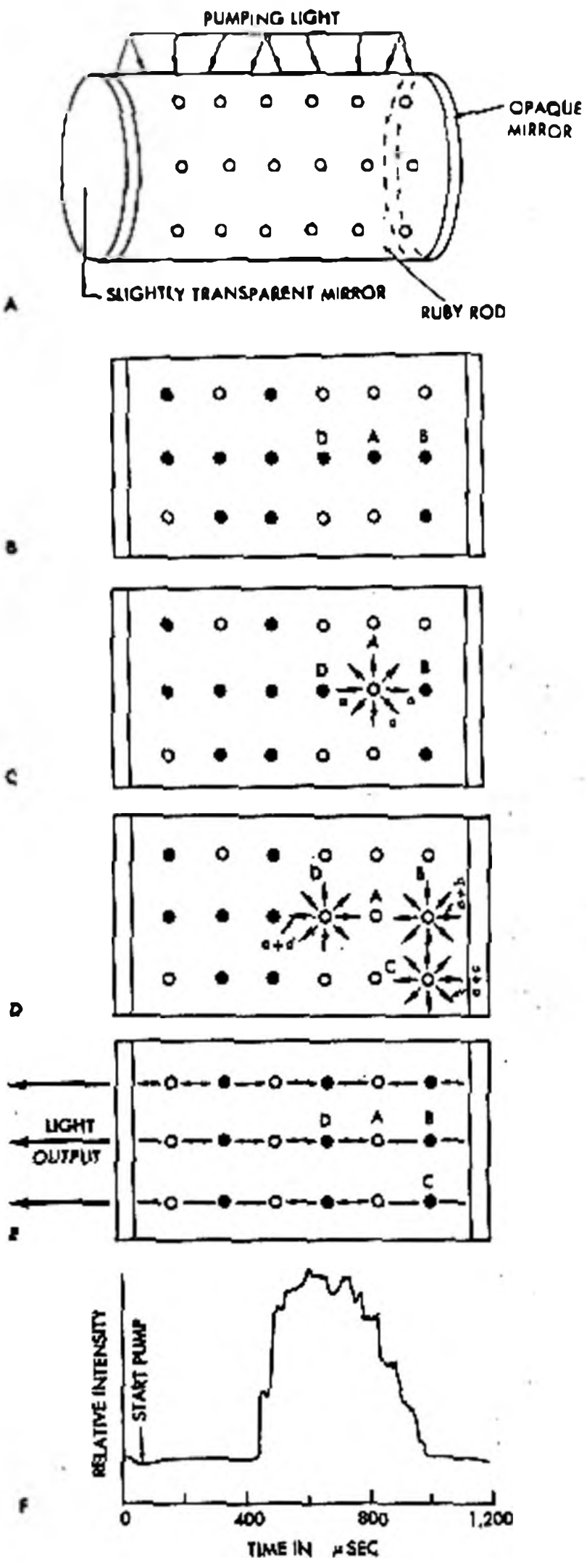


Figure 11-16. Emitted Radiation Sequence of Laser Beam

have been pumped up to level E. Some Cr+++ ions drop to level 1, radiating photons that have various wavelengths centered about the central wavelength of R_1. (Consider a photon as a bundle of light energy that has wave-like properties.) Figure 11-16C shows ion A, which is dropping to level 1, spontaneously emitting radiation. In these simplified sketches, ion A is the first ion (and the only one that is shown) to emit R_1 radiation spontaneously; that is, without being stimulated by R_1 radiation. The radiated photon tends to stimulate radiation of the same wavelength from other Cr+++ ions of level E that are in its path. This is indicated in figure 11-16 C and D. Assume that incident radiation a from ion A has the same wavelength of the strongest R_1 emission. Incident photon a is reinforced by stimulated photons b, c, and d in a precise phase relationship, as indicated by light rays a + b, a + c, and a + d in figure 11-16D. The opaque mirror reflects a + b back into the ruby cavity, but a + c passes through the side wall and is lost. The cavity enhances radiation propagated parallel to the axis of the ruby rod, and minimizes radiation going in other directions. Because of the amplification caused by photons of the same wavelength, rays composed of photons of the center wavelength of the R_1 line, which is the strongest (that is, most prevalent) wavelength, become predominant over other R_1-wavelength rays. This action makes the laser output highly, but not completely, monochromatic.

Since photons traversing paths other than in the direction of the long axis of the crystal escape from the sides of the rod (figure 11-16D), the laser output beam is highly directional. Photon streams reflect back and forth between the end mirrors and emerge from that end of the mirror which is slightly transparent. Figure 11-16E indicates the cohering effect of the light field in the cavity used to stimulate emission by the ions. To maintain the stimulated-emission process, losses such as those caused by the escape of radiation from the crystal and losses caused by mirror reflection must be overcome by photon amplification. The beam angle of the ruby laser is on the order of 0.01 radian.

Although vast numbers of ions within the ruby crystal are individual radiators of photons, laser action causes the ions to radiate their energy in step (coherently). The key action of the laser process that produces in-step radiation is the triggering of an ion by an incident photon, in order to emit a photon in phase. Since the individual radiators radiate nearly in step, and since these radiators produce radiation of approximately the same wavelength, the laser's output beam has space and time coherence.

Typical PRF rates have been on the order of several pulses per minute. Each output pulse contains large-amplitude spikes (figure 11-16F) which result from the inability of the pump to supply energy level 2 at a rate fast enough to keep up with the rate at which these ions drop from level 2 during the stimulated-emission process. CW operation of a ruby laser has been accomplished by Dr. V. Evtuhov of Hughes Research Laboratory.

One type of gas laser provides continuous operation at several infrared wavelengths — 11,180, 11,530, 11,600, 11,990, and 12,070 Angstroms, the strongest being the 11,530-Angstrom output line. The main difference between the way the gas laser functions and the way solid-state lasers function is the method by which they are excited and pumped into an inverted-population condition. Helium, at a pressure of 1 millimeter (mm) of mercury, and neon, at a pressure of 0.1 mm, are sent into the laser tube, whose ends contain flat, parallel, and semitransparent mirrors. A 30 MHz RF generator produces an electrical discharge through the gas mixture, thus raising the energy of ground-state (0-energy) between atoms to the 2^3S energy level (figure 11-17). These helium

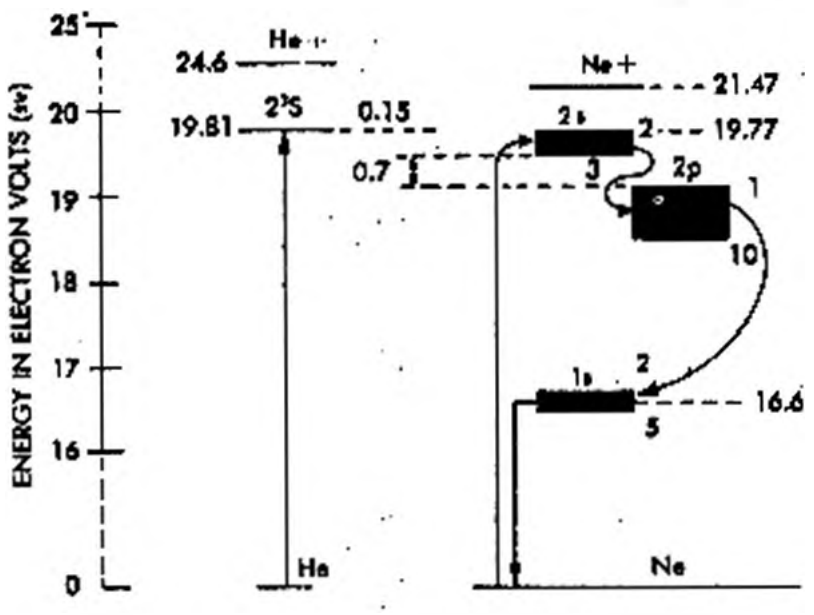


Figure 11-17. Energy Level Diagram of Helium and Neon Gases

atoms collide with neon atoms that are in the ground state and energy exchanges between helium and neon atoms take place. Because of the collision, the internal energy of the ground-state neon atom increases to the 2s level, which comprises four sublevels, and the internal energy of the 2^3S helium atom drops to zero. This collision process produces 2s-level neon atoms rather than 2p or 1s-level neon atoms, because the energy-exchange process is such that the least possible amount of change in the total internal energy of the colliding atoms occurs. Since the energy of the 2s levels is nearly equal to that of the 2^3S helium level, a 2^3S helium atom readily transfers its internal energy to a 2s neon atom, with little of the 2^3S helium atom's internal energy being converted to kinetic energy. Building up the 2s neon population produces a sufficiently large inverted population between levels 2s and 2p to achieve laser action.

Soon after optical pumping raises the population of the 2s level above the population of the 2p level, the light field that is built up inside the laser-tube cavity becomes coherent, and stimulates coherent emission of 2s level atoms throughout the tube.

Level 2s is composed of four sublevels and level 2p is composed of 10 sublevels. There are 30 possible transitions in which atoms in level 2s may drop to level 2p and radiate infrared, but only five of these transitions have thus far been stimulated emissions. The $2s_2$ to $2p_4$ energy change produces 11,530 Angstroms, the strongest stimulated emission.

LASER CLASSIFICATION

Lasers have largely been classified according to the nature of the material (see tables 11-1 and 11-2) placed between the two reflecting surfaces. If this material is a solid, such as a rod of ruby crystal, the device is referred to as a solid-state laser. If the material is a gas, such as a mixture of helium and neon enclosed in a quartz tube, the device is termed a gas laser. If the action takes place at the junction of two semiconductor materials, such as a gallium-arsenide (GaAs) diode, the device is termed a semiconductor diode laser. If the material is a liquid or a plastic, such as doped chelates in alcohol, or vinylic resin, the device is designated a liquid or plastic laser, respectively. These latter devices are sometimes referred to as Raman lasers when they are used to study or apply the light-scattering phenomenon known as Raman spectra. These are the extra spectral lines appearing near the prominent lines of the spectrum obtained when a strong light passes through a transparent liquid or gas.

Table 11-1. Arbitrary Classification of Laser Types

OPTICALLY PUMPED SOLID DIELECTRIC	Pulsed—low energy; high energy; high repetitive rate—high average power Continuous Crystalline host (sometimes called solid state) Glass host Rare-earth host
GAS LASER	Pulsed Continuous—low power; medium power Arrays Ionized gas—pulsed Continuous—low power; high power R F excitation (pump) DC excitation (pump) Optical excitation (rare)
DIODE LASER	Pulsed—low energy Continuous—low power; high power Array Room temperature Large volume diode laser Electron beam pumped semiconductor laser (not diode)
STORAGE LASER	Giant pulse Q-switched, etc. Passive Q switch (saturation absorption) Storage diode laser
SPECIAL LASERS	Frequency doubled Frequency tripled Raman laser Liquid laser Plastic laser
AMPLIFYING LASER (special cases of basic types)	Oscillator—amplifier for pulsed power Traveling wave laser (nonresonant cavity) Dump amplifier—saturation amplifier Ring laser(s)

Table 11-2. Laser Materials

HOST	ION	WAVELENGTH (microns)	TEMP. (degrees K)
INJECTION LASER DIODES			
GaAs		0.8400	77
Ga $(As_{1-x}P_x)$		0.84—0.71	77
InP		0.908	77
InAs		3.1	4.77
SiC		0.9	300
ORGANICS			
Benzophenone—naphthalene		0.47	77
Eu^3+—TTA chelate		0.6130	77
GASEOUS LASERS			
He—Ne		0.6328—3.2913	24 or more lines
He		2.0603	
Ne		1.1523—17.88	4 lines
A		1.6180—12.14	5 lines
Kr		1.6900—7.06	9 lines
Xe		2.0261—12.913	2 lines
$A—O_2$		0.8446	
$Ne—O_2$		0.8446	
Cs		7.1821	

Figure 11-18 illustrates some relative relationships between various laser outputs and the visible electromagnetic spectrum.

A summary of important properties of typical lasers is presented in table 11-3. The lasers are essentially listed by types in most cases with a variation of laser material.

Laser materials and construction are under intensive and continual development in scores, perhaps hundreds, of laboratories. Table 11-4 lists materials and data that apply to typical lasers that have worked. The only working gas laser listed is the helium-neon laser. All of the other lasers that are listed use solid-state laser materials.

LASER CONSTRUCTION

The resonating structures (cavities) of most lasers have parallel reflecting surfaces at the ends of the cavities, as shown in figure 11-19A. The confocal resonator shown in figure 11-19B is formed by two spherical reflectors separated by their equal radii of curvature. A confocal resonator has certain advantages over a parallel-plane cavity. It has lower diffraction losses, and requires less pumping power. Optical alignment of parallel

Table 11-3. Summary Data of Important Properties of Typical Lasers

LASER MATERIAL	TYPE	MEANS OF EXCITATION	EMISSION WAVELENGTH	OPERATION MODE	EFFICIENCY OR GAIN	TYPICAL OUTPUT POWER
Ruby	Solid	Optical pumping	6943 A	Pulsed	.2%	1 to 10 megawatts
Helium—neon	Gas	Electrical discharge	6328 A 1.15 μ 3.39 μ	cw	.05 .1 %/m 5	10 to 100 mw
Helium—xenon	Gas	Electrical discharge	3.5 μ 2.026 μ	cw	1%/m 15%/m	100 mw
Xenon	Gas	Electrical discharge	3.5 μ	cw	5%/m	10 mw
Cesium—helium	Gas	Optical pumping	7.18 μ	cw	1%/m	1 to 10 mw
Argon		Electrical discharge	4880 A	cw	15%/m	1 to 10 watts
Neodymium in glass	Solid	Optical pumping	1.06 μ	cw	.05%	1 watt
Gallium arsenide	Diode	Injection current	8450 A	cw, Pulsed	10 to 30%	.1 to 1 watt

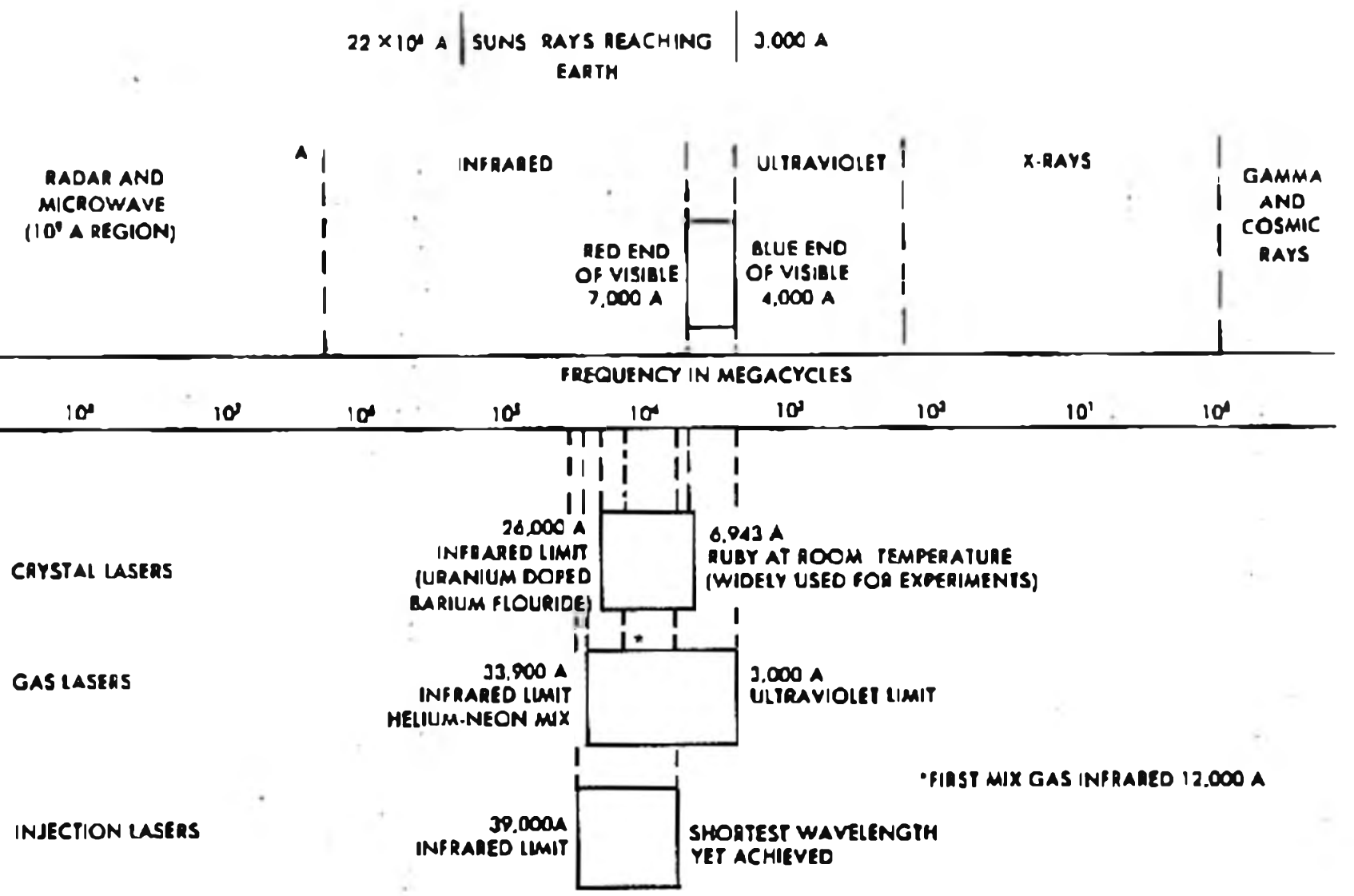


Figure 11-18. Some Relative Relationships between Various Laser Outputs and Visible Electromagnetic Spectrum

reflectors is critical, whereas optical alignment of spherical reflectors is not. Figure 11-19C shows a cavity geometry that is designed to attain total internal reflection.

Silver mirrors and mirrors with dielectric coatings have been used as end reflectors in both solid-state and gas lasers. A dielectric coated mirror contains a number of dielectric layers, having different optical matching characteristics. Silver has a lower reflectance than a dielectric coating, and causes relatively higher losses, particularly when used in ruby lasers which have high peak powers. Silver coatings deteriorate with time and use. After several hundred output pulses, the exact number depending on operating conditions, silver mirrors have to be replaced because their deterioration begins to reduce the output significantly. Dielectric coatings do not suffer such deterioration.

Heat affects solid-state lasers more than gas lasers. Tests and observations of extended periods of operation have shown that a ruby laser does not function when it is overheated. Internal heating of the laser material is one of the factors that thus far have prevented CW laser operation of solid-state materials, in spite of the use of such external coolants as liquid nitrogen. Since heat transfer requires time, heat may be trapped within

Table 11-4. Typical Laser Materials and Output Data

LASER MATERIAL (DOPING MATERIAL IS SHOWN IN PARENTHESES)	OUTPUT SPECTRAL LINES (angstroms) (OUTPUT WAVELENGTH VARIES WITH TEMPERATURE)
He—Ne	11,180, 11,530, 11,600, 11,990, 12,070
Ruby (0.05% Cr ion)	6,943
Ruby (0.5% Cr ion)	7,009, 6,934, 7,041
CaF_2 (0.05% U ion)	25,000
CaF_2 (0.1% Sm)	7,082
BaF_2 (U ion)	27,000
$CaWO_4$ (Neodymium ion)	10,600

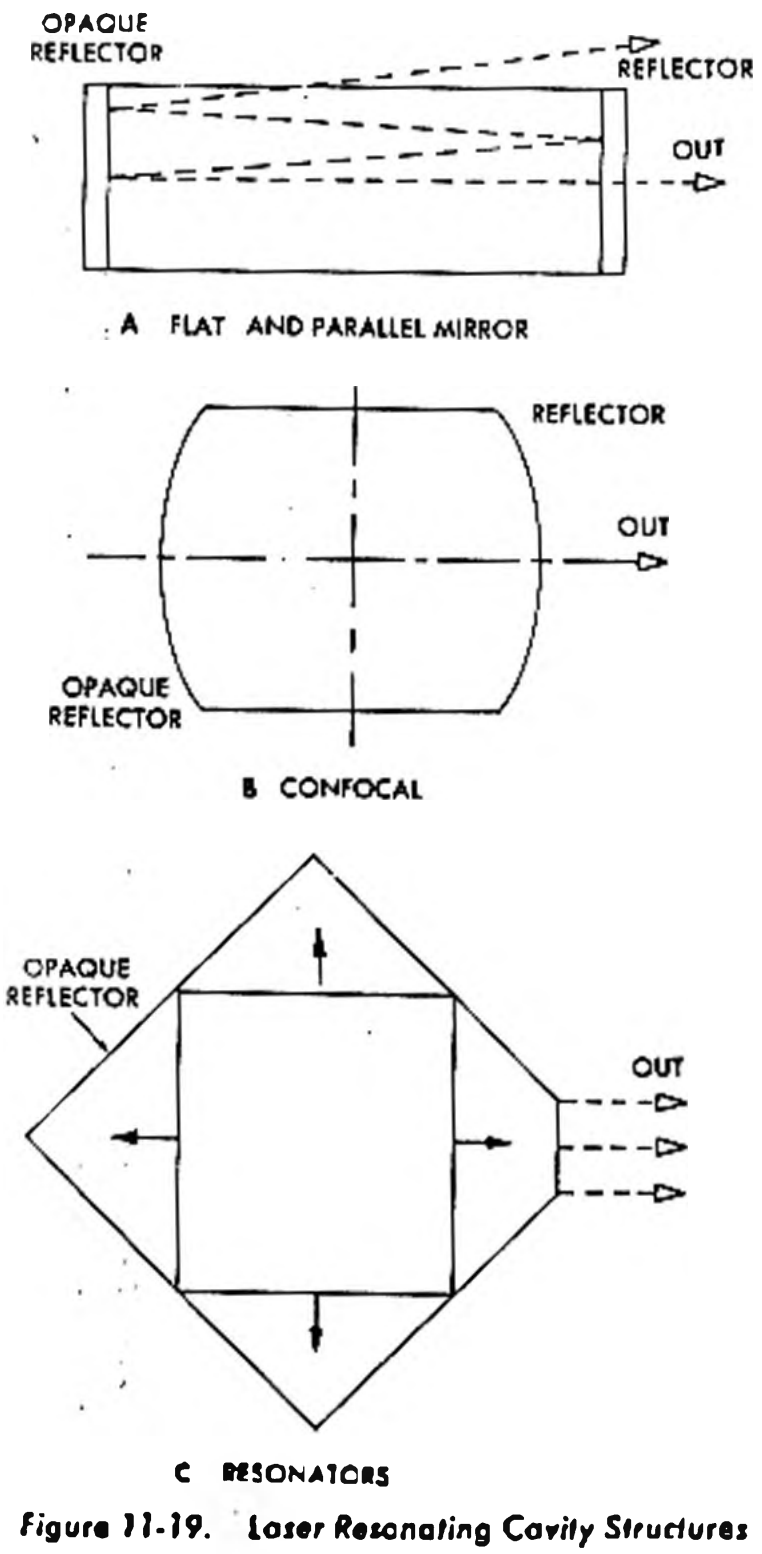


Figure 11-19. Laser Resonating Cavity Structures

the laser crystal for a long enough period to block continuous laser action. Heating tends to broaden the widths of radiated spectral lines.

The helium-neon gas laser shown in figure 11-20 operates at lower power levels than the solid-state lasers developed up to now. Typically, RF excitation power is 50 watts, although it can be as low as 10 watts, and has been reported as high as 80 watts. If too much RF power is applied, laser action will not occur. The minimum discharge-tube length that is necessary for laser action is about 20 cm. The RF exciter of the laser runs at 28 MHz (30-MHz excitation has also been reported). The tube is filled with a mixture of helium at 1.0-mm pressure and of neon at 0.2 mm. The Pyrex discharge tube (quartz tubes have also been used) is mounted on a low-expansion Invar rod frame. Each of the two end assemblies shown in figure 11-20B contains a flexible metal bellows. A differential screw at each end brings the dielectric-coating mirrors into parallel alignment, the differential-screw movement providing a resolution of 1 second of arc. Mirror faces are spaced 1.0 meter apart. Uniaxial flexible studs allow slight positioning adjustments of the end plates. Output windows are sealed with copper gaskets. The window at the end that is not shown in figure 11-20B is a quartz flat. A spring-loading mechanism (not shown) provides a vernier adjustment of parallelism. Copper-brazed stainless steel is used throughout the construction.

INJECTION LASERS

The newest and smallest member of the family of lasers is a tiny block of semiconductor crystal called the injection laser. This represents a giant step toward practical realization of the great potential in laser light. The injection laser is closely related to the optically pumped laser crystal. In 1960, the optically pumped crystal gave a new form of light to science and industry. For the first time, light possessed the useful properties of radio waves and microwaves; it retained its phase for a large number of cycles.

As with radio waves, light consists of electromagnetic oscillations propagated through space. Prior to the laser, light waves had not been produced with the spectral purity that characterizes radio waves. The laser changed this situation by producing light that is extremely well-defined in frequency, and spatially uniform.

The most striking feature of the injection laser is the simplicity with which its electrons are pumped, or excited, into high energy states. Ultimately, the energy used to pump almost all lasers is electrical, but this electrical energy must be converted into energy of excited electrons in a laser.

The simplicity of injection laser pumping can be seen easily by comparing it to two other types (see figure 11-21). In the case of the optically pumped laser, electrical current operates a lamp that emits radiation in the pumping band of the laser. Radiation in the pumping band produces

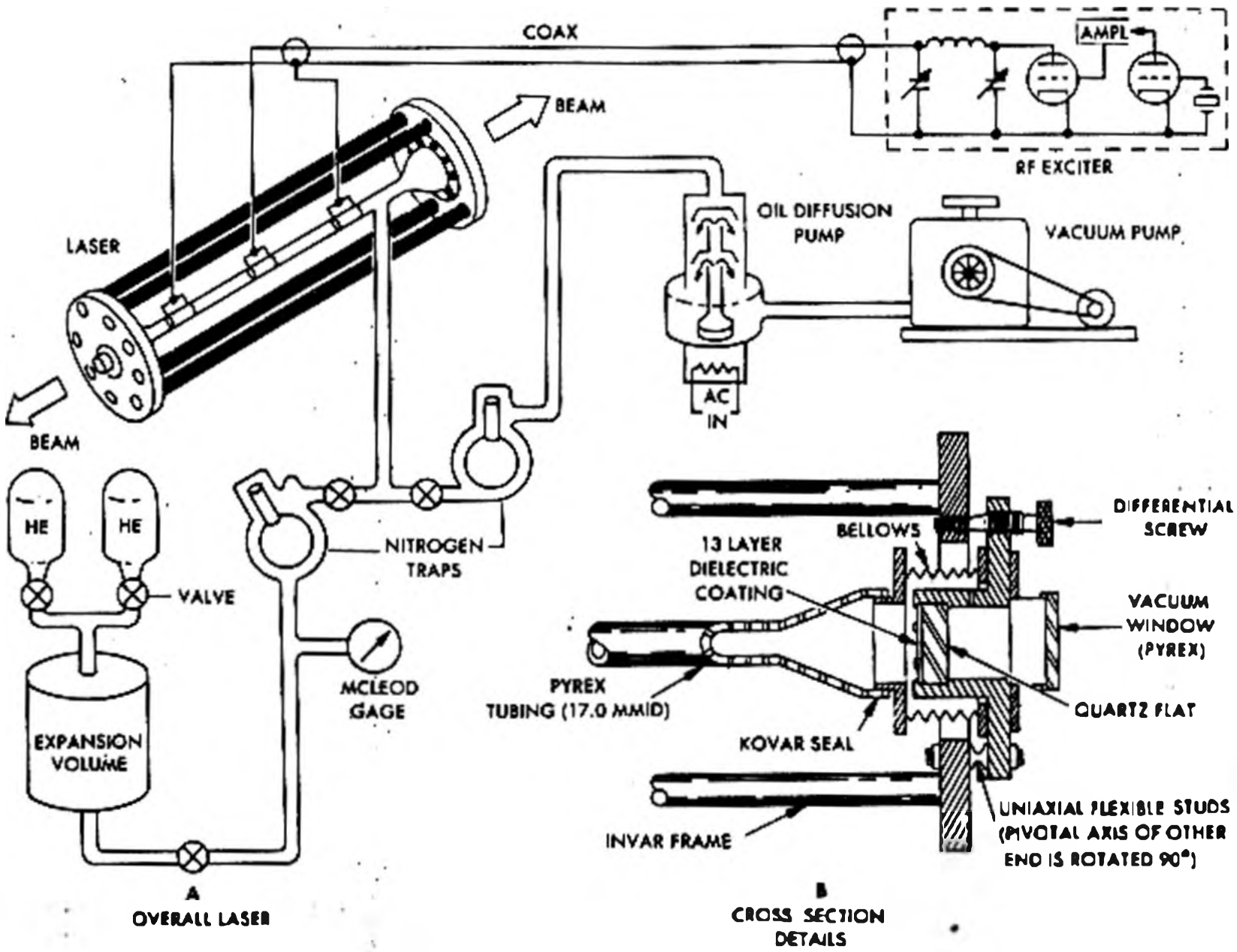


Figure 11-20. A Typical Gas Laser Diagram

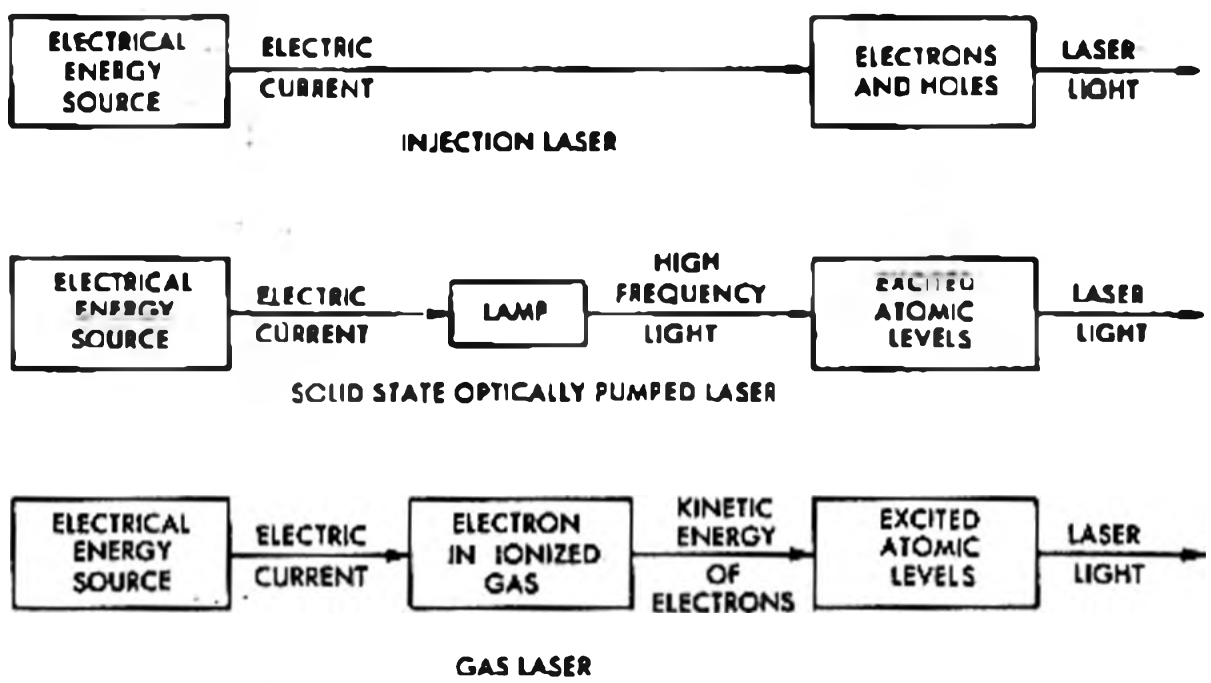


Figure 11-21. Block Diagrams of Basic Lasers

excited atoms. Similarly, in the gas laser the electrical energy is converted first into kinetic energy of electrons. The high-energy electrons then excite atoms of the gas by colliding with them. Injection lasers use electrical energy directly, pumping electrons to high energy states by injecting electrons and holes across a P-N junction.

Eliminating an intermediate stage in the transformation of electrical energy to light has two important advantages. First, although the conversion of energy from one form to another is usually inefficient, the injection laser's direct transformation process brings the efficiency to a relatively high level. Efficiency of the HeNe laser is typically as follows: injection type 20 percent, optically pumped 1 percent, and gas 0.01 percent. A CO_2 is typically 15 percent efficient and CO has been reported to be in excess of 40 percent.

The second advantage is the simplified equipment used in pumping electrons directly with electrical energy. Lasers that have to be pumped through an intermediate form of energy need bulky, expensive equipment for the transformation. The injection laser eliminates this auxiliary apparatus. This second advantage of the injection laser is especially apparent when compared with the optically pumped laser, which is itself a small crystal of fluorescent material. This small crystal must be surrounded by a much larger array of lamps and reflectors when in operation.

The phenomenon of P-N junction electroluminescence was first noted in 1923. It was found that light was often emitted when point electrodes were placed on certain silicon carbide crystals and current passed through them. Modern semiconductor theory explains this phenomenon. If electrons are injected into P-type material, or positive holes injected into N-type material, they recombine spontaneously with the majority carriers existing in the material, and may emit radiation in the recombination process.

Previously, P-N junction electroluminescence had never been a technologically useful phenomenon because its efficiency was low. Most electrons do not emit a photon in recombining with a hole, but rather convert recombination energy to heat. The function of junction electroluminescence in science and technology was altered rapidly in the early 1960s.

In 1960, gas lasers and optically pumped solid lasers stimulated research into the possibility of a semiconductor junction laser. The discovery in 1961 and 1962 that the efficiency of recombination radiation was very high in gallium arsenide focused attention on this compound as a potential laser material. In gallium arsenide, most of the electrons emit a photon during the recombination energy as heat through some nonradiative process.

Laser action is observed when large forward currents are applied to special gallium arsenide diodes maintained at very low (cryogenic) temperatures. The start of laser action occurs at a definite current threshold, indicating that a population inversion is involved. These observations confirm the existence of this new type of laser, the injection laser.

Figure 11-22 shows the experimental setup required to observe injection laser operation. Pulses of high current density are passed through the diode while it is immersed in liquid nitrogen. The light which emerges from the plane of the junction is usually in the near infrared region and may only be viewed indirectly through an infrared image converter (also called a "snooper-scope"). Alternately, by passing the light through a spectrometer (an optical frequency analyzer), the spectrum of the emitter radiation may be determined.

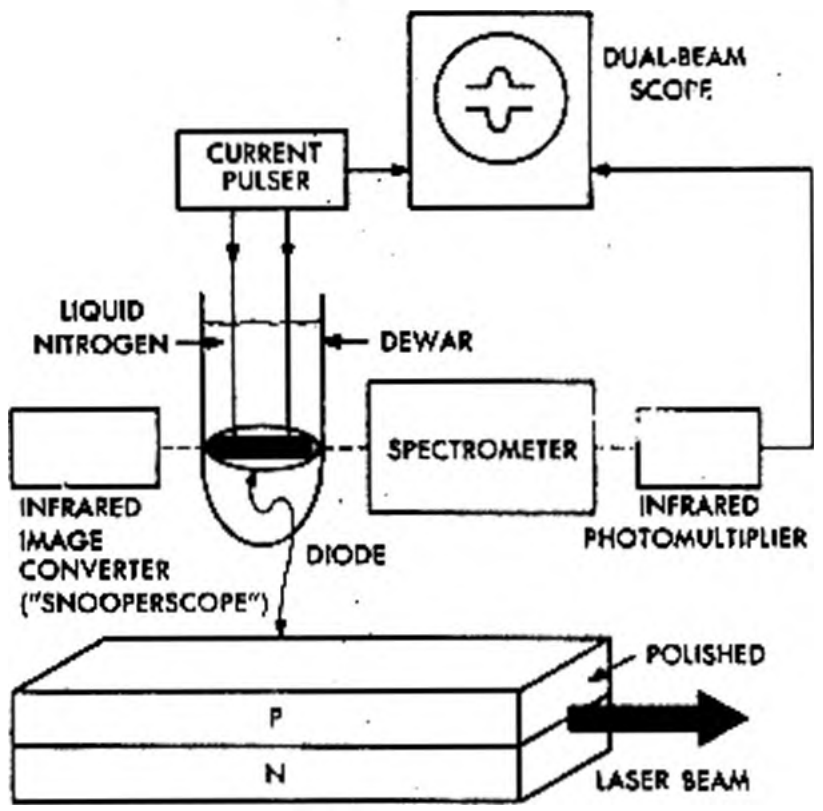


Figure 11-22. Experimental Setup for Observing Injection Laser

This spectrum shows a marked peaking coincident with the onset of laser action. To understand the operation of the injection laser, you should take note of the close analogy between the band structure diagram, familiar to those who work with semiconductors, and the energy level diagram which represents the behavior of the atoms.

Of particular interest is the fact that recombination of an electron-hole pair at a semiconductor junction is frequently accompanied by the emission of a photon. The photon's energy may be expressed as $E = hf = \Delta E_G$, where ΔE_G is the energy band gap. The energy gap is the difference in energy between the lowest level of the conduction band, and the highest level in the valence band.

At normal temperatures, recombination of carriers (conduction electrons and holes) may occur without any radiation. These recombinations occur at sites where the periodic crystal structure is disturbed by the presence of impurities and defects in the crystal. Such sites are called *traps*. The lifetime of an electron in the semiconduction band is strongly limited by the traps. For this reason, the conduction band is generally unsuitable as the metastable state, which is so crucial in obtaining population inversion.

When the temperature of the diode is lowered, the effectiveness of the traps is reduced, and lifetimes are considerably extended. In certain semiconductors at extremely low temperatures, the radiative process dominates recombination. Under this condition, the situation is analogous to laser transition levels in the ruby crystal. In order to pump when this method is used, a region must be created in which conduction electrons and holes abound. The most elementary considerations show that this exists in the immediate neighborhood of the junction when a forward bias is applied. At extremely low temperatures and sufficiently strong forward bias, the region of the junction should be an amplifier for radiation satisfying $hf = \Delta E_G$. By cleaving and polishing the diode into a rectangle with faces perpendicular to the plane of the junction, a resonant cavity is formed, and oscillation can be observed.

The overall efficiency of the diode laser is 10 to 30 percent. The junction region is necessarily quite small, and the total power which may be radiated is limited. Because the conduction and valence bands are considerably broader than the discrete atomic energy levels, the emission contains a comparatively broad band of frequencies. These facts tend to offset the obvious advantages gained by the ability to pump the diode laser directly with wide electrical energy, and in being able to modulate the output at high frequencies (about 1000 MHz) by modulating the pumping current.

LASER PUMPING

Pumping action of the different types of lasers is accomplished by several different methods, depending on the type of laser.

Operation of a gas laser (pumped by electron collision) requires a means of creating an electrical discharge in a tube containing low-pressure gas. This is done by either one of two methods, RF or DC pumping. In one case of RF pumping, an RF generator, capable of supplying 50 to 100 watts of power, is used. This supply is coupled to an electrode, or electrodes, spaced around the outside of the gas-filled tube. Coupling is provided by a matching coil specially designed for maximum power transfer to the electrodes (see figure 11-23). For an approximately 1-mw output laser (typical helium-neon operation), pumping power levels run from 20 to 50 watts at about 30 MHz.

Alternately, the discharge may be maintained by supplying a sufficiently high DC field between electrodes contained in the laser tube. Heating the cathode with a filament supplies electrons for the collision process. The necessary potential depends on the separation between the electrodes and the tube diameter. A potential drop of 4 kilovolts is typical in a laser tube 4 mm in diameter and 1 meter long.

Pumping a solid laser rod, such as ruby, is accomplished optically. To obtain sufficient light power for inversion, it is usually necessary to fire a high-power xenon-filled flash tube. The flash tube is optically coupled by mirror geometry to the laser rod. Operation of the flash tube is similar to that of a thyratron. A high-voltage trigger breaks down the xenon gas between two electrodes which have been maintained at different potentials. Upon breakdown, the lamp resistance drops to a low value of 1 ohm, and high peak currents flow.

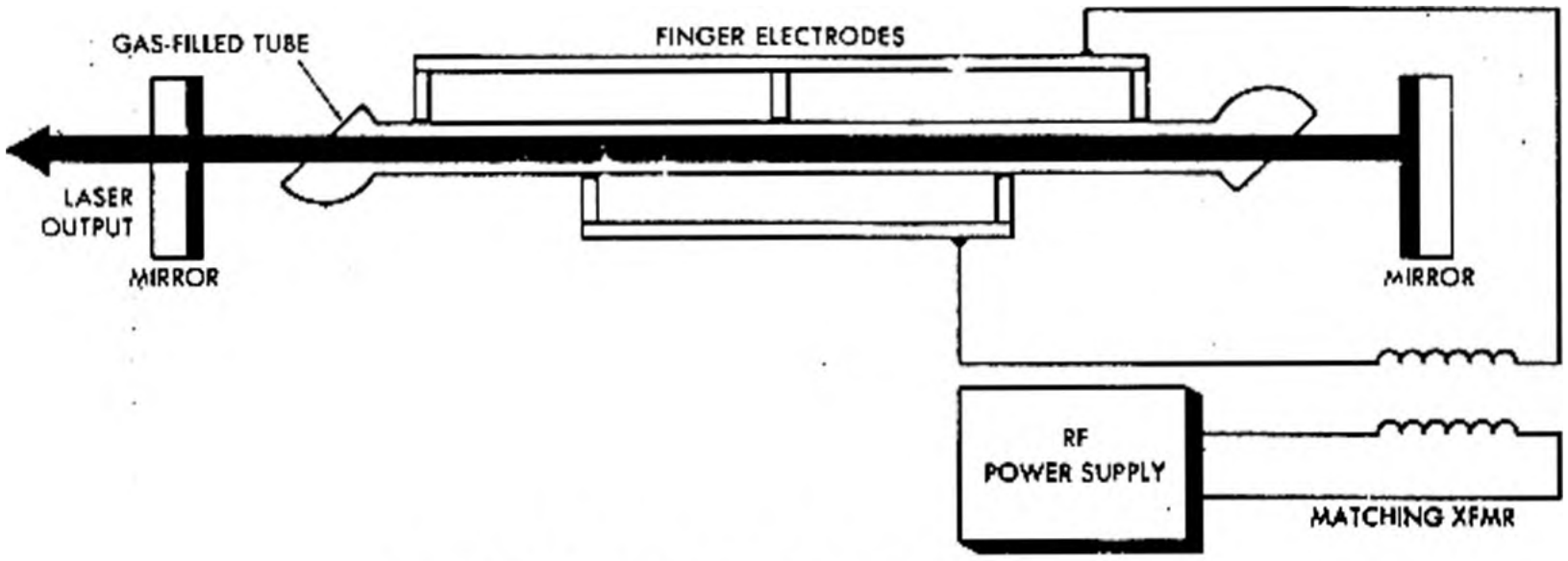


Figure 11-23. Method of RF Pumping in Gaseous Laser

The magnitude and duration of these currents depend upon the capacity and voltage of the supply; that is, the peak current is $I_p = V/R$ and the duration, T, is equal to RC where V is the supply voltage, C is the capacitance, and R is the resistance of the flash tube. Flash tubes are generally rated in maximum watt-seconds (or joules) per current pulse. Typically, these values are quite high. For example, 250 watt-seconds are necessary for a small ruby laser, and up to 12,000 watt-seconds may be needed for a large ruby laser. Considerable energy must therefore be stored in large capacitor banks before discharge. The large values of stored energy at high voltages make safety an important consideration in handling the power supplies for even small ruby lasers. Figure 11-24 shows a typical power-supply schematic for a small ruby laser.

Operation of the injection laser on a pulsed basis (typical usage) is similar in many respects to firing a flash lamp. High peak currents of 70 amperes are necessary to create inversion, and the forward resistance of the diode is considerably less than 1 ohm. There is no provision for triggering the current. A transistor or a silicon-controlled rectifier (SCR) must be placed in series with the diode. Direct modulation of forward current, and hence laser output, may be applied by coupling via a tee from an RF generator to the pumping supply. Modulation rates of up to 1000 MHz are possible in this manner.

LASER MODULATION AND DEMODULATION

Since the methods used for modulation and demodulation are different than those used for conventional electronic systems, a brief description is provided in the following paragraphs.

Modulation

The laser, by virtue of its coherence properties, has offered communications systems the possibility of using carrier frequencies of 10^{14} to 10^{15} Hertz for communication. The significance of this contribution is established by communication theory, in which it has been shown that the information capacity of a communications channel is proportional to its bandwidth. The modulation capabilities of a carrier wave are proportional to its center frequency.

Theoretically it is possible to modulate a 100-MHz carrier with approximately 200 20-KHz audio channels without overlapping. A laser channel at a center frequency of 10^{15} Hertz will permit 10 million times that number. A single laser is therefore capable of simultaneously emitting 30 billion 20-KHz audio channels. To develop the laser to its full potential as a communications carrier, means of coherent modulation and demodulation which are compatible with the tremendous allowable bandwidth must be provided.

