MASSACHUSETTS INSTITUTE OF TECHNOLOGY

APOLLO

GUIDANCE AND NAVIGATION

These papers were presented by the authors at Wiesbaden, Germany and Brussels, Belgium in June of 1965 under the sponsorship of NATO's Advisory Group for Aeronautical Research and Development.

R-500

SPACE NAVIGATION GUIDANCE AND CONTROL

Volume 2 of 2

JUNE 1965

by Dr. C. S. Draper
Dr. W. Wrigley
D. G. Hoag
Dr. R. H. Battin
J. E. Miller
D. A. Koso
Dr. A. L. Hopkins
Dr. W. E. Vander Velde

MIT INSTRUMENTATION LABORATORY

CAMBRIDGE 39, MASSACHUSETTS

COPY #

ACKNOWLEDGMENT

This report was prepared under DSR Project 55-238, sponsored by the Manned Spacecraft Center of the National Aeronautics and Space Administration through Contract NAS 9-4065.

The publication of this report does not constitute approval by the National Aeronautics and Space Administration of the findings or the conclusions contained therein. It is published only for the exchange and stimulation of ideas.

Massachusetts Institute of Technology.
Published by the Instrumentation Laboratory of the
Massachusetts Institute of Technology.
Printed in Cambridge, Massachusetts, U.S.A., 1965

TABLE OF CONTENTS

Volume II of II

TABLE OF CONTENTS (Continued)

Part III

EXPLICIT AND UNIFIED METHODS OF
SPACECRAFT GUIDANCE
Dr. Richard H. Battin

Volume II

Part IV

INERTIAL MEASUREMENT UNITS AND PULSE TORQUING
John F. Miller

Part V

OPTICAL MEASUREMENTS AND NAVIGATION PHENOMENA
D. Alexander Koso

TABLE OF CONTENTS (Continued)

Part VI

GUIDANCE COMPUTER DESIGN
Dr. Albert L. Hopkins, Jr.

Part VII

SPACE VEHICLE FLIGHT CONTROL
Dr. Wallace E. Vander Velde

PREFACE

The material in this book was assembled to support a series of lectures to be given by the authors in Europe in June 1965, under the sponsorship of the Advisory Group for Aerospace Research and Development, an agency of NATO.

The general subject of Space Vehicle Control Systems is the subject of discussion with particular application to the present Manned Lunar Landing Program. The man-machine interaction along with requirements of the mission are first described. These mission requirements in terms of specific hardware along with the performance requirements and underlying reasons for choice are next explained. Lastly, the theoretical background, the system analysis and the derivation of the control functions to integrate the hardware into a precision guidance, navigation and control system are discussed. The book is organized into seven sections following the pattern of the lectures.

Section I provides historical background to the fundamental problems of guidance and navigation. The basic physical phenomenon and associated instrument techniques are discussed.

Section II continues with background information going more specifically into the problems and approach of the guidance, navigation and control of the Apollo manned lunar landing mission. This section illustrates some of the basic philosophy and approaches to the Apollo tasks, such as the success enhancing decision to provide equipment that will perform all necessary operations on-board and using all ground based help when available.

Section III concerns in detail the analytic foundation for performing on-board calculations for navigation and guidance. The achievement of a unified and universal set of equations provides an economy in on-board computer program to perform all the various mission tasks.

Section IV covers in detail the mechanization of the inertial sensor equipment of the Apollo guidance and control system.

Section V provides the same visability into the optical navigation sensors and measurement techniques.

Section VI provides background and specific techniques in the mechanization of on-board digital computers. Application to the Apollo mission illustrates several problems of interest such as the method for providing reasonable and straightforward astronaut data input and readout.

Section VII concerns the specific problems and solutions of vehicle attitude control under conditions both of rocket powered flight and the free-fall coast conditions. The Apollo mission provides a diversity of examples of this area of technology in the control schemes of the command and service module, the lunar landing vehicle, and the earth entry return configuration.

The general problems of Space Navigation, Guidance, and Control requires a great variety of discipline from the engineering and scientific fields. The successful completion of any one space mission or phase of a space mission requires a team effort with a unified approach. Of equal importance are the software deliveries and performance with the hardware. This lecture series is an attempt to integrate many of the disciplines involved in creating successful and accurate space vehicle control systems.

These lectures represent, on every ones part, an interplay between equipment and theory. While in each case emphasis may be on one or the other, in the whole equal emphasis is applied.

All sections may be treated as separate entities however in the case of Section III through VII it is helpful to have the background of Section II. There is cross reference between sections to avoid unnecessary duplication.