Although the energy levels appearing on the energy level diagram for an atom are quite well defined and narrow, the stimulated emission is not strictly monochromatic. The spreading of the spectrum of radiation is due to several causes, the most significant of which is simple Doppler shifting. This phenomenon is essentially the same

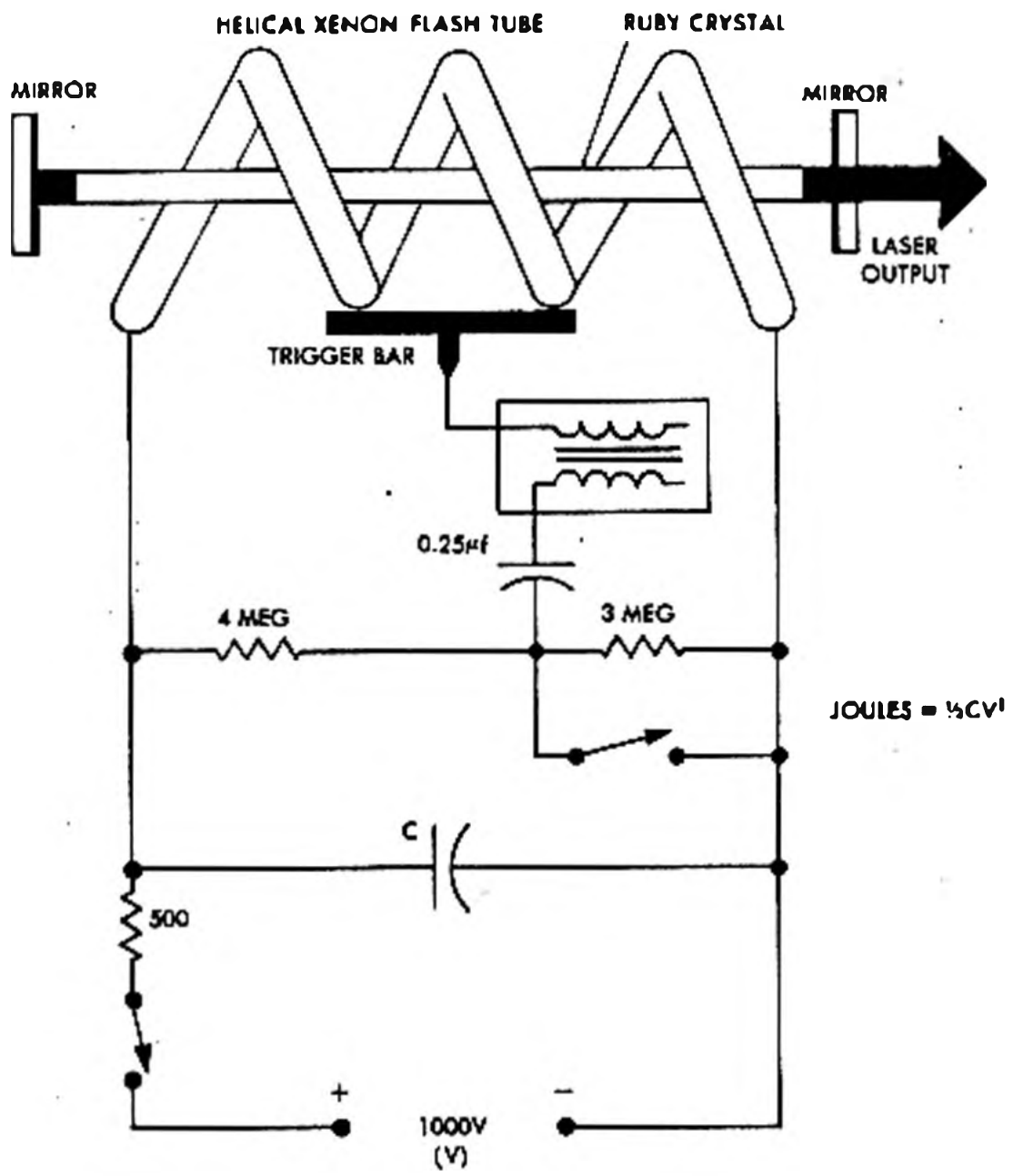


Figure 11-24. Ruby Laser Pumped with Xenon Flash Tube

as that encountered in sound waves. The motion of the source relative to the receiver causes a shift in the received frequency proportional to the relative velocity. The atoms in a gaseous discharge have random velocities because of their thermal motions, which may be as high as 10^6 cm/sec. These random velocities shift the emission and absorption frequencies of the atom over a 100-MHz frequency band.

The cavity modes for the laser oscillator are quite closely spaced because the cavity is many times larger than the wavelength. Frequency separation, f, for two adjacent axial modes in a long, narrow cavity of length, L, is given by $f = C/2L$, where C is the velocity of light. For L = 1 meter, the frequency spacing between adjacent modes is 150 MHz. The spacing between resonant frequencies is then much narrower than the spread of frequencies caused by the Doppler shift. The result is that the laser oscillator output consists of several frequencies spaced in accordance with cavity geometry.

The width of cavity resonances, as in microwaves, is determined by the cavity Q. Cavity Q refers to the fraction of energy lost per cycle. The frequency spread resulting from these losses is f_0/Q, where f_0 is the center frequency. Because the gain of a laser amplifier (expressed in percent/meter) is quite low (generally smaller than 1 percent), cavity losses must be proportionally small to achieve oscillation (gain smaller than 1). Therefore, the Q is made high, and the frequency spread of the cavity resonances is consequently small as compared to the space between them. There are also off-axis resonant modes of the cavity with sufficient gain to cause oscillation.

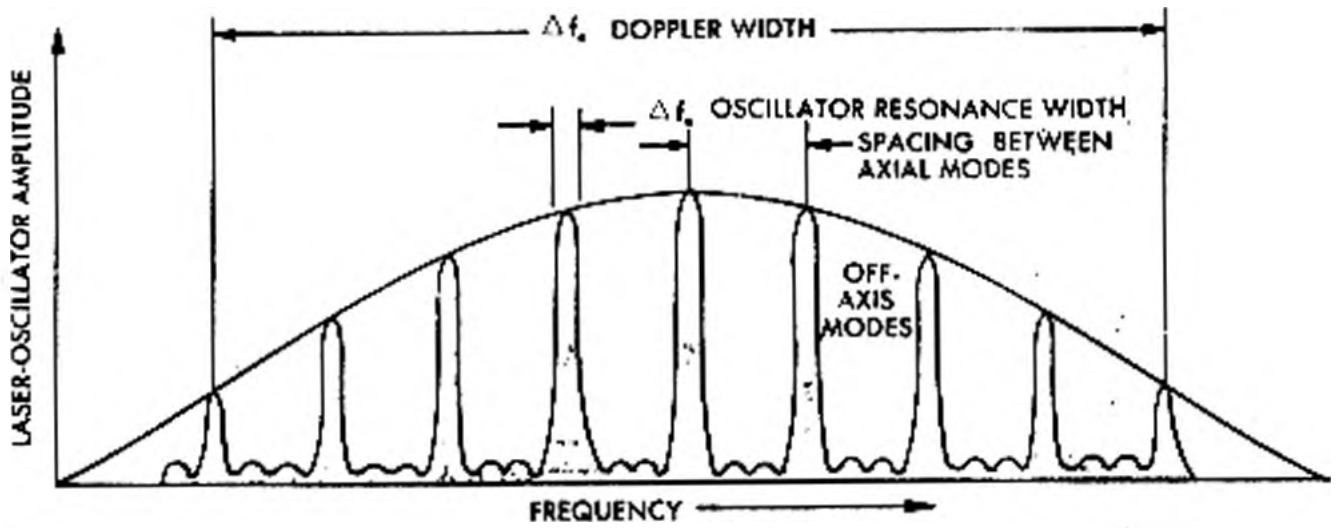


Figure 11-25. Spectrum of Oscillating Helium-Neon Laser

Figure 11-25 shows the typical spectrum of a helium-neon laser. Note that although the individual oscillator resonances are narrow, the laser oscillator is operating at several frequencies simultaneously. To achieve a single oscillator frequency, the output must be passed through a narrow bandpass optical filter. This is done by coupling to another high-Q cavity of a different length, which is resonant for one of the oscillator outputs. Since the lengths are different, the spacing between resonant modes is different, and the second cavity eliminates the other modes. The laser output, after passing through the narrowband filter, contains only a single frequency.

Applying this laser output to communications requires a means of wide-band modulation and demodulation. Direct modulation of the pumping source (flash lamp or gas discharge) in gaseous or solid lasers, which is limited by thermal inertia to a few kHz is obviously insufficient. Other techniques to be considered can be divided into two groups. These are internal modulations in which the modulation is impressed upon an element of the feedback loop, and external modulations in which the modulation is impressed directly upon the oscillator (light-beam) output.

Various potential wide-band modulation schemes have been proposed, several of which have been successful over several GHz. The following paragraphs consider only amplitude modulation by means of varying the oscillator cavity length.

To obtain a fast shutter which may be directly operated by a voltage, a Kerr cell can be used. A Kerr cell consists of a special clear liquid (for example, nitrobenzene) between two electrodes. When voltage is applied to the electrodes, the liquid acts to rotate the plane of polarization of light passing through it. The amount of rotation is proportional to the length of the cell and the square of the applied voltage. By modulating the voltage across the two electrodes, a linearly polarized beam will become polarization-modulated. To convert polarization modulation to amplitude modulation, the light is passed through a polarizer. The transmission, T, of a linear polarizer for linearly polarized light is simply $T = T_o\cos^2\theta$, where θ = the angle between directions of polarization, and T_o is the transmission when they are perfectly aligned.

Figure 11-26 shows the operation of a Kerr cell with a laser transmitter. Modulation by this means is limited to a few MHz by the time constant of the Kerr cell itself. To achieve higher modulation frequencies, the interaction distance must be increased by forming a transmission line imbedded in a time-varying dielectric, where the Kerr cell liquid acts as the dielectric.

Frequency modulation of the laser output may be accomplished by changing the cavity length, thereby varying the frequency of the resonant modes. For a helium-neon laser, a change of cavity length of one-half wavelength (to the next resonant mode) corresponds to a frequency shift of 150 MHz. A half wavelength for a helium-neon laser is only 3×10^{-5} cm. The change of length necessary for wide-band modulation is, therefore, quite small. Such small displacements may be conveniently provided by a piezoelectric material.

A piezoelectric material, such as quartz, is a

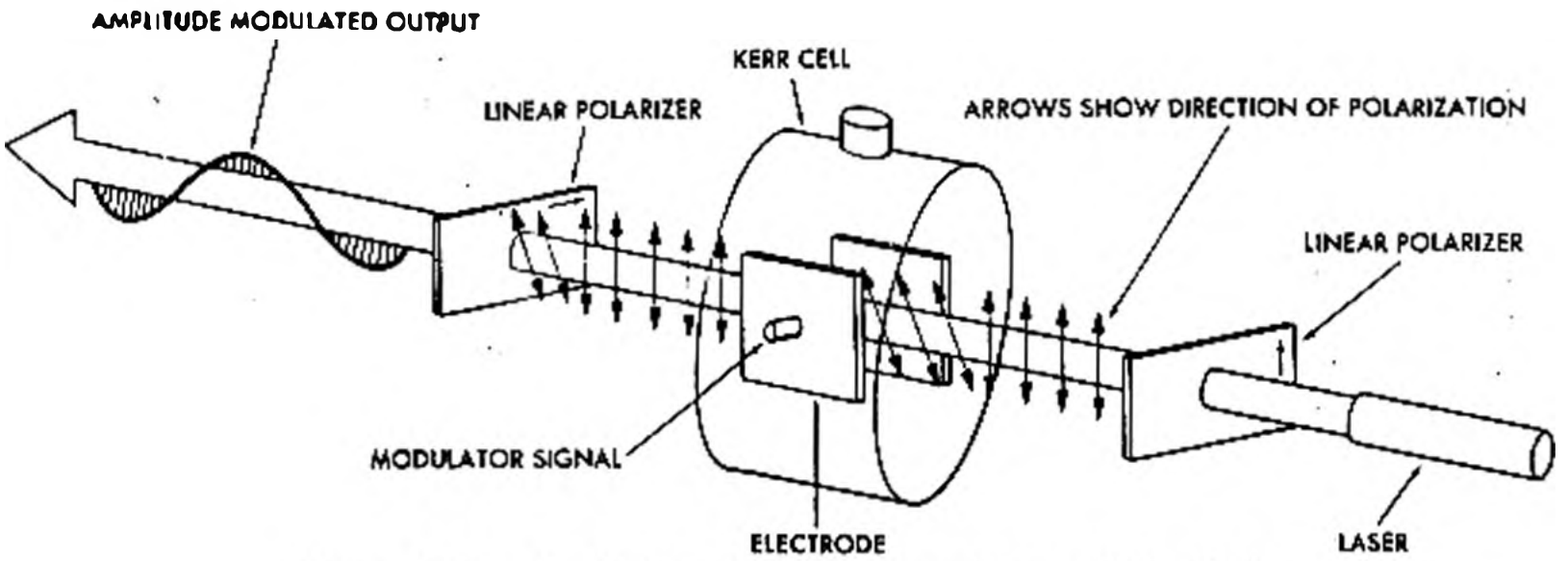


Figure 11-26. Amplitude Modulation of Laser Beam with Kerr Cell

material which contracts and expands along a dimension perpendicular to the direction of an applied voltage. If the applied voltage is a varying voltage, and one face of a thin slab of piezoelectric material is secured firmly to a stationary object, the exposed surface will undergo a vibration. This vibration can be transferred to a variation in cavity length by mounting one of the cavity mirrors on this surface. The motion of the mirror modulates the cavity resonance, and thus frequency-modulates the laser output. This method of modulation is suitable up to modulation rates of 1 MHz at which point the inertia of the mirror becomes the limiting factor. It is not practicable to consider the extension of such a method to a very high frequency, since the range of modulation is limited by the Doppler width of emission line.

To detect coherently a modulated carrier, the received signal must be compared to a local-oscillator signal in a nonlinear element (mixer). In the case of a laser communications link, the local oscillator is provided by a second laser at the receiver, operating at the same wavelength as the transmitter. The nonlinear element may be any material with nonlinear response at the carrier wavelength. It may be the surface of an ordinary photodetector. All photodetectors are essentially square-law devices; that is, the voltage developed is proportional to the intensity of light at the sensitive surface, which, in turn, is proportional to the square of the electric field.

If the two inputs are mixed at the sensitive surface, the resultant response contains a term whose amplitude is proportional to the product of the local oscillator signal and incoming signal strength. The frequency is equal to the difference in frequency between the output of the local oscillator and the signal. By frequency-locking the local oscillator to the transmitter, the carrier is suppressed and the information (AM or FM) appears directly as voltage at the terminals of the detector.

Note that the signal band-widths concerned must exceed 1000 MHz effective use of the laser carrier frequency. This places a requirement on the frequency response of the photodetector which can be met only by means of specialized construction of the detector element.

Microwave Modulation of a Gallium Arsenide Injection Laser

Gallium arsenide (GaAs) has a minority carrier recombination lifetime on the order of 10^{-9} sec. Since the radiative recombination lifetime decreases in the stimulated emission mode, it is theoretically possible to modulate gallium-arsenide lasers at the microwave frequencies.

A block diagram of the modulation and demodulation arrangement is shown in figure 11-27. A 0.1-μsec, 12.5-ampere video pulse is used to drive the gallium-arsenide diode into the laser action mode. A 1-mw CW signal from the microwave generator is gated by the microwave modulator to produce 0.1-μsec pulses synchronized with the video pulses. The microwave pulses are amplified in the traveling-wave tube to a peak level of 1 watt, and added to the video pulses in the DC monitor tee. The sum is fed to the laser mounting termination.

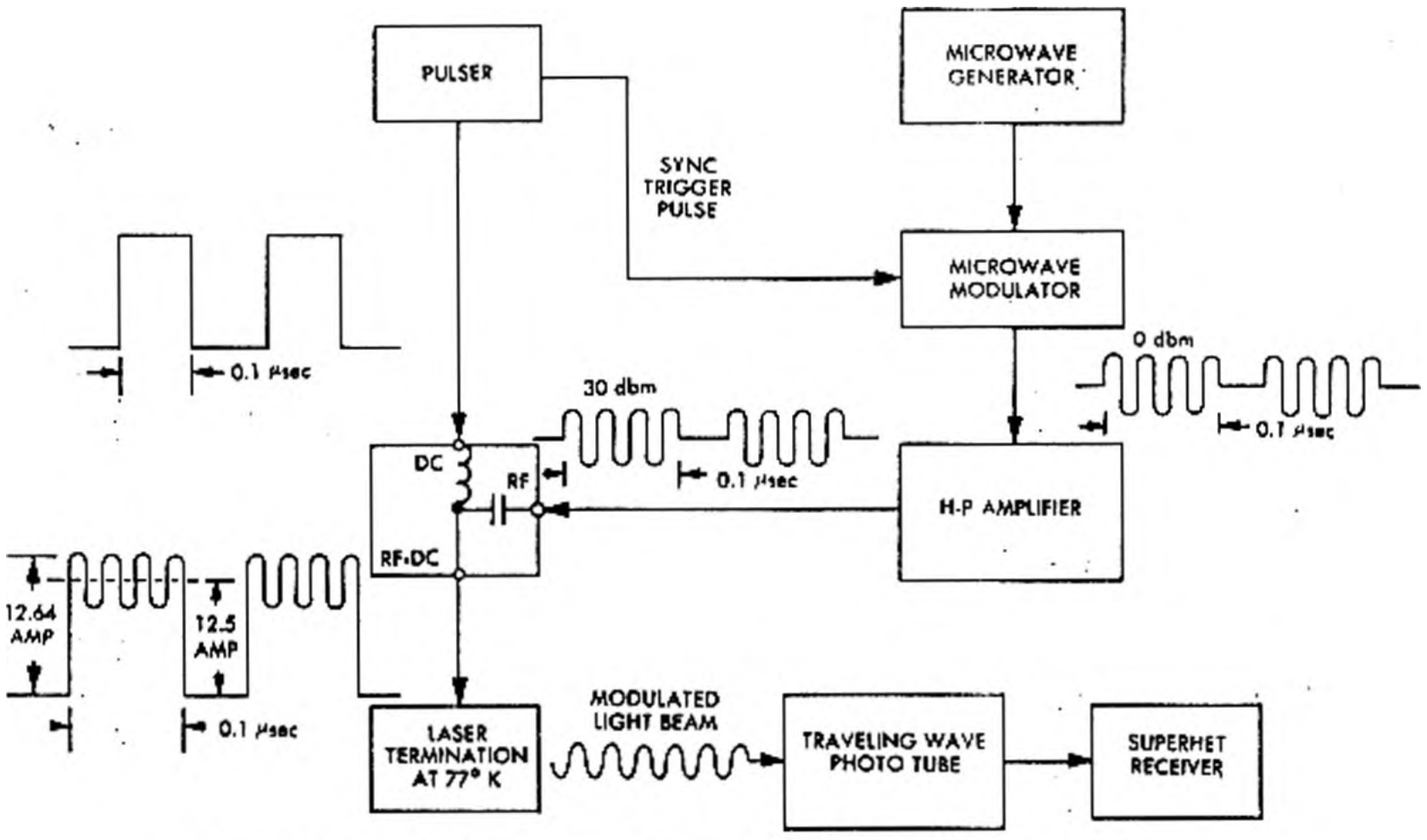


Figure 11-27. Block Diagram of Modulation and Demodulation Arrangement

The termination is shown in figure 11-28. A 50-ohm carbon composition resistor is used as a series resistive match. The laser diode is pressed between the resistor end cap and the copper cooling block, which is immersed in liquid nitrogen. The spring loading maintains constant contact pressure under thermal contraction of the diode. The laser light beam is emitted through the hole in the outer conductor.

The laser is constructed in the low-inductance pill configuration shown in figure 11-29. The crystal is made by diffusing zinc into tellurium-doped gallium arsenide with a net impurity concentration of $3 \times 10^{-17}/cm^3$. The crystal is cleaved

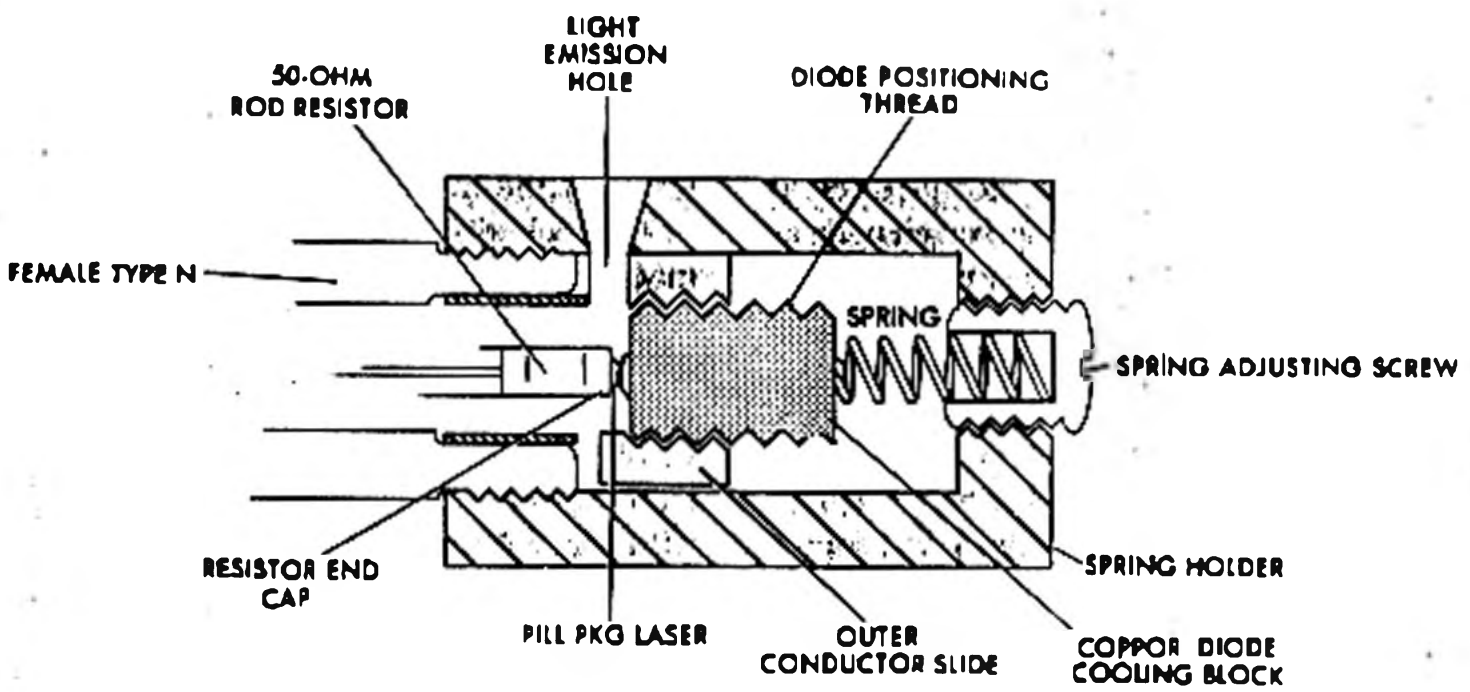


Figure 11-28. Termination for Microwave Modulation

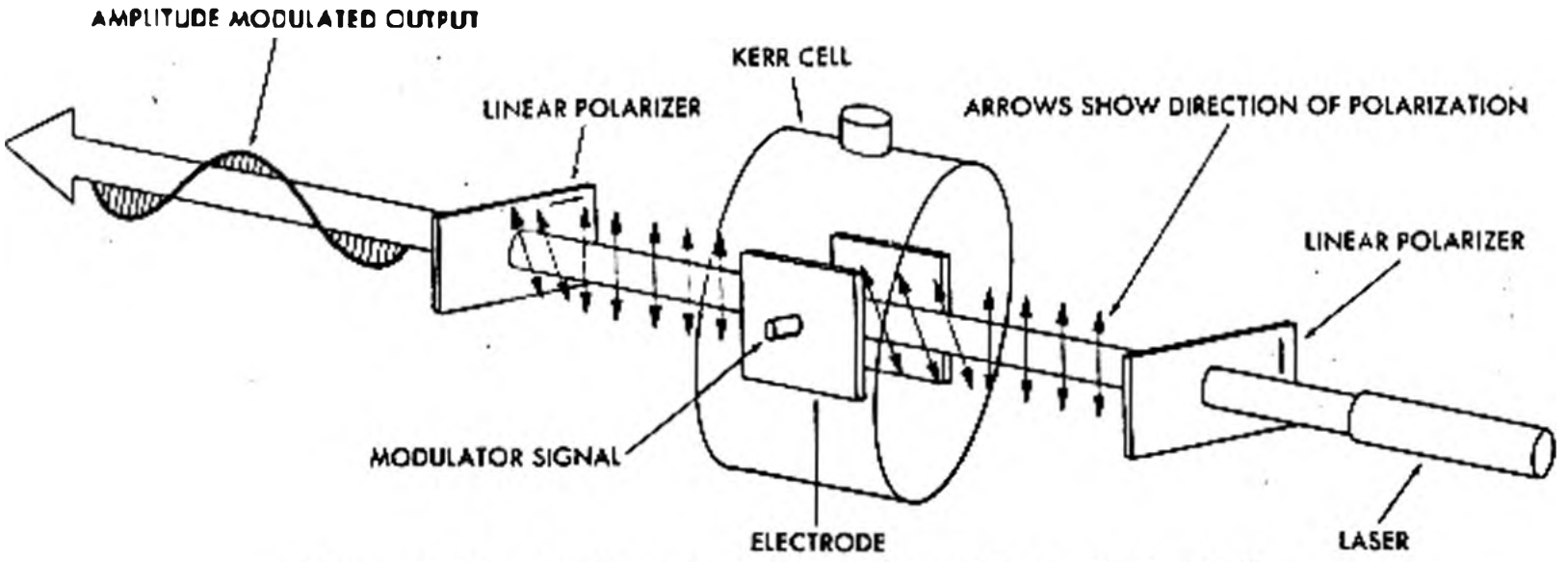


Figure 11-26. Amplitude Modulation of Laser Beam with Kerr Cell

material which contracts and expands along a dimension perpendicular to the direction of an applied voltage. If the applied voltage is a varying voltage, and one face of a thin slab of piezoelectric material is secured firmly to a stationary object, the exposed surface will undergo a vibration. This vibration can be transferred to a variation in cavity length by mounting one of the cavity mirrors on this surface. The motion of the mirror modulates the cavity resonance, and thus frequency-modulates the laser output. This method of modulation is suitable up to modulation rates of 1 MHz at which point the inertia of the mirror becomes the limiting factor. It is not practicable to consider the extension of such a method to a very high frequency, since the range of modulation is limited by the Doppler width of emission line.

To detect coherently a modulated carrier, the received signal must be compared to a local-oscillator signal in a nonlinear element (mixer). In the case of a laser communications link, the local oscillator is provided by a second laser at the receiver, operating at the same wavelength as the transmitter. The nonlinear element may be any material with nonlinear response at the carrier wavelength. It may be the surface of an ordinary photodetector. All photodetectors are essentially square-law devices; that is, the voltage developed is proportional to the intensity of light at the sensitive surface, which, in turn, is proportional to the square of the electric field.

If the two inputs are mixed at the sensitive surface, the resultant response contains a term whose amplitude is proportional to the product of the local oscillator signal and incoming signal strength. The frequency is equal to the difference in frequency between the output of the local oscillator and the signal. By frequency-locking the local oscillator to the transmitter, the carrier is suppressed and the information (AM or FM) appears directly as voltage at the terminals of the detector.

Note that the signal band-widths concerned must exceed 1000 MHz effective use of the laser carrier frequency. This places a requirement on the frequency response of the photodetector which can be met only by means of specialized construction of the detector element.

Microwave Modulation of a Gallium Arsenide Injection Laser

Gallium arsenide (GaAs) has a minority carrier recombination lifetime on the order of 10^{-9} sec. Since the radiative recombination lifetime decreases in the stimulated emission mode, it is theoretically possible to modulate gallium-arsenide lasers at the microwave frequencies.

A block diagram of the modulation and demodulation arrangement is shown in figure 11-27. A 0.1-μsec, 12.5-ampere video pulse is used to drive the gallium-arsenide diode into the laser action mode. A 1-mw CW signal from the microwave generator is gated by the microwave modulator to produce 0.1-μsec pulses synchronized with the video pulses. The microwave pulses are amplified in the traveling-wave tube to a peak level of 1 watt, and added to the video pulses in the DC monitor tee. The sum is fed to the laser mounting termination.

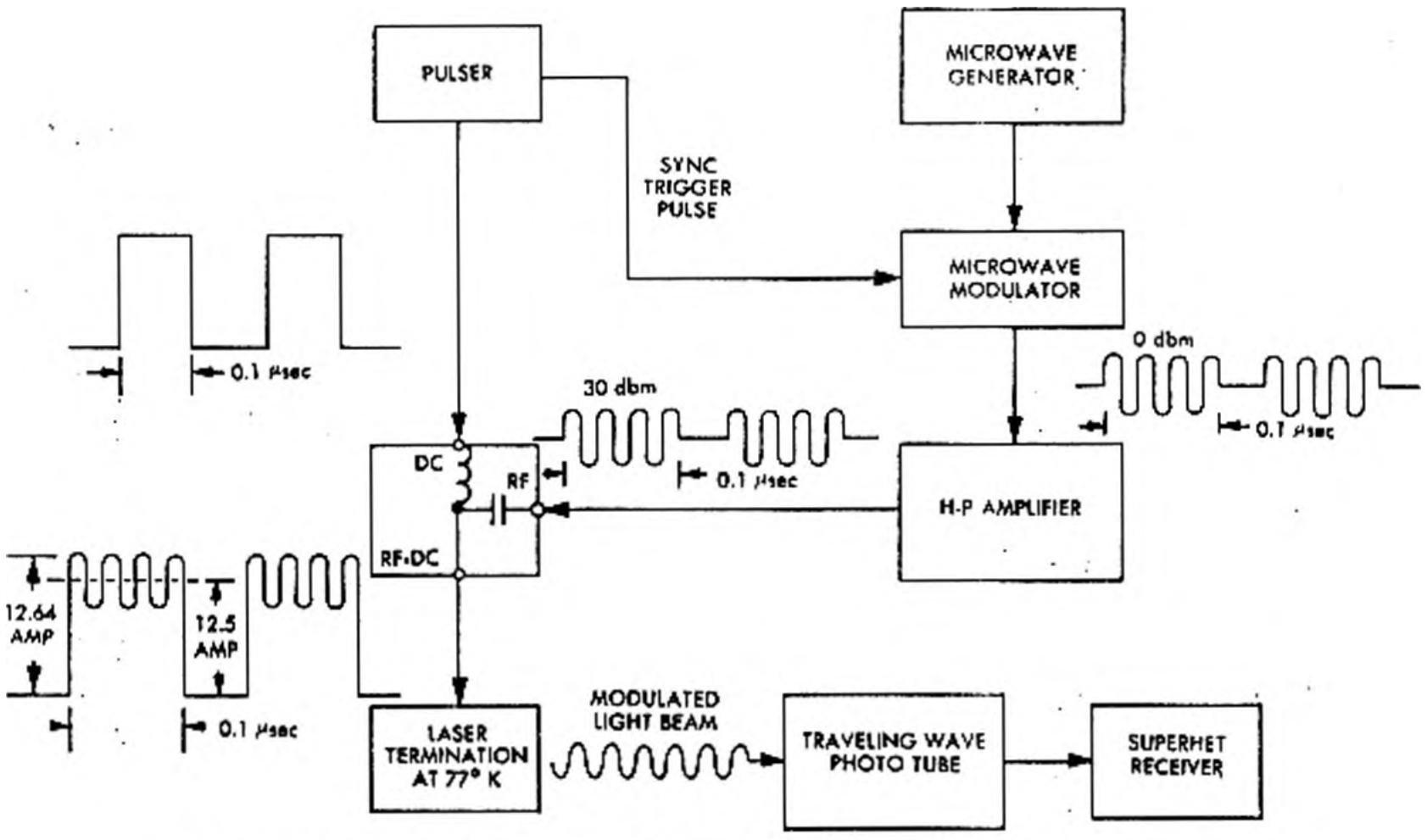


Figure 11-27. Block Diagram of Modulation and Demodulation Arrangement

The termination is shown in figure 11-28. A 50-ohm carbon composition resistor is used as a series resistive match. The laser diode is pressed between the resistor end cap and the copper cooling block, which is immersed in liquid nitrogen. The spring loading maintains constant contact pressure under thermal contraction of the diode. The laser light beam is emitted through the hole in the outer conductor.

The laser is constructed in the low-inductance pill configuration shown in figure 11-29. The crystal is made by diffusing zinc into tellurium-doped gallium arsenide with a net impurity concentration of $3 \times 10^{-17}/cm^3$. The crystal is cleaved

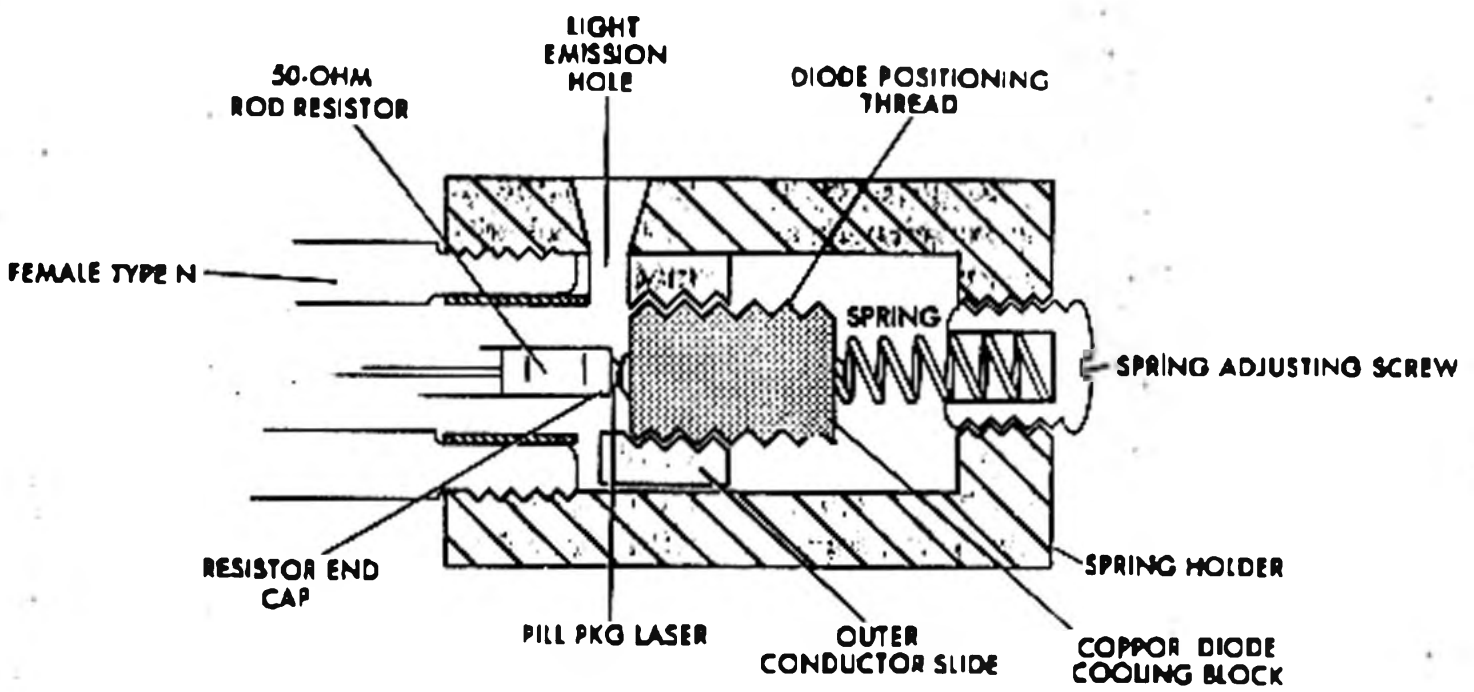


Figure 11-28. Termination for Microwave Modulation

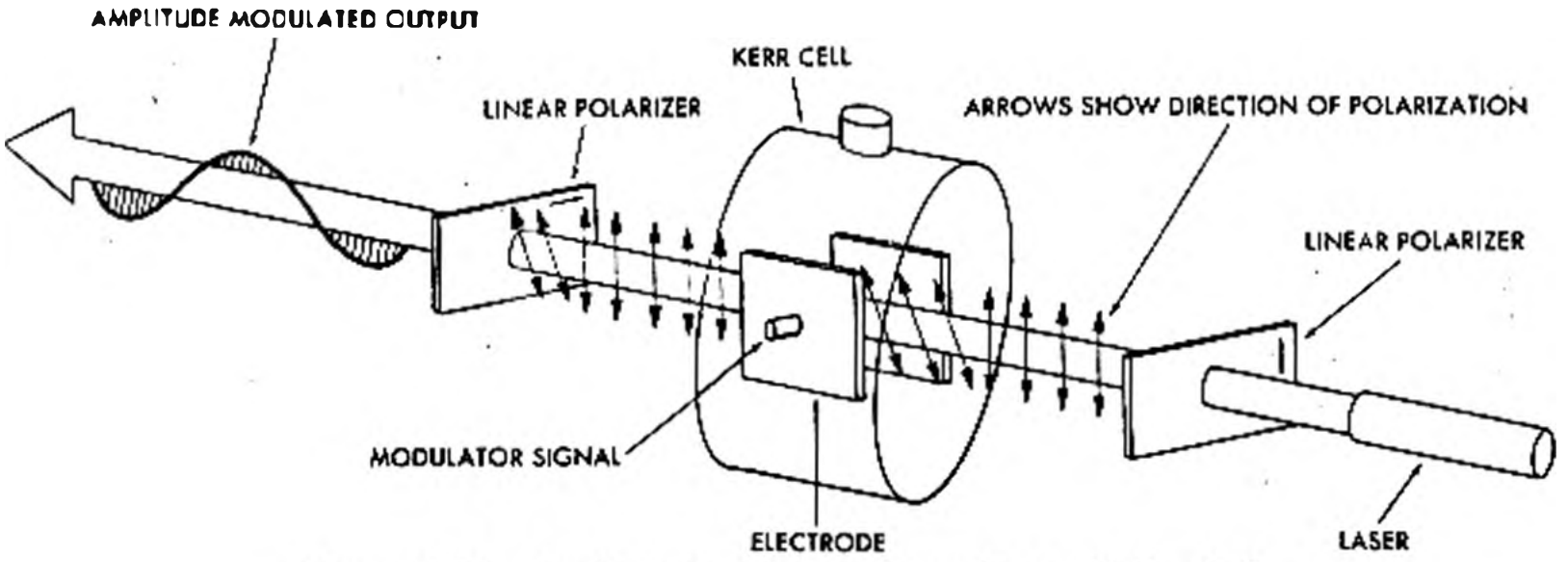


Figure 11-26. Amplitude Modulation of Laser Beam with Kerr Cell

material which contracts and expands along a dimension perpendicular to the direction of an applied voltage. If the applied voltage is a varying voltage, and one face of a thin slab of piezoelectric material is secured firmly to a stationary object, the exposed surface will undergo a vibration. This vibration can be transferred to a variation in cavity length by mounting one of the cavity mirrors on this surface. The motion of the mirror modulates the cavity resonance, and thus frequency-modulates the laser output. This method of modulation is suitable up to modulation rates of 1 MHz at which point the inertia of the mirror becomes the limiting factor. It is not practicable to consider the extension of such a method to a very high frequency, since the range of modulation is limited by the Doppler width of emission line.

To detect coherently a modulated carrier, the received signal must be compared to a local-oscillator signal in a nonlinear element (mixer). In the case of a laser communications link, the local oscillator is provided by a second laser at the receiver, operating at the same wavelength as the transmitter. The nonlinear element may be any material with nonlinear response at the carrier wavelength. It may be the surface of an ordinary photodetector. All photodetectors are essentially square-law devices; that is, the voltage developed is proportional to the intensity of light at the sensitive surface, which, in turn, is proportional to the square of the electric field.

If the two inputs are mixed at the sensitive surface, the resultant response contains a term whose amplitude is proportional to the product of the local oscillator signal and incoming signal strength. The frequency is equal to the difference in frequency between the output of the local oscillator and the signal. By frequency-locking the local oscillator to the transmitter, the carrier is suppressed and the information (AM or FM) appears directly as voltage at the terminals of the detector.

Note that the signal band-widths concerned must exceed 1000 MHz effective use of the laser carrier frequency. This places a requirement on the frequency response of the photodetector which can be met only by means of specialized construction of the detector element.

Microwave Modulation of a Gallium Arsenide Injection Laser

Gallium arsenide (GaAs) has a minority carrier recombination lifetime on the order of 10^{-9} sec. Since the radiative recombination lifetime decreases in the stimulated emission mode, it is theoretically possible to modulate gallium-arsenide lasers at the microwave frequencies.

A block diagram of the modulation and demodulation arrangement is shown in figure 11-27. A 0.1-μsec, 12.5-ampere video pulse is used to drive the gallium-arsenide diode into the laser action mode. A 1-mw CW signal from the microwave generator is gated by the microwave modulator to produce 0.1-μsec pulses synchronized with the video pulses. The microwave pulses are amplified in the traveling-wave tube to a peak level of 1 watt, and added to the video pulses in the DC monitor tee. The sum is fed to the laser mounting termination.

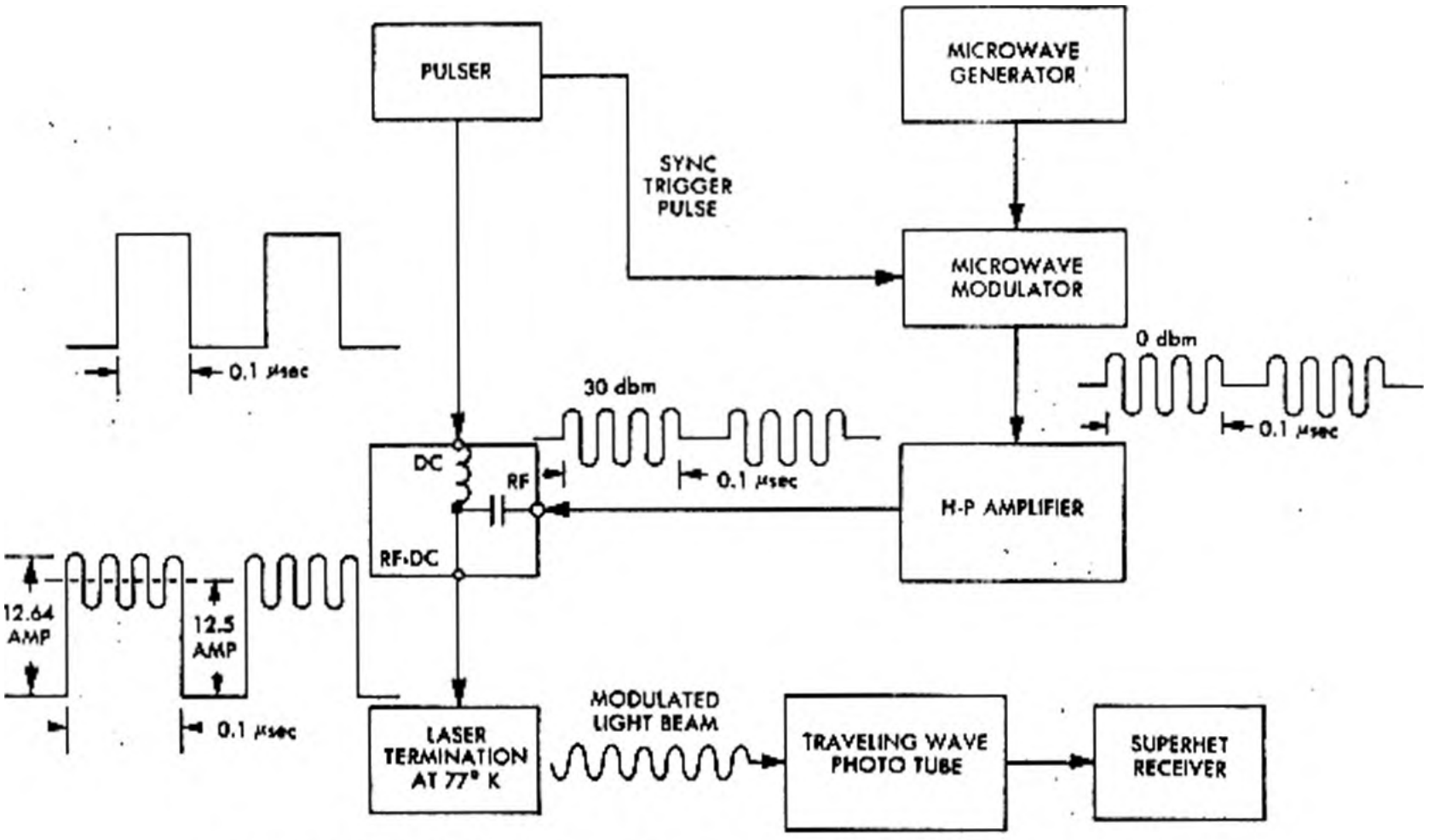


Figure 11-27. Block Diagram of Modulation and Demodulation Arrangement

The termination is shown in figure 11-28. A 50-ohm carbon composition resistor is used as a series resistive match. The laser diode is pressed between the resistor end cap and the copper cooling block, which is immersed in liquid nitrogen. The spring loading maintains constant contact pressure under thermal contraction of the diode. The laser light beam is emitted through the hole in the outer conductor.

The laser is constructed in the low-inductance pill configuration shown in figure 11-29. The crystal is made by diffusing zinc into tellurium-doped gallium arsenide with a net impurity concentration of $3 \times 10^{-17}/cm^3$. The crystal is cleaved

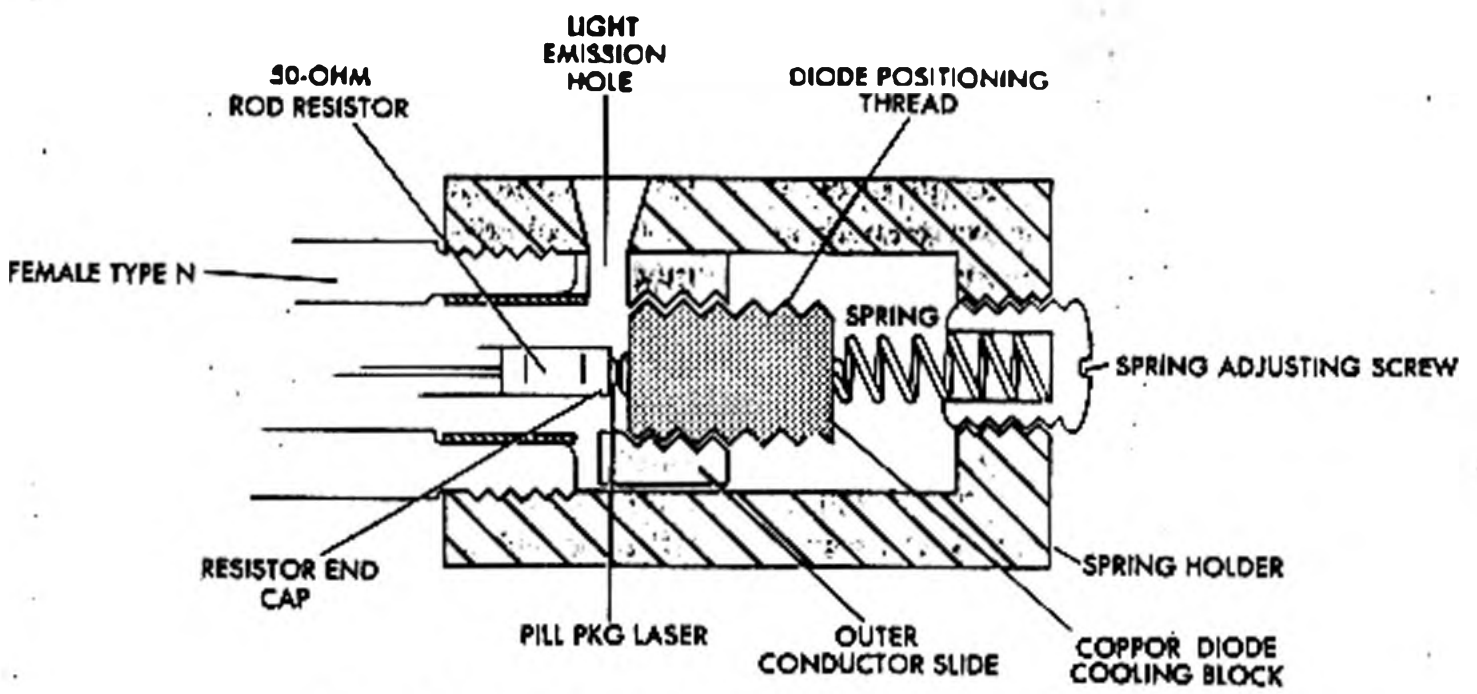


Figure 11-28. Termination for Microwave Modulation

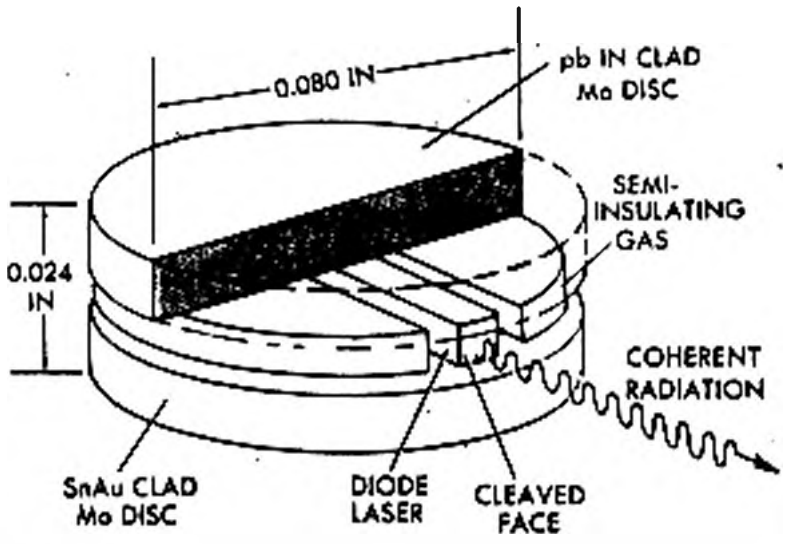
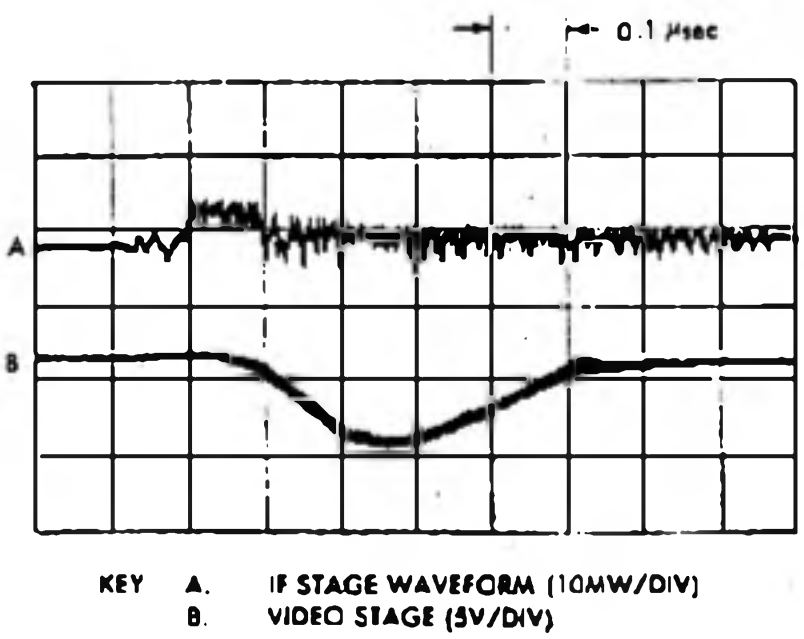


Figure 11-29. Pill Laser Diode Configuration

along the (110) plane. The junction area is $2 \times 10^{-3}/cm^2$, and the laser action threshold for this particular laser is 10.5 amperes (5.25×10^3 amperes/cm^2).

Demodulation

Demodulation can be accomplished with a microwave traveling-wave phototube with a photo surface. The laser beam is focused on the photocathode. The detected microwave signal is fed to a microwave superheterodyne receiver. Figure 11-30 shows the waveform of the detected 2 GHz signal at the receiver IF and video stages. This waveform disappears when the laser beam is blocked or when the local oscillator is detuned. The laser output radiation is modulated with a modulation index of about 7 percent and produces a peak photocathode current of 0.5μa. The detected microwave signal level is -73 dbm. The phototube calibration data indicate that this is the obtainable output power for a cathode current of 0.5μa and a 7 percent modulation index.



KEY A. IF STAGE WAVEFORM (10MW/DIV)
B. VIDEO STAGE (5V/DIV)

Figure 11-30. Demodulated Waveform

LASER RADAR

Although still in the early stages of development, optical radar already provides a coherent light beam 100 times narrower than beams of ordinary microwave radar equipment. Radar pictures made with laser radar are exceptionally sharp, revealing far more detail than previously obtainable. For example, an optical radar demonstrated recently can distinguish between two adjacent 10-foot objects at five miles. Enhanced resolution is obtained with equipment of small size and light weight, because light waves can readily be focused by lens-and-mirror arrangements. Compare this with ordinary radar microwaves that require immense antennas to focus the radio energy into comparably narrow beams.

Laser radars should be especially valuable for use in space, where size and weight are critical, but where there is no absorption or scattering of light rays by rain, clouds, or by the atmosphere itself. It follows that in many space applications, stray sunlight, which might overwhelm the light pulse detected by the laser radar receiver, can be avoided. With the exceptionally narrow light beam practical with laser methods now, it will soon be entirely feasible to illuminate a spot only 10 miles in diameter on the moon's surface. Far more accurate mappings of lunar topography is not far away from actuality.

Principles of Operation

For radar operation, a short burst of red light from the laser is aimed or directed at a target to be detected. Light reflected back from the target is gathered by a telescope and detected by a light-sensitive phototube. Distance to the target is determined by measuring the time required for the light burst to make the round trip. Elements of this method are shown in the block diagram of figure 11-31. The electrical signal resulting from the light pulse returned to the receiver is displayed on a cathode-ray oscilloscope. Time delay for distance measurement is obtained by comparing the received signal with a sample of the transmitted light burst as monitored by a reference phototube.

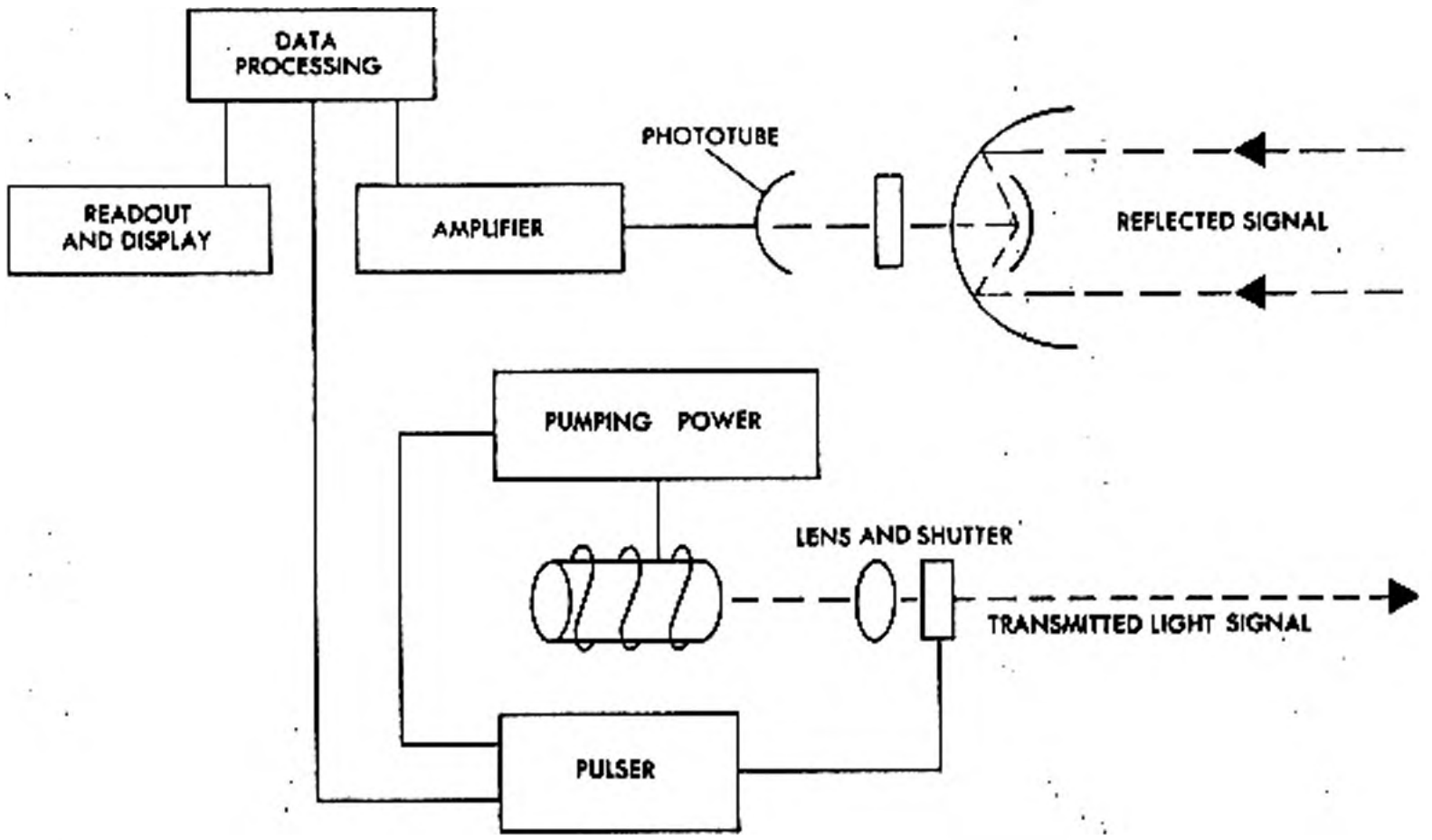


Figure 11-31. Block Diagram of Laser Radar Set

Figure 11-32 illustrates the experimental laser receiver.

The key element of the new optical radar is the ruby-laser light source, sometimes called an optical maser. The ruby crystal is made the driver by a flash of white light from a powerful xenon flash tube similar to those used in a type of electronic photoflash gun. When illuminated or irradiated in this way, the ruby produces an intense burst of pure-red light in an extremely narrow beam.

Although lasers eventually should provide beams as narrow as 0.004 degree without the aid of lenses, the laser used in the experiments emitted a 0.7-degree beam because of inherent imperfections in the ruby crystal. To reduce the beamwidth further, a telescope with a 6-inch mirror was employed and provided a forty-fold improvement. Use of the telescope resulted in a 0.023-degree beam. When this equipment was directed at a building 3½ miles distant, the spot from the beam of light was only 7 feet in diameter.

The present laser transmitter operates with peak powers up to several kilowatts at a wavelength of 6943 Angstroms, or 0.00006943 centi-

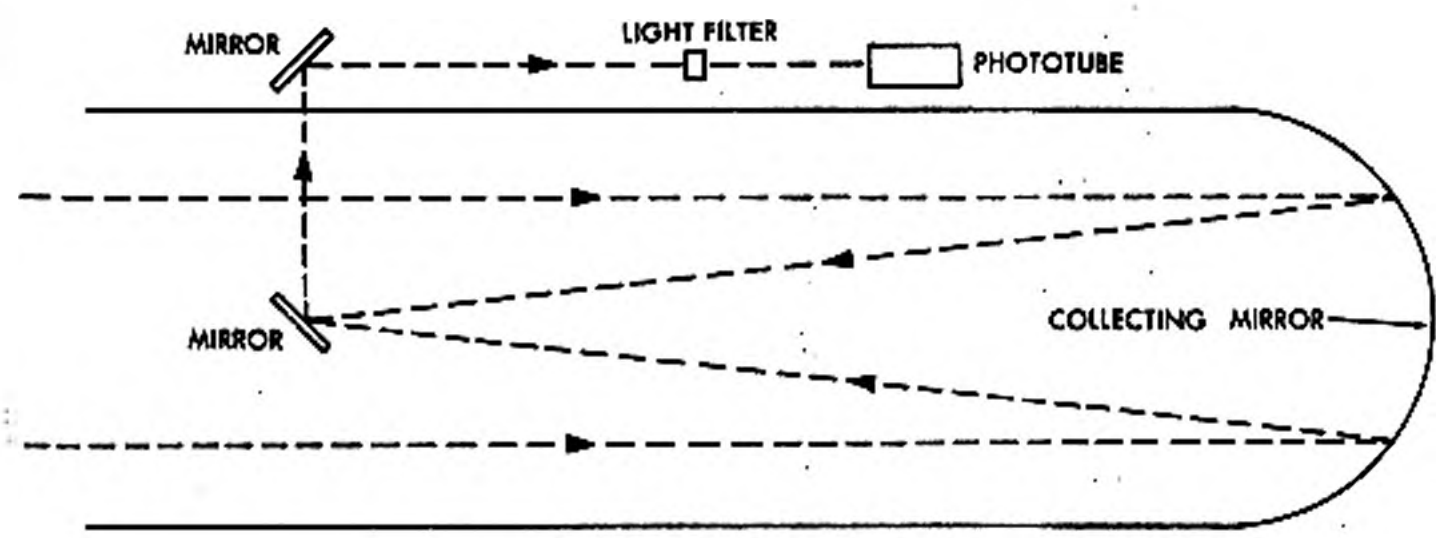


Figure 11-32. Laser Radar Receiver

meter. The laser output is extremely monochromatic; that is, composed of one color or hue. All of the output energy is concentrated in a 0.1-Angstrom-wide spectral band at the wavelength mentioned, thus permitting a narrow spectral filter to be used at the receiver.

For ideal microwave radar performance, the transmitter output should be composed of a single short pulse. The pulse preferably should last for a few microseconds, or a few millionths of a second.

One of the principal components of the laser ranging equipment is the receiver. Here light is gathered by another telescope, with a 5-inch lens. The larger the diameter of the telescope, the more sensitive the receiver. Light entering the telescope is first passed through a red filter which does not impede the 6943-Angstrom radiation from the laser, but which does reject 99.9 percent of the light from all other ordinary sources, such as sunlight, light from incandescent lamps, etc. Light passing through the filter is then directed at a very sensitive phototube where it is converted into an electrical signal for display on the oscilloscope. A photomultiplier-type of phototube provides a very sensitive detector.

Rough calculations indicate that only two or three orders of magnitude improvement in the system would be required to obtain detectable pulses from the surface of the moon. Refinements already well under way include high power lasers operating with short bursts (to reduce problems from back-scatter from dust and water particles in the air), more sensitive receivers, and improved optical systems.

Noise Effects

As with any radar system, the basic limitation in performance that can be achieved by the laser radar is determined by the noise level through which the signal must be detected. Three types of noise affect the system. These are internal or receiver noise, external noise caused by unwanted light entering the receiver through the red filter, and fluctuation or shot noise in the signal itself.

Receiver noise is caused by dark current (current flowing in the absence of any light) in the phqtotube. This current is produced by the thermionic emission from the sensitive photosurface and can be essentially eliminated through refrigeration of the photocell. Dark-current noise at room temperature is low enough to be neglected with present-day systems.

Secondly, for daytime operation on the earth's surface, that part of the scattered sunlight which can pass through the red filter imposes the limiting noise for the equipment. This has been reduced by using a 155-Angstrom interference-type filter. Moreover, similar filters with spectral bandwidths as narrow as 1 Angstrom are feasible today.

The third type of noise comes from signal fluctuation—a new kind of noise not encountered in ordinary radio—which would still be an ultimate limitation for a laser radar even if the other two sources were completely eliminated. Fundamental in nature, it is caused by the random rate of arrival of the light photons (basic unit of light energy) which make up the received signal. Noise of this kind can be understood in a qualitative way by recognizing that at least one photon must be detected at the receiver during the time interval when a signal is expected. It represents an ultimate minimum detectable signal power, and is likely the phenomenon which will be the working limitation for some of the optical methods of the future.

Typical Gaseous Laser Radar

This device combines a gas-laser light source with a scanning principle used in the early days of television. The laser radar can scan targets angularly and produce a C-type oscilloscope display, resembling a television image. The primary advantage of a laser light source over conventional incoherent sources is the higher resolution, together with higher energy density available at the target. The narrow optical bandwidth enables suppression of ambient light, thus improving the noise level.

The present experimental stage of development includes a helium-neon laser. As shown in figure 11-33, the beam is reflected successively by two rotating mirrors which give it a scanning ability in the x and y directions, and the mirrors then focus the beam onto the target. The reflected light is picked up from the target by a multiplier phototube, amplified, and displayed on a cathode-ray oscilloscope. Two oscillators provide the x and y scanning signals for both the rotating mirrors and for the CRT display.

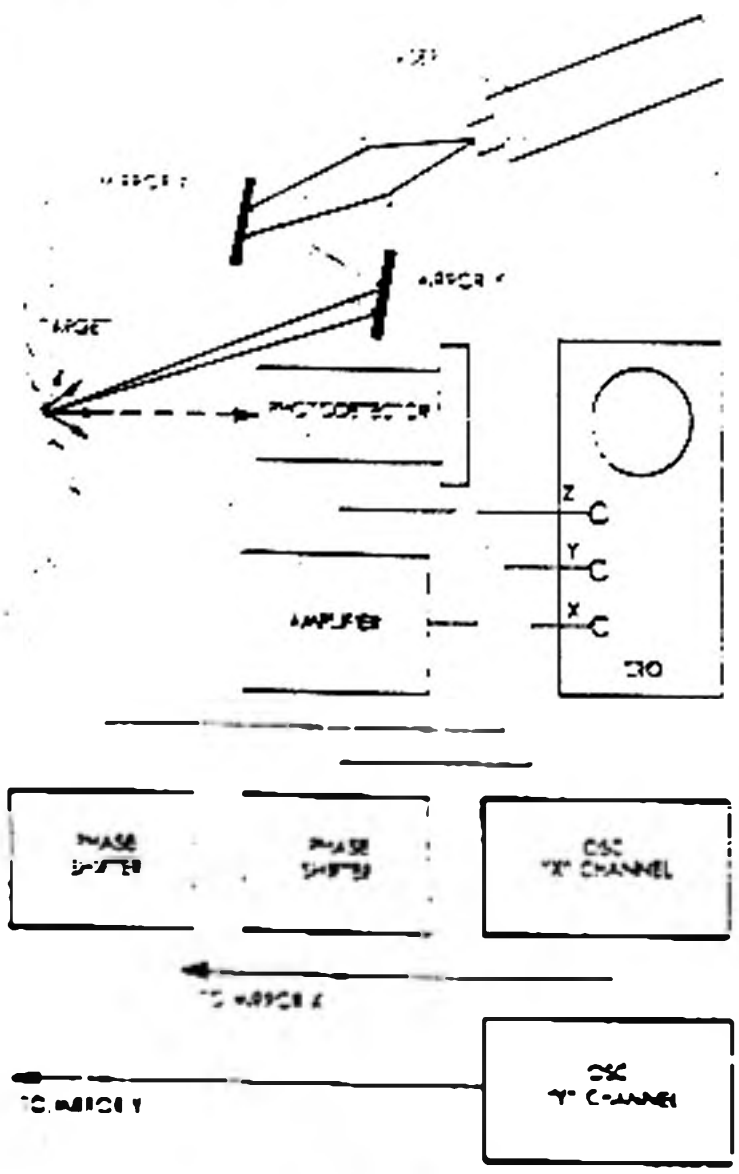


Figure 11-33. Block Diagram of Gaseous Laser Radar Set

The transmitter laser is an unmodulated helium-neon laser, radiating at 6238 Angstroms, providing a CW output of under 1 milliwatt. The beam is taken from a small circular spot on the plane mirror of the laser (operating in the hemispherical configuration), and focused by a lens to converge on the target. The beam diameter at the lens is 1 mm, narrow enough to be completely intercepted by the small beam-deflecting mirrors which measure 3 x 4.5 mm. Spot size at the target is about 2 mm.

Each mirror is mounted on a galvanometer element, giving an angular displacement of ±1 degree at 500 Hertz. The two mirrors are oriented so that the laser beam scans horizontally because of the motion of mirror x, and vertically because of the motion of mirror y. Each mirror is driven by its own sine-wave generator; consequently, the beam traces out on the target a Lissajous figure appropriate to the two movements.

The deflated beam is bounced off a variety of targets, including a photograph and an electric motor, at a distance of about 2 meters giving a target area of 50 cm².

After reflection from the target, the beam is filtered by a 6238-Angstrom interference filter to suppress ambient light. The horizontal and vertical sweeps for the oscilloscopic display are supplied by the same oscillators used for deflecting the mirrors, with a phase shifter in each line to compensate for the mechanical lag of each galvanometer movement.

As the laser beam scans the target, the backscattered light becomes intensity-modulated, depending on the average reflectivity of the particular target spot. After collection and amplification, this signal is used to intensity-modulate the z-axis of the oscilloscope.

Both oscillators run at about 500 Hertz, but, because their outputs are sinusoidal rather than linear, they provide nonuniform illumination of the target. Range contrast is limited by multiplier phototube saturation, and possibly by nonlinear response of the oscilloscope's P20 phosphor to variations in beam intensity.

LASER GYROSCOPE

Counter-rotating coherent light beams using a newly developed technique provide sensing of rotation rate with respect to an inertial frame of reference. The light beams are produced in a traveling-wave laser that forms a closed circuit. Called a ring laser, the device holds promise for inertial guidance methods that would be more stable, have less drift, be simpler, cheaper to produce, and be more sensitive than present methods using gyros.

If the ring laser were to be mounted in a space vehicle that is not rotating relative to the stars, the photodetector would see no difference in frequency between the two light beams. Even the tiniest rotation of the vehicle would cause one beam to travel slightly farther than the other around the ring. The frequency of the two beams would then vary slightly, and the difference would be proportional to the rate of rotation. Measurement of the rate could then be used to redirect

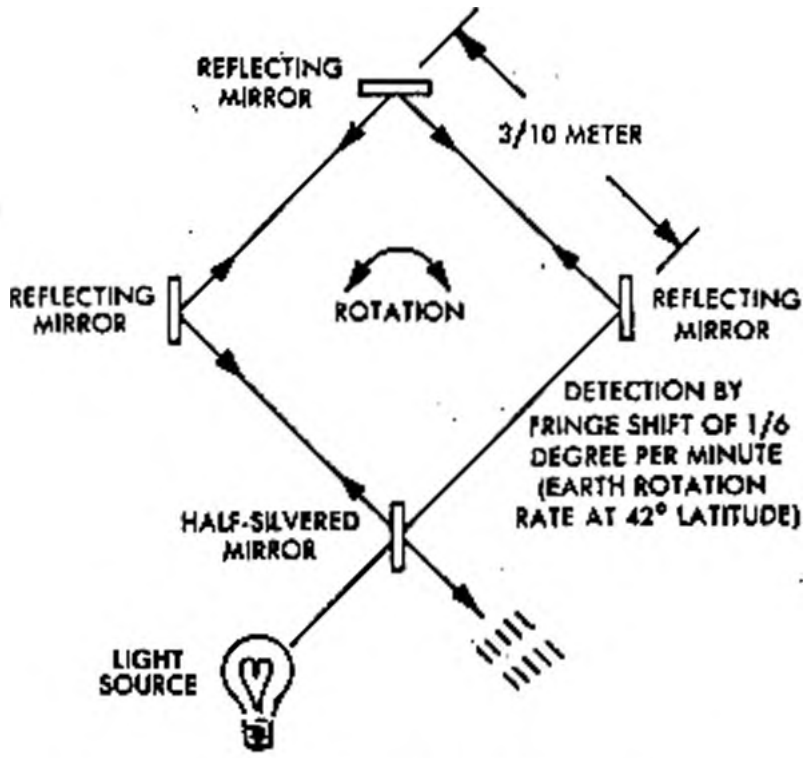


Figure 11-34. Sagnac Experiment Diagram

the attitude of the vehicle to maintain its correct flight path. The angular displacement of the vehicle, or how far it has moved from its correct flight path, would also be measured by the ring laser.

Development of the ring laser has roots in optical experiments conducted a half-century ago. In 1914, the French physicist, G. Sagnac, used a ring configuration approximately one meter square to form an interferometer (see figure 11-34). His device has poor sensitivity, and was barely able to detect a rotation rate even as fast as 43,000 degrees per minute. His interferometer formed fringes, and he discovered that the fringes shifted slightly sideways when the loop was rotated.

In 1925, two American experimenters set up a loop that had better sensitivity, but was physically huge. (See figure 11-35.) The loop, employing light-tight, evacuated pipes, measured two-fifths by one-fifth of a mile. The two experimenters were able to detect the rotation of the earth, approximately one-sixth of a degree per minute, by observing the fringe shifts which were proportional to the rotation of the earth. The loop was stationary on the surface of the earth and could not be speeded up or slowed down. To obtain an independent reference of fringe shift, the physicists set up a smaller loop inside the larger loop, and compared the magnitude of the two fringe shifts with the areas of the two loops to show that the constant proportionality was caused by the earth's rotation.

The ring laser consists of four helium-neon gas tubes, each about three feet long. These are positioned to form a square, with Brewster angle windows and four external-corner mirrors. The square ring composes the cavity of the laser so that light beams traveling around the circumfer-

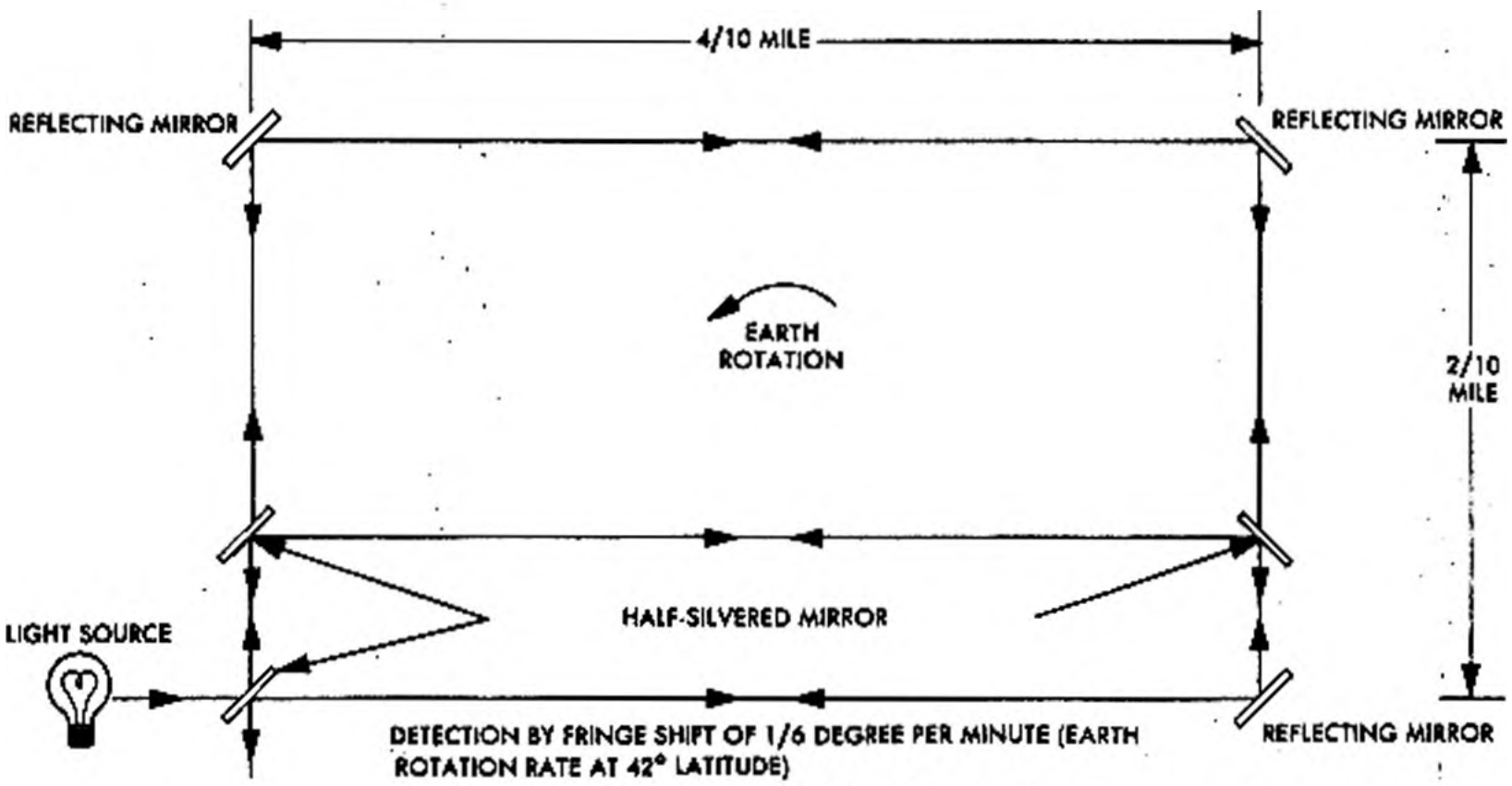


Figure 11-35. Optical-Loop Experiment Diagram

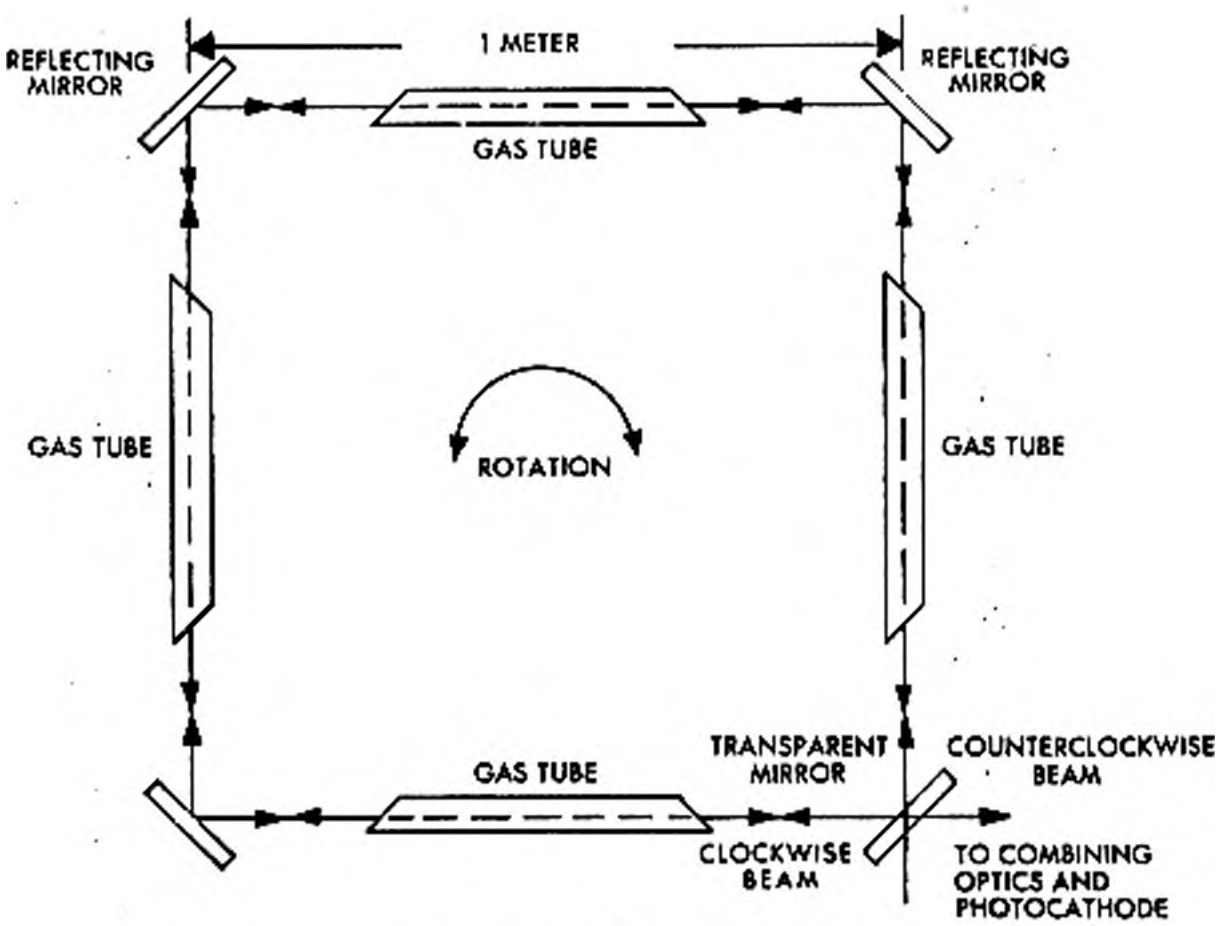


Figure 11-36. Ring Laser Rotation Rate Sensor Diagram

ence oscillate independently of each other. Their frequencies are determined by the length of the paths they travel.

When the ring (or its mounting) is rotated slightly, one light beam travels a slightly greater distance than the actual circumference to close on itself; the frequency of this beam is lowered. The opposite beam must travel a slightly shorter distance, so its frequency is raised. The two beams are mixed by optical heterodyning in a silver-cesium-oxide photodetector. The resultant beat frequency is proportional to the rotation rate of the ring.

Output of the photodetector can be integrated by circuitry that clips the sine waves, differentiates them, and then counts cycles. Thus rate output is a measure of angular displacement, accurate to within 0.015 degree.

The equipment illustrated in figure 11-36 uses a meter-square ring laser oscillating at the helium-neon 1.153-micron wavelength. Excitation of the gas is accomplished by 27-MHz, RF energy. Beat frequencies ranging from 500 Hz to better than 150 kHz were observed in the first experiments, with corresponding rotation rates of 2 degrees per minute to 600 degrees per minute. These are not the physical limits of the sensing effect, but are the limits set by equipment and techniques in the first demonstration setup.

Ultimate limits of sensitivity have not yet been determined, but developers are working to bring the 500-Hz sensitivity down as close to zero Hz, or as near zero drift as possible. The range of sensitivity is adjustable by varying the size and form of the ring.

The beat frequency for a square ring $\Delta f = \Omega L/\lambda$, where Ω is the angular velocity of rotation, L is the ring perimeter, and λ is the wavelength. Maximum sensitivity is obtained when the axis of rotation of the sensor is perpendicular to the plane of rotation of the ring.

Present self-contained rotation sensors are gyroscopic, and depend on conservation of the angular momentum of a rotating mass. This mass is subject to acceleration and gravitational forces. The laser rotation-rate sensor is free of these mass effects. It is a motionless strapped-down unit that requires neither bearings nor other mechanical moving parts. The simple construction of the sensor promises a low unit cost on a production basis.

Since the output of the ring laser sensing unit is basically digital, rather than analog like the output of a gyroscope, instrumentation is simple. The

beat frequency can easily be clipped, differentiated, and counted. Instantaneous readout of both angular velocity in Hz and angular displacement in total cycles is available. The output is much more adaptable to instantaneous readout than the angular-shaft rotation output of a gyroscope.

LASER SAFETY HAZARDS

The introduction and ever increasing use of laser equipment has brought with it a number of new and hidden dangers not previously recognized. Information on these dangers to health must be disseminated to all personnel associated with laser equipment.

Biological Effects of Laser Radiation

The body organ most sensitive to radiation is the eye. The ability of the eye to refract the near ultra-violet, visible, and near infra-red radiation is the most important physiological characteristic contributing to the laser hazard. The energy density of the radiation incident on the cornea may be increased by as much as several hundred thousand times at the retina by the focusing power of the eye. This can cause a major cataract hazard. If one looks directly into a laser beam, eyesight could be permanently destroyed.

Organs lying close to the skin are also susceptible to damage by laser radiation. Needless exposure of the skin to laser radiation should be avoided regardless of power density or energy density levels.

Protective Equipment

Protective equipment should provide a safe environment for personnel with minimum interference to the accomplishment of the mission. Each particular laser system will require personnel protection specifically correlated to the operational parameters of the system.

Current laser filters are designed to filter a narrow band of wavelengths and usually offer protection for only one laser system.

EYE PROTECTION. Filters can be worn in spectacle frames, goggles, and visors. They can also be fitted in windows and viewing ports. The application will dictate the weight and thickness of filter allowable. Four-millimeter-thick glass is generally considered maximum for most goggle applications.

All protective eye wear should be identified with the laser wavelength for which it is intended to be used, the optical density of that wavelength, and the visible light transmission.

SKIN PROTECTION. Should skin protection be required, heavy white cloth, such as a standard laboratory smock, reduces the exposure by at least a factor of 100 over the spectrum from the near infrared to the near ultra-violet. Gloves should be used to protect hands. Shielding and complete enclosure of the beam is preferable to personal protective devices and should be accomplished whenever possible.

Medical Surveillance Program

The medical service is responsible for the initial and periodic physical examination of military and federal civilian employees who work within or approach nearer than the SEED (safe eye exposure distance) of an operating laser system.

General Safety Precautions

There are some hazards common to all laser systems and other hazards that are specific to certain types.

EYE EXPOSURE. Do not look into primary laser beams or at specular reflections. Special precautions should be taken to avoid accidental viewing of direct beams or specular reflections when power or energy densities exceed the PEL (permissible exposure level).

BEAM TERMINATION. Laser beams should be properly terminated by suitable targets or backstops that have no specular reflections and are fire resistant.

HIGH VOLTAGE. Many lasers employ high voltages. Proper safety precautions should be taken for high voltages as you would with any electronic equipment.

TOXIC CHEMICALS. Toxic gases are often produced as a result of high energy laser beams ionizing air or disintegrating the target. Procedures for dealing with toxic materials are detailed in AFMs 127-100 and 127-101 (specifically chapter 5).

CRYOGENIC MATERIALS. There are many hazards associated with cryogenic liquids. Contact with

the skin will produce a burn. Asbestos gloves should be worn when filling or pouring from Dewar and thermos bottles. The bottles should be protected by an outside container should the bottle implode.

BUDDY SYSTEM. As with all hazardous systems, no personnel should enter a laser hazard area unaccompanied. Use the "Buddy System."

Only a few of the safety topics have been brought out. Persons involved in laser systems should read and adhere to AFM 161-8, Laser Health Hazards Control.

CHAPTER 12

Instrumentation

A simple instrumentation system might be a voltmeter, a piece of paper, and a pencil. In this case, you could record a voltage level and its polarity. The voltmeter measures the amount of voltage in an electrical component to which it is attached and displays this information to the operator as a pointer deflection. To preserve the information for later use, the operator makes a permanent record of the voltage measurement by recording the measurement on paper.

As shown above, the voltmeter performs two functions: it measures a quantity and displays the measurement. When a measurement must be made in a hazardous or inaccessible place, these two functions can be accomplished separately. In these cases, the actual measuring is done by an instrument or device known as a sensor or transducer. The output of a transducer is a signal which is a known function of the physical quantity being measured.

A large and complex ballistic missile has an instrumentation system incorporated within its airframe because a wealth of data is needed for study and correction of its operating characteristics. Since outward visual observations of its in-flight behavior are not enough and because personal observations are dangerous and inadequate, a system of various instruments is designed to measure different kinds of missile data. Such measurements change too rapidly for the human mind to recall and record. Besides, the accuracy of human observations leaves much to be desired since it cannot be matched against a time-base scale.

In simplest terms, instrumentation can be defined as instrument engineering. Such engineering entails the inventory, designing, making, and applying of instruments or instrument systems to observe physical phenomena. Observations which result in scientific analyses to discover the cause or properties usually require a wealth of information, primarily in the form of measurements. The measurable quantities or properties of a phenomenon can include mechanics, matter, sound, heat, electricity, and light.

These properties are the basic and classical divisions of physics. Because of the natural differences of these properties, various kinds of instruments or devices are needed to obtain phenomena information or data. By incorporating additional instruments into the data collection system, various data signals can be measured, monitored, conditioned, transmitted, received, converted, and recorded. Such a system is known as a radio telemetry system or simply as telemetry.

Telemetered raw data, when received at a missile range ground tracking station, are changed to usable forms suitable for immediate visual display. Thus, various kinds of preflight, inflight, and terminal data can be shown on different colored lamps, meters, or scopes. This type of data is known as real-time data; that is, information about an event as it is occurring at the moment. Real-time data are essential for monitoring the trajectory of a test missile by range safety personnel and for predicting the missile impact point. Provisions are also available at missile test ranges to store telemetered data for use at a later time when exhaustive and precise analyses are made of a missile test. Raw telemetry data are stored by means of magnetic tape recordings, or other methods of data storage such as plots or core storage, which make it possible to retrieve the desired information later.

During various phases of missile research and

development programs, a vast amount of data is needed in such areas as prelaunch, launch, meteorology, flight, and guidance. Once the missile has been launched on its predetermined trajectory, measurements by means of transducers and telemetry systems may be divided into the general classes which follow.

- Propulsion.
- Flight sequence.
- Environmental.
- Attitude and trajectory.
- Special.

All of the above measurement data are transmitted to ground stations along the missile range by telemetry systems. Many of the measurements are interrelated. Some require continuous samplings along with a high order of time resolution, while for others only a few samplings per second are necessary. For a large missile such as the Saturn, the telemetry system must be capable of handling and transmitting an extremely large amount of varied data each second of its flight.

To illustrate this, look at the information-gathering capability of the instrumented ship stations being used on the Eastern Test Range. A large ballistic missile that can achieve a speed of 20,000 mph is in receiving range of the ships for about 2 minutes. In these 2 minutes from 10 to 20 million bits of information in digital form are expected to be received and recorded. After being recorded, these millions of information bits are relayed back to Cape Kennedy.