It is observed that the authors have emphasized the Apollo mission and hardware as examples in their treatment of the subjects. This is partially because of their intimate familiarity with Apollo in the development work at the Instrumentation Laboratory of MIT and partially because Apollo provides, in an existing program, an excellent example in its multiple requirements and diversity of problems. Because Apollo is currently under development, no particular attempt has been made to make reference only to the latest configuration details. Indeed the authors have utilized various stages of the Apollo development cycle without specific identification in every case as they provide the guidance, navigation, or control technique example desired.

The authors wish to express their appreciation to NASA for the opportunity to participate in the lecture series and for permission to use material from the research and development contracts NAS 9-153 and NAS 9-4065. They also recognize that this does not constitute approval by NASA of this material. In addition, they wish to thank the many members of M. I. T. 's Instrumentation Laboratory; who are working on the Apollo system, for their inspiration and generation of material.

N66 28760

PART IV

INERTIAL MEASUREMENT UNITS AND PULSE TORQUING

by

John E. Miller

JOHN E. MILLER

Deputy Associate Director, Instrumentation Laboratory
Massachusetts Institute of Technology

John E. Miller, Deputy Associate Director of Instrumentation Laboratory, Massachusetts Institute of Technology, heads the Laboratory group that is developing the inertial measurement unit for the Project Apollo spacecraft guidance system.

Mr. Miller was born April 30, 1925, at Aberdeen, S. D., and was graduated from high school there in 1943. He served in the U. S. Army from 1943 to 1945 prior to appointment to the United States Military Academy, West Point, N. Y., where he was graduated in 1949.

Mr. Miller served in the U. S. Air Force from 1949 to 1958 and during this period was sent to M. I. T. for graduate study. He was awarded the M. S. degree in instrumentation in 1953. Also while serving as an Air Force officer, Mr. Miller taught inertial navigation and automatic control at the Air Force Institute of Technology, Dayton, O.

He joined the Instrumentation Laboratory in 1959 and initially worked on the development of accelerometers used in the Mark II guidance system the Laboratory designed for the Navy's Polaris missile.