SECTION A

External Instrumentation Systems

The largest and most complex external instrumentation systems are found at the missile ranges. The external instrumentation includes the various equipments whose primary functions are to track a missile along its trajectory.

Tracking of guided missiles undergoing tests is generally accomplished by optical and electronic instruments and devices which are not usually installed as regular missile components. However, to aid ground installations in tracking missiles along their trajectories, it may be necessary to install radar beacons or transponders, smoke generators, or photoflash cartridge ejectors within the test missile. All the devices are considered part of an external instrumentation system even though they are carried aloft by the missile.

Typical external instrumentation for a missile range, such as Atlantic Fleet Weapons Range, is shown in figure 12-1. Note that both fixed ground installations as well as ocean-going instrumented ships and instrumented aircraft are used as missile tracking stations. The enormous length of the Atlantic range together with the possible ranges of various missiles require mobile tracking stations in such ships. The versatility of these ships enhances the effectiveness of a missile test range in a number of ways. First is their ability to extend range operations by tracking guided missiles with intercontinental range. Another is their aid in nose cone recovery operations as well as confirming the location of the impact area. Finally, these ships have the ability to furnish tracking data in areas between widely-spaced ground stations. With their comprehensive instrumentation systems, the ships furnish meteorological data for the geographical areas in which they may be assigned. They complete the weather picture along the intended trajectory of a test missile.

Ground stations along the range also gather meteorological data in their immediate areas and transmit them to the launch area. Here the data are considered and entered into the computations of a guided missile trajectory. Meteorological data also aid in determining the actual launching time through the prediction of good weather so important for good tracking results from optical instrumentation. Weather data also are used at a later time when the performance of a test missile is evaluated.

In chapter 8, some atmospheric conditions that affect missile trajectories are described. The most important of these conditions is the winds which are produced by differences in atmospheric pressures and which are influenced by temperatures. A particular type of wind, known as the *jetstream*, must be taken into consideration when computing a missile trajectory. Weather units, through their instrumentation, can detect the presence of a jetstream. It is important to detect, evaluate, and monitor this stream of air (whose speed is known to have reached 310 miles per hour) because the jetstream represents a rapidy moving force which can cause a long-range missile to drift away from

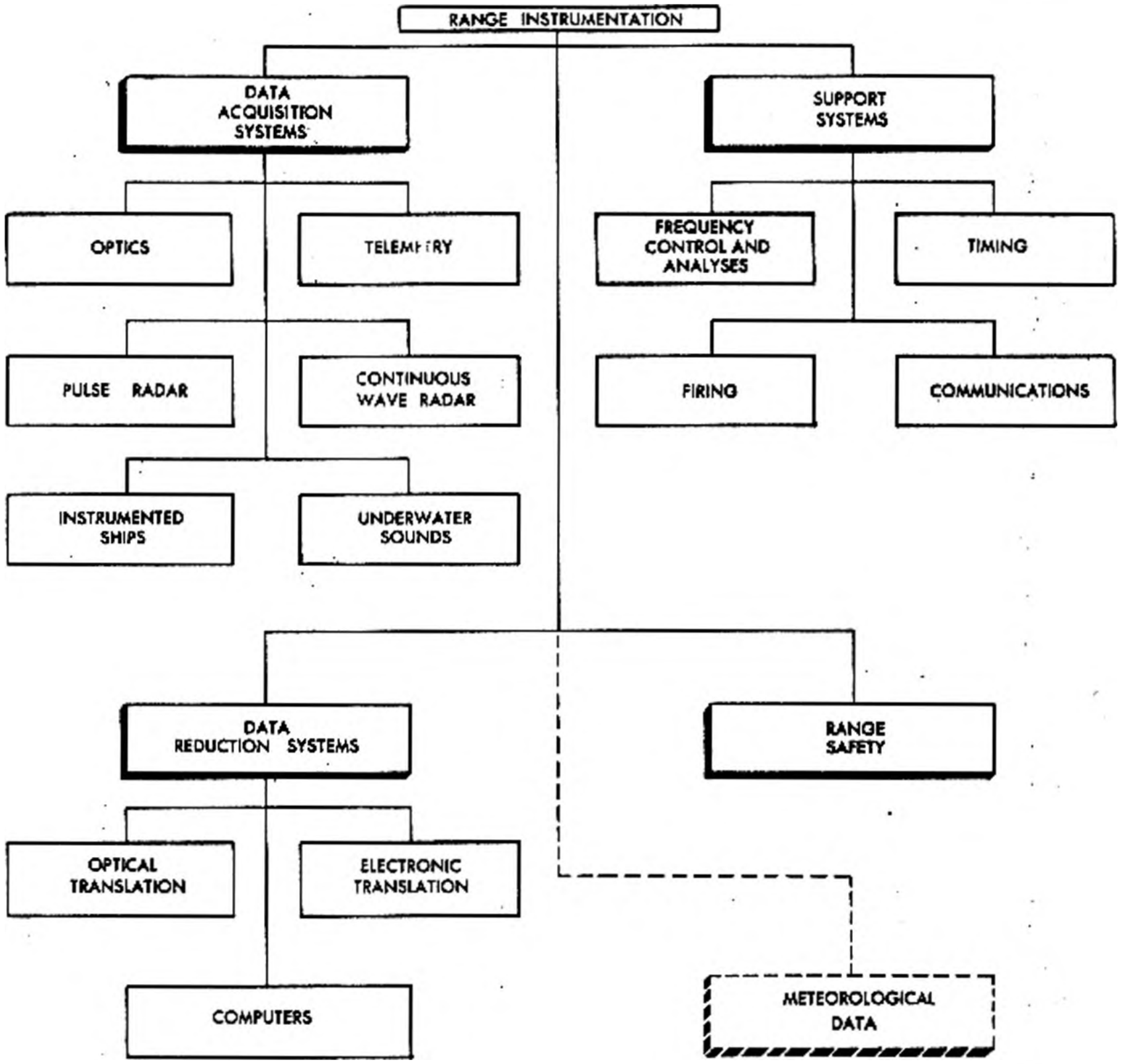


Figure 12-1. Typical Missile Range Instrumentation

its planned trajectory. To a lesser degree, the same situation is true as the missile passes through winds that are present at other altitudes.

Optical instrumentation is ideal to record trajectory drift by observing the trail of smoke leaving the missile or by photographing the reentry path of the missile nose cone. At predetermined intervals along the missile trajectory, a photoflash cartridge can be ejected upon command from a ground station. Once ejected, the cartridge produces a brilliant flash of light which can be seen over considerable distances. From a series of photographed light flashes, the missile trajectory, attitude, velocity, and altitude can be determined. Precise measurements are possible when optical data and electronic data are coordinated with each other, such as a radar beacon or transponder determining missile velocity and distance through Doppler radar principles.

OPTICAL EQUIPMENT

At a missile range, three basic metric optical instruments are used to obtain much of the neces-



Figure 12-2. Cinetheodolite

sary information about a missile trajectory. These instruments are the cinetheodolite, ballistic camera, and ribbon-frame camera. The first is an extremely accurate tracking instrument which is a combined surveyor's theodolite and motion picture camera. The other cameras are fixed installations; that is, they do not move to track the missile as it rises from its launching pad. As fixed cameras, the same field or area is repeatedly photographed.

The optical instruments are used as triangulation instruments to obtain the exact position in space of a test missile as it flies along a trajectory at specific instances of time. From this position data, it is possible to arrive at missile velocity and acceleration information. Both the fixed cameras and cinetheodolite can be used to obtain data about the missile attitude while in flight.

Cinetheodolite

Figure 12-2 shows the two operators of a cinetheodolite—one of whom tracks a test missile in azimuth (direction) while the other tracks in elevation. The camera portion of this instrument is pulse operated. Pulses from central timing stations activate a system of several cinetheodolites to assure that all theodolites "readout" simultaneously.

A theodolite consists essentially of a telescope mounted on standards so that is can be rotated vertically. The horizontal axis of the vertical rotation passes through the center of a vertical circle scaled for elevation-angle measurements.

The standards by which the telescope is supported are mounted on a calibrated base plate which can be rotated horizontally through 360°. The vertical axis of the horizontal rotation passes through the center of a horizontal circle which constitutes the azimuth scale of the theodolite.

In figure 12-3, you can see the optical and elec-

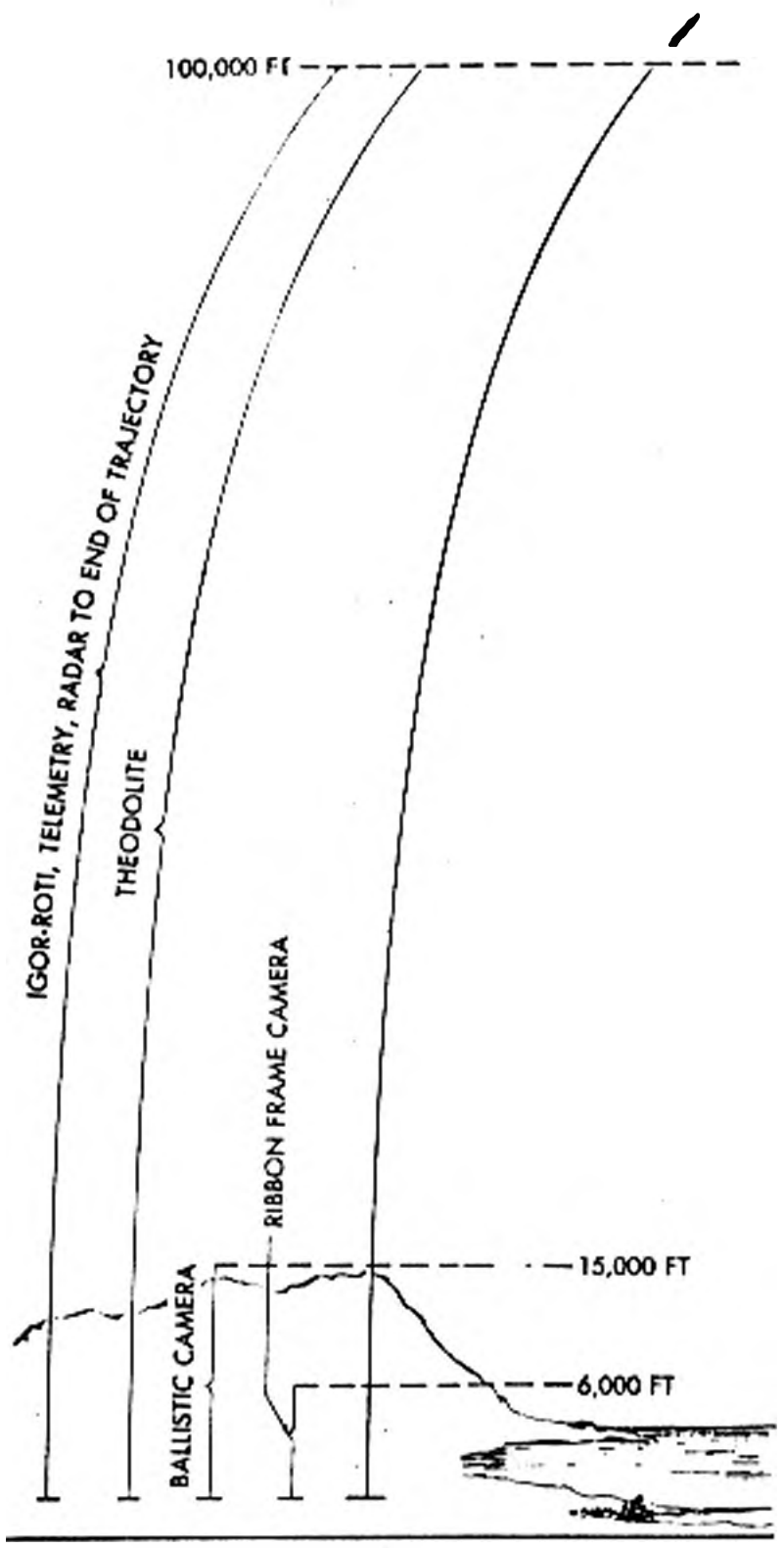


Figure 12-3. Typical Optical and Electronic Data Acquisition During Ballistic Missile Launch

tronic data acquisition systems coverage of a typical ballistic missile launch. Note that the cinetheodolite system provides secondary coverage for the fixed camera (ribbon-frame camera) system, and primary coverage of missile position and velocity data between the 2000- to 100,000-foot altitudes, during good photographic conditions.

Each exposure of the double 35-mm format (24- x 36-mm) records marks to indicate the optical center of the cinetheodolite 24-inch lens, and the azimuth and elevation dials of the instrument. In addition, range timing marks (in binary code) and frame count are shown in each exposure.

The cinetheodolite system is very accurate and reliable and permits coverage of missile trajectories up to 20 miles in good weather. Electronic systems have the advantage of being all-weather systems. In addition, instruments such as the beacons and radars are capable of giving instantaneous positions of the guided missile along its trajectory. The optical system can obtain precise information on a missile in the early portions of its trajectory where electronic systems have difficulty in tracking because of missile antenna configuration, flame effects, glint, and multipath.

Ribbon-Frame Camera

A ribbon-frame camera is one of the two fixed, metric camera systems. Unlike the cinetheodolite, which is a tracking optical instrument, the fixed or ballistic cameras record a field or area through which the trajectory of a test missile will be apparent. As with the cinetheodolite, two or more of these cameras are needed to photograph the same field from different angles. This is important since each camera must give the direction of a ray in space from the camera to the missile.

The fixed cameras are gimbal mounted on trailers or permanently fixed in towers. Like the cinetheodolite, the ribbon-frame camera uses rollfilm (5½ inches wide), but it is exposed at faster frame rates (30, 60, 90, or 180 frames per second). These cameras can be used at night, although for maximum accuracy of missile acceleration data, a special missile-borne, continuous light is necessary.

Ballistic Camera

The ballistic camera is more accurate with respect to position data. This optical instrument is used at night when a star background is first recorded to fix the trace of a missile exhaust or airborne flashing light with respect to position. When the required trajectory has been recorded, the camera shutter is closed until a post-calibration signal is received. Once again the shutter is opened to make the necessary star exposure. To the average person, the resulting glass plate negative would be meaningless since all that can be seen is a series of point exposures made by the flashing airborne light and short unbroken lines representing star tracks. Precise timing along with measured angles to the various stars (there are about 26 or 30 stars recorded on each glass plate) are used to determine the missile position in space at any desired time. The exposed glass plate negative represents raw data which must be reduced to a usable form by complex mathematical processes. This is where electronic computers enter the picture to maintain the extreme accuracy possible with the ballistic camera system.

The ballistic camera is able to obtain more accurate trajectory data than other instrumentation systems can obtain, and this fact is put to good use in calibrating such systems. This optical instrument is available with lenses of different focal lengths. The longest lens (300-mm, approximately 12 inches) has photographed trajectories pinpointed by brilliant flashes of light at ranges in excess of 1000 miles. This was made possible by photoflash cartridges ejected from the missile.

Other Optical Instruments

Although the three basic optical instrument systems just described obtain much of the needed missile data at a test range, other optical instruments are required to furnish radar backup and special data and coverage. These include tracking telescopes, recorder, infrared, and high-speed cameras. Figure 12-4 shows an example of such special coverage which was obtained by ROTI (Recording Optical Tracking Instrument). It is basically a tracking telescope with a lens whose focal length can be extended from 100 to 500 inches automatically. A 70-mm camera is attached to the telescope which is mounted upon a 5-inch naval gun turret platform.

Frame 1 of the sequence in figure 12-4 was taken of a Titan missile about 25 miles above

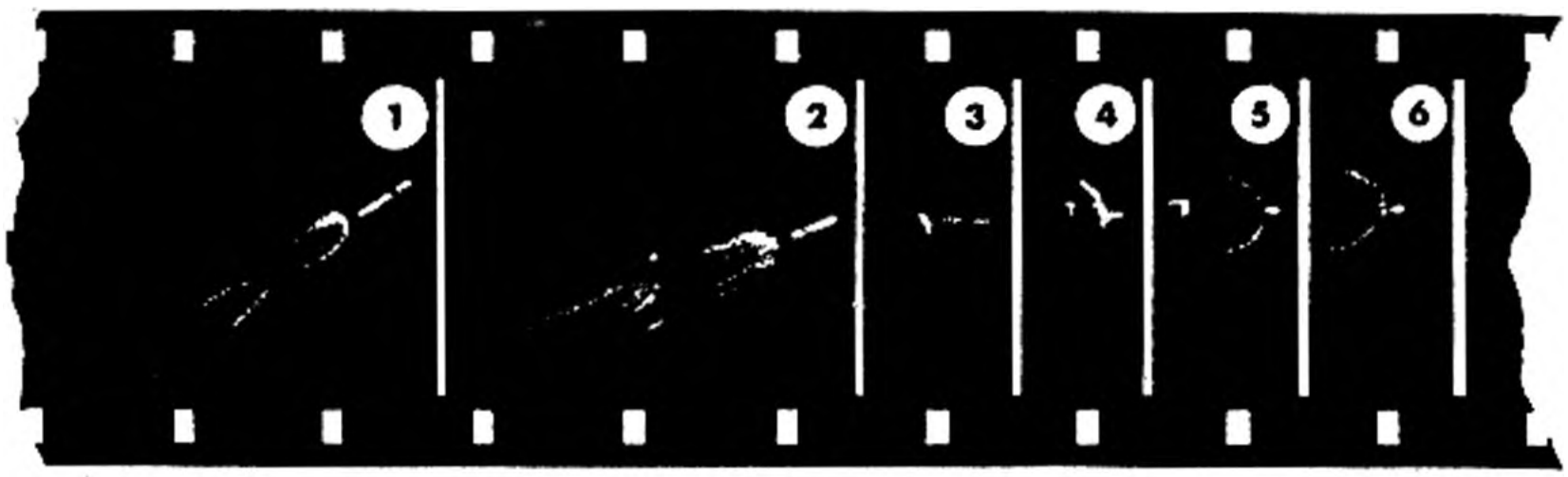


Figure 12-4. Recording Optical Tracking Instrument (ROTI) Sequence

Cape Kennedy, while frame 6 shows the last stage more than 52 miles away. Frame 2 shows the missile reaching its first stage burnout. Stage separation is recorded in frame 3, while frame 4 shows the ignition of the Titan's second stage.

During long-range flight observations, more than one optical observation station is required. For accurate data determination, the ground stations must be synchronized so that the data collected at separate stations can be consolidated into a single continuous recording.

RADAR EQUIPMENT

Radar equipment used in external instrumentation systems consists of several different kinds of both pulse and continuous wave radars. The principles of the two kinds of radar were given previously in chapter 9. Associated equipment includes a target acquisition system and radar data recorders which use magnetic and paper tapes.

Radar provides tracking information to meet range safety requirements. The range of the radar will vary in accordance with the two kinds of aerospace vehicle tracking—"skin" and "beacon." In the former, tracking is done by reflecting pulses from the outer skin of the vehicle. In the latter, the vehicle is tracked through signals emitted from a beacon (sometimes called a transponder or "parrot") installed inside the airborne vehicle. Beacon tracking increases the effective range of a radar with the amount of range increase being a function of beacon-radiated power. Radar provides range safety requirements in the form of real-time (instant information about an event), present position, and impact prediction data. It can also provide metric position of the missile from launch through burnout and, in some tests, throughout the entire missile trajectory. Figure 12-5 shows how tracking is done with the different radars available. Note the region of radar coverage overlap that occurs between station 1 at Cape Kennedy and station 3 at Grand Bahama Island.

External instrumentation of missiles requires a radar tracking and plotting system network which is capable of tracking any missile throughout its flight path and of presenting azimuth, elevation, height, and range position at any instant. Within the tracking and plotting system, there must be a data-transmission system to convey the tracking information to computers and/or plotters. The tracking data are resolved into permanent recordings of the flight path and flight characteristics of the missile by the computing and plotting system.

Radar systems used in missile instrumentation generally employ automatic tracking antennas which use some form of conical scanning to develop the positioning data voltages required for their operation. When the target is not in the center of the beam, error voltages proportional to the deviations in elevation and azimuth are developed in the antenna-positioning system.

These error voltages, or position-data voltages, when compared in phase or polarity with a reference voltage representing a fixed azimuth or elevation angle of the antenna, provide the azimuth and elevation data which indicate the instantaneous position of the target missile with respect to the radar antenna.

Typical Radar Set

A typical radar set which has been used in a

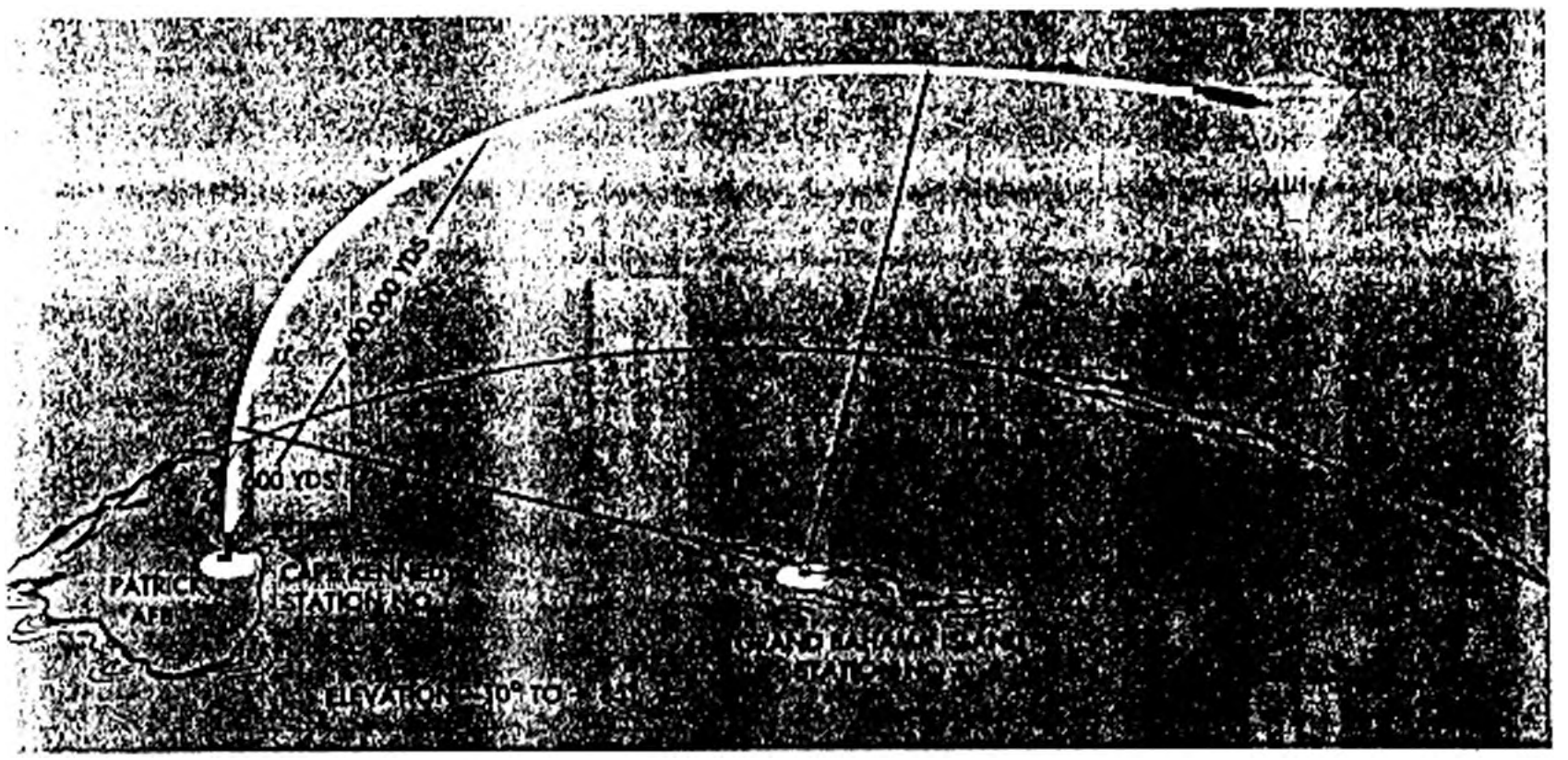


Figure 12-5. Typical Radar Tracking Coverage

modified form in missile instrumentation is the SCR-584. This set, as shown in the diagram of figure 12-6, is divided into the following seven systems: transmitting, radiofrequency, receiving, range, PPI, antenna, and data transmission.

The *transmitting system* generates RF pulses 0.8 microsecond (μsec) wide at a frequency of approximately 2800 MHz and with a peak power of 250 kw. The system receives a trigger pulse from the range unit every 586 microseconds, shapes the trigger (giving it the required form, width, polarity, and amplitude), and uses it to apply a DC voltage of 22 kv for 0.8 μsec to a magnetron.

The magnetron oscillates to produce RF pulses. The pulses pass through the RF system to the antenna and are radiated into space. The transmitting system consists of a driven unit which shapes the pulse and triggers a modulator. The modulator, in turn, acts as an electronic switch which applies the high DC voltage to the magnetron. A rectifier in the transmitting system produces the DC voltage from a 115-volt, 60-Hz, AC supply, and from the magnetron.

The *radiofrequency system* conducts the RF pulse from the magnetron to the antenna; and it radiates RF energy, receives target echoes, and conducts the echo signals from the antenna to the receiving system.

The system consists of an RF transmission line, the T-R box, antenna, and the reflector. The RF output of the magnetron in the transmitting system is fed into a coaxial transmission line which conducts the RF pulse to the T-R box. The T-R box prevents any appreciable amount of the pulse from entering the receiver, and it passes the pulse toward the antenna. The reflector gives this radiated energy its directional property by concentrating it into a narrow beam. The returning echo pulse enters the antenna and passes through the RF line to the T-R box, which shunts it to the crystal mixer of the receiving system. The spinner motor and the reference generator are in the antenna-positioning system.

The *receiving system* amplifies and detects the target echoes received by the RF system. The system consists of a crystal mixer, a local oscillator, a preamplifier, a superheterodyne radio receiver, and a remote video amplifier. The echo signal from the target is conducted through the RF system to the crystal mixer, where it is mixed with a signal from the local oscillator. The resultant intermediate-frequency (IF) signal is amplified in the preamplifier and the IF section of the receiver.

The IF signal then divides and enters the range and servo channels of the receiver. The signal is amplified in the range channel, then it is fed to

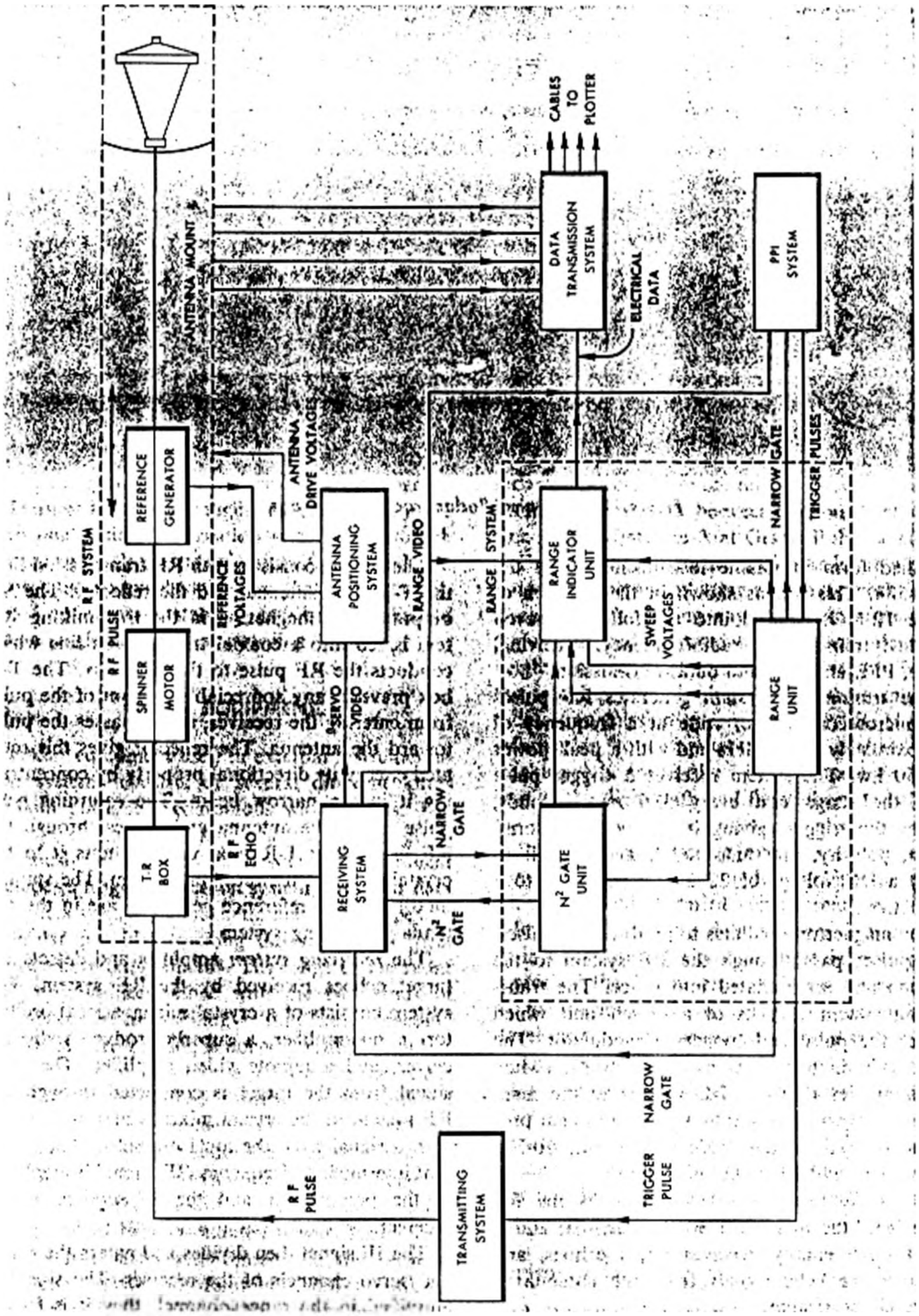


Figure 12-6. Typical Radar System

the scopes in the range and PPI systems. The servo channel is gated by the pulse from the N^2 (narrow-narrow) gate unit. As a result, the channel is opened up only long enough to receive the echo from the target being followed. In the servo channel, the error voltages corresponding to the echo being tracked are passed on to the antenna-positioning system.

The *range system* furnishes the timing and gating pulses and visual range data on oscilloscopes. The system also supplies range potentiometer and selsyn data to the data-transmission system. The range system consists of a range unit, a range indicator unit, and an N^2 gate unit.

The range unit generates the trigger pulses and sweep voltages for the set. It furnishes synchronized triggers to the transmitting system, the PPI system, and the other components of the range system. It also supplies a narrow gate to the PPI system, as well as sweep voltages and illuminating gates to the range indicator. Finally the master timing frequency and the narrow gate are supplied to the N^2 gate unit by the range unit.

The range indicator unit consists of two range scopes, tracking and slewing handwheels, and an added tracking mechanism. The range scopes are J-scan (circular sweep) scopes; their sweep voltages are generated in the range unit. All targets within range of the RF beam are displayed on the 32,000-yard scope (coarse range) which is illuminated for the duration of the wide gate. The 2000-yard scope (fine range) expands any 2000-yard portion of the 32,000-yard scope for greater tracking accuracy. Two gates are applied to the 2000-yard scope: the narrow gate and the N^2 gate. The narrow gate illuminates a sector on the 2000-yard scope, which may be manually varied by the narrow-gate width control from 250 yards to 1800 yards. The N^2 gate is applied to the 2000-yard scope so that the operator may track a specific target. Range potentiometers and selsyns supply slant range data to the data-transmission system.

The N^2 gate unit generates the N^2 gate from the master timing frequency after being triggered by the narrow gate. This N^2 gate is fed to the servo channel of the receiver and also to the range indicator. It appears on the 2000-yard scope as a slight increase in intensity, 50 yards wide, and can be seen at the top of the echo signal being tracked. The range operator tracks the target by keeping the range hairline at the leading edge and the N^2 gate at the center of the target being tracked.

The *PPI system* is used for general searching and provides the means of initially locating the target. The system includes a PPI unit, the PPI scope, and the PPI selsyn. All targets within the RF beam are displayed on the PPI scope together with the narrow gate and range-marker circles. By the proper adjustment of handwheels, the target is placed within the tracking gate.

The *antenna-positioning system* controls the positioning of the antenna. This system includes the antenna-position-control unit, the automatic tracking unit, the azimuth and elevation tracking unit, the azimuth and elevation motor-generators, the spinner motor, and the reference generator (two phase).

In automatic tracking, the error voltage is combined with the reference voltages, producing the desired control voltages for the drive motors. When manually positioning the antenna, artificial error signals, created by moving the azimuth and elevation handwheels, are combined with a reference voltage to produce the same result. A single group of circuits serves for all methods of antenna control, with the difference in the operation of the circuits depending on the nature and source of the error and reference voltages.

The *data-transmission system* converts the range, azimuth, and elevation information on target position into potentiometer (voltage) and selsyn (phase-position) data, and the system supplies the data to plotting boards. The system includes an antenna-position indicator; azimuth, range, and elevation potentiometers; selsyns; an attitude data unit; and an altitude converter. Coarse and fine selsyns transmit angular data for azimuth and elevation. Potentiometers develop range, altitude, and angular data for azimuth and elevation. The selsyns and potentiometers respond instantly to all changes in target position. Transmission of data is electrical through the use of control voltages developed across the potentiometers, or mechanical by means of the selsyns. Accuracy of the data supplied to the gun director or plotter depends primarily on the accuracy with which the set tracks the targets.

The mechanical (selsyn) and electrical (potentiometer) data obtained from the tracking radar

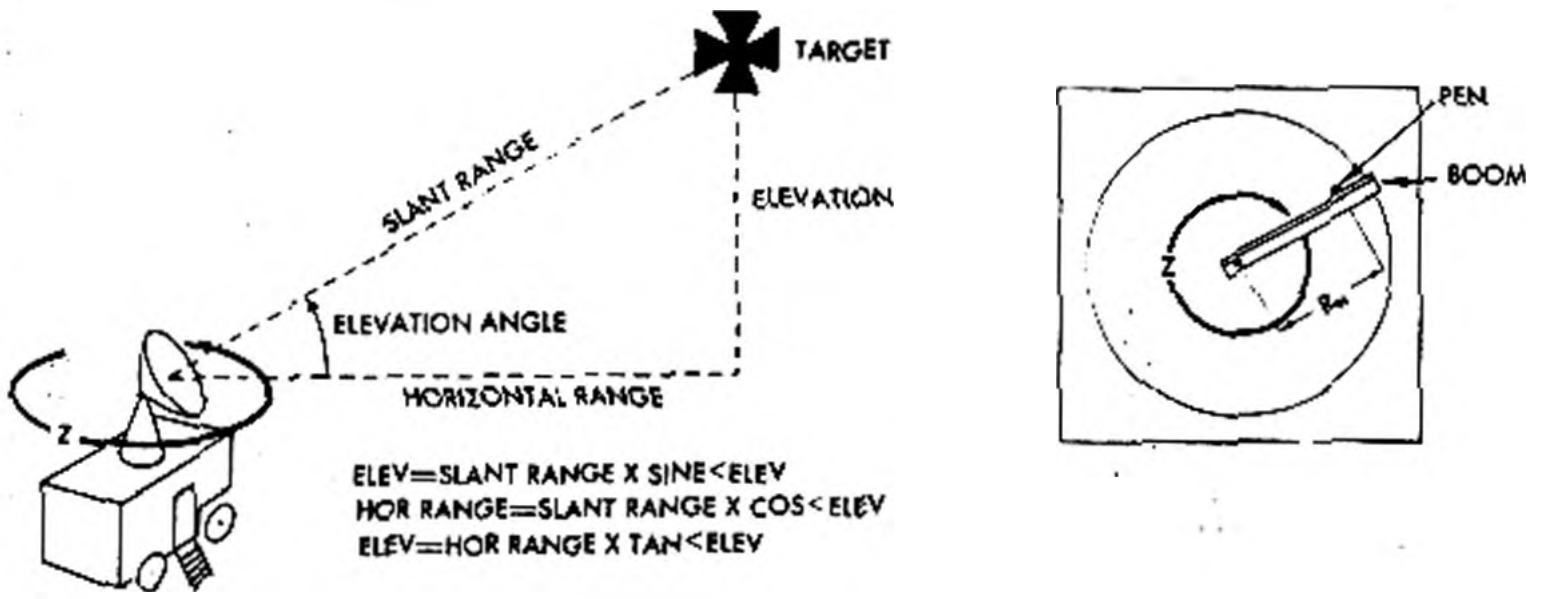


Figure 12-7. Basic Plotting System

are resolved into "X" (range) and "Y" (elevation) components. These components are proportional to the variations of the angular error voltage, "Z" (azimuth angle) data, with respect to the reference voltages. The "X" and "Y" components are converted into graphic form in the plotting system.

Basic Plotting System

An example of a basic plotting system is illustrated in figure 12-7. The plotting boom revolves around a center axis and is positioned by the azimuth (Z) data. The tracing-pen trolley is positioned by horizontal range data. The axis of the boom represents the radar location, and the pen position indicates the instantaneous location of the target. The horizontal range data used to position the tracing-pen trolley is derived from the slant range data in terms of the cosine of the elevation angle. Elevation data is in terms of the sine of the elevation angle and slant range data or in terms of the tangent of the elevation angle and horizontal range. If desired, the data are marked on the plotting chart at each target-position point.

Many improvements and modifications have been made in the basic radar tracking and plotting equipment to meet the requirements of the missile program. Automatic tracking over full range of the radar and greater flexibility of the antenna-positioning system have been provided in some modifications. Larger and more versatile plotters with integral auxiliary indicating and recording devices have been developed to meet the requirements for tracking high-velocity, long-range missiles. The same principles of operation which prevail in the new modified equipment are embodied in the basic radar tracking and plotting systems.

SECTION B

Internal Instrumentation Systems

An internal instrumentation system uses pickoff devices and a data transmitter located within the missile, and receiving and recording equipment located in a ground station. Internal systems are capable of supplying a much greater variety of data than can be obtained from the external systems and are not so restricted by conditions of weather, visibility, or terrain.

Whenever it is desired to collect data of a physical or mechanical nature and transmit it from a missile, the data must be converted into an electrical form. By making this conversion, data can be transmitted in the form of modulation of an RF carrier to a receiver located at the ground station.

This process is referred to as *telemetry*. The word is of Greek origin. It means measurement from a distance, and it may be defined as a method of measuring a quantity, transmitting the result to a distant station, and there indicating or recording the quantity measured. Telemetry includes conversion of quantities to be studied into electrical signals, the transmission of these signals over a radio link, their reception, and their presentation by indicators or permanent recordings.

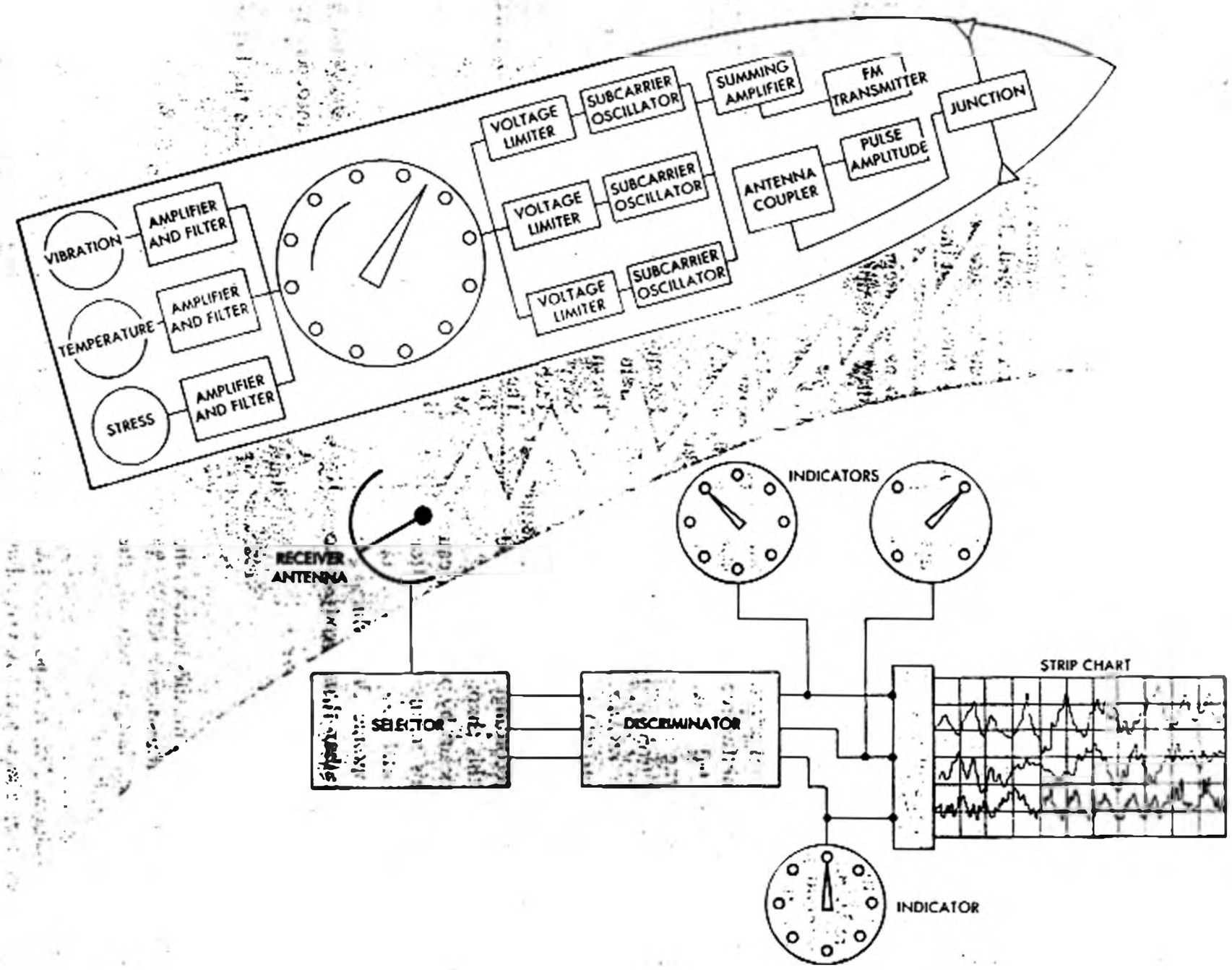


Figure 12-8. Radio Telemetry System

The quantities in this final form of presentation can be studied at leisure.

A simplified diagram of a radio telemetry system is given in figure 12-8 where components comprising an internal instrumentation system are shown. The jagged line, representing a stream of inflight data in the form of radio signals, serves as the connecting link between the internal instrumentation system of a test missile and the external instrumentation system of a ground station. At the ground station, telemetry data are changed into a usable form and shown on indicators or strip charts.

Another method of gathering data is through photographic or other recordings that are made aloft, but this method is not considered telemetry since there is no distance involved between the instruments and the recorders.

Formerly, the photographic recording of data collection could not be used effectively. This was true because of the doubtful recovery of photographic or tape recorded data. Now, instrumented capsules containing such recorded data, upon a command signal from a ground station, can be ejected from the missile and retrieved. This method of aerial retrieval was successfully used for recovering capsules which were ejected from the Discoverer series of earth satellites. Photographic and tape recordings are described in greater detail later in this chapter.

Television is another means by which data can be obtained; however, this means is more often used to collect environmental data; that is, cloud, terrain, storm, and astronomic data which are visible from a satellite rather than aerodynamic data of its carrier vehicle.

TRANSDUCERS

Internal instrumentation of missiles includes the use of data pickups or transducers which are attached to, or form an integral part of, the component under test within the missile. A transducer converts mechanical or physical variations into a desirable voltage. This voltage is used to modulate the carrier to a degree corresponding to the amount of change in the quantity being measured.

The transducers, if not part of an electrical circuit, are supplied with a reference-voltage source which is preset to some output value corresponding to the normal position or operating characteristic of the component under test. Any deviation from normal provides an output voltage which varies from the normal value by an amount corresponding to the amount of the component variation.

A transducer measures or collects the data at its source. The nature of the data to be collected from the source determines the type of pickup or transducer employed. The terms pickup, pickoff, and transducer are used more or less interchangeably throughout the missile field. Whenever possible, a transducer is built into the component under test as an integral part to avoid external loading, drag, or shunt effect of the pickoff device on the component.

Small, lightweight, highly accurate transducers have been especially designed for telemetry systems. A few of these transducers which are commonly employed in missile telemetry systems include:

- Potentiometers (position data).
- Strain gages.
- Thermistors.
- Selsyns.
- Tachometers.
- Thermocouples.
- Thermostats.
- Manometers.
- Bourdon tubes (airspeed).
- Diaphragms or bellows.
- Vanes.
- Gyros.

A few of the more common transducers are illustrated in figure 12-9.

The transducers are used to obtain the following telemetered data:

- Altitude.
- Static differential pressure.
- Low airspeed.
- High airspeed.
- Vertical acceleration.
- Electronic chamber pressure.
- Hydraulic pressure.
- Plenum-chamber pressure.
- Nozzle pressure.
- Attitude.
- Control surface position.

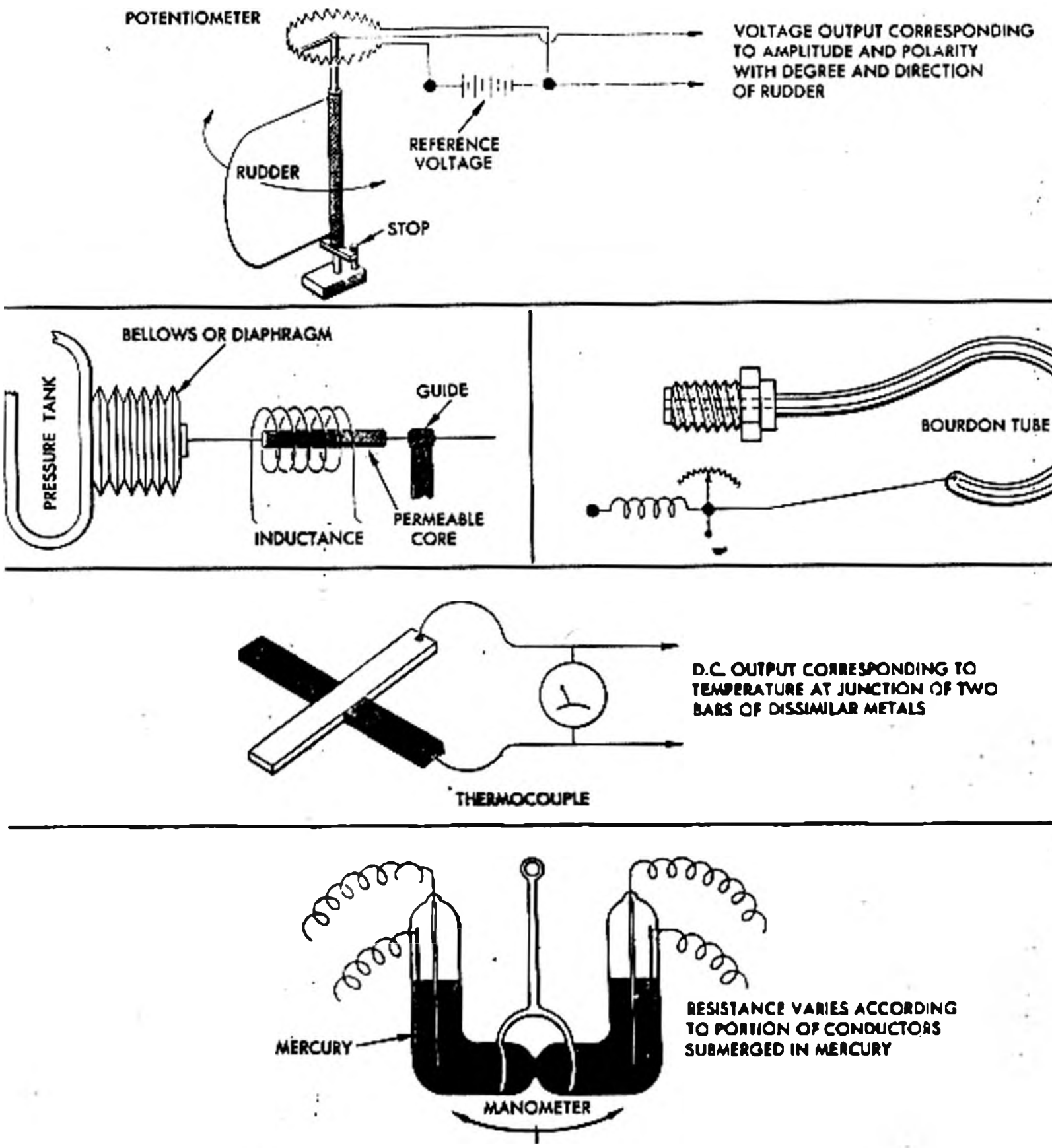


Figure 12-9. Common Transducers

Missiles and satellites require transducers of all types to collect in-flight data. Figure 12-10 illustrates how some transducers are used and some of the locations of their installation. Many of the transducers are made more rugged to withstand the tremendous extremes of acceleration, temperature, pressures, stresses, and vibrations usually encountered in missiles.

The multitude of transducers in use at the present time allows only a general coverage of the field. Transducers may be classified according to whether they are modulating or generating in action. Examples of the modulating type of transducers are the variable inductance, variable capacitance, and variable resistance. Variable-resistance transducers include potentiometers, re-

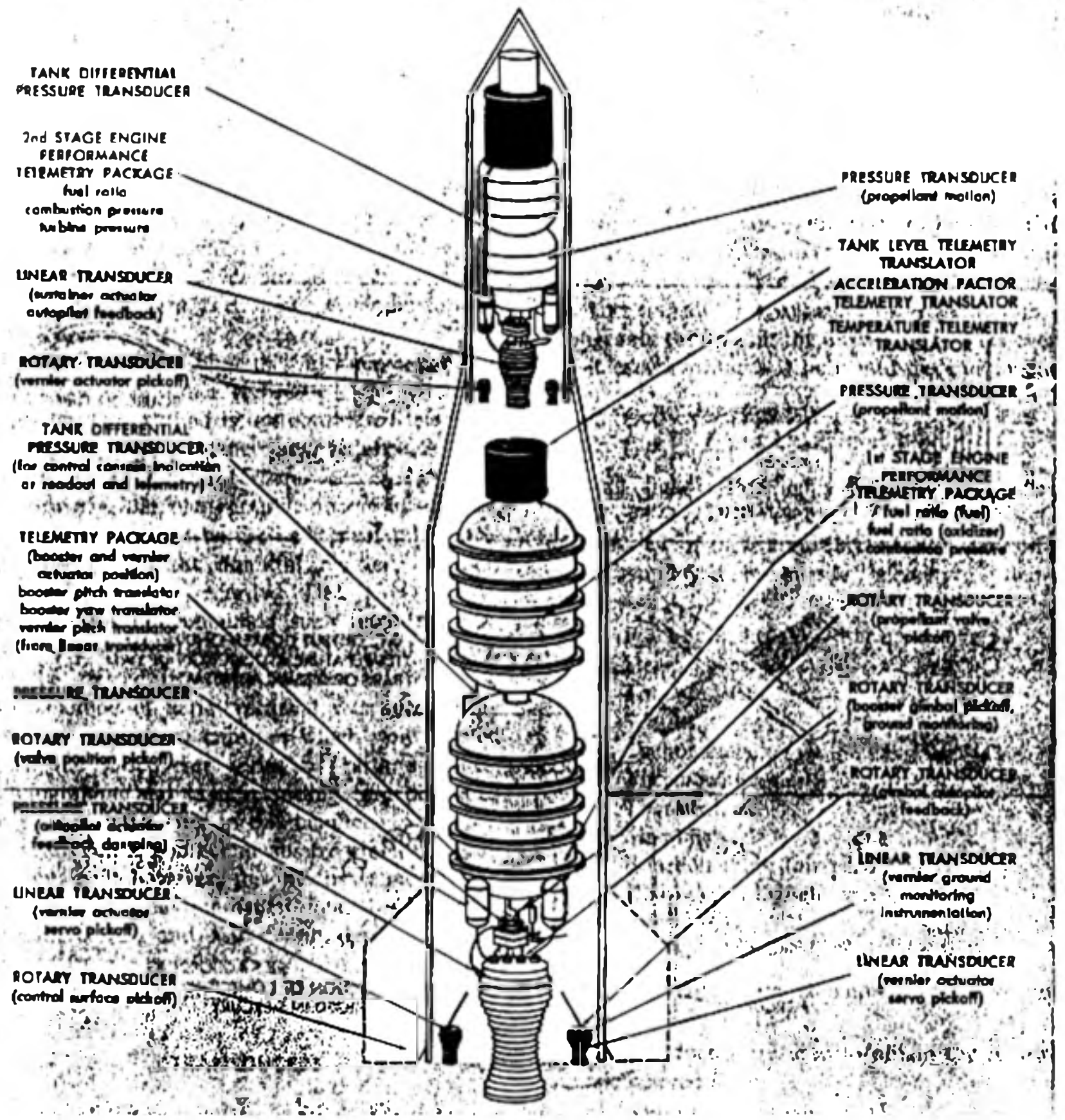


Figure 12-10. Representative Applications of Transducers

sistance strain gages, thermistors, transistors, and electron tubes.

Generating types of transducers include photoelectric, thermoelectric, and piezoelectric devices. The generating type is not always capable of completely modulating the designated circuit of the transmitter, nor is the modulating type always capable of completely performing this function. In such cases, their outputs are further amplified.

Examples of the stages and circuits which are necessary to make pickup transducers measure data adequately are shown in figures 12-11 and 12-12. These examples of pickup units may be considered as simple but complete instruments of which hundreds are needed in a test missile.

An essential step is the calibration of each in-

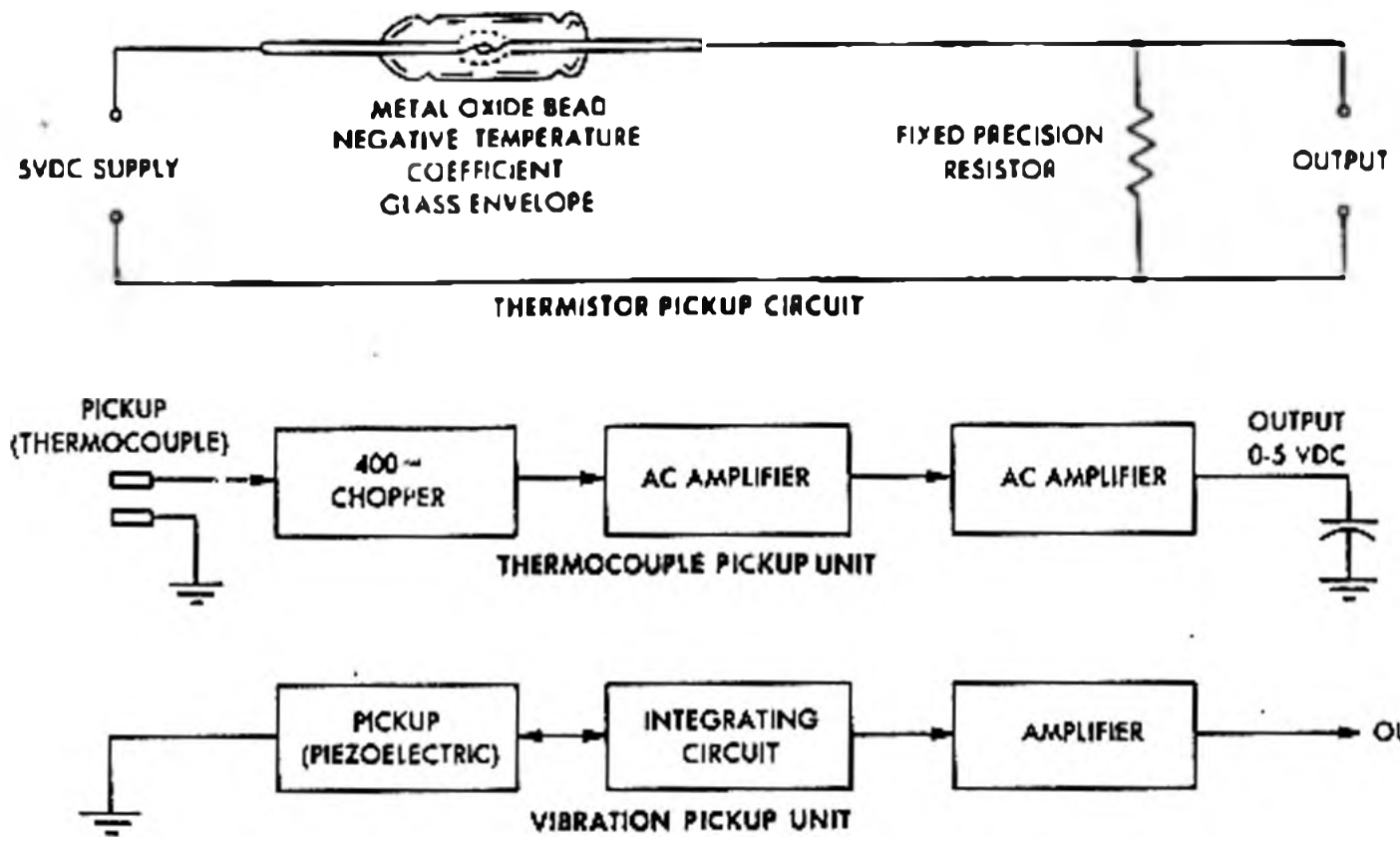


Figure 12-11. Examples of Various Pickup Units

strument with its attached transducer. Frequently, standard components such as bridges, potentiometers, and other instruments are combined with the appropriate pickup or transducer. In such cases, the transducer must not impair the operation of the instrument. Where an accurate indication of certain data measurement is desired, the combined instrument and transducer unit must be specifically designed for this purpose.

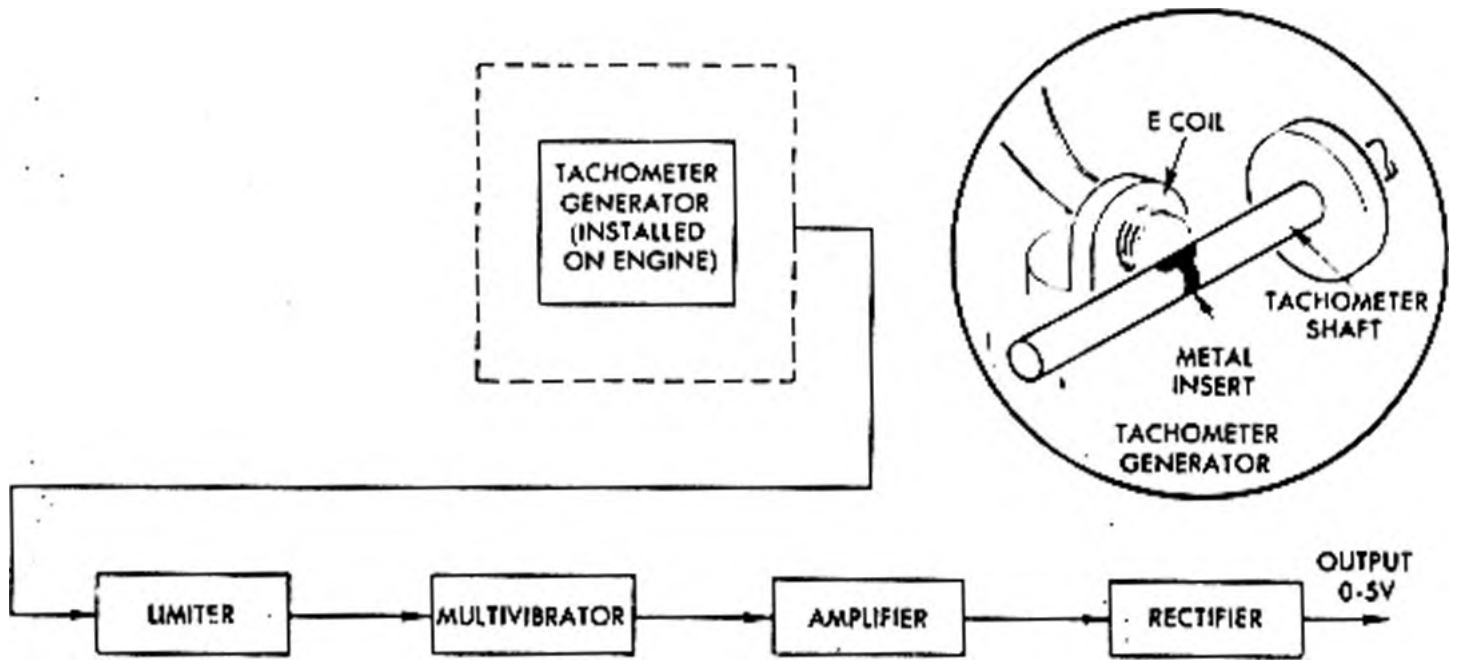
Another method is to couple a variable-reluctance type of transducer to the instrument by attaching a small piece of magnetic material to the needle or moving part. The torque then necessary to drive the needle is extremely low. This setup is best fitted to work into a frequency-changing arrangement such as results when the reluctance of a frequency-controlling inductance is caused to vary with motion. Usually, this is the tank coil of an audiofrequency oscillator. It is also possible to obtain the change in frequency from a large change in airgap with increased spacing of the metal. Such transducers are excellently adapted for attachment to such instruments as accelerometers, altimeters, airspeed indicators, pressure gages, tachometers, and various displacement meters.

One type of transducer easily adapted for use with tachometers consists of a segment of mu metal (metal of very high permeability) imbedded in the rotating shaft of the instrument and so located that as the shaft rotates the mu metal moves in and out of the field of an inductor which is a component of an oscillatory circuit. An example of this type of transducer is shown in figure 12-12.

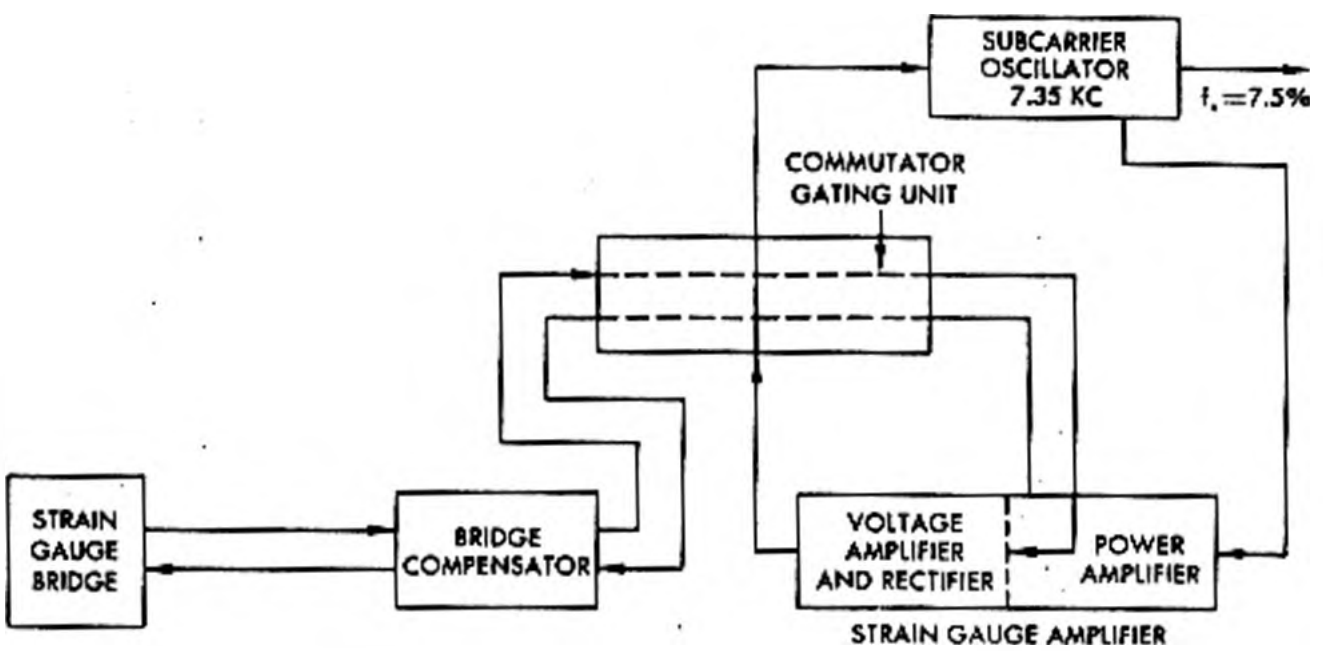
This arrangement varies the inductance of the coil and produces a change in the frequency of the oscillatory circuit at a rate corresponding to the rpm of the shaft. Mu metal is used in this application because of its freedom from hysteresis (lag in magnetic change), which is common in most magnetic materials and which might introduce errors.

The mu metal and E-coil transducer offers little drag on the rotating device. It can be used as a miniature alternator whose frequency and amplitude of output voltage depend on the rpm of the shaft and the spacing between the metal segment and the coil.

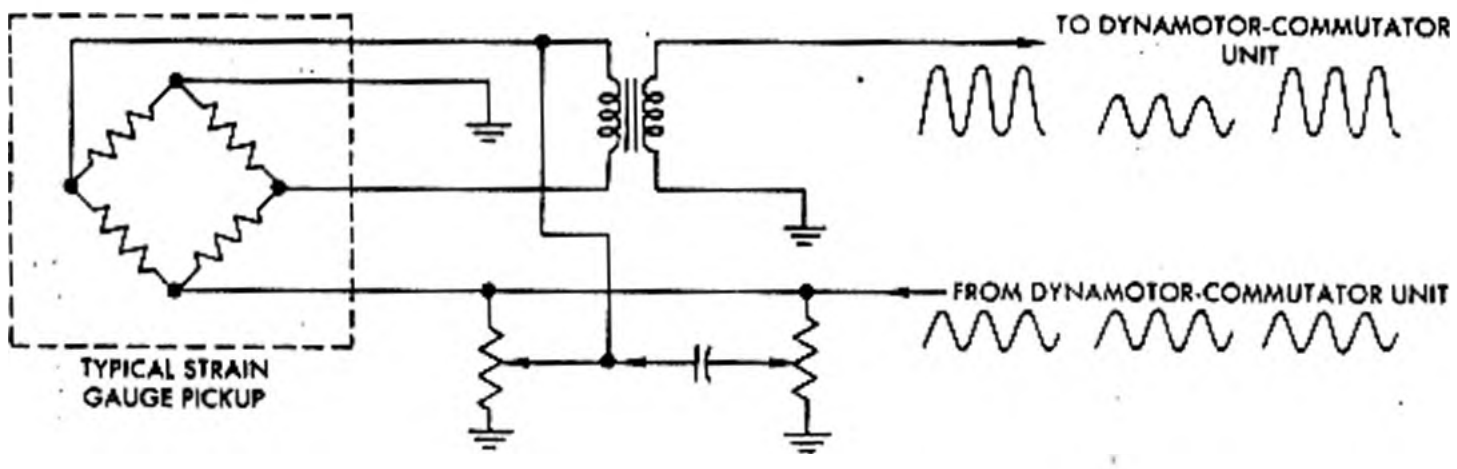
Another typical transducer assembly consists of linking the mechanical-data source to the wiper arm of a linear potentiometer which is supplied with a reference voltage from a reference battery. In operation, an output equal to the reference voltage represents maximum mechanical or pressure action, and an output of 0.0 volts represents minimum action. As mechanical action is initiated or pressure developed, the output voltage rises



TACHOMETER AMPLIFIER UNIT



COMMUTATED STRAIN GAUGE SYSTEM



BRIDGE COMPENSATOR

Figure 12-12. Types of Transducer Units

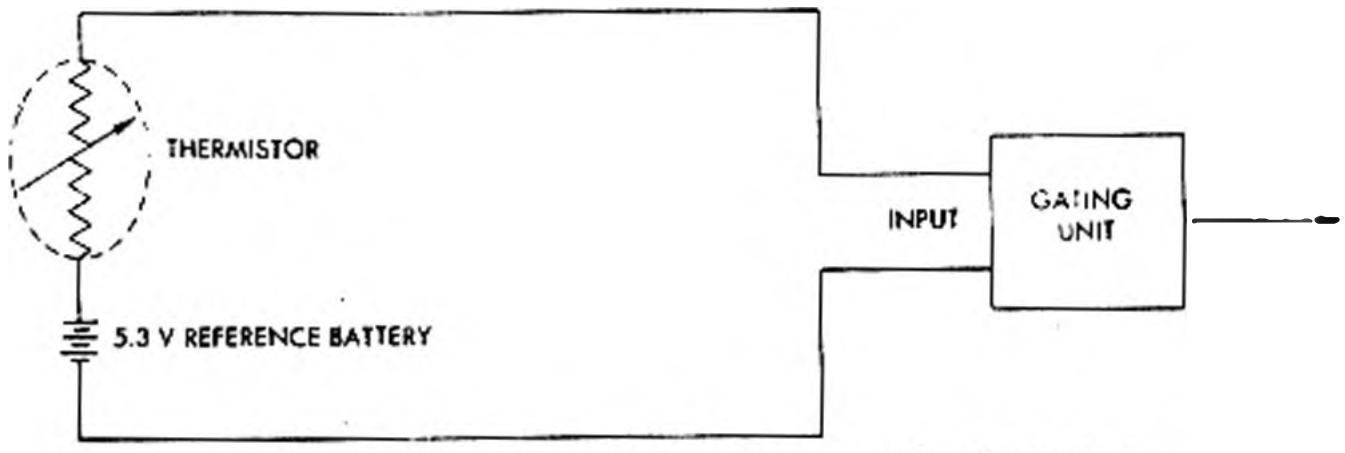


Figure 12-13. Thermistor Circuit Connected to Gating Unit

from 0.0 volts to a positive voltage in proportion to the mechanical action.

Thermistors, consisting of combinations of ceramic materials and various metallic oxides, are commonly used to detect and respond to fluid- or air-temperature changes.

Air-temperature thermistor elements, generally located on the center and aft equipment shelves in a missile, are mounted in such a manner that air can circulate freely around them. The thermistor element is connected in series with a reference battery and its assigned gating-unit contact. The diagram in figure 12-13 shows a thermistor circuit and its connection to a gating unit. As the thermistor temperature varies, the resistance varies; therefore, the voltage output to the gating unit is linear with the temperature variations. An increase in temperature decreases the resistance of the thermistor, producing a smaller voltage drop across it. This action, in turn, causes a higher voltage to appear across the unchanging impedance of the input to the gating unit.

Thermistors are particularly useful in missile fueling systems where narrow ranges of measurements at extreme temperatures must be covered. This coverage normally varies from −365° to −338° F. At these extreme temperatures, thermocouples or resistance thermometers do not have enough output voltage even after amplification.

Strain gages are variable-resistance elements so designed that any mechanical movement or stress applied to them causes a corresponding variation in their resistance. They are made in many forms and types to suit application requirements. A common type of strain gage consists of an adhesive tab, shown in figure 12-14, in which one or more loops of fine resistance wire are imbedded.

FM/FM TELEMETRY SYSTEMS

A widely used telemetry system currently employed in missile instrumentation is the frequency-modulated subcarrier, frequency-modulated carrier type, referred to as FM/FM.

This system is commonly used because it can be designed with inherent freedom from external signals (noise) which are amplitude variations. Also, conventional equipment and circuitry can be used; space, weight, and power requirements are not extreme. Ruggedness and simplicity are attainable with a relatively high degree of accuracy

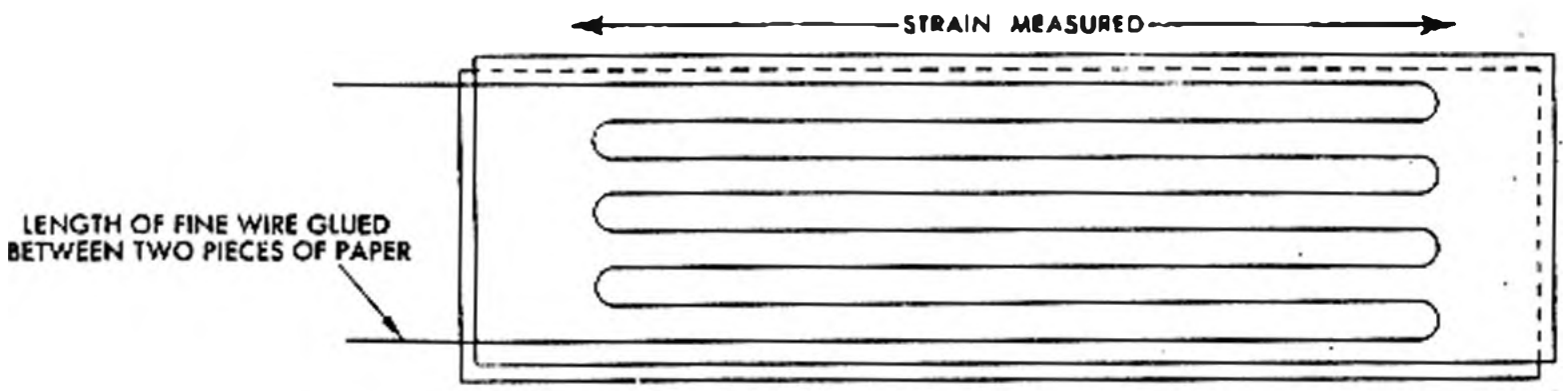


Figure 12-14. Strain Gage

and adaptability to missile telemetering requirements. Ease of calibration and a minimum amount of ground-station equipment are important factors, as well as versatility in handling many different types of data.

The underlying theory of operation is similar to FM broadcasting. Frequency modulation of the subcarriers also has an advantage in minimizing interference from other electrical circuits within the missile. In an FM/FM system, amplitude variations from the data pickoffs are converted into frequency variations by the transmitting system. These frequency variations are converted into voltage data of varying amplitude in the receiving system. The frequency deviations of an FM carrier are not affected by noise or static from sources outside the system because such types of interference, while tending to vary the amplitude of the carrier, affect its frequency only slightly.

In designing an FM/FM telemetry system, emphasis is placed on frequency stability of the sub-carrier oscillators even though ambient conditions such as temperature, humidity, acceleration, or power-supply voltage vary as the flight progresses. Remember that any drift in sub-carrier frequency would appear as a change in the function being telemetered.

A typical FM/FM telemetry system includes the following airborne components:

Transducers.
Subcarrier modulator units (one or more).
Transmitter.
Frequency-converter units.

The ground station for an FM/FM system generally includes the following components:

Receivers.
Subcarrier discriminators.
Recorders.
Test equipment.
Calibration equipment.
Maintenance equipment.

When an airborne transmitter must collect data from a large number of sources in the missile, additional subcarrier modulators can be added to the transmitter to accommodate additional data channels. For data channels requiring continuous sampling or a high degree of time resolution, several additional subcarrier modulators are needed. For data of the slowly changing type, a single subcarrier modulator with its input commutated from a number of end instruments or channels may suffice. Additional commutators and subcarrier modulators can be added if required. If any are added, additional subcarrier filters and discriminators must be provided at the ground station.

If a sufficient number of research missiles are available, the best practice is to telemeter a small number of channels, thus insuring the highest degree of accuracy. When many data sources must be sampled from a single missile, the possibility of interference by cross-modulation in data channels necessitates careful shielding of cables and components and careful calibration of both transmitting and receiving equipment. Even so, the varying loads on the power-supply and transmitter components may introduce false indications.

Generally, when feasible, certain types of data are collected from one missile and other types from additional and identical missiles. Some data may be in simple form which requires only a few samples per second to produce a complete record of operations under various flight conditions.

Many types of intelligence are gathered when a test missile is fired; and after careful evaluation of all data, specific types of intelligence are filtered out, analyzed, and either set up for repetition or replaced by more desirable information from the next missile.

Some of the types of information that can be telemetered in missile research are as follows:

Throttle setting.
Engine rpm.
Ambient temperatures.
Compressor inlet temperature.
Nozzle temperature.
Structure temperature.
Altitude.
Airspeed.
Induction air pressure (total).
Induction air pressure (static).
Nozzle pressure (total).
Nozzle pressure (static).
Fuel pressure.
Fuel-flow rate.
Power-supply voltages.
Hydraulic pressure.

Groundspeed.
Generator voltage supply.
Airspeed signal input.
Yaw error.
Pitch error.
Roll error.
Missile control-surface position.
AGC voltages.
Fuze arming.
Radio-control relays.
Terminal-dive signals.
Airspeed and altitude relays.

The data above are only samples of the types of information which can be transmitted from an airborne missile. While the types of data may vary from test to test, the number of channels that are available to transmit data is limited to the number of data channels in the telemetry system being used.

Some transducers used in FM/FM systems and examples of the intelligence they provide are as follows:

Potentiometers—rudder position.
Selsyns—roll signal.
Tachometers—engine rpm.
Thermocouples—nozzle temperatures.
Bourdon tubes—altitude.
Direct connection—dive signal.

An FM/FM telemetry system uses an FM RF carrier modulated by a number of FM subcarriers. These subcarriers may be continuously modulated by varying information or by commutated (sampled) information or a combination of these methods. The oscillator outputs are combined and amplified and then used to modulate the FM RF carrier of the transmitter.

The equipment is used for any application desired so long as its use is consistent with the design of the equipment. For example, transducers can be used to modulate subcarrier frequencies continuously with varying information. This permits the information from one pickup per subcarrier to be telemetered. Also, a number of subcarrier frequencies can be modulated using continuous information from pickups of any of the three types, while one or more of the higher subcarrier frequencies are modulated by commutated data inputs (each of a slowly varying nature) from other transducers.

Consider a typical missile FM/FM telemetry system. A functional block diagram of such a system is presented in figure 12-15. The system illustrated is designed to transmit data on six subcarrier frequencies. The sixth (highest-frequency) channel can be commutated at the rate of five samples per second for each of 27 data pickoffs.

Channels are modulated continuously when the intelligence is expected to vary more rapidly or when a high degree of data resolution is required. Systems employing higher commutation rates are available if it is desired to commutate information varying more rapidly.

The system in figure 12-15 includes the following components:

Subcarrier-oscillator unit.
Power-supply unit.
Fuel totalizer; engine-tachometer switching unit.
Cam calibrator.
Signal converter unit.
Voltage-controlled subcarrier oscillators.
Subcarrier-oscillator mounting unit.
Commutator-gating unit.
Transmitter unit; RF power-amplifier unit.
Antenna.

The order of presentation is not indicative of the signal sequence.

Pickoff data may be suitable for direct application to a voltage-controlled subcarrier oscillator, or may involve one or more of the other components before being applied to the voltage-controlled subcarrier oscillator.

A detailed treatment of these components and their operational functions follows.

Subcarrier-Oscillator Unit

The subcarrier-oscillator unit, illustrated in figure 12-15, is composed of voltage-controlled oscillators. The center frequencies, as shown in the diagram, are 2.3 kHz, 3.0 kHz, 3.9 kHz, 5.4 kHz, 7.35 kHz, and 14.5 kHz. The 6-channel subcarrier-oscillator output is coupled to the transmitter/exciter unit which drives the final RF amplifier.

In this FM/FM system, the data signals are coupled to the commutator-gating unit or the subcarrier-oscillator. Data expected to vary at a low rate are coupled to the gating unit. AC signals received from the missile controls and guidance

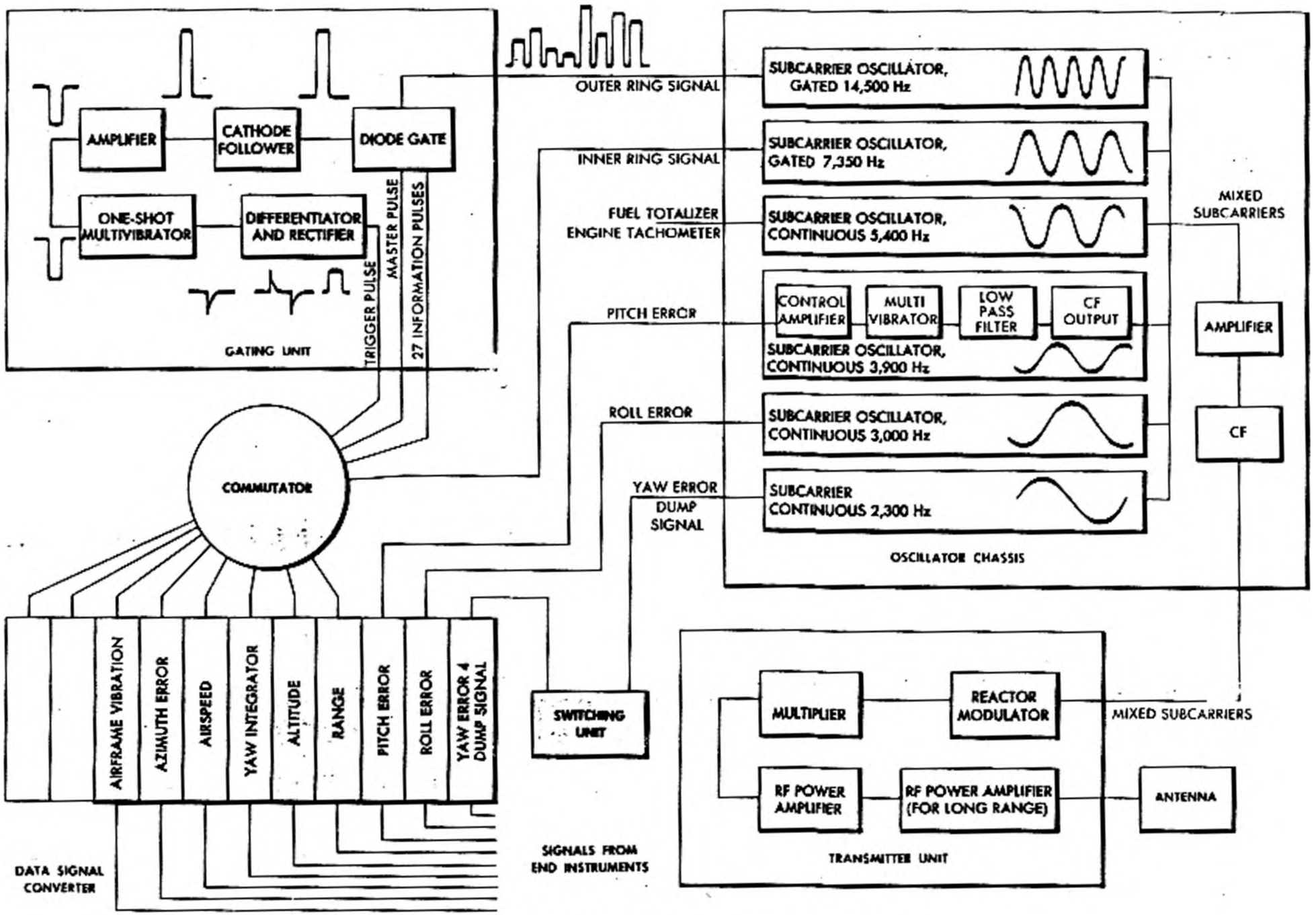


Figure 12-15. Functional Diagram of FM/FM Telemetry System

units are coupled through the signal converter's nine channels for conversion to DC signals.

Power-Supply Unit

The primary purpose of the power-supply unit is to furnish regulated voltages. The power supply varies in design depending on the particular instrumentation package used. A fairly simple power-supply unit such as would be used in the block diagram may consist of a number of batteries, while the power supply in the Apollo program is much more complicated. The Apollo capsule besides having a complicated power supply system also has backup power supplies should the main system fail.

As the power supply gets more complicated than a bank of batteries, voltage regulators, DC to AC converters, and current regulators are added. Some equipment needs steady DC voltage to operate properly. Because of the various loads imposed on the power supply, the DC from the battery will vary. Small DC motors can make the battery voltage change by as much as a volt or more. This voltage change is very serious, especially when the power supply voltage is only 6 to 12 volts as in many transistorized packages.

The DC to AC converters are needed whenever part of the instrumentation package must use AC power. Current regulators are becoming increasingly important in solid state devices. Transistors are current-operated devices, and many of the other solid state devices must also have the proper amount of current supplied. Within certain limits, the current is more important than the voltage.

Fuel Totalizer; Engine-Tachometer Switching Unit

The engine tachometer information is normally sent to the 5.4 kHz subcarrier-oscillator. Whenever a predetermined amount of fuel is used, the tachometer information is interrupted. When these data are read out on the ground, the interruptions of the tachometer information indicate the amount of fuel used. When designing this type of circuit, care must be taken that the information from the tachometer is readily recognizable from the information from the fuel totalizer.

There are several ways the information can be separated. The one used will depend on items available, designer choice, etc. If the output from the tachometer is a constant voltage level, the fuel totalizer signal could be a burst of two or three cycles from the oscillator. If the output from the tachometer is a sine wave, then the fuel totalizer burst could be a level input or even just an open input.

Cam Calibrator

The calibrator circuit consists of a motor-driven cam which actuates two microswitches for applying 100% information and 0% information signals alternately as flight calibration for telemetering. The motor is actuated either by a signal from the ground or at a sequence determined by the internal program of the package. Either way the calibrating signals are sent to all the ungated subcarrier-oscillators. On the gated subcarrier-oscillators, there are pulses set aside for calibration on every cycle of operation.

On many instrumentation packages, the calibration is entirely an electronic operation. In some instances, the calibration command triggers relays which stop the flow of information to the subcarrier-oscillators and substitute the calibration signal. In others, a diode gate is used to switch the subcarrier-oscillator input from information to calibration. No matter which way the calibration is accomplished, it is very important to the ground station. With the calibration sequence you can determine if the package is still calibrated to the original specifications. If the calibration has changed, the calibration sequence tells how much the calibration is off and how the apparent information must be changed to obtain the correct information.

Signal Converter Unit

The purpose of the signal converter unit is to convert transducer outputs to a proper signal for the subcarrier-oscillator. On many subcarrier-oscillators the proper input is from 0 to 5 volts. Also the subcarriers are quite limited to frequency input. The outputs from transducers are often too low or too high in voltage and too high in frequency.

For transducer outputs which are too high or too low in voltage, the signal conditioner is simply a voltage attenuator or amplifier. For transducers with too high of a frequency output, the frequency output is changed to a DC voltage. An example of this is the 400 Hz power supply monitor. The

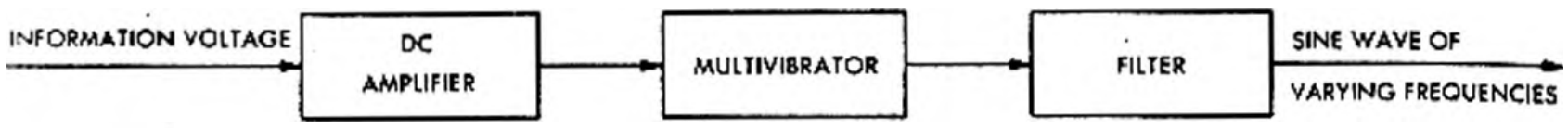


Figure 12-16. Subcarrier Oscillator

output of the signal converter may have an output of 0 to 5 volts for a frequency of 370 Hz to 430 Hz. In this case, an output of 2.5 volts would indicate the 400 Hz power supply is operating normally as far as frequency is concerned.

Voltage-Controlled Subcarrier Oscillators

The voltage-controlled subcarrier oscillators are voltage-sensitive. Their purpose, as indicated by the block diagram in figure 12-16, is to receive information in the form of a varying DC voltage and transform this voltage into a sine wave AC voltage, varying in frequency. The frequency variation is dictated by the information voltage.

As shown in the figure, the subcarrier-oscillator is basically a multivibrator whose frequency is controlled by the input signal. The output of the multivibrator is fed through a filter to get rid of the harmonics present in the multivibrator output. Subcarrier-oscillators for one system are the same for each subcarrier for the multivibrator frequency determining components and the output filter.

On instrumentation packages which conform to the Inter-Range Instrumentation Group (IRIG) standards, the subcarrier oscillators have definite center frequencies ranging from 400 Hz to 165 kHz. Depending on the center frequency, the channels can be deviated ±7½%, or on the higher frequencies ±15%. Because of the different deviations on different channels, this system is called the proportional bandwidth system. If the same bandwidth is needed for all channels, a constant bandwidth system can be used. In the constant bandwidth, the deviation is the same number of cycles regardless of the center frequency. Since changes occur frequently, consult the current IRIG Document 106-69, Telemetry Standards, for the latest specifications.

Subcarrier-Oscillator Mounting Unit

The block diagram in figure 12-17 shows the oscillators mounted on the subcarrier-oscillator mounting unit. The unit has two main functions in addition to serving as a mounting base: (1) to supply input power and connections for the six subcarrier oscillators and (2) to mix and amplify the FM output of the oscillators to frequency-modulate the RF transmitter.

The mounting unit provides sockets for firm seating of the oscillator units. It also provides a metal coverplate for clamping down the units securely to reduce the effects of vibration and acceleration.

The oscillator units are plugged into the sockets. An external cable from one pin of each socket provides for connecting each oscillator unit to its end instrument or data pickup.