Part IV

INERTIAL MEASUREMENT UNITS AND PULSE TORQUING

CHAPTER IV-1

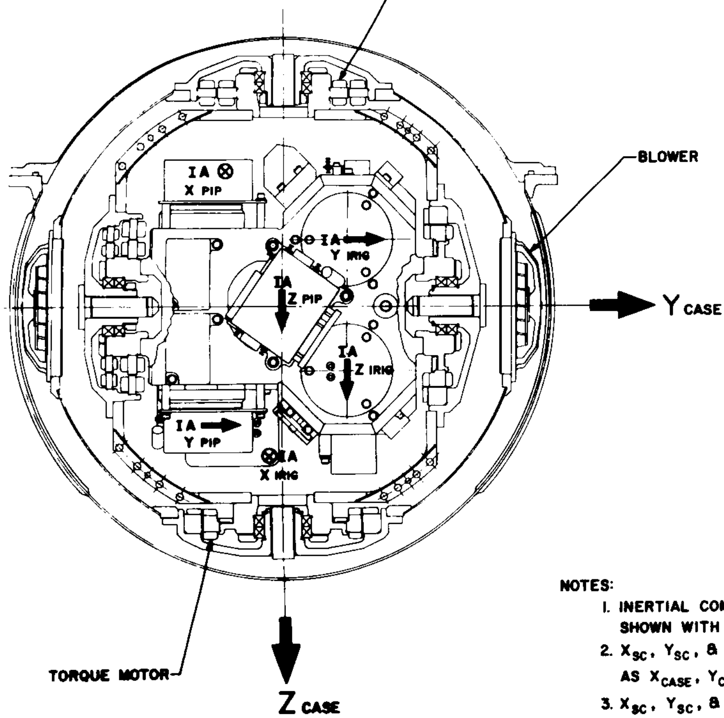
THE APOLLO INERTIAL MEASUREMENT UNIT

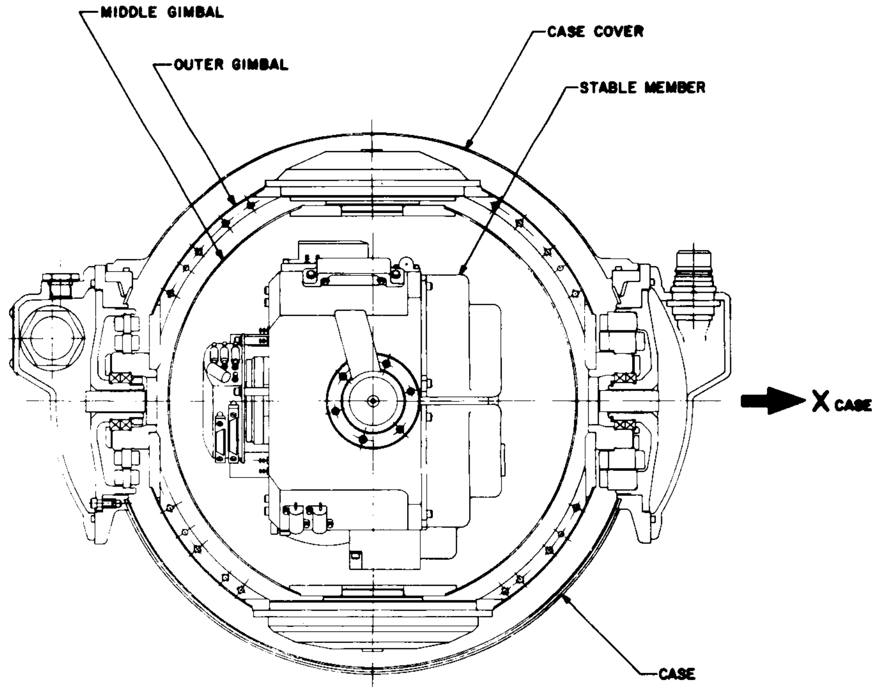
INTRODUCTION

The Inertial Measurement Unit, IMU, provides three fundamental functions in a space guidance system. It established a non-rotating computing coordinate frame oriented to a desired inertial reference such that acceleration is measured in that frame. It provides information to a guidance computer of the orientation of that coordinate frame with respect to the control member, the spacecraft. It provides a measure of the specific force of the spacecraft (usually the integral of the specific force) in suitable form to the guidance computer.

The Apollo IMU is a three degree of freedom gimbal system utilizing integrating gyroscopes to detect angular deviations of the stable member with respect to inertial space, and to provide along with their servo electronics the establishment of a non-rotating member. On this stable member are three accelerometers in an orthogonal triad. The accelerometers are single degree of freedom pendulums with a digital pulse restraining system. Angle information as to the orientation of the computing coordinate frame, stable member, with respect to the navigation base is derived from a two speed resolver system mounted on each axis of the IMU. This resolver system provides knowledge of the stable member to both the astronaut and to the computer: to the astronaut by the means of a ball indicating system with resolvers servo controlled to follow the resolvers on the IMU. The ball has the same gimbal order as the IMU. The same resolver system by means of a Coupling Data Unit, CDU, provides to the computer quantized angle increments for changes in gimbal angles. The CDU couples angle information to and from the guidance computer, performing both analog to digital and digital to analog conversion.

Thermal and Mechanical - The Apollo Inertial Measurement Unit (IMU) is a three degree of freedom gimbaled system. The reasons for using this inertial measurement configuration have been previously stated. The IMU is described from the stable member to the case using Figure IV-1. The inertial components are (3) three single-degree-of-freedom Inertial Reference Integrating Gyroscopes





NOTES:
1. INERTIAL COMPONENT INPUT AXIS DIRECTIONS SHOWN WITH GIMBALS IN ZERO POSITION
2. X_{SC}, Y_{SC}, & Z_{SC} AXES ARE IN THE SAME DIRECTION AS X_{CASE}, Y_{CASE}, & Z_{CASE} AXES
3. X_{SC}, Y_{SC}, & Z_{SC} ARE SPACECRAFT AXES WITH X_{SC} ALONG THRUST DIRECTION

Fig. IV-1 12.5 IMU

(IRIG) and (3) three single-degree-of-freedom Pulsed Integrating Pendulums (PIP). These components are mounted on the inner member which is referred to as the stable member. It is the member which will be non-rotating with respect to the fixed stars except by the drift of the gyroscopes. The stable member also contains necessary electronics. Three degrees of freedom are obtained by coupling the stable member to a middle gimbal by means of an intergimbal assembly to provide one axis of rotation. Likewise a degree of freedom is provided between the middle and the outer gimbal and the outer gimbal and the case. Mounting pads are used as a means of precision alignment of the case onto the navigation base.