The block diagram of the mounting unit shows the output signals from all of the oscillator units are fed through individual output potentiometers. The potentiometers are used for preemphasis adjustment. From here, the signal passes through RC decoupling networks to one common-output potentiometer for mixing. This common-output

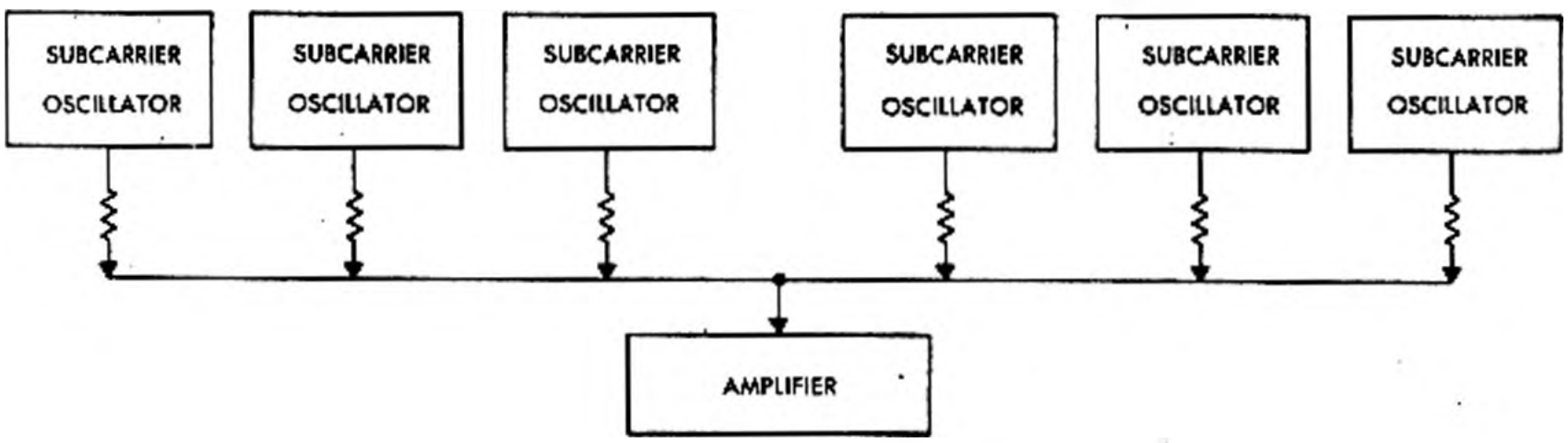


Figure 12-17. Subcarrier-Oscillator Mounting Unit

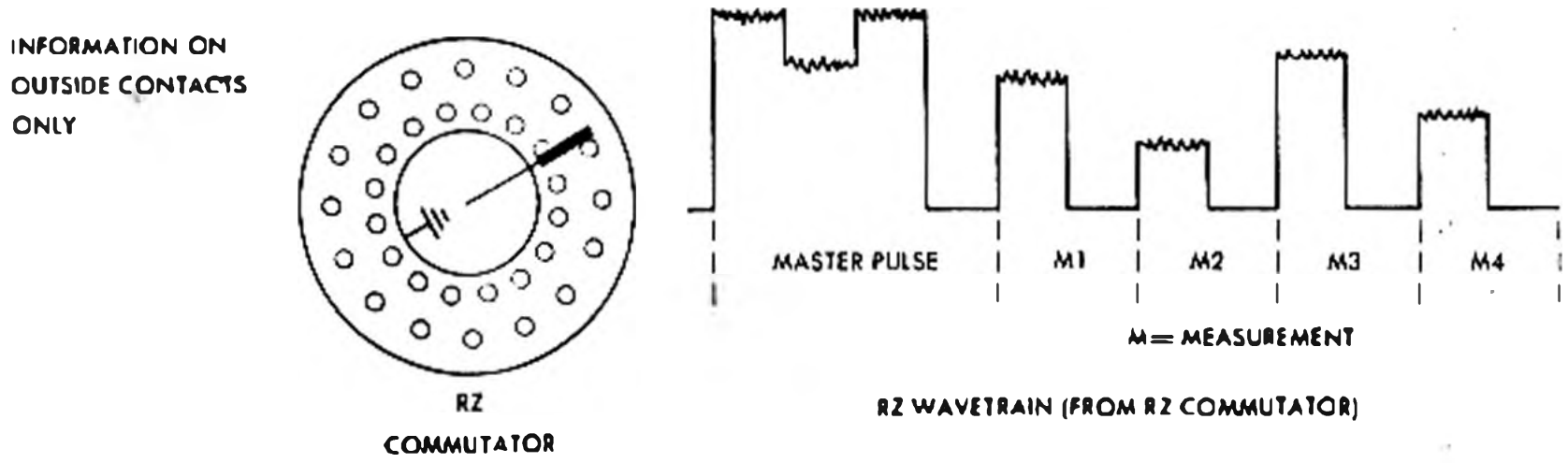


Figure 12-18. Return-to-Zero Commutation

potentiometer acts as the master gain control for amplifier stages of the mounting unit and as the deviation control for the RF transmitter.

Commutator-Gating Unit

The commutator-gating unit, depending on the program involved, is either mechanical or electronic. Either one is acceptable for use as long as the desired information is not distorted too much.

A mechanical commutator is similar in operation to an automobile distributor. The outputs from various transducers are connected to contacts mounted on a disc. Revolving around the disc is a brush, or wiper arm, which is driven by a motor. As the brush touches each contact, the information on that contact is sampled. The commutator is connected in series with the transducers and the remainder of the instrumentation system. This allows only one sample of information to pass at any one specific time.

There are two types of mechanical commutators: RZ (return to zero) and NRZ (non-return to zero).

The outside contacts of an RZ commutator carry the information while the inside contacts are grounded (figure 12-18). As the wiper arm passes over the outer contacts, it receives a bit of information, then returns to zero as it passes the ground contacts. The RZ commutator produces pulses which are more easily read than the NRZ commutator.

NRZ information is obtained from all contacts so the wavetrain does not return to zero (figure 12-19). As the wiper arm passes over a contact, it receives that bit of information. When it moves to the next contact, it momentarily engages both contacts and results in a spike which actually is the summation voltage of both contacts.

The NRZ method allows more information to be commutated, but it requires special circuits to: (1) reshape the data samples to more usable forms, (2) mix the master pulse with data pulses, and (3) insert a pedestal voltage. NRZ is extensively used in pulse code modulated systems where it is necessary to know only whether a pulse is ON or OFF, and the exact amplitude is not critical.

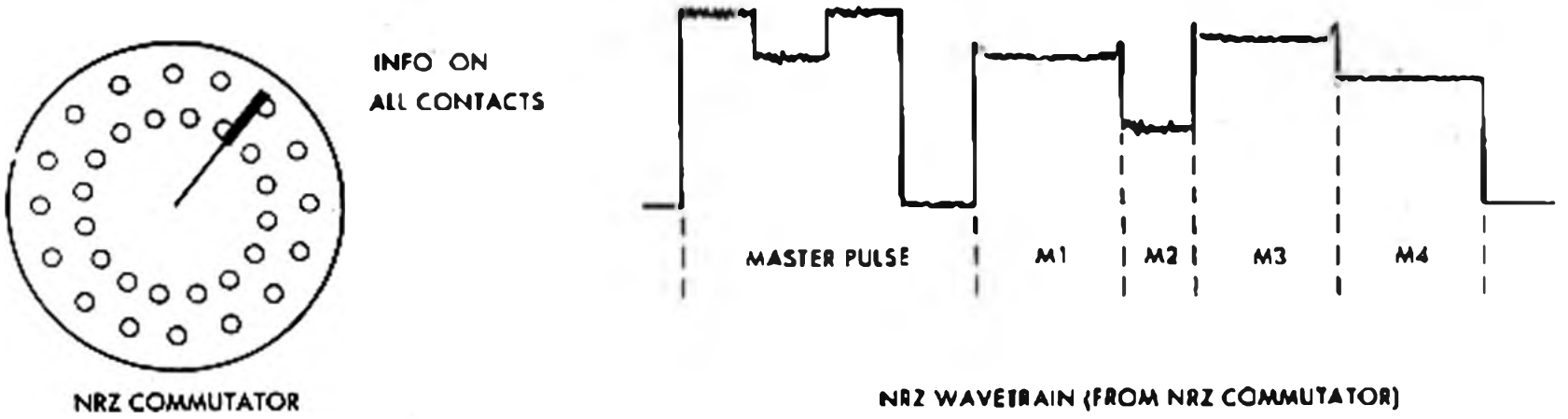


Figure 12-19. Non-Return-to-Zero Commutation

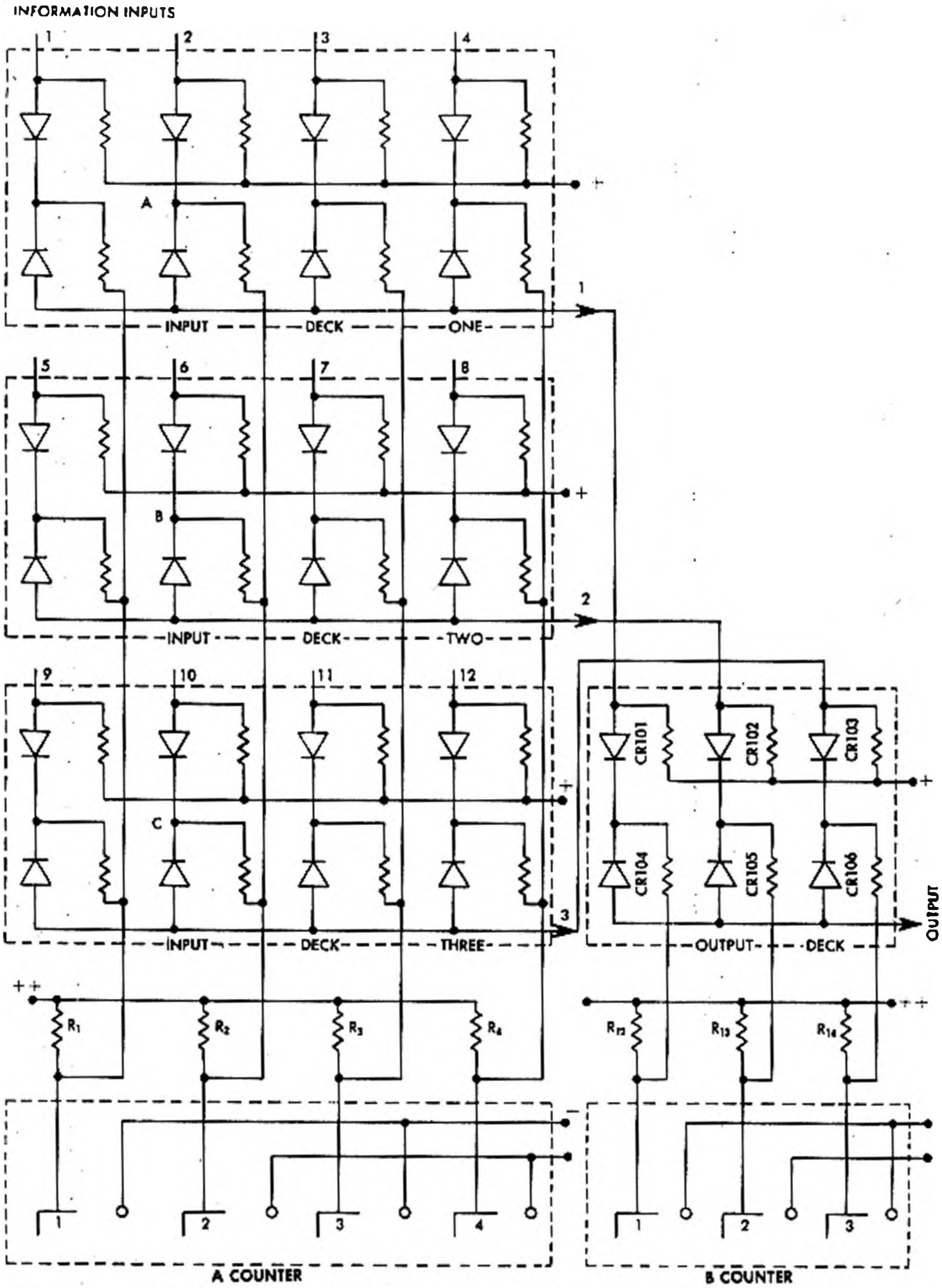


Figure 12-20. Basic Electronic Computer—Simplified Schematic

Electronic commutators were developed because of the limitations of mechanical commutators. Primarily, these limitations are an inability to operate for extended periods of time, heavy power requirements, noise factors, and limited data-handling capabilities.

Electronic commutators have no moving parts, can operate almost indefinitely, use a very small amount of power, are not as susceptible to external noise, and handle large amounts of data. They are considerably more expensive than their mechanical counterparts, but the performance obtained is usually well worth the monetary consideration.

The basic parts of an electronic commutator are: a series diode switching arrangement, units counter, tens counter, clock timer, and a synchronization pulse generator. The main factor limiting the speed of an electronic commutator is the time necessary for the diode switches to operate. Units can be designed to operate at almost any sampling rate, but for use on the test ranges they must conform to IRIG standards.

For a signal to pass through the commutator and on to the telemetry system, the diode switches in its path must be closed. The simplified schematic (figure 12-20) shows three input decks and one output deck. A 90-channel electronic commutator would most likely have ten input decks with nine inputs on each. *NOTE: There are two diodes in each network on both the input deck and the output deck.* The information signal is connected to the top of the uppermost diode. At this same point, a high + potential is applied through load resistors and is felt on each diode network. Though there are 12 inputs to this particular commutator, the description which follows is concerned only with inputs 2, 6, and 10.

Assume that there is a low voltage at input 2 of the A counter, and at input 1 of the B counter. This causes a voltage drop across resistor R2, which results in a negative voltage that causes a difference in the potential felt at points A, B, and C, and makes conduction through these diodes possible. When the three diodes conduct, the information signals waiting on inputs 2, 6, and 10 are allowed to pass through the first set of diode switches in their path.

The output deck has the same circuitry as the three input decks. The output signals, which are now inputs to the output deck, are applied to the diode networks labeled CR101, CR102, and CR103. The only way that these inputs may pass through the diode networks is to have a change of potential across the first series of diodes. When input 1 on the B counter is at a low voltage, a voltage drop across resistor R12 is sufficient to allow diode CR101 to conduct. This allows the information waiting on its input to enter the deck and be felt across diode CR104, and then to the output.

The series of pulses out of the commutator must be in the proper order and of a definite time length. To insure this, a clock timer is used. The clock timer usually consists of an astable multivibrator and logic circuits. The multivibrator controls the time for each information pulse and the logic circuits make sure that the proper commutator output is triggered for output.

Transmitter Unit; RF Power-Amplifier Unit

Figure 12-21 shows that a transmitter unit uses the composite output of frequency-modulated subcarrier oscillators to frequency-modulate an RF signal produced by the transmitter. The modulated signal is then multiplied in frequency and amplified for transmission. The center frequency of the RF carrier is adjustable within the range of 1435-1540 MHz and 2200-2300 MHz.

Most RF transmitter units are low powered (about 3 watts). The output of the 3-watt power-amplifier stage can be fed directly into a load such as a quarter-wave antenna, or it can be used to drive an amplifier of higher wattage output. A

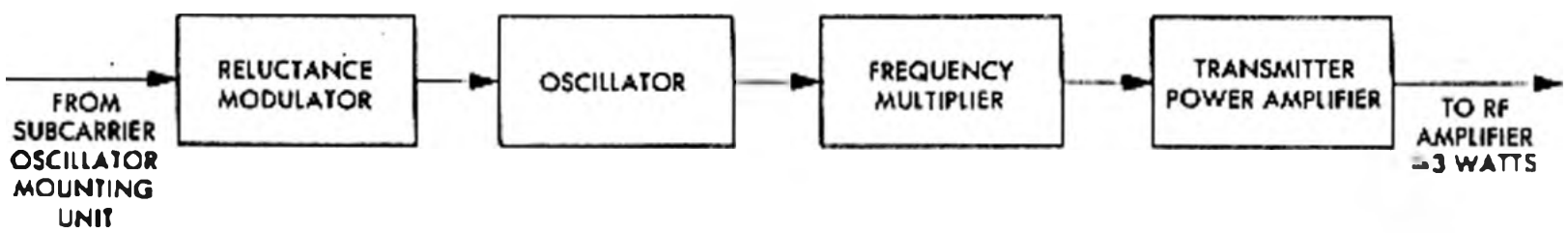


Figure 12-21. Transmitter Unit

filter box surrounding the cable that contains all inputs and power leads reduces the possibility of RF leaking back to the other units of the system.

Antenna

The antennna generally used with an FM/FM telemetry system is a quarter-wave stub antenna. It is suitably mounted beneath the missile to radiate power in an unobstructed path to the ground station receiver.

AMPLITUDE-MODULATED (AM) TELEMETRY SYSTEMS

Amplitude-modulated (AM) RF-carrier telemetry systems are used in some industrial applications in which the data requirements are not critical and in which the information samples can be averaged over relatively long periods of time. By averaging over long periods of time, the effects of noise and interference can be minimized.

Such data as water levels in reservoirs, fuel levels in storage systems, YES and NO data (indicating action or lack of action), and opening and closing of electrical circuits can be telemetered by AM systems.

However, in telemetering data from research missiles, many data samples of widely differing characteristics must often be taken in a brief period of time and under widely differing conditions of temperature, pressure, acceleration, etc. These data must be transmitted with the highest degree of accuracy even though the sampling times are usually short and physical condition measurements are complex. To insure the greatest accuracy of such data, FM/FM systems or pulsed radar systems are used rather than AM systems.

The intelligence conveyed by the RF carrier in an AM system is represented by the amplitude of the carrier. Any change in carrier amplitude produced by static or from sources outside the system might be mistaken for changes taking place at the data-pickoff points. This results in erroneous interpretations of raw telemetry data at a ground receiving station.

Circuitry of AM telemetry systems follows conventional lines and resembles that of the usual AM radio-communication systems. Some synchronized switching and timing system must be employed so that the amplitude variations of the received signal at any given instant may be identified with the channel or data source being sampled at that particular time. For this purpose, a commutator unit similar to that used in an FM/FM system is employed.

Received data are recorded on tapes or charts by recording meters which employ time-division roll charts. A functional block diagram of an AM telemetry system appears in figure 12-22.

TELEVISION TELEMETRY SYSTEMS

The visual presentation of certain forms of data collected and monitored during in-flight tests was originated for testing high-speed aircraft. Television cameras could be mounted to take pictures of aircraft instrument panels and record the dial indications of important instruments. In

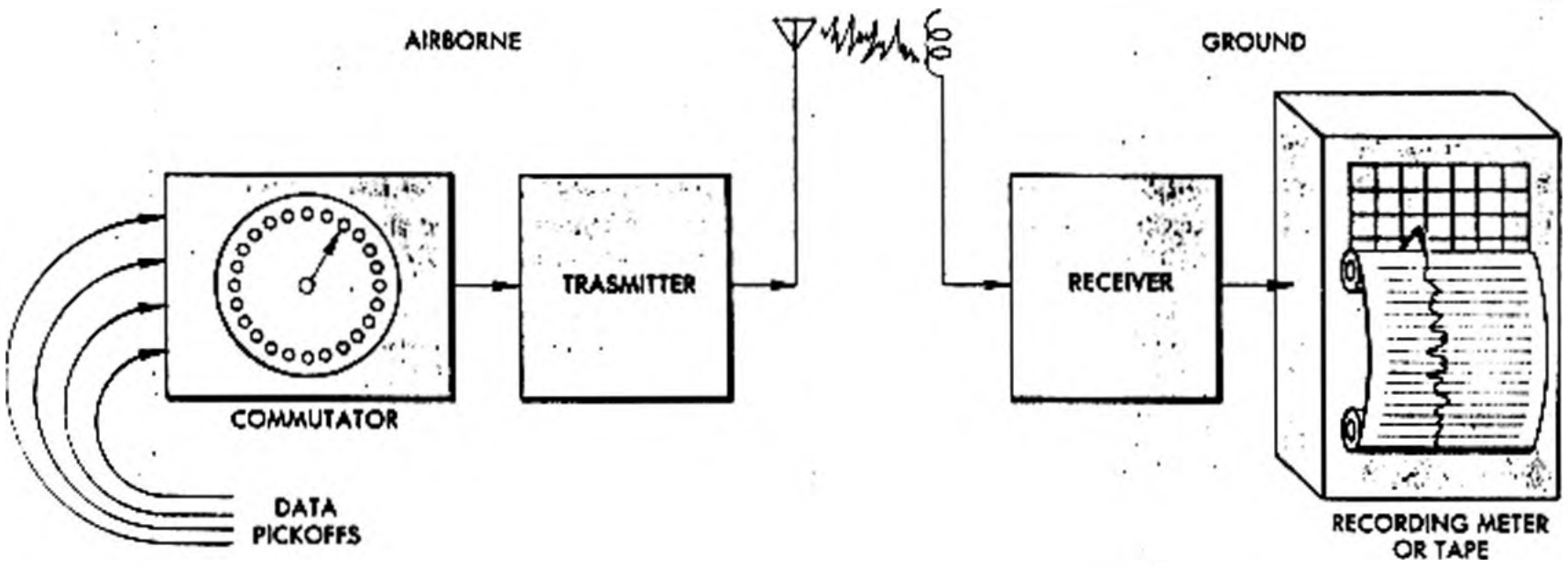


Figure 12-22. Functional Diagram of AM Telemetry System

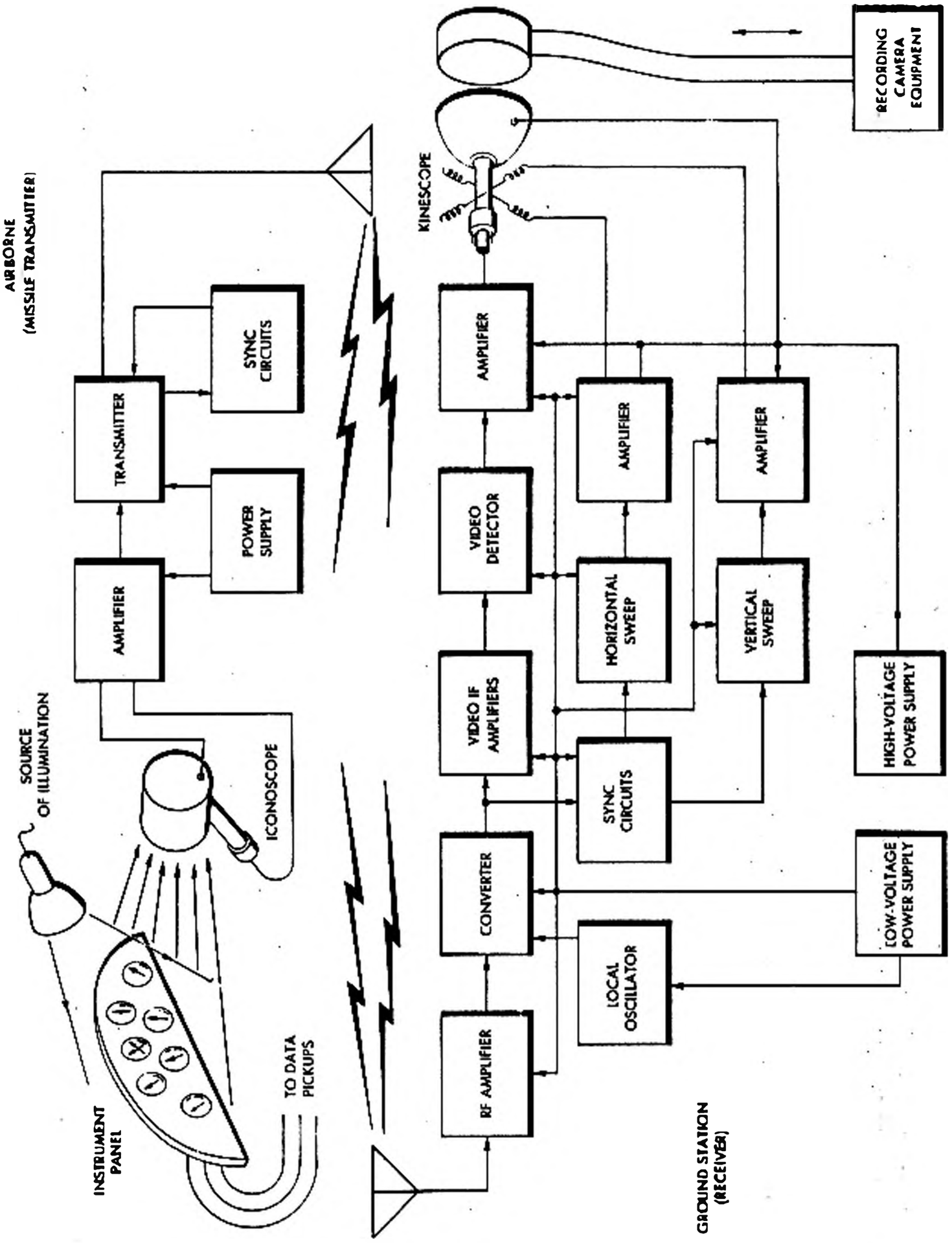


Figure 12-23. Television Telemetry System

addition, cameras could be focused upon important control surfaces to detect flaws in their operation during various maneuvers of the aircraft. Continued reductions in the weight and size of the cameras make them an ideal data collection system for the larger ballistic and aerodynamic missiles.

An excellent example of television telemetry is the system used during the flight testing phases of NASA's Centaur missile. This missile is intended to be orbited about the earth as well as to place instrumented capsules on the moon or other planets. Since the Centaur is the first missile to be fueled with liquid hydrogen, a television camera was mounted on the fuel cell bulkhead to study fuel reaction. In this position, video data is gathered every 2 seconds and recorded on magnetic tape. The video signal is relayed back to the ground station monitoring the Centaur and converted to television and motion pictures for a detailed study of liquid hydrogen. A 100-kilowatt strobe light is used for each television exposure in the vehicle.

Missile manufacturers have used a form of television coverage in their landline or hardwire instrumentation for some time in static testing their missiles. Studies of the exhaust flame of the propulsion unit with respect to flame length, shape, luminosity, and shock-wave pattern can be made by closed television circuits. Results of such studies aid in determining the effectiveness of engine nozzles, fuel ignition, thrust control, and propellants. The use of television units permits closeup views of propulsion units running at maximum thrust levels which would be dangerous to human observers because of th accompanying high temperatures, light, and no: :.

Television telemetry systems have also been used in satellite vehicles such as the Tiros series. Primarily a weather satellite, Tiros television cameras have taken thousands of pictures of the clouds that continually cover the earth's surface. Studies of such clouds are made to detect the formation of hurricanes and to monitor the paths of fully-developed hurricanes from their places of origin. Ordinarily, television systems for aerospace vehicle instrumentation do not include a sound channel; therefore, only the components for video transmission and reception are needed. A simplified block diagram of a typical television telemetry system is shown in figure 12-23.

Present-day video systems use an amplitude-modulated carrier, which is susceptible to interference of an AM nature. If the system is used to televise meters and panel instruments, distortion of the picture is not a critical factor as long as the meter readings are visible. If structural surfaces are being televised, distortion might be interpreted erroneously as variations of the structure itself. For example, a "jittery" picture might be interpreted as a view of vibrating surfaces, and nonlinearity of the picture might be interpreted as the effect of strain on the surface being viewed.

These latter defects can be overcome by use of positive reading gages and indicators whose absolute readings are shown on the viewing screen rather than a view of the structure or surface itself.

SECTION C

Reduction of Telemetry Data

This section is concerned with data reduction or, in other words, the methods used for changing telemetered raw data into usable form. Some types of equipment used for this are antennas and receiver calibration equipment.

TELEMETRY GROUND STATIONS

A block diagram of an FM/FM telemetry system, greatly simplified, is shown in figure 12-24. The diagram includes both transmitting and receiving sections of the system. In FM/FM telemetry systems, the frequencies of a number of audiofrequency subcarrier oscillators are varied in accordance with the variations of the quantities being measured. These frequencies or tones are mixed and then used to frequency-modulate a high-frequency transmitter. This arrangement is represented by the upper half of the diagram.

The signal thus produced is received at the ground station, illustrated in the lower part of the diagram, by a superheterodyne FM receiver which reproduces the mixed subcarrier audiofrequencies and then separates them by means of the discriminators. The individual subcarrier frequencies are converted into DC currents that vary in accordance with the variations of the original quantities being sampled by the pickoffs and transducers in the research missiles.

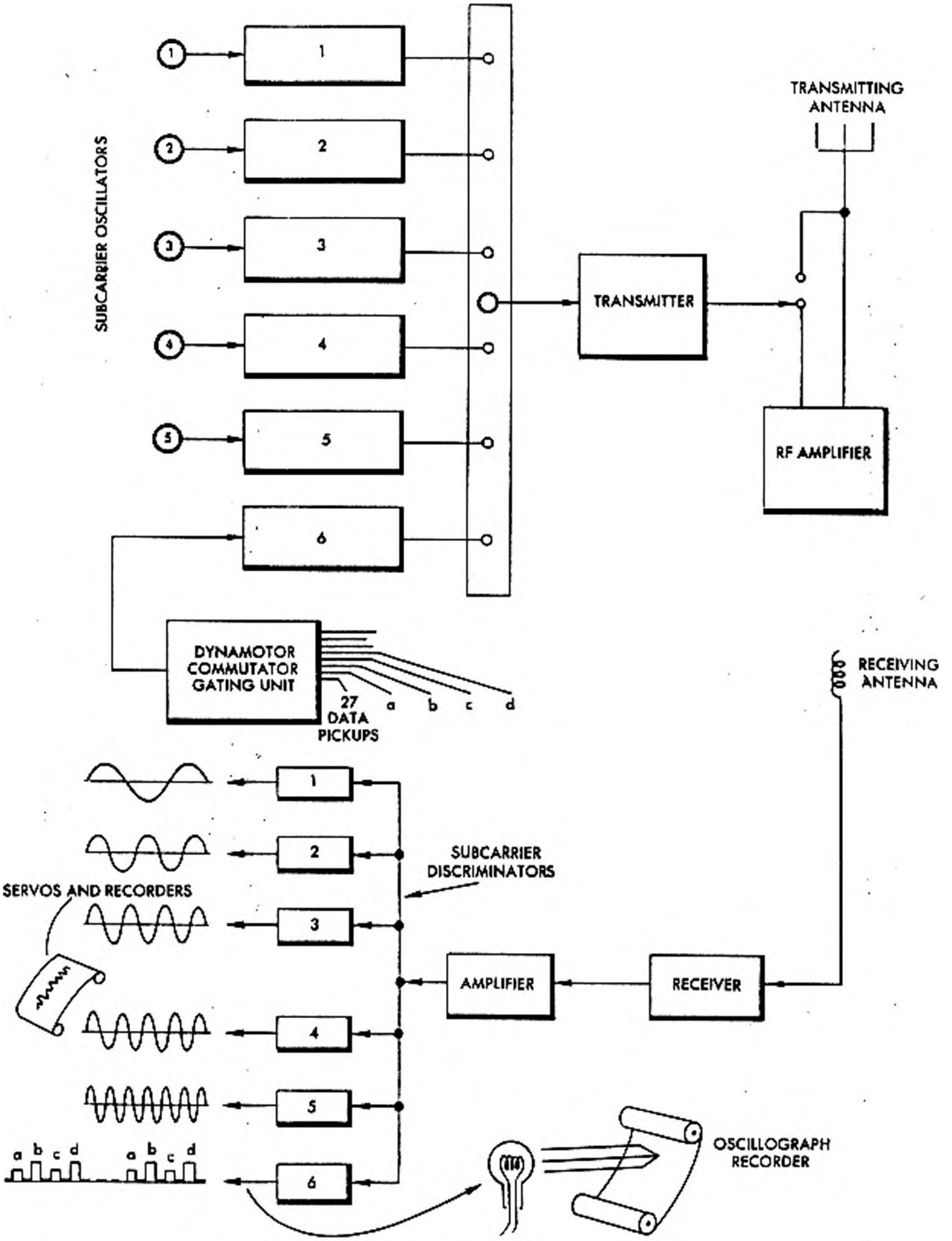


Figure 12-24. Typical Data-Separation System

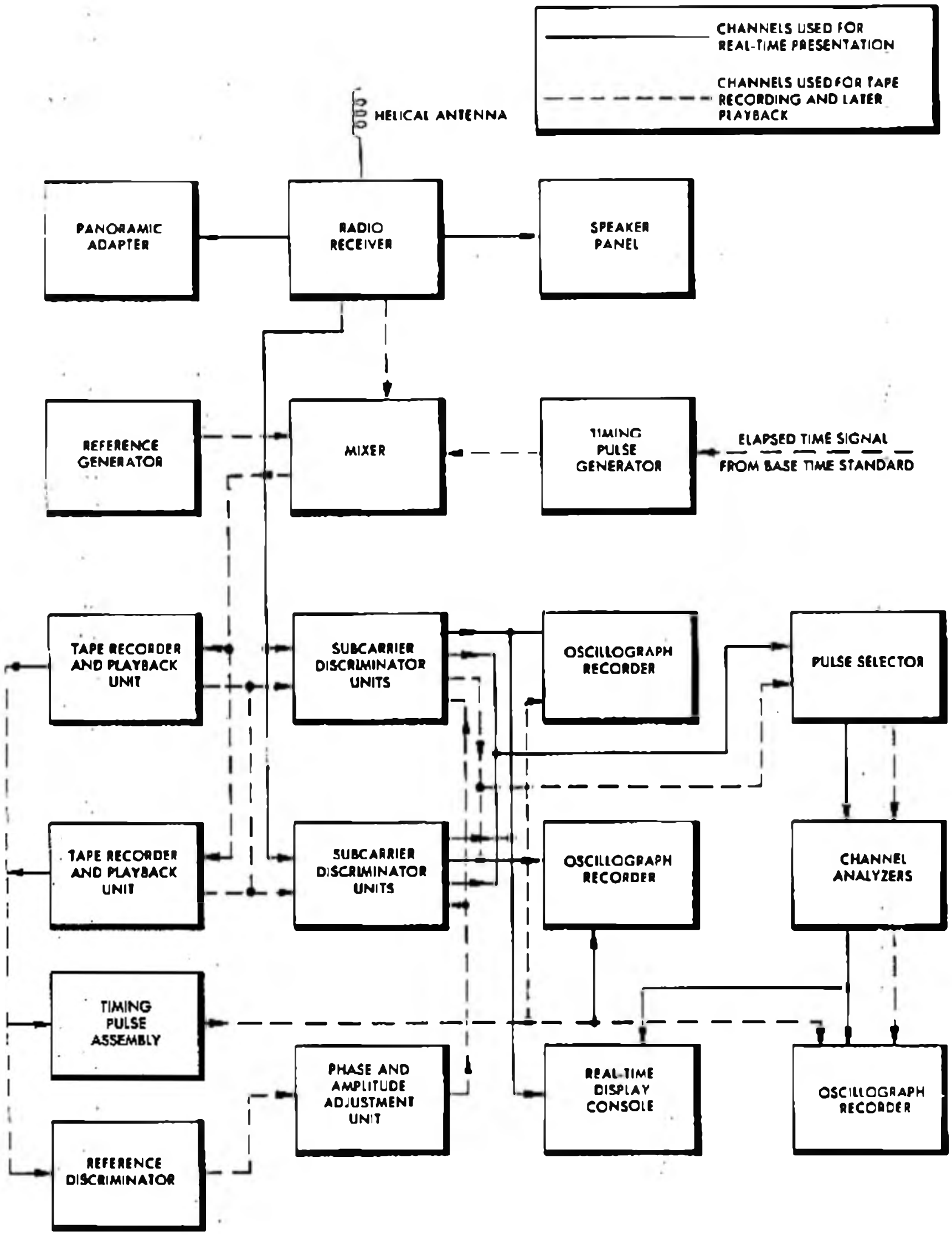


Figure 12-25. Functional Diagram of Telemetry Receiving and Recording Station

These currents, in turn, operate galvanometers in a recording oscillograph or other devices which produce a permanent record. The mixed output of the receiver is also recorded on a tape recorder for future study or for insurance against failure of the recording oscillograph. Nominal radio frequencies used are 1435 to 1540 MHz and 2200 to 2300 MHz.

The audio subcarrier-oscillator frequencies used are in the range of 400 to 165,000 Hz. As many as 21 subcarriers are used. The deviation of each subcarrier is limited to ±7½ percent of the center frequency to insure stability of operation of the subcarrier oscillators. On the 8 highest subcarrier-oscillator frequencies, the deviation allowed is ±15%. In some instances, two telemetry systems are employed, each operating on a different RF frequency.

Figure 12-25 shows in more detail, and from a functional standpoint, the major components of a typical ground installation for receiving and recording the telemetry signals.

The signal from the transmitter is picked up by an antenna and conducted to the receiver; there it is amplified and demodulated. Then it is passed on to the system's subcarrier-discriminator units as mixed audio tones. An alternate signal of mixed tones of known frequency from the calibration oscillator (test equipment not shown in the illustration) can be switched into the system in place of the receiver output to calibrate the data reduction and recording system.

The mixed tones from either the receiver or the calibration oscillator are connected through the mixer to a tape recorder for raw-data storage and to a bank of subcarrier discriminators, one for each subcarrier band used.

Before launching a test missile, all downrange tracking stations and their instruments are synchronized with respect to time. After synchronization, coded timing signals that are repeated at certain rates are transmitted from the master station. Basically, six types of time codes are available, and all are presented serially in a form which may be recorded by a magnetic recorder or photographic recorder. Figure 12-26 shows in a simplified form the manner in which timing serves a missile test range. Note how the firing

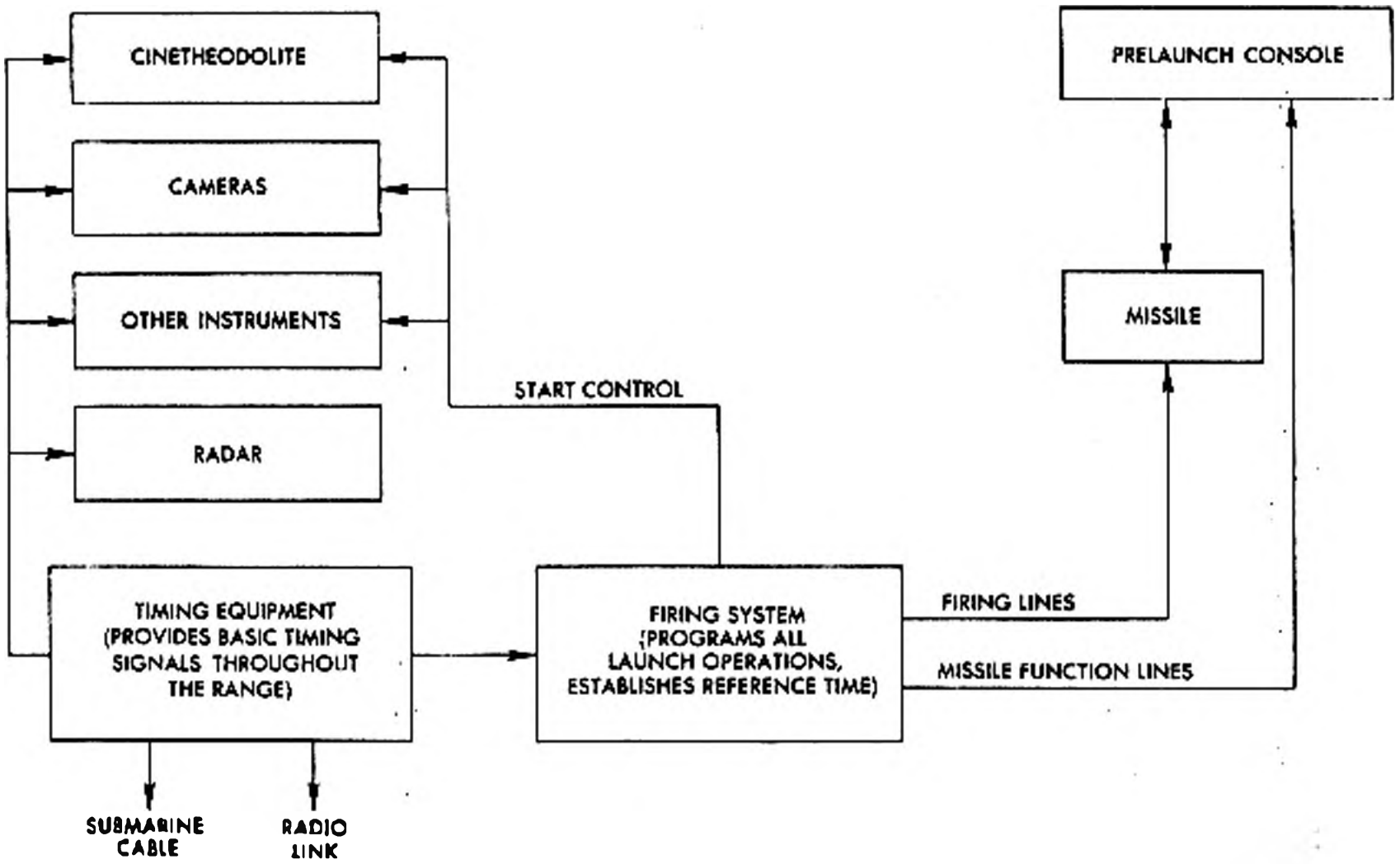


Figure 12-26. General Timing and Firing System Block Diagram

system is integrated within the timing system since a firing signal, transmitted through a start control line, activates many of the tracking instruments.

Synchronizing timing signals are inserted through the timing-pulse generator, providing the recording equipment with accurate time-interval reference. The discriminators separate the channels and convert frequency variations of the bands into DC currents. The DC currents, in turn, operate the recording instruments. The raw data stored on the tape recorder can, at any future time, be played back through the discriminators to produce an oscillograph record.

Outputs of the discriminator units consist of data which are continuous in nature and data which are commutated. One output is a combination of both types. The combination output is desirable for aiding in the selection and separation of a desired tone signal. Note that both continuous and commutated signals eventually reach and are displayed on the operator's console.

Superheterodyne FM receivers, similar to that shown in figure 12-27, are used at ground stations. The input impedance of these receivers is set at 52 ohms for the specific operating frequency. The incoming signal is amplified in a 2-stage RF amplifier, converted to 105 MHz in a mixer by combination with the local oscillator voltage, and then amplified in the 3-stage IF amplifier. Local oscillator tuning only is employed.

The 105 MHz signal is amplified and clipped in the 2-stage limiter, and the constant-amplitude,

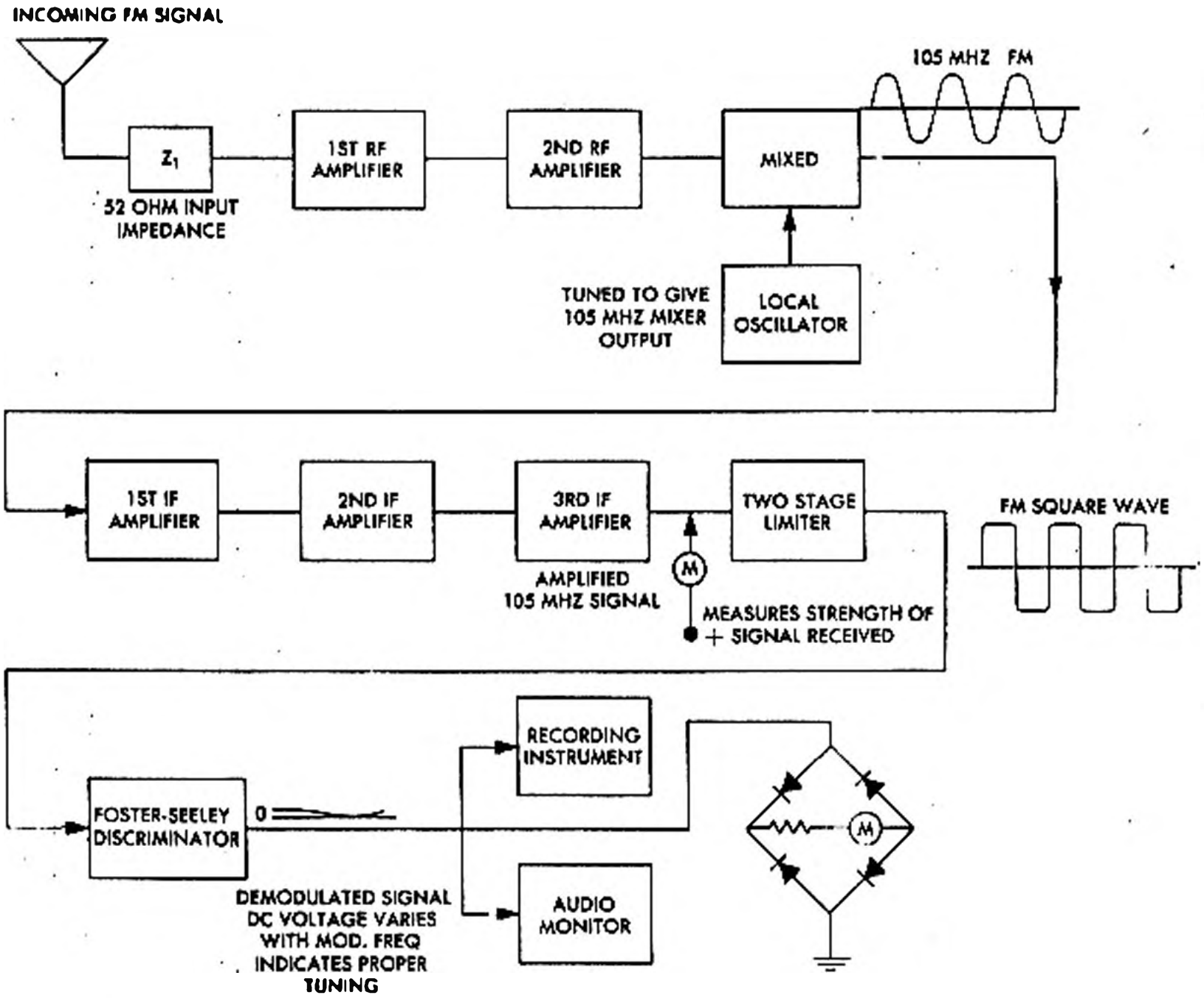


Figure 12-27. Typical FM Telemetry Receiver

frequency-modulated square wave is demodulated in a discriminator.

The first limiter is metered to determine the strength of the received signal. The DC voltage output of the discriminator indicates correct tuning of the receiver.

The demodulated signal from the discriminator is fed to two outputs—one output is connected to recording instruments, and the other to headphones for aural monitoring. Each output has an independent gain control. A third output is connected to the rack output and operates a meter rectifier bridge to permit continuous monitoring of the receiver level.

ANTENNAS USED WITH GROUND RECEIVER UNITS

Various types of antennas are used in different ground station installations, depending somewhat on the type of antenna used by the transmitter and the angle of coverage of the radiated signal.

In figure 12-28 you see a *helical-beam antenna* which is quite efficient for FM/FM carrier reception and which is widely used with mobile ground receiver units. The text here is centered on this type.

This antenna was designed at the New Mexico College of Agriculture and Mechanic Arts. It consists of a helix which operates in the axial mode of radiation; that is, the helix is directed so that its axis is in line with the source of radiation (transmitter antenna). In this mode of operation, the field is at maximum in the direction of the helix axis and is circularly polarized, or nearly so. The gain of the antennas is about 11 decibels (dbs) when receiving from a circularly polarized antenna and about 8 dbs when receiving from a linearly polarized antenna. The beam width between half-power points is about 36°.



Figure 12-28. Helical Beam Antenna

When higher antenna gain is needed, three or four helix antennas can be used on a single base. These antennas are called respectively tri- and quad-helix antennas. The helix antennas are lightweight and good for mobile tracking stations where space and weight come at a premium price. For long range packages the helix antenna does not have enough gain, and some sort of parabolic antenna is usually used.

Three types of *parabolic antennas* are commonly used. All depend on the physical action of a parabolic surface focusing high frequency waves into a narrow beam. Figure 12-29 illustrates three common types of parabolic antennas. Of the three shown, the Cassagrain seems to be the best. The simple parabolic has some spillover of energy, and putting a preamplifier at the antenna feed point introduces structural problems. The horn, while very efficient, has to be quite large. The larger the antenna the more it will weigh, which makes it harder to move. Many of the parabolic antennas used for satellite instrumentation are 60 and 85 feet in diameter. Even at this size, they are expected to follow the satellite on its track. These large antennas which weigh over 100 tons, are still expected to make up to 360° around in 2 minutes and track a satellite with an accuracy of 1/500th of a degree.

RECEIVER CALIBRATION EQUIPMENT

Calibration of a telemetry receiving system is accomplished by substituting (for a received signal of mixed audiofrequencies) a number of signals of known frequency from the calibration equipment.

An oscillator for each band is designed to produce any one of three frequencies. The desired frequency can be selected by a switching arrangement. The three frequencies used are the channel-center frequency and the upper- and lower-limit frequencies for the channel (center frequency ± 7½ percent).

When the frequencies are recorded in time sequence, a 3-point calibration is obtained. The

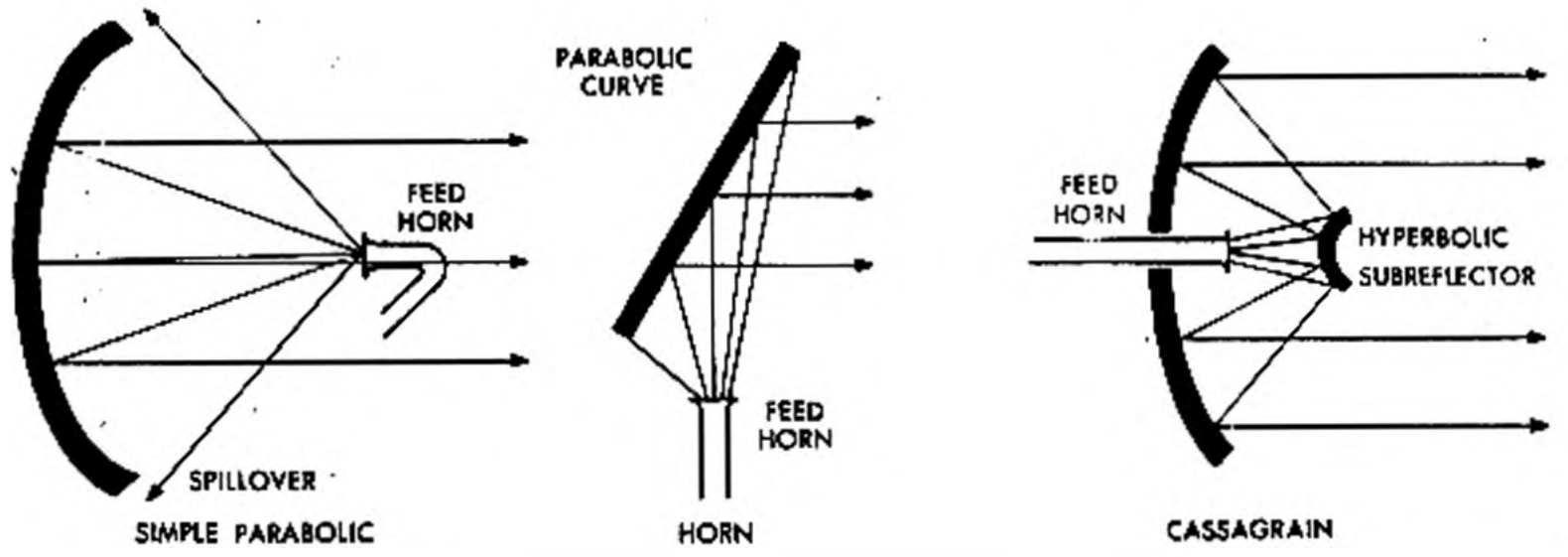


Figure 12-29. Parabolic Antennas

telemetry equipment produces data that are linear enough so that no other calibration is required. By means of ganging the switches from all calibration oscillators, simultaneous calibration of all channels is obtained.

The output of each calibration oscillator is fed into a combination mixer and amplifier stage, as shown in figure 12-30.

Calibration Oscillators

In telemetry calibration, the calibration oscillator chassis provides space for plug-in oscillators, as well as for circuitry for mixing the outputs of all the oscillators. The calibration oscillator for each channel is built on its own plug-in chassis. Electrically and mechanically, the oscillators are identical except for a tuned circuit. Each tuned circuit is a plug-in unit of its respective oscillator. This design provides for a quick change of units in case of a failure. A spare unit for each channel is always kept available, making it unnecessary to disable the entire ground telemetry equipment to repair a single oscillator.

The 5-position calibration switch is located in the center of the control panel. It is used to switch the frequencies of each oscillator between HI, MID, and LO. All oscillators are switched simultaneously. In addition, this switch connects the receiver to the system. The receivers are connected when the switch is in either end position. In the three middle positions, a calibration oscillator is connected to the system and supplies either the high, middle, or low frequencies as required.

Discriminator Units

A discriminator, or frequency-detector, is necessary to convert the mixed audiofrequencies from the receiver output or from the calibration oscillators into currents that vary in accordance with the changes in frequency in each channel.

Each discriminator, with its associated components, is mounted in one plug-in unit. Each plug-in unit contains an isolation amplifier, a bandpass filter, a limiter stage, a specially designed discriminator, and an output filter.

The discriminator units are identical both mechanically and electrically, with the exception of the frequency-sensitive components (bandpass filter, two discriminating condensers, and lowpass output filter) which are separate plug-in units. As in the case of oscillators, this design permits rapid replacement of defective units without shutting down the entire system. Also, a discriminator plug-in unit can be readily adapted to any channel by inserting the proper frequency-sensitive components.

An additional advantage of having separate bandpass and output filters for a subcarrier discriminator is that channel playback characteristics can be changed rapidly. By changing channel characteristics, information can be retrieved that would not be available otherwise. Sometimes the subcarrier-oscillator in the instrumentation package gets off center frequency. Changes in the input bandpass filter can correct this. If the subcarrier gets too noisy, changing the output filter to one with a lower frequency range or one with slightly different frequency characteristics will make data reduction much more accurate. If it were not for

the plug-in units of the discriminators, ground stations would have to have a very large number of subcarrier discriminators, or else be limited in the number of channels they could process.

RECORDING DATA

Collecting and recording data from a guided missile can be readily accomplished by means of recording-type meters or by photographic means. Direct recording is the simplest and most accurate method for obtaining flight data.

In some recoverable research missiles and capsules, direct-reading instruments using tapes for graphic or magnetic recordings are employed as auxiliaries to telemetry equipment. Photographic systems are also used. The data recorded on tape or film are used to check the data received from a telemetry system. In this way, not only can the flight characteristics of a missile be determined but also the accuracy of its telemetry system.

For some types of data which are subject to very rapid changes, suitable direct-recording devices have not been perfected; consequently, some kind of conversion link must be used between the data sources and the recording instruments. Frequency dividers or down counters are examples of such intermediate links for slowing down data to some definite proportionate rate which is within the range of a recording instrument.

Pen Recorder

A typical pen recorder, suitable for use in either a missile or a ground station, consists essentially of four main components: the case, the chart drive, the writing system, and the electromagnet assembly.

Electrically actuated, the instrument may have from 5 to 20 pens simultaneously recording the outputs from a like number of data pickoffs. Data are taken from sources whose information is of the ON-OFF, YES-NO type, indicating the times and durations of the data events and how many

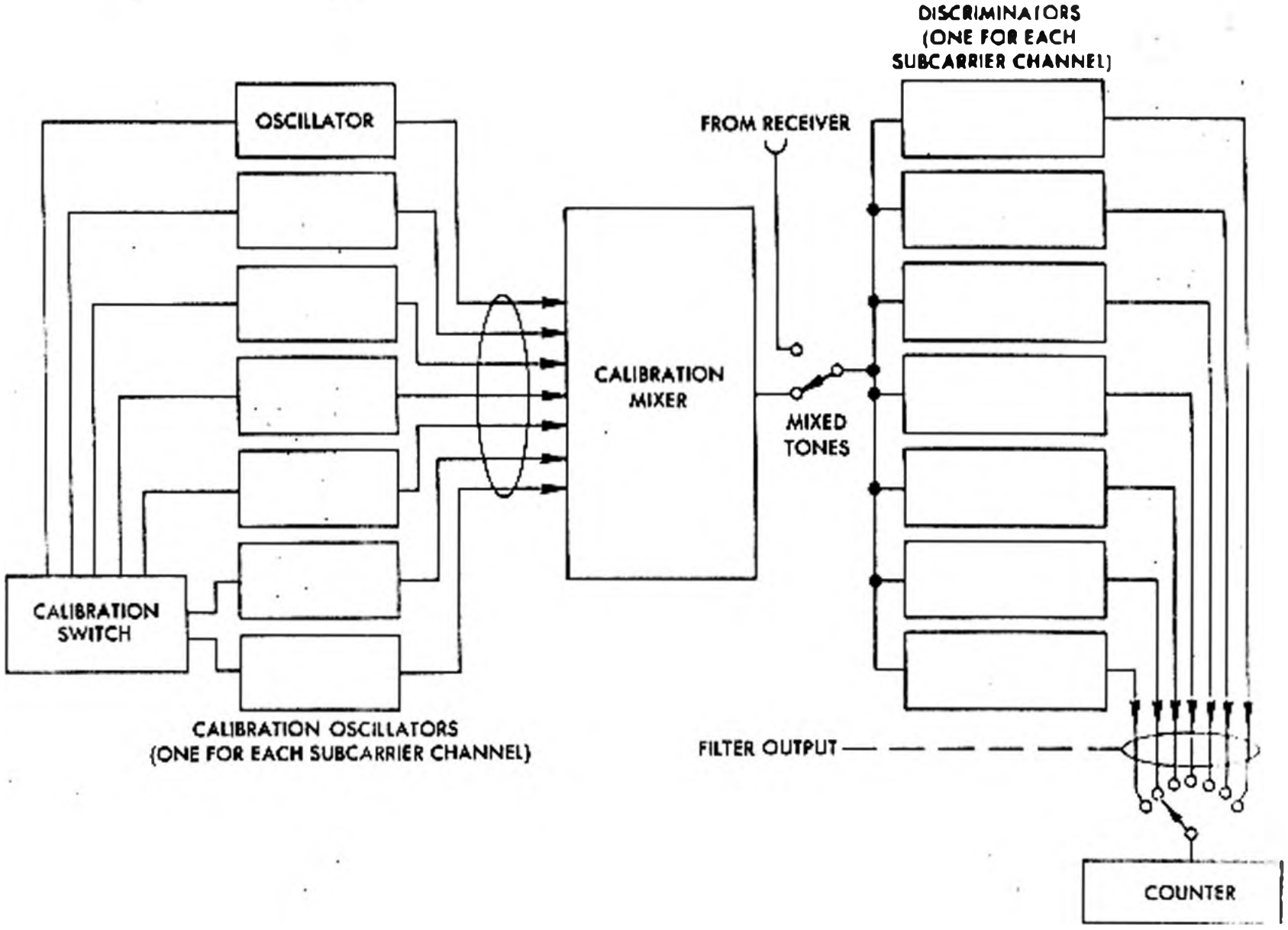


Figure 12-30. Telemetry Receiver Calibration Equipment

operations or changes occur. The recorder charts are rectilinear—a characteristic which makes it easy to compare the times of various recordings.

Pen recorder charts are rolls made of special grade paper. This chart paper can be obtained with or without time-calibration marks along its edges. The chart may be propelled by a self-contained synchronous motor clock, an external motor or timing machine, or by a combination of these devices.

Where time intervals must be measured accurately to less than 1 second, a chart speed of at least ¾ inch per second is used. Speeds from ¾ inch to 3 inches per second are made possible by an external motor drive.

Three types of record patterns are obtained on the chart, depending on the relation of the length of time the pen is energized, the time the pen is deenergized, and the speed of the chart. These types are the rectangle, band, and impulse patterns which are illustrated in figure 12-31.

Since the drag of the pens on a strip chart is very light, the chart driving mechanism has ample power to overcome any drag which might produce a time error. Movement of the pens is laterally across the chart, and each pen is actuated by an electromagnet which receives its energy from the pickoff source.

Each electromagnet element can follow as many as 10 complete ON-OFF cycles per second, provided the ON and OFF periods are substantially equal. There must be at least 0.05 second of energized time to allow the pen to complete its stroke reliably. If the pen is normally energized, the deenergizing impulse must also last for at least 0.05 second. The simplest method for actuating the electromagnet elements is by the direct application of voltage.

For data measurements involving changes of more than 10 cps, or very wide ranges of voltage variation, pen recorders are not suitable; thus, devices such as recording oscillographs are employed.

Recording Oscillographs

A typical recording oscillograph contains up to 18 recording galvanometers. Each galvanometer deflects a beam of light across a roll of photosensitive paper in accordance with the output current of the data pickoffs or the output of receiver-data channel discriminators. Note the cutaway section of a recording oscillograph in figure 12-32. The photosensitive paper is rolled past the galvanometer light beam at a constant rate, producing a permanent record pattern on the paper, as shown in figure 12-33.

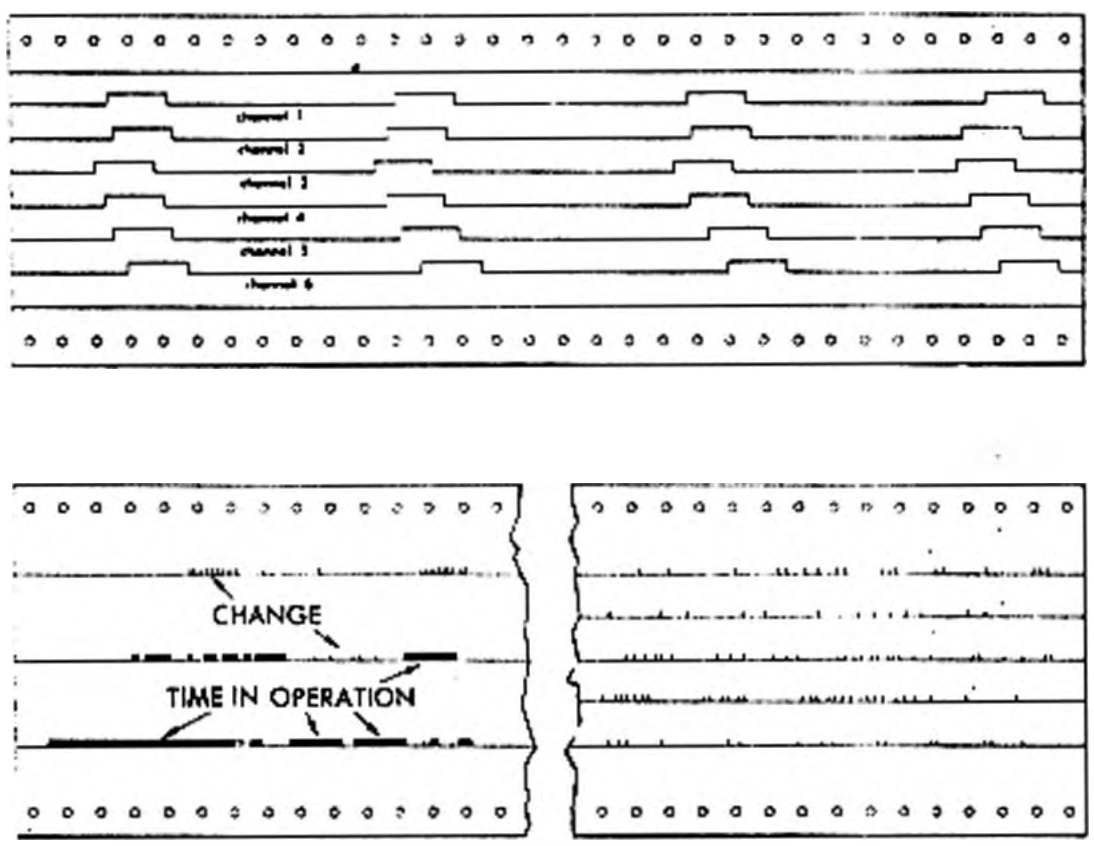


Figure 12-31. Types of Record Patterns used for Strip Charges

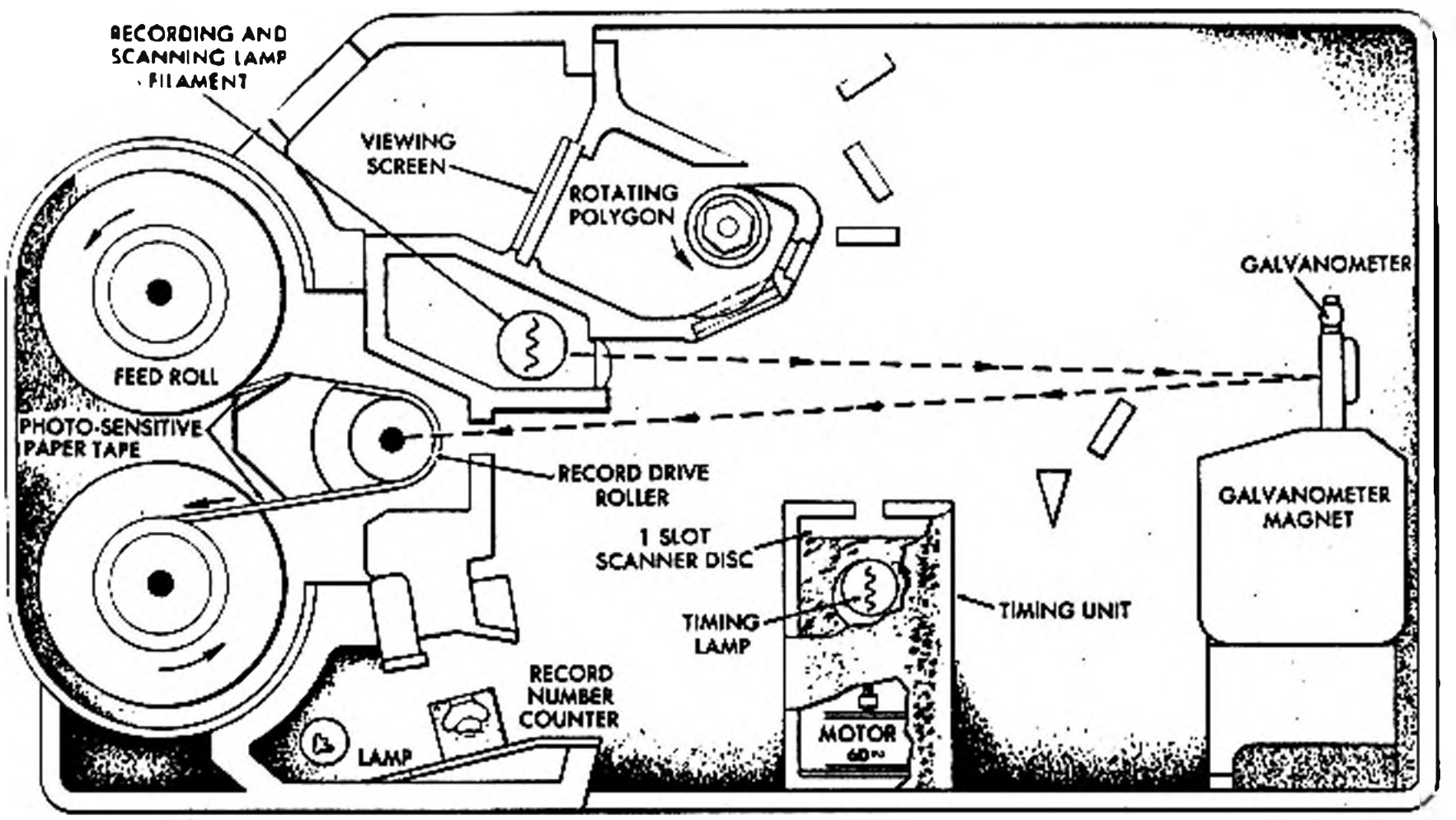


Figure 12-32. Cross Section of Recording Oscillograph

The recording oscillograph is a highly flexible instrument and is well suited to recording data from telemetry systems.

The time base for the recording is the speed at which chart paper is run through the recorder. The paper speed is continuously variable over a wide range, and a large selection of galvanometers is available, thus permitting the recording of many types of data.

Tape Recorder

In conjunction with a recording oscillograph, a supplementary recording device is used. The most commonly used type is a tape recorder.

This machine is used to record directly the output of the receivers and the calibration oscillators. The process is sometimes termed raw-data recording.

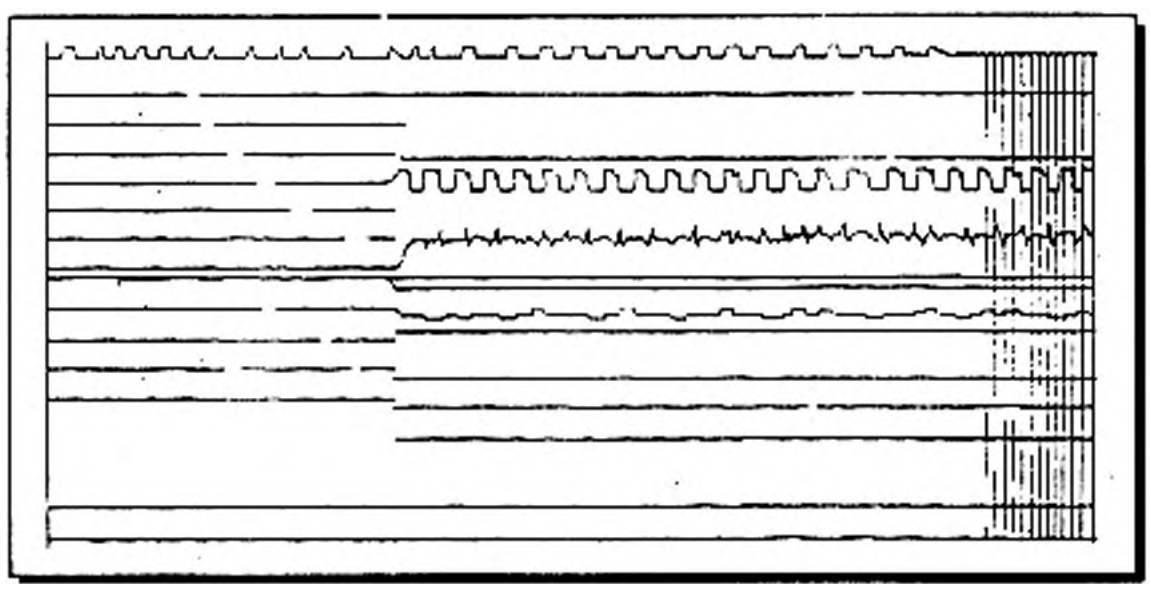


Figure 12-33. Oscillograph Record Patterns

Raw-data recording possesses many advantages. If the oscillograph should fail, the tape-recorded data are still available. If a frequency response which is different from that originally recorded on the oscillograph is desired, the original datum is available. When desired, different time bases (which call for different paper speeds) can be obtained by this process. Frequently, large numbers of crossovers of galvanometers are present on the original oscillograph recording. By repositioning the galvanometers and playing the tape-recorder data back through the system, any confusion of traces can be eliminated.

One of the most useful purposes of a tape recorder is the playback, to the oscillograph, of only selected channels. The playback is sufficiently accurate to permit data reduction of the resulting record with greater ease than from the original oscillograph record.

The accuracy of recording and playback can be improved by using an external capstan amplifier. This insures constant tape speed (uniform time base) when recording equipment is supplied power from a gasoline-driven generator which may have poor frequency stability. A capstan amplifier also improves recording results when the equipment is operated from a commercial power source. Although accurate over long average periods of time, commercial power is subject to considerable variation and fluctuations for short time intervals.

An accurate reference tone is usually recorded on the tape along with the data. Variations in tape speed either in recording or in playback can be easily detected by noting any deviations in this tone. These variations can be detected by comparing the frequency of the tone on playback against the tone generated. Accurate time lines can also be transferred from the tape to the oscillograph record by feeding the recorded reference tone to the oscillograph. This output may also be used (with appropriate amplification) to drive the oscillograph time-line motor, or it may be used to drive the pen recorder chart-drive motor.

Instrumentation recorders use tape speeds of 1⅞, 3¾, 7½, 15, 30, 60, and 120 inches per second (ips) for recording information on tape. These tape speeds were chosen to provide the maximum interchangeability between different recorders and for guides when specifying airborne recorders. Notice, for instance, that all the tape speeds are multiples of those used on standard home recorders. The record/reproduce bandwidth of a recorder is stated at the highest recording speed because the reproducible bandwidth of a recorder is directly proportional to the tape speed. A recorder capable of handling 500,000 Hz at 120 ips will handle only 250,000 Hz at 60 ips.

Recorders are grouped into three groups according to their bandwidth. The three bands are:

1. Low band. Direct record response to 100,000 Hz at 60 ips.
2. Intermediate band. Direct record response to 250,000 Hz at 60 ips or 500,000 Hz at 120 ips.
3. Wide band. Direct record response to 1.5 MHz at 120 ips.

Most instrumentation recorders are used to record a variety of signals. Some are slowly changing in amplitude. Some change so slowly that they cannot be recorded because the recorder does not respond to a frequency which is that low. For other signals, the leading and trailing edge of a pulse is significant while the interval between is relatively useless. One amplifier is not sufficient to record all these signals equally well. To overcome this problem, the input and output amplifiers of recorders are plug-in units which can be switched around to provide the best recording characteristics for that particular signal.

The direct record method of recording signals is not only the basic way but is also the one with which you are most familiar. Home recorders use this method. The recording of the signal in a direct record system is accomplished by introducing the signal into a record amplifier. The basic requirement of the amplifier is that it be able to accept the wide frequency spectrum of the composite signal.

For some signals it is not desirable to record direct. For these signals, a technique called FM recording is used. The FM technique uses an oscillator that is FM modulated by the incoming information. To reproduce the information the reproduce amplifier must demodulate the FM carrier.

When recording pulses, the times of the leading and trailing edges of the pulse are the most important because no information is carried in the amplitude of the pulse. The record amplifier is not only an amplifying circuit but also a differentiating circuit. A block diagram is shown in figure

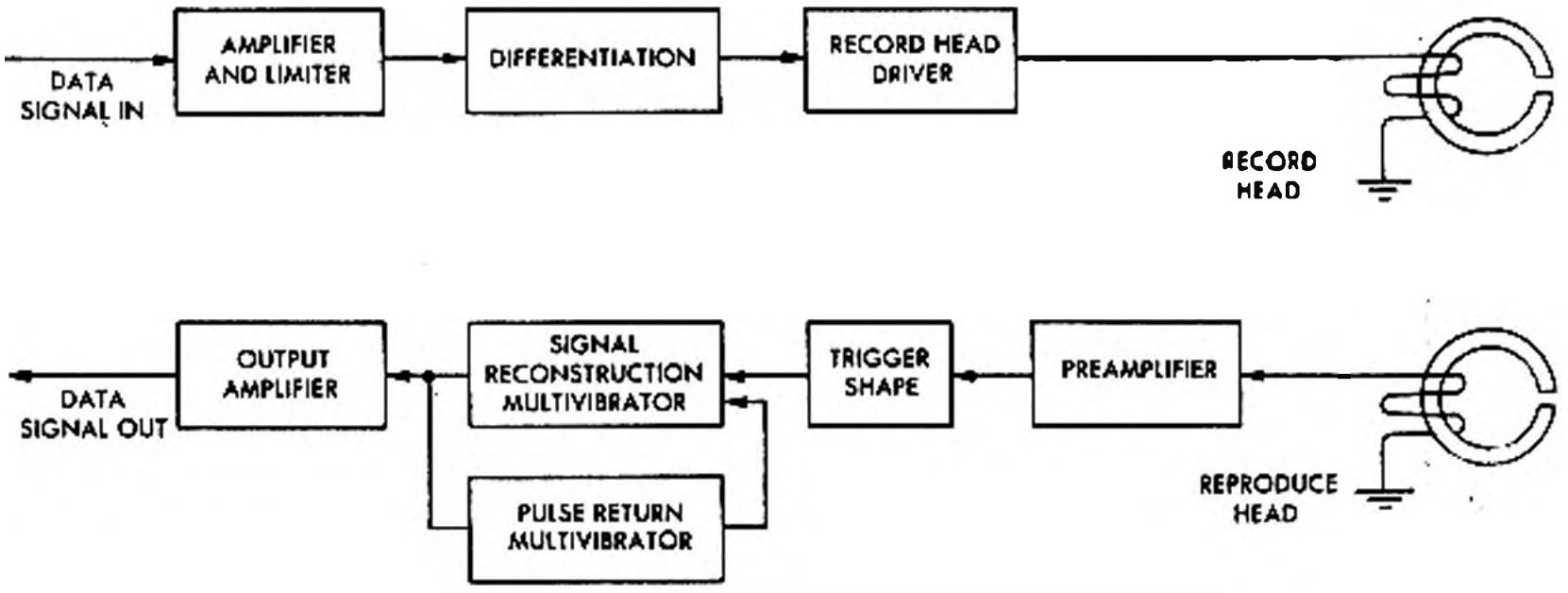


Figure 12-34. Pulse Recording and Reproducing

12-34. The record amplifier differentiates the leading edge of the pulse to a sharper pulse of less duration and records it. The same is done to the trailing edge. The result is that on the tape there are two pulses for every information pulse. The reproduce amplifier has a multivibrator that is turned on and off by these two pulses. The output of the reproduce amplifier is a pulse of the same duration as that of the original signal.

Auxiliary Equipment for Data Reduction

In addition to the galvanometer recording oscillographs and tape recorders, auxiliary equipment and circuits for data reduction and evaluation are employed.

Determining Field Strength of Received Signal. Knowing the field strength of a received signal is useful in determining if the system is functioning properly. For this purpose, the automatic gain control (AGC) current is available. This current is passed through a milliammeter (mounted on the antenna pedestal or near the antenna) so that an indication of the meter can serve as a guide in positioning the antenna.

After proper attenuation, the current is passed through a galvanometer in the oscillograph for recording along with other data.

Calibration of a field-strength trace is accomplished by inserting known levels of signals from a suitable signal generator into the antenna plug of a receiver. The galvanometer attenuator is set to give adequate deflection with high signal level. Then, various levels are recorded to provide deflection vs signal strength data.

Providing Timing Signals. A central ground station at a missile test range usually provides timing signals for use by all participating stations. These signals are recorded by each recorder used. The timing signals used are controlled by IRIG standards although there is some difference from range to range of the particular timing codes available.

At the Air Force Eastern Test Range, the Timing System is a complex of independent central timing systems located at instrumentation stations throughout the Eastern Test Range. These sites fall into the following categories: Ground Stations, Instrumentation Aircraft, Ocean Range Vessels, Sub-Central Installations, and Mobile Installations.

Range time is synchronized to transmissions of the East Coast Loran Chain which is steered by the U.S. Naval Observatory or with transmissions of the National Bureau of Standards station WWV which is periodically synchronized with time information from the U.S. Naval Observatory. The basic oscillator frequency is subdivided into various repetition rates, and both 24-hour and long recycle codes are accumulated.

The correlation accuracy at instrumentation sites within 1000 feet of the timing generator is equal to the central timing accuracy plus a measurable delay of less than 10 microseconds. Most remote sites are served through the timing distribution systems and timing terminal units. Signals

sent through the audio cable distribution system provide end accuracy greater than 1 millisecond at Cape Kennedy, Grand Bahama, Eleuthera, Grand Turk, Trinidad, and Antigua; and in excess of 6 milliseconds at Ascension and Pretoria. Where the UHF radio timing distribution system is used, the end signal accuracy is equal to the timing generator accuracy plus a delay of 10 microseconds when a terminal unit is used. Where relay closures are involved, an additional delay of 20 milliseconds or less is incurred.

The specific time code formats available are:

- IRIG time code formats A, B, D, C, and E.
- NASA codes for Apollo ships and aircraft only.
- Continuous Pulse Repetition Rates: Thirty-three rates, ranging from 1 pulse per hour to 1 mpps, are used for synchronization of instrumentation with range time and for interpolation of time-coded signals.
- Continuous Sine Waves: Ten continuous sine wave frequencies ranging from 1 Hz to 5 MHz, whose zero crossover points are phase locked with other time codes and repetition rates.

PATCH BOARDS. Even though ground stations are large and complex, they are very versatile in the different types of support that they can give. The flexibility comes from the independence of each piece of equipment. All equipment in a ground station is designed for certain types of signals in and certain types of signals out. Depending on the type of program being run, the various needed types of equipment will be connected together by means of patch panels.

Figure 12-35 shows a typical patch panel that could be found at a ground station. This patch panel controls two receivers (RCVR #1 and RCVR #2), a tape recorder (Recorder A), a telemetry indicator (TM IND), an oscillographic recorder (Light Beam RCDR), a pen recorder

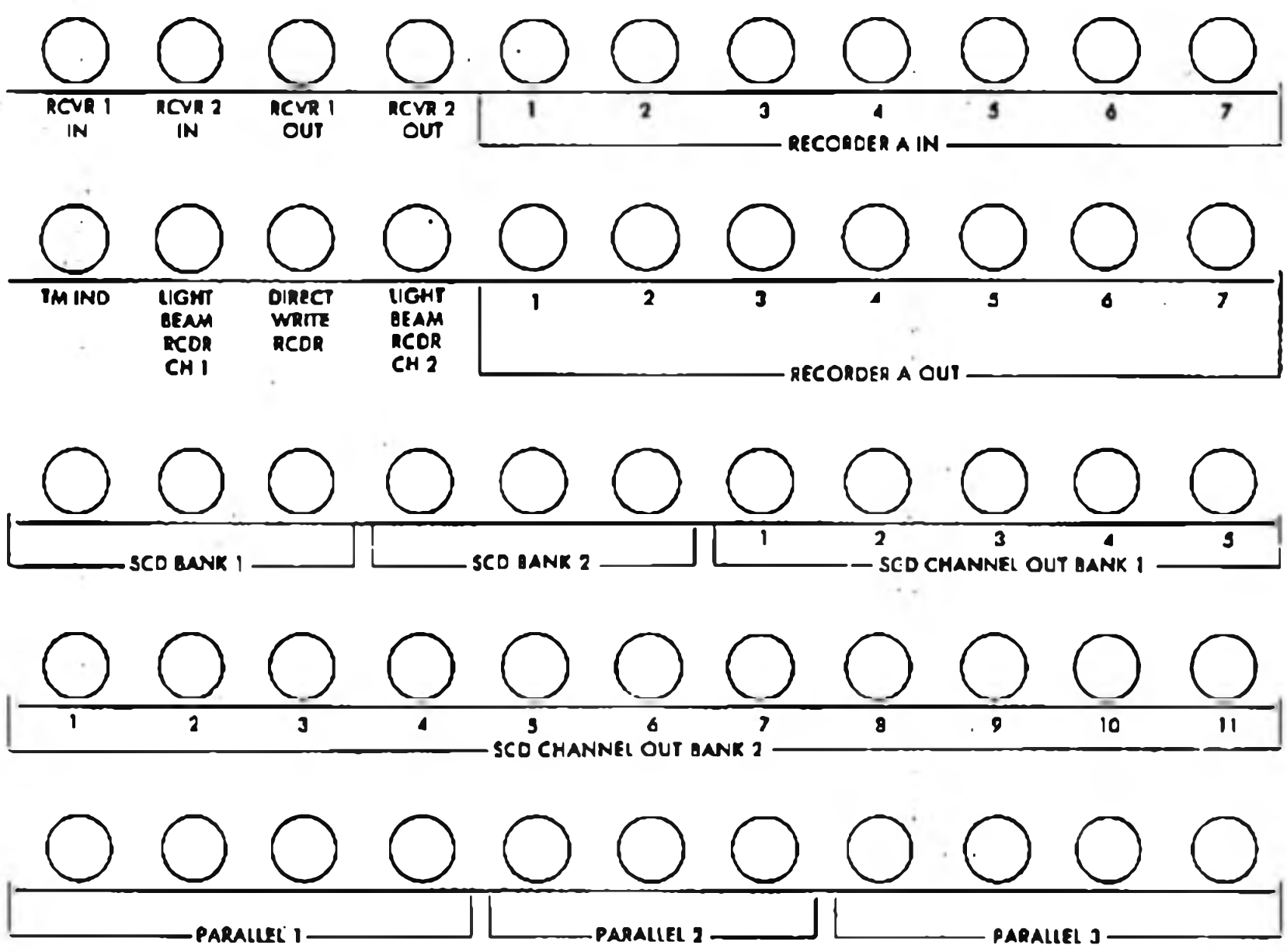


Figure 12-35. Patch Panel

Table 12-1. Telemetry RF Configuration

RF LINK	CONTAINS
#1	SCO Channels 3, 9, 12, D, H
#2	SCO Channels 1, 4, 6, 8, 11, 13, B, D, F, H

Table 12-2. Station Configuration

SCD BANK #1	CHANNEL
Pos 1	3
Pos 2	9
Pos 3	H
Pos 4	D
Pos 5	12

SCD BANK #2	CHANNEL
Pos 1	F
Pos 2	13
Pos 3	D
Pos 4	H
Pos 5	6
Pos 6	8
Pos 7	8
Pos 8	1
Pos 9	11
Pos 10	4

ON TAPE RECORDER TRACK	RECORD
1	RF Link #1
2	RF Link #2 SCD Ch H
3	RF Link #1 SCD Ch D
4	RF Link #2 SCD Ch B
5	RF Link #2
6	RF Link #1 SCD Ch H
7	Blank

ON LIGHT BEAM RECORDER	RECORD
Ch 1	RF Link #1 SCD Ch D
Ch 2	RD Link #2 SCD Ch F

ON DIRECT WRITE RECORDER	RECORD
	RF Link #2 SCD Ch 4

(Direct Write RCDR), two sets of subcarrier discriminators (SCD Bank #1, SCD Bank #2, SCD Channel Out Bank #1, and SCD Channel Out Bank #2). The parallel outputs at the bottom make possible the routing of information to more than one destination.

All the input and output markings on the patchboard are self-explanatory except for the SCD banks. Each SCD bank contains several positions for SCDs; for example, SCD Bank #1 in figure 12-35 has five spaces. The SCD channel that is in any particular position is entirely dependent on which SCD channel you place in that slot. For the test set up on the patchboard in the following paragraphs, slot number 1 contains SCD channel 3, slot number 2 contains SCD channel 9, slot number 3 contains SCD channel H, slot number 4 contains SCD channel D, and slot number 5 contains SCD channel 12. Also notice that there are three inputs to the SCD banks. These are parallel inputs to the whole SCD bank.

There is no correct way to make the connections on a patchboard, except that when the board is patched the signal flow should be as specified by the test conductor. Table 12-1 is a list of signals from a missile. Table 12-2 is an example of ground station configuration and signal flow specifications that may be requested by a test engineer. Table 12-3 is a list of patches that you could use to set up the specifications in table 12-2.

Table 12-3. Patches Needed to Configure Ground Station to Specifications in Tables 12-1 and 12-2

FROM	TO
RCVR #1 Out	SCD Bank #1
RCVR #2 Out	SCD Bank #2
SCD Bank #1	Recorder A In #1
SCD Channel Out Bank #2 Pos 4	Recorder A In #2
SCD Channel Out Bank #1 Pos 4	Parallel #1
Parallel #1	Recorder A In #3
SCD Channel Out Bank #2 Pos 6	Recorder A In #4
SCD Bank #2	Recorder A In #5
SCD Channel Out Bank #1 Pos 3	Recorder A In #6
Parallel #1	Light Beam RCDR Ch 1
SCD Channel Out Bank #2 Pos 1	Light Beam RCDR Ch 2
SCD Channel Out Bank #2 Pos 10	Direct Write RCDR

BY ORDER OF THE SECRETARY OF THE AIR FORCE

OFFICIAL

JOHN D. RYAN, *General, USAF*
Chief of Staff

DWIGHT W. COVELL, *Colonel, USAF*
Director of Administration

Summary of Revised, Deleted, or Added Material

Chap 1: Introductory paragraphs revised to present current events; classification of guided missiles updated.

Chap 2: Unnecessary information deleted and modern concepts added.

Chap 3: Detailed operation of pulse jet deleted and more information on newer propulsion systems added.

Chap 4: Treatment of transistor theory updated.

Chap 5: Pneumatic operated control system deleted; control systems used in ballistic missiles added.

Chap 6 and 9: More solid state circuitry presented in guidance systems.

Chap 9: Several current guidance systems presented; actual system used in AIM-4A Falcon explained from a signal flow analysis; radar and infrared systems as applied to AIM missiles presented; actual inertial guidance system presented to show various functions of components and how they depend on each other—principles can be applied to all inertial type guided missiles; obsolete chanicle guidance system deleted although principles of using hyperbolic lines is retained.

New chap 11: Introduces development and basic principles of laser operation; explains uses of lasers and their application to missiles.

Chap 12 (former chap 11): Revised to include developments on collection and interpretation of test data in missile and satellite programs.

New Atch 1: *Provides* ready reference to more important missiles in combined services inventory.

ATTACHMENT 1

Department of Defense Guided Missiles

This attachment lists in numerical order many, but not all, of the guided missiles used by the combined services. Most of the better known missiles are listed with the exception of those that are obsolete or obsolescent. Not listed are some of the later missiles that are classified, and those that are in limited use.

RIM-2E, F (Terrier)
Navy (Surface-to-air)

Designed for shipboard use, the missile has a solid rocket booster and solid propellant rocket sustainer. Guidance is by a semiactive radar beam-rider homing system. The two-stage missile measures 57 feet in length.

AIM-4A (Falcon)
Air Force (Air-to-air)

The AIM-4A is one of the family of Falcon intercept missiles. Propulsion is by a solid propellant rocket of very short burning time. Guidance is by a semiactive radar homing system.

AIM-4F (Falcon)
Air Force (Air-to-air)

The AIM-4F is an improved intercept supersonic missile with solid propellant and semiactive radar homing system.

AIM-4G (Falcon)
Air Force (Air-to-air)

The AIM-4G is an improved intercept supersonic missile with solid rocket propellant and passive infrared homing system.

RGM-6 (Regulus I)
Navy (Surface-to-surface)

The Regulus is an air-breathing missile, using a turbojet engine. The missile uses inertial guidance. This 34-foot long missile was designed for shipboard launching from a submarine. A later version of the Regulus is the BQM-5C, capable of being launched from more than one environment.

AIM-7D, E (Sparrow III)
Navy/AF (Air-to-air)

The Sparrow is a solid fuel supersonic missile approximately 12 feet long. Guidance is by a semiactive CW homing system. Range is about 8 nautical miles.

YAIM-7F (Sparrow)
Navy/AF (Air-to-air)

The 7F model is an advanced solid state version of the 7E and has greater range, performance, reliability, and larger warhead.

RIM-7H (Sparrow III)
Navy (Surface-to-air)

The RIM-7H is the surface-to-air version of the AIM-7E. It is launched by solid propellant booster and guides to target by semiactive homing. The RIM-7H has improved guidance, folding wings, and clipped fins.

RIM-8F, G (Talos)
Navy (Surface-to-air)

The Talos is a supersonic missile designed for long-range fleet air defense. It has two stages with

an overall length of 30 feet. The Talos has a solid propellant booster with a ramjet sustainer. Guidance is accomplished by beam-riding to the target area, then semiactive homing in terminal phase.

RGM-8H (Talos (ARM))
Navy (Ship-to-surface)

This missile is similar to the RIM-8G except it uses an advanced terminal guidance system. Propulsion is by solid propellant booster with ramjet sustainer.

AIM-9B (Sidewinder)
Navy/AF (Air-to-air)

The Sidewinder is a supersonic, solid propellant missile. The missile length is 9.5 feet and weight is about 185 pounds. Infrared passive homing system is used.

CIM-10B (Bomarc)
Air Force (Surface-to-air)

The Bomarc is a long-range intercept missile ground-controlled by SAGE. Propulsion is by solid propellant booster with ramjet sustainer. Radar homing is used for terminal guidance.

CQM-10A (Bomarc)
Air Force (Surface-to-air)

The CQM-10A is the Bomarc missile modified to drone configuration.

AGM-12B, C (Bullpup)
Navy/AF (Air-to-surface)

The Bullpup is a short range missile used against small defended targets. These 11- and 13-foot missiles use a prepackaged liquid-propellant engine. Radio-link command guidance is used.

CGM-13B (Mace-B)
Air Force (Surface-to-surface)

The Mace-B is a long-range missile designed to be launched from a hardened site. An all-inertial guidance system is employed. This missile measures 44 feet in length with a wing span of 22.5 feet.

MQM-13B (Mace-B)
Air Force (Surface-to-surface)

The MQM-13B is the CGM-13B modified for use as a drone.

MIM-14B, C (Nike-Hercules)
Army (Surface-to-air)

The Nike-Hercules is a high altitude missile with primary ground-to-air mission and secondary surface-to-surface mission. This large intercept missile uses a solid propellant and is guided by command control (target-tracking radars).

RGM-15A (Regulus II)
Navy (Ship-to-surface)

The Regulus II is a large inertial guided missile using a solid rocket booster and turbojet engine.

ADM-20C (Quail)
Air Force (Air-launched)

The Quail is an air-launched, bomber decoy missile. Used to simulate the B-52, the Quail is guided by a gyroscopic autopilot and powered by a turbojet engine.

XMIM-23B (Hawk)
Army (Surface-to-air)

The Hawk is a low-altitude defense intercept missile with semiactive radar homing and a dual-stage, solid-propellant propulsion system.

RIM-24B, C (Tartar)
Navy (Surface-to-air)

The Tartar is a ship-launched missile used for fleet defense. Guidance is accomplished by a radar semi-active homing system. The dual-burn solid propellant provides a speed of Mach 2 over a range exceeding 10 nautical miles.

LGM-25C (Titan II)
Air Force (Surface-to-surface)

The Titan II is a silo-launched ICBM with a range of over 6,300 nautical miles and speed of over 16,000 mph. The Titan II uses liquid propellants, two engines for first stage and one engine for second stage. Guidance is all-inertial.

AIM-26A, B (Falcon)
Air Force (Air-to-air)

The AIM-26 is a supersonic missile with solid rocket propulsion and radar semiactive seeker guidance. The missile is designed for a nuclear warhead.

UGM-27A, B, C (Polaris)
Navy (Surface-to-surface)

The Polaris is a strategic missile fired from submerged submarines. The missile carries a nuclear warhead, has solid propellant propulsion, and is guided by an all-inertial system.

AGM-28A, B (Hound Dog)
Air Force (Air-to-surface)

The Hound Dog is a supersonic strategic medium-range missile. Range is over 600 nautical miles. Guidance is by an all-inertial system. Propulsion is by a turbojet engine.

MGM-29A (Sergeant)
Army (Surface-to-surface)

The Sergeant is a single stage, field artillery type ballistic missile. Its length is 34.5 feet and its weight 10,000 pounds. Propulsion is by solid propellant. Speed is supersonic over a range of 25 to 75 nautical miles. The missile uses an all-inertial guidance system.

LGM-30A, B, F, G (Minuteman)
Air Force (Surface-to-surface)

The Minuteman is a three-stage, solid propellant, silo-launched ICBM. The guidance is all-inertial. Range is 6000-7000 nautical miles.

XMGM-31A (Pershing)
Army (Surface-to-surface)

The Pershing is a two-stage, solid propellant, ballistic missile. The missile travels at supersonic speed, guided by an inertial system. Missile is launched from four-tracked vehicles for mobility. Range may be selected between 100 and 400 nautical miles.

MGM-32A (ENTAC)
Army (Surface-to-surface)

The ENTAC is a solid propellant, wire-guided anti-tank missile. Speed is about 180 mph with a range of 6,600 feet.

XBQM-34E (Firebee)
Navy (Drone)

The Firebee is a turbojet, high-altitude, supersonic aerial target. It provides a realistic high performance aerial target capable of simulating enemy aircraft.

XFIM-43A (Redeye)
FIM-43C
Army (Surface-to-air)

Using an infrared seeker and an electromechanical guidance device, the Redeye seeks out and destroys low-flying enemy aircraft. The weapon is carried and launched by one man.

UUM-44A (SUBROC)
Navy (Underwater-to-underwater)

The SUBROC is a supersonic, anti-submarine solid rocket with inertial guidance. It is fired from submarine torpedo tubes, leaves the water in flight, reenters the water and sinks to a pre-determined depth to detonate.

AGM-45A (Shrike)
Navy/AF (Air-to-surface)

The Shrike is a solid propellant, short-range missile designed for firing by attack aircraft against enemy radar-controlled air defenses and radar installations. This anti-radiation missile uses a passive radar-homing guidance system with the target furnishing the guidance information.

XLIM-49A (Nike-Zeus)
Army (Surface-to-air)

The Nike-Zeus is a three-stage, solid propellant, command-guided anti-missile. It is designed to intercept and destroy hostile ballistic missile warheads through the coordinated function of ground guidance and ground support systems.

MGM-51A, B, C (Shillelagh)
Army (Surface-to-surface)

The Shillelagh is a direct fire, command to line-of-sight guided missile powered by solid propellants. The missile is designed for infantry support against armor.

XMGM-52B, C (Lance)
Army (Surface-to-surface)

The Lance is a division support missile using a simplified inertial guidance system and pre-packaged, storable liquid rocket for propulsion. Range is from 3 to 30 miles.

AGM-53A (Condor)
Navy (Air-to-surface)

The Condor is a long-range electro-optical guided missile using a pre-packaged liquid propellant for propulsion.

AIM-54A (Phoenix)
Navy (Air-to-air)

This is a single-stage, solid propellant missile. The Phoenix has extensive range, altitude, and all-weather capabilities. Guidance by radar has both active and semi-active modes.

ZAGM-65A (Maverick)
Air Force (Air-to-surface)

The Maverick is an electro-optical homing missile to be used against small hard targets.

YRIM-66A (Standard (MR))
Navy (Surface-to-air)

This is a surface-to-air supersonic missile with surface-to-surface capability for shipboard use. It uses a dual-thrust solid-propellant rocket motor. The continuous-wave semiactive homing system is fully solid state.

YRIM-67A (Standard (ER))
Navy (Surface-to-air)

The YRIM-67A is also a surface-to-air supersonic missile with surface-to-surface capability for shipboard use. It uses a two-stage solid propellant; a booster stage which separates after boost and sustainer stage. The CW semiactive homing system is fully solid state.

AGM-69A (SRAM)
Air Force (Air-to-surface)

The SRAM is a supersonic, solid fuel missile capable of several flight trajectories. Design details are classified.

XMIM-72B
Army (Surface-to-air)

The XMIM-72B is a supersonic, IR seeker missile accomplishing attack on low flying aerial targets.

ZUGM-73A (Poseidon)
Navy (Surface-to-surface)

The ZUGM-73A is an inertially-guided, solid-propellant ballistic missile, launched from a submarine at enemy land targets. Range is approximately 2,500 nautical miles.

AGM-78A (Standard Arm)
Navy (Air-to-surface)

The AGM-78A is a tactical anti-radiation missile. Propulsion is by pre-packaged liquid propellant. Guidance is electro-optical.

ATTACHMENT 2

Index

A

Aberration of light rays.....4-35
Absorption 11-7
Absorption spectrum.....4-30
Acceleration 4-3
Accelerometer 6-15, 9-16
  angular 4-3
  compensation computer..9-38
  integrating 6-18
  servo loops.....9-24
Acceptor atom.....4-44
Accumulator 5-71
Acoustic sensor 6-13
Active homing systems.....9-74
Actuator 7-21
  combination-type 5-74
  constant-speed 7-21
  electric 5-72, 7-21
  hydraulic 5-66
  mechanical linkage.....5-75
  pneumatic 5-72
  purpose 5-66
  units 5-66, 6-77
  variable-speed 7-21
Advanced propulsion
  systems 3-37
Aerial target
  considerations 10-5
Aerodynamic control.....2-19
Aerodynamic loads.....2-19
Aerodynamics, ballistic
  missile 2-28
Aerodynamics of
  supersonic flight....2-12, 2-37
Aerodynamics, propulsion....2-28
Aerospace ground
  equipment 10-14
Aerospace vehicles
  classification letters.....1-9
  mission 1-12
  types 1-13
Afterburners 3-29
Aileron 2-21
AIM-4A Falcon missile.....9-70
AIM-4D Falcon missile.....9-63
Air-breathing propulsion
  systems 3-4, 3-18
Airflow 2-17, 2-31
Airfoil arrangements 2-19
Airframe 2-29, 5-52
Air log 6-65
Airspeed transducer.....5-18
Altimeter, pressure.....5-16
Altimeter, radio.....5-18
Altitude computer.....9-30
Altitude programmer.....9-38
American Rocket Society.....1-4
AM telemetry.....12-26
Amplidyne 5-64
Amplifier
  circuits 7-20
  feedback 6-38
  frequency-selective 6-70
  magnetic 5-60
  mixing 6-73
  paraphase 5-56
  power-type 5-59
  summing 6-41
  transistor 5-57
  units 5-53, 6-69, 12-25
  voltage 5-53
Amplitude modulation 4-56
Analog computer.....6-24, 6-36
Aneroid barometer.....4-17
Angstrom unit...4-30, 11-2, 11-10
Angular acceleration.....4-7
Antenna 4-40, 6-6
  arrays 6-8
  dipole 6-8
  FM/FM reception.....12-33
  half-wave 6-7
  Hertz 6-7
  Marconi 6-7
  polarization 6-6, 6-23
  radar 6-7
Artillery-type trajectory.....8-6
Aspect ratio.....2-10
Atmospheric pressure.....4-16
Atomic theory.....11-2
  Bohr atom model.....11-4
  complex atoms.....11-6
  definition 4-42, 11-2
  electrons 11-2
  energy gap.....11-5
  energy levels.....11-4
  excited state 11-5
  ground state.....11-5
  hydrogen atom.....11-3
  ions 11-2
  metastable state.....11-6
  neutrons 11-2
  nucleus 11-2
  Pauli's exclusion
    principle 11-6
  permissible orbits.....11-4
  photons 11-4
  Planck's hypothesis.....11-4
  planetary electrons.....11-3
  protons 11-2
  quantum physics.....11-4
  shells and subshells.....11-6
  spontaneous emission....11-6
Automatic celestial
  navigation (ACN).....9-47
Automatic long-range
  Loran 9-51
Autopilot 7-20
Autosyn 6-34
Auxiliary control systems....7-24
Avogadro's law.....4-19
Axes, transformation of.....8-3
Axial loads.....2-34
Axis
  lateral 2-12
  longitudinal 2-12
  pitch 2-12
  reference 8-1
  reference programmer...7-20
  roll 2-12
  vertical 2-11
  yaw 2-11

B

Balancing adjustments....7-27
Ballistic camera....12-5
Ballistic missile
  aerodynamics....2-28
  airframe environment...2-29
  airframe structure....2-29
  axial loads....2-34
  base drag....2-32
  bending loads....2-35
  dynamic loads....2-36
  effects....2-32
  flight loads....2-36
  forces....2-30
  friction drag....2-32
  fundamentals....2-30
  heating....2-33
  high-speed....2-37
  internal pressure....2-36
  loads....2-36
  parameters....2-31
  progress reports....2-38
  structure design and
    environment....2-33
  thermal loads....2-36
  trajectory
    characteristics....2-33, 8-4
    wind tunnels....2-37
Ballistic missile
  control systems....7-11
Ballistic-type trajectory....8-6
Ballistite....3-8
Bandpass filter....6-14
Barometer....4-16
Base drag....2-32
Basic inertial
  guidance system....9-15
Basic radar
  plotting system....12-10
Beam-rider course....8-4
Beam-rider
  guidance system..6-3, 9-12
Bending loads....2-35
Bernoulli's theorem....2-4, 4-15
Bias effects....9-47
Bias voltages....4-53
Binary numbers....6-42
Blast-effect warheads....10-8
Bohr atom model....11-3
Bolometer....6-11
Bomarc missile....10-16
Boolean algebra....6-44
Booster assembly....3-38
Boron atom....11-3
Boundary layer....2-5
Boyle's law....4-18
Brake-clutch....9-20
Bridge circuits....5-25, 6-20
Burble point....2-9

C

Calibration
  guidance phase....9-2
Calibration oscillators....12-34
Cam calibrator....12-21
Camera
  ballistic....12-5
  cinetheodolite....12-4
  ribbon-frame....12-5
  tubes....4-41, 6-12
Canard structure....2-25
Candlepower....4-22
Capacitance pickoff....5-27
Carrier waves....4-54
Cartesian coordinates....6-35
Casualty agents....10-9
Cathedral....2-7
Celestial coordinates....8-14
Celestial guidance
  systems....6-3, 9-43
Celestial latitudes....8-14
Celestial navigation....6-3, 9-43
Centaur missile....12-28
Center of gravity....4-12
Center of pressure....2-8
Center of oscillation....4-14
Centrifugal force....4-5
Centripetal force....4-5
Channel interconnection....7-7
Checkout facility....10-14
Chemical warhead....10-9
Chopper, mechanical....5-53
Chopper, vacuum-tube....5-53
Chlorine pentafluoride,
  trifluoride....3-14
CIM-10A missile....10-16
Cinetheodolite....12-4
Classification of
  guided missiles....1-9
Closed-loop system....7-24
Coding and decoding units...6-24
Coherence....11-7
  spatial....11-8
  temporal....11-8
Coincidence tube....6-27
Collinear slot arrays....6-8
Color....4-20
Combination-type
  actuators....5-74
Combustion chamber....3-16
Command guidance
  systems....6-2, 9-3
Commutator....5-55, 6-21
Commutator, telemetry....12-23
Compass....9-55
Complex atoms....11-6
Composite guidance
  systems....9-80
Composite propellants....3-8
Compressibility of gases....4-18
Computer....5-27, 9-26
  analog....6-24, 6-36
  coding and decoding....6-24
  digital....6-25, 6-41
  functions....5-45
  integrators....5-15
  mixers....5-27, 6-41
  navigation....9-35
  rate components....5-40
  units....5-27, 6-24
Conical scanning....6-21, 9-58
Conservation of matter....4-10
Console, test....7-27
Constant dive-angle system...9-76
Constant-speed motor....7-21
Constant true bearing (CTB)..9-61
Control, jet....2-26
Control matrix....9-81
Control, starting speed....2-26
Control surface
  aerodynamic....2-19
  canard....2-25
  dual-purpose....2-24
  external....2-19
  primary....2-19
  secondary....2-21
  slots....2-24
  spoilers....2-24
  tabs....2-21
  variable-incidence....2-25
Control systems
  auxiliary....7-24
  ballistic missile....7-11
  checkout....7-25
  components....5-1
  definition....7-1
  electric....7-21
  energy sources....7-1
  hydraulic-electric....7-3
  pneumatic....7-11
  pneumatic-electric....7-11
  power sources....7-1
  requirements....5-1, 7-1
  types....7-1
Controller units...5-61, 6-75, 7-21
Coordinates
  applications....8-16
  Cartesian....6-35
  celestial....8-14
  cylindrical....8-15
  geometric....8-15
  guidance....8-1
  polar....6-36
  rectangular....8-15
  spherical....8-16
Coriolis force....4-9, 8-9
Counting, computer....6-47
Course terms....8-12
Critical angle of incidence...4-28

Cryogenic materials...3-10, 11-35
Curves, trajectory.....8-4
Cylindrical coordinates.....8-15

D

Dalton's law.....4-18
Damper 6-15
Data reduction
auxiliary equipment....12-39
Decoding,
computer units....6-24, 6-46
Delay circuits.....6-29, 6-50, 6-56
Demodulation, laser.....11-29
Development of
guided missiles.....1-8
Differential analyzer....6-25, 6-37
Differential selsyn.....5-22, 5-33
Differentiator, rate
components 5-40
definition 5-42
RC amplifier.....5-44
Digital computer.....6-25, 6-41
Dihedral 2-7
Dipole radiators.....6-8
Direct-recording
instrumentation 12-35
Directional gyro.....5-15
Discriminator 6-26
phase comparator.....6-33
pulse-recurrence
frequency 6-32
pulse-spacing 6-30
pulse-width 6-26
units.....12-34
Displacement gyro.....5-12, 7-20
Donor atom.....4-44
Doppler 4-70, 9-82
Double-base propellants.....3-7
Drag 2-3
base 2-32
coefficient 2-9
determination 2-9
friction 2-32
induced 2-10
parasite 2-10
profile 2-11
reduction 2-15
Dual-purpose controls.....2-24
Ducted engines.....3-4
Dynamic loads.....2-36
Dynamic servo
response checks.....7-29

E

Eccles-Jordan multivibrator..6-48
Eddies 2-3
Electric actuation checkout...7-26
Electric actuators.....5-72
Electric control systems.....7-21
Electric-hydraulic
control systems.....7-3
Electric pickoff.....7-3
Electric-pneumatic
control systems.....7-11
Electrical point-to-point
checkout 7-27
Electrical propulsion 3-38
Electrical timer.....5-48
Electromagnetic sensor.....6-6
Electromagnetic
theory of light.....4-20
Electromagnetic waves.....4-39
Electronic mixer.....5-28
Electronic rate circuits.....5-43
Electrons 4-42, 11-2
Elevator 2-21
Emission
characteristics 11-13
induced 11-7
spectrum 4-30
stimulated 11-7
Energy
characteristics 11-7
definition 4-10
gap 11-5
kinetic 4-10
levels 11-4
potential 4-10
sources 7-2
transfer 5-76
Engine
ducted system.....3-4
pulsejet 3-5, 3-20
ramjet 3-5, 3-21
turbofan 3-29
turbojet 3-25
turboprop 3-29
Engine-tachometer
switching unit.....12-21
Equilibrium 4-10
Erection, gyro.....5-12
Error signals.....7-29
Ether 4-21
Ethyl alcohol 3-11
Exhaust nozzle 3-17, 7-15
Exhaust velocity 3-7
External guidance reference..6-53
External instrumentation
system 12-2

F

Falcon missile.....9-61
Feedback amplifier.....6-38
Feedback units 5-77, 6-78
Fission 10-11
Fixed camera 12-4
Flat trajectory 8-7
Flight loads 2-36
Flight terminology
airfoil 2-5
attitude 2-6
axis 2-7
camber 2-6
cathedral 2-7
chord 2-6
dihedral 2-7
drag 2-5, 2-9
lift 2-5
stability 2-6
streamlines 2-5
wing span 2-6
Fluorescence 4-32
Fluorine 3-13
Flutter 2-24
Flux gate compass.....9-55
Flux valve 5-14
FM loran 9-52
FM receiver 12-32
FM/FM telemetry systems..12-17
antenna 12-26, 12-33
cam calibrator 12-21
commutator-gating
unit 12-23
engine-tachometer
switching unit.....12-21
fuel totalizer 12-21
power-supply unit 12-21
RF power-amplifier
unit 12-25
signal-converter unit ...12-21
subcarrier oscillator
mounting unit 12-22
subcarrier oscillator
unit 12-19
transmitter unit 12-25
voltage-controlled
subcarrier oscillator ..12-22
Followup 7-11
Followup generator 5-79
Followup loops 5-78
Followup units 5-77
Foot-candle 4-22
Force
acting 2-2
centrifugal 4-5
centripetal 4-5
Coriolis 4-9, 8-9
definition 2-3
Foucault's method 4-25
Fragmentation warhead 10-8
Free gyro 5-9
Frequency changers 6-54
Frequency detector 12-34
Frequency modulation (FM)..4-61
Frequency-selective
amplifier 6-70
Friction, boundary layer 2-5

Friction drag 2-32
Fuel, liquid 3-10, 3-31
Fuel totalizer 12-21
Fusion 10-11
Fuzes for missile warheads..10-12

G

Gallium-arsenide
laser 11-17, 11-27
Gas
atmospheric pressure ...4-16
Avogadro's law 4-19
Boyle's law 4-18
compressibility 4-18
Dalton's law 4-18
definition 4-15
helium 3-14
high pressure 3-14
kinetic theory 4-16
laser 11-11
mechanics of 4-15
nitrogen 3-15
specific gravity 4-18
viscosity 4-19
Gaseous laser radar 11-31
Gates, computer 6-47
Gating unit, commutator ...12-23
General laser
safety precautions 11-35
Generator
amplidyne 5-64
followup 5-79
Geometric coordinates 8-15
Germanium crystal 4-44
Goddard, Robert H. 1-3
Goniometer 6-36, 6-60
Grain, propellant 3-8
Gravity
apparent 8-9
center of 4-12
effects of, on trajectories..8-9
Great circle route 8-12
Ground stations, telemetry..12-28
Ground support
equipment (GSE) 10-14
Guidance coordinates,
trajectories 8-1
Guidance methods 6-2
Guidance references 6-53
Guidance systems 9-1
actuator units 6-77
AIM-4A missile 9-70
AIM-4D missile 9-63
amplifier units 6-69
beam-rider 6-3, 9-12
celestial 6-3, 9-43
classification 9-2
command 6-2, 9-3
composite 9-80
computer units 6-24
controller units 6-75
feedback units 6-78
gyro-stabilized 9-15
homing 6-4, 9-57
hyperbolic 6-3, 9-10, 9-49
inertial 6-4, 9-15, 9-43
long-range 9-15
magnetic 9-55
methods 6-2
mapmatching 9-54
preset 6-2, 9-3
reference units 6-53
sensor units 6-2
short-range 9-2, 9-10
target approach 9-57
terminal 9-57, 9-76
terrestrial 6-2, 9-52
Guided missile
classification 1-9
configuration 2-15
control 2-19, 7-1
designators 1-9
desirable characteristics...1-1
development 1-8
history 1-3
operations 10-1
propellants 3-6
stability 2-27
trajectories 8-4
Guillemin line 6-64
Gyro stabilized
inertial guidance 9-15
Gyroscope 4-9, 5-4, 9-18
air bearing 5-11
directional 5-15
displacement 5-12, 7-20
drift 5-7
erection and
slaving circuits 5-12
error 5-10
floated unit 5-10
flux valve 5-14
gimbal lock 5-7
horizontal 5-14
integrating 5-11
laser 11-32
pitch control 7-5
precession 4-8, 5-4
rate 5-7, 5-45, 7-20
rigidity 5-4
roll control 7-4
S-1 directional control ..5-15
stable platform 9-19, 9-21
types 4-9
uses 5-8
vertical 5-13
Gyrosyn compass 9-55

H

Harmonic motion 4-12
Heading demand circuit 9-35
Heat barrier 2-13
Helical beam antenna 12-33
Helical scanning 6-22
Helium gas 3-14
Helipot 5-24
High pressure gases 3-14
History of guided missiles 1-3
American rocketry 1-3
German rockets 1-4
jet engines 1-4
World War II 1-5
Homing guidance
systems 6-4, 9-57
Horizontal gyroscope 5-14
Horn-type radiators 6-8
Hydraulic accumulator 5-71
Hydraulic actuation
checkout 7-26
Hydraulic actuator .....,5-66
Hydraulic-electric
control system 7-3
Hydraulic mixer 5-35
Hydraulic pump 5-69
Hydrazine 3-11
Hydrocarbon fuels 3-12
Hydrogen atom 11-3
Hydrogen, liquid 3-12
Hydrogen peroxide 3-14
Hyperbolic guidance 6-3
system curves 8-4
systems 9-10, 9-49
Hypersonic speed 2-13

I

Iconoscope 4-41
Ignition system 3-17, 3-31
Impact fuze 10-12
Impulse, specific 3-6
Impulse, total 3-6
Incendiary warheads 10-9
Induced drag 2-10
Induced emission 11-7
Inertia 4-2
Inertia sensors 6-15
Inertial guidance
systems 6-4, 9-15
Inertial guidance, terminal ..,9-76
Infrared image converter ...11-22
Infrared pickoff 6-20
Infrared radiation 6-10
Infrared sensor 6-10
Injection laser ..11-2, 11-20, 11-27
Injector 3-16, 3-36
Input units, computer 6-45

Instrumentation 12-1
classification 12-4
definition 12-1
direct coupling 12-35
external 12-2
internal 12-10
optical equipment 12-3
radar equipment 12-6
Integrated circuits 4-54
Integrating accelerometer 6-18
Integrator 5-35, 9-16
action 7-7
integrating gyro 5-40
Miller 6-72
RC 5-38
RL 5-40
thermal 5-40
use 5-36
variable-speed motor 5-37
Intensity of light 4-21
Intermittent engine 3-5
Internal guidance references..6-53
Internal instrumentation
systems 12-10
Internal pressure 2-36
Inverse square law 4-23
Ions 11-2

J

JB-2 engine 3-20
Jet control systems 2-26
Jet propulsion systems 3-4
air-jet engines 3-18
basic components 3-16
basic formulas
and laws 3-1
classification 3-4
combustion chamber 3-16
definition 3-1
ducted engines 3-4
rocket motors 3-27
Jetstream 12-2
Jet vanes 2-26, 7-15
Joules 11-4

K

Kerosene 3-12
Kerr cell 11-26
Kinetic energy 4-10
Kinetic theory of gases 4-16

L

Lambert 4-23
Laser
application 11-1
biological effects 11-35
classification 11-17
coherence 11-7
construction 11-18
cryogenic materials 11-35
definition 4-53, 11-1
demodulation 11-29
energy and emission
characteristics 11-13
gas 11-11
gaseous laser radar 11-31
gyroscope 11-32
history 11-1
injection 11-2, 11-20
modulation 11-24, 11-27
noise effect 11-31
population inversion ...11-7
pulsed ruby 11-1, 11-9
pumping 11-8, 11-23
radar 11-29
Raman laser 11-17
resonant cavities 11-8
ring laser 11-32
ruby laser 11-9
safety hazards 11-35
stimulated emission 11-7
types 11-2, 11-17
Lateral axis 2-12
Launch environment symbols.1-11
Launch guidance phase 9-2
Launchers
air-launch 10-14
ground-launch 10-14
monorail 3-43
multistage 3-43
silo 3-39
tower 3-43
zero-length 3-39
Launching methods 3-38
Launching facilities 10-14
Laws of motion,
Newton's 2-3, 4-1
Lead, defined 7-29
Lead-angle course 8-4
Lead circuit 5-42
Lead homing 9-61
Lenses 4-35
LGM-25 missile 10-17
LGM-30 missile 10-19
Lift 2-3
coefficient of 2-9
determination of 2-9
effectiveness 2-16
Light
brightness 4-23
candlepower 4-22
color 4-20
electromagnetic
theory of 4-20
ether 4-21
fluorescence 4-32
flux 4-21
intensity 4-21
lambert 4-23
measurement 4-24
phosphorescence 4-32
polarization 4-31
rectilinear,
property of 4-21, 4-37
reflection 4-26
refraction 4-26, 9-47
spectra 4-29
ultraviolet 4-20
velocity 4-25
waves 4-20, 4-39
Light-sensitive sensor 6-11
Line of position (LOP) 8-12
Liquid propellant 3-10
bipropellant 3-10
classification 3-10
comparison 3-14
cryogenic 3-10
engines 7-19
fuel 3-11
hydrogen 3-12
kerosene 3-12
monopropellants 3-10
oxidizer 3-13
oxygen 3-13
storable 3-10
thrust cutoff control 7-19
Liquid rockets 3-31
LITV control 2-26, 7-17
Lobe switching scanning 6-22
Logical unit, computers 6-47
Longitudinal axis 2-12
Long-range guidance
system 9-15, 9-47
Loran 6-6, 9-51
Lumen 4-21

M

Mach angle 2-13
Mach number 2-13, 2-31
Magnetic amplifier 5-60
Magnetic compass 9-55
Magnetic field 8-10, 9-55
Magnetic guidance
systems 9-55
Magnetic sensor 6-14
Magnetic-tape pickoff 6-66
Magnetometer 6-14
Maintenance facility 10-14
Manometer 6-17
Mapmatching
references ...6-68, 9-54, 9-75
Map projection 8-11
Maser 11-1
Matrix, control 9-81
Measurement of light 4-21
Mechanical chopper 5-53

Mechanical filter 9-19
Mechanical linkage,
actuator 5-75
Mechanical mixer 5-34
Mechanical systems
digital computer 6-51
Mechanical timer 5-47
Mechanics of gases 4-15
Medical surveillance
program (laser) 11-35
Memory unit,
digital computer 6-50
Mercury column 4-16
Mercury tank sonic
delay line 6-50
Meteorological effects
on trajectories 8-8
Microphones 6-13
Microsyn 5-24
Midcourse guidance phase ...9-2
Miller integrator 6-72
Minuteman missile 10-19
Mirrors 4-33
Missile
air-launched 10-14
course terms 8-12
ground-launched 10-15
ground support
equipment 10-14
launching methods 3-38
launching facilities 10-14
range instrumentation ...12-2
selection 10-6
terminal guidance 9-57
trajectory curves 8-4
warheads 10-7
Mixers 5-27
differential selsyn ..5-22, 5-33
electronic 5-28
mechanical 5-34
parallel mixing 5-28
pneumatic and
hydraulic 5-35
requirements 5-27
resistance bridge 5-30
series mixing 5-30
transformer mixing 5-31
transistorized 5-33
types 5-28
vacuum-tube 5-32
Mixing amplifier 6-73
Mixture ratio 3-7
Modulation
amplitude 4-56
definition 4-54
Doppler principle ..4-70, 9-82
frequency 4-61
index 4-62
laser 4-53, 11-24
percentage 4-58
phase 4-64
power distribution 4-59
pulse 4-64
rate of deviation 4-61
single sideband 4-60
Modulator, rectifier 5-55
Moment of inertia 4-2
Monochromaticity 11-2
Monocoque design 2-30
Monopropellant 3-10, 3-31
Monorail launcher 3-43
Motor
actuator 7-21
servo 6-77
timer 5-48
Mu metal 12-15
Multistage launching 3-43
Multivibrator
bistable 6-55
Eccles-Jordan 6-48
start-stop 6-58

N

Navier-Stokes equations 2-31
Navigation
celestial 6-3, 8-13, 9-43
computer 9-35
long-range
hyperbolic 6-3, 9-15
practical 8-11
short-range
hyperbolic 6-3, 9-10
Newton's laws
of motion 2-3, 3-1, 4-1
Neutrons 4-42
Nitric acid 3-13
Nitrogen gas 3-15
Nitrogen liquid 3-15
Nitrogen tetroxide 3-13
NPN junction transistors 4-47
Nuclear
fission 10-11
fusion 10-11
propulsion 3-38
warheads 10-10
Nucleus 11-2

O

Open-loop system 7-24
Optical devices 4-33
Optical equipment 12-3
Optical spectrum 11-8
Optics 4-20
Oscillator 6-53
calibration 12-34
subcarrier, unit ..12-19, 12-22
Oscillograph, recording 12-36
Output unit, computer 6-51
Oxidizer, liquid 3-13
Oxygen, liquid 3-13

P

Parabolic antennas 12-33
Parabolic reflectors 6-8
Parallel mixing 5-28
Parameters, aerodynamic 2-31
Paraphrase amplifier 5-56
Parasite drag 2-10
Passive homing systems 9-61
Pauli's exclusion
principle 11-6
Pendulum 4-13
Pen recorder 12-35
Phantastron circuit 6-58
Phase comparator 6-33
Phase modulation 4-64
Phosphorescence 4-32
Photomultiplier tube 6-12
Photometer 4-24
Photons 11-4
Phototube 6-11
Physical inspection
checkout 7-27
Physical principles 4-2, 4-19
Physical references 5-52
Physics of flight 2-1
Pickoff 5-19
capacitance 5-27
commutation principles ..6-21
electric 7-3
infrared sensing 6-20
magnetic-tape 6-66
potentiometer 5-23
probe 6-19
radar system 6-19
reluctance 5-26
selsyn 5-20
signal 6-18
thermistor 6-20
transformer 6-19
Pitch 2-21, 7-5
Pitch axis 2-12
Planck's hypothesis 11-4
Planetary electrons 11-3
PN junction transistors 4-45
PNP junction transistors 4-49
Pneumatic actuator 5-72
Pneumatic control systems ...7-11
Pneumatic-electric
control systems 7-11
Pneumatic mixers 5-35
Pneumatic timer 5-50
Point-contact transistors 4-50
Polarization 4-31, 6-6, 6-23

Population inversion 11-7
Potential energy 4-10
Potentiometer pickoff 5-23
Power sources 7-2
Power supply computer 9-42
Power supply systems check ..7-26
analysis of results 7-30
Power supply unit 12-21
Precession 4-8, 5-4
Predictor circuit 5-42
Preset guidance systems ..6-2, 9-3
Pressure altimeter 5-16
Pressure-relief valve 5-71
Primary control surfaces 2-21
Printed circuits 4-54
Prism 4-29, 4-34
Profile drag 2-11
Programmers 6-65, 7-20, 9-27
Propellants 3-6
high-pressure gases 3-14
liquid 3-10
solid 3-7
Propellant performance terms
exhaust velocity 3-7
mixture ratio 3-7
specific impulse 3-6
specific fuel consumption 3-7
specific thrust 3-7
total impulse 3-6
Proportional guidance 9-61
Propulsion aerodynamics 2-28
Propulsion systems
advanced 3-37
air breathers 3-18
electrical 3-38
liquid propellant 3-10
nuclear 3-38
solid propellant 3-7
Protons 4-42
Proximity fuzes 10-13
Pulse-forming circuit 6-62
Pulsejet engine (JB-2) ...3-5, 3-20
Pulse modulation 4-64
Pulse-recurrence frequency discriminator 6-32
Pulse-spacing discriminator ..6-30
Pulse-time multiplex 4-65
Pulse-width discriminator ...6-26
Pulsed ruby laser 11-1, 11-9
Pump, hydraulic 5-69
Pumping (laser) 11-8, 11-23
Pursuit curve 8-4
Pursuit homing 9-60
Push-over arc 9-77

Q

Quanta 11-3
Quantum physics concepts ...11-4

R

Radar
command guidance 9-8
Doppler 9-82
equipment 12-6
laser 11-29
mapmatching guidance .9-54
pickoffs 6-19
plotting system 12-10
sensor 6-7
telemetry system 12-10
tracking system 12-6
Radio altimeter 5-18
Radio propagation 8-11
Radio sensor 6-6
Radius of gyration 4-2
Raman laser 11-17
Ramjet engine 3-5, 3-21
Range computer 9-32
Rate circuit 5-42
Rate components 5-40
Rate gyro 5-8, 5-45, 7-20
Raw data recording 12-35
RC amplifier differentiator ...5-44
RC integrator 5-38
Reactor, saturable 5-59
Real-time data 12-1
Receiver calibration
equipment 12-33
calibration oscillator ..12-34
discriminator units 12-34
Receiver, command 9-6
Receiver, FM 12-32
Recessed antennas 6-9
Recorder
oscillograph 12-36
pen 12-35
tape 12-37
Recording optical tracking instrument (ROTI) 12-5
Rectangular coordinates 8-15
Rectifier commutator 5-56
Rectifier modulator 5-55
Reference
axes programmer 7-20
concepts for trajectories ..8-1
external guidance 6-53
internal guidance 6-53
navigational, stars 8-14
physical 5-52
terrestrial 9-52
time 5-46
voltage 5-45
Reflection of light 4-26
Refraction of light 4-26, 9-47
Relativity of motion 2-1
Relaxation time 2-13
Relay, controller units 6-75
Relay switch 5-62
Reluctance pickoff 5-26
Research and development ...1-8
Resistance bridge mixer 5-30
Resonant cavities 11-8
Reynolds number 2-13, 2-31
RF power-amplifier unit ...12-25
Rhumb line 8-12
Ribbon-frame camera 12-5
Rigidity, gyroscope 5-4
Ring laser 11-32
RL integrator 5-40
Rocket motor 3-27
classification 3-30
propulsion units 3-37
liquid 3-31
solid 3-30
Rocket systems 3-4
Rocket trajectories 8-7
Roll 2-21, 7-4, 7-18
Ruby laser 11-9
Rudder 2-21

S

S-1 directional gyro 5-15
Safety precautions, laser ...11-35
beam termination 11-35
buddy system 11-36
cryogenic materials 11-35
eye exposure 11-35
eye protection 11-35
high voltage 11-35
skin protection 11-35
toxic chemicals 11-35
Sagnac ring laser 11-33
Saturable reactor 5-59
Sawtooth scanning 6-21
Scanning methods 6-21, 9-58
Scintillation 4-28
Schuler tuning 9-20
SCR-584 radar 12-7
Secondary control surfaces ...2-21
Selsyn, differential 5-22, 5-33
Selsyn pickoffs 5-20
Selsyn, two-speed 5-21
Semiactive homing system ...9-70
Semiconductor laser 11-17
Sensors 6-2
acoustic 6-13
altimeter 5-16
electromagnetic 6-6
gyroscope 5-4
inertia 6-15
infrared 6-10, 6-20
light sensitive 6-11
magnetic 6-14
pickoffs 5-19, 6-18
polarization methods 6-5, 6-23

radar 6-7
radio 6-6
scanning methods 6-21
transducers 5-18
units 5-4
Series mixing 5-30
Servomotor 6-77
Servo multiplier 6-39
Sextant 9-45
Shaped charge warhead 10-9
Shells and subshells, atom ...11-6
Shock wave 2-12
normal 2-14
oblique 2-15
Shoran 6-6
Short-range guidance
systems 9-2
Signal converter unit 12-21
Signal, followup 5-77
Signal, pickoff 6-18
Signal resolution 9-21
Signal simulation 7-30
Silo launching 3-39
Simulator 6-37
Slaving, gyro 5-12
Slot 2-24
Smokes, warheads 10-9
Snell's law 4-26
Snooper-scope 11-22
Solar spectrum 4-31
Solenoid 5-61
Solid propellant 3-7
ballistite 3-8
classification 3-7
comparison 3-9
configuration of grains ..3-8
engines 7-19
limitations 3-9
thrust cutoff control 7-19
types 3-7
Solid rocket 3-30
Solid state devices 4-42
Sonar 6-13
Sonic speed 2-13
Specific gravity of gas 4-18
Specific impulse 3-6
Specific fuel consumption 3-7
Spectrum 4-29
absorption 4-30
emission 4-30
solar 4-31
Speed classification 2-13
Spherical coordinates 8-16
Spiral scanning 6-21
Spoilers 2-24
Spontaneous ignition 3-17
Stability, definition of 2-6
Stability of missiles 2-11, 2-27
Stabilization gyro com-
pensation computer 9-41
Stabilized
platforms 6-68, 9-19, 9-21
Stalling angle 2-9
Standard candle 4-22
Standby guidance system 9-81
Stars, navigational
reference 8-14
Star-tracking telescope 6-12
Static equilibrium 4-10
Static servo
response checks 7-28
Stellar supervised inertial
autonavigator (SSIA) ...9-43
Stimulated emission 11-7
Strain gage 12-17
Subcarrier-oscillator unit ...12-19
Subshells (atom) 11-6
Subsonic speed 2-13
Summing amplifier 6-41
Superheterodyne receiver ...12-32
Supersonic flight
aerodynamics 2-37
Supersonic flight terminology
heat barrier 2-13
hypersonic speed 2-13
Mach angle 2-13
Mach number 2-13
normal shock wave 2-14
oblique shock wave 2-14
relaxation time 2-13
Reynolds number 2-13
shock wave 2-12
sonic speed 2-13
subsonic speed 2-13
supersonic speed 2-13
Supersonic speed 2-13
Surface-target selection 10-1
Sweepback 2-11
Switching unit 12-21
Symbols, missile designator
launch environment 1-11
mission 1-12
status prefix 1-10
vehicle type 1-13
Synchronous mechanical
switch commutator 5-55

T

Tabs 2-21
Tachometer switching unit ..12-21
Tape recorder 12-37
Tapes, computer input 6-45
Targets
aerial 10-5
classification 10-1
considerations 10-2, 10-6
detection and evaluation .10-5
groups 10-6
location on maps 8-11
post-attack analysis,
aerial 10-6
post-attack analysis,
surface 10-5
priority assignment 10-4
selection 10-1
surface 10-1
systems 10-2
terminal guidance 9-57
Telemetry
AM 12-26
data reduction 12-28
definition 12-10
FM/FM 12-17
ground station 12-28
radar 12-10
receiver calibration
equipment 12-33
television 12-26
Telescopes 4-37, 6-12
Television 6-3
Television telemetry systems .12-26
Terminal guidance
systems 9-57, 9-76
Terrestrial guidance
systems 6-3, 9-52
Test console 7-27
Theodolite 12-4
Thermal integrator 5-40
Thermal loads 2-36
Thermistors 6-20, 12-17
Thermocouple 6-11
Thermoelectric timer 5-49
Thermonuclear 10-12
Thrust 3-2
Thrust cutoff control 7-19
Thrust horsepower 3-4
Thrust vector control 7-
gimballing 7-
jet vanes 7-
LITV 7-17
swivelling exhaust
nozzles 7-15
Time delay fuzes 10-13
Time reference 5-46
Timers
application 5-47
characteristics 5-46
electrical 5-48
mechanical 5-47
pneumatic 5-50
Timing controls 6-53
Titan II missile 10-17
Torque 4-7
Torsion pendulum 4-14
Total impulse 3-6
Tower launcher 3-43
T-R box 6-10
Track computer
and programmer 9-35

Trajectory characteristics 2-34
Trajectory curves 8-4
ballistic type 8-6
beam-rider course 8-4
flat 8-7
guidance coordinates 8-1
hyperbolic 8-4
influence of natural factors on 8-8
lead-angle course 8-4
pursuit curve 8-4
reference concepts 8-1
rocket 8-7
transformation of axes ...8-3
Transducer 5-18, 12-12
airspeed 5-18
definition 5-18, 12-12
strain gage 12-17
thermistor 12-17
types 12-12
Transfer valve 5-62
Transformation of axes 8-3
Transformer mixing 5-31
Transformer pickoff 6-19
Transistor 4-42
amplifiers 5-57
bias voltages 4-53
definition 4-42
NPN and PNP junction..4-47
physics 4-42
PN junction 4-45
point-contact 4-50
structure of atoms and crystals 4-42
Transmitter, command 9-5
Transmitter unit 12-25
Transonic region 2-13
Triode vacuum tube 4-52
Turbofan engine 3-29
Turbojet engine 3-5, 3-25
Turboprop engine 3-5, 3-29
Two-speed selsyn 5-21
Types of lasers 11-2, 11-17

U

Unsymmetrical dimethyl-hydrazine (UDMH) 3-11

V

Vacuum-tube, comparison to transistor 4-52
Vacuum-tube chopper 5-54
Vacuum-tube commutator ...5-56
Vacuum-tube mixer 5-32
Valve, pressure relief 5-71
Valve, transfer 5-62
Variable incidence control ...2-25
Variable reluctance pickoff...5-26
Variable-speed motor 7-21
Variable-speed motor integrator 5-37
Vector representation 2-2, 4-2
Veeder counter 6-65
Velocity 4-2
Velocity damping Doppler radar 9-82
Velocity of light 4-25
Vertical axis 2-11
Vertical dive system 9-76
Vertical gyroscope 5-13
Vibrating accelerometer 6-17
Viscosity 4-19
Voltage amplifier 5-53
Voltage, bias 4-53
Voltage division 6-40
Voltage multiplication 6-38
Voltage reference 5-45
Voltage summation 6-40

W

Warheads, missile 10-7
characteristics 10-8
detonation points 10-13
fuzes 10-12
types 10-8
Wave
electromagnetic 4-39
reinforcement 6-8
shock 2-12
theory of light 4-20
Wind tunnel 2-37
Wing design 2-18
Wingtip vortex 2-10

X

Xenon flash tube 11-23

Y

Yaw 2-21, 7-6
Yaw axis 2-11

Z

Zeroing adjustments 7-27
Zero-length launcher 3-39
Zero-lift inertial system 78
Zero termination antenna 8

★ U. S. GOVERNMENT PRINTING OFFICE: 1972 O - 469-481