The gyroscopes and accelerometers are temperature controlled. The case of the IMU provides a hermetically sealed environment containing air at atmospheric pressure with leak rates less than 10^{-5} CC of He/sec equivalent. There is in the case an integral coolant passage for a heat sink. This integral coolant passage is formed by placing a pattern on a sheet of aluminum. This sheet of aluminum is roll bonded to another sheet except in the area of the pattern where no bonding takes place. The piece or case is formed and machined to the proper dimensions. The coolant passage is then inflated using air to form a coolant passage where the pattern has been placed. This provides a means of making a leak free integral coolant passage over a spherical surface. Water and glycol at 7^{o}C from the spacecraft coolant system is circulated at a rate of 15 Kg/hour and the pressure drop in the IMU is less than 703 Kg/M^2. The required operation of the IMU is for ambient pressure of zero for 14 days and environmental extremes from -18^{o}C to $+65^{o}$C. It may see these ambient structure temperatures depending upon orientation of the spacecraft with respect to the sun. The temperature of the gyroscopes and accelerometers is controlled by means of a thermostatic control system. Heat flow is from the stable member to the case and coolant. Around the stable member is a dead air space. The gyroscopes are controlled to 57^{o}C and the accelerometers to 54^{o}C. The major source of power is in the gyroscope wheels with smaller amounts in electromagnetic components such as torquers. A pair of blowers on the outer gimbal provide heat transfer by moving air over the middle gimbal and the case. There is insulation over the case that is shown in Figure IV-2 primarily because the case will be below the dew point. This is particularly true at the launch at Cape Kennedy. A mercury thermostat senses the stable member temperature by being mounted on the appropriate position of the stable member. A single thermostat is used to control the temperature of six inertial components. The thermostat energizes an end-mount heater on the pendulum, and end-mount heaters on each end of each gyro, and two stable member heaters. The inertial components have a vacuum shroud around the cylindrical wall to prevent circumferential gradients. All heat flow is thus axial which is

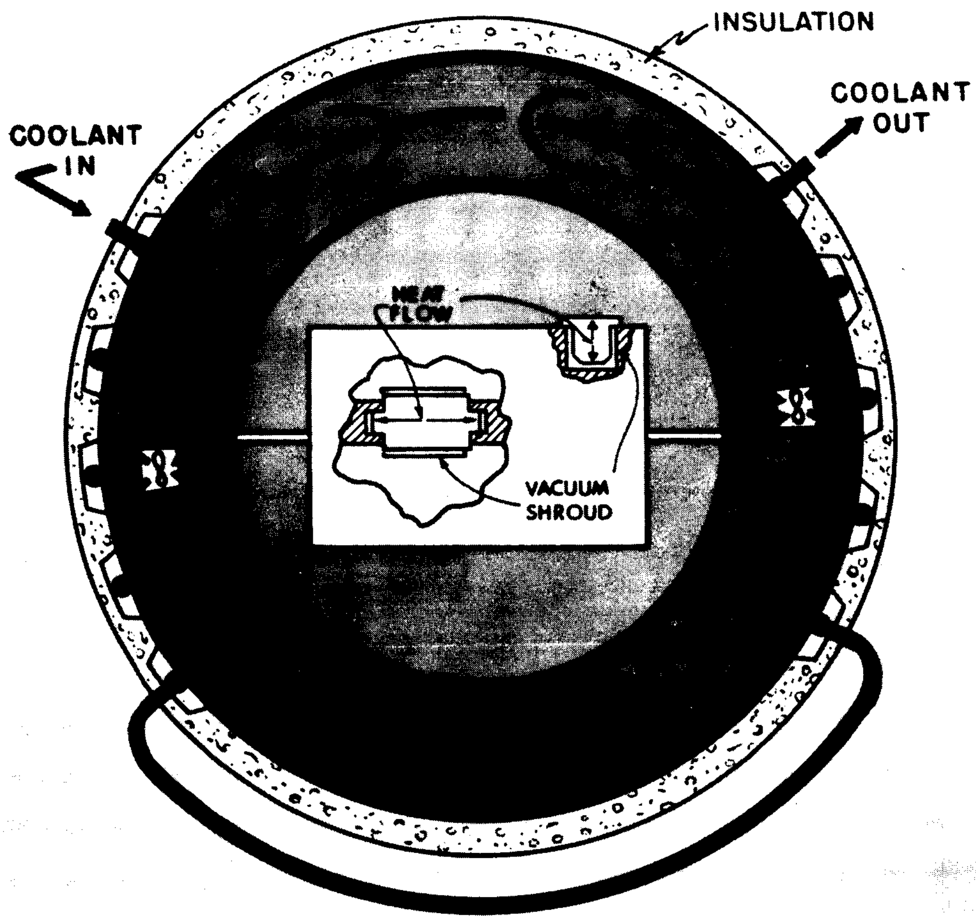


Fig. IV-2 IMU Heat Transfer Diagram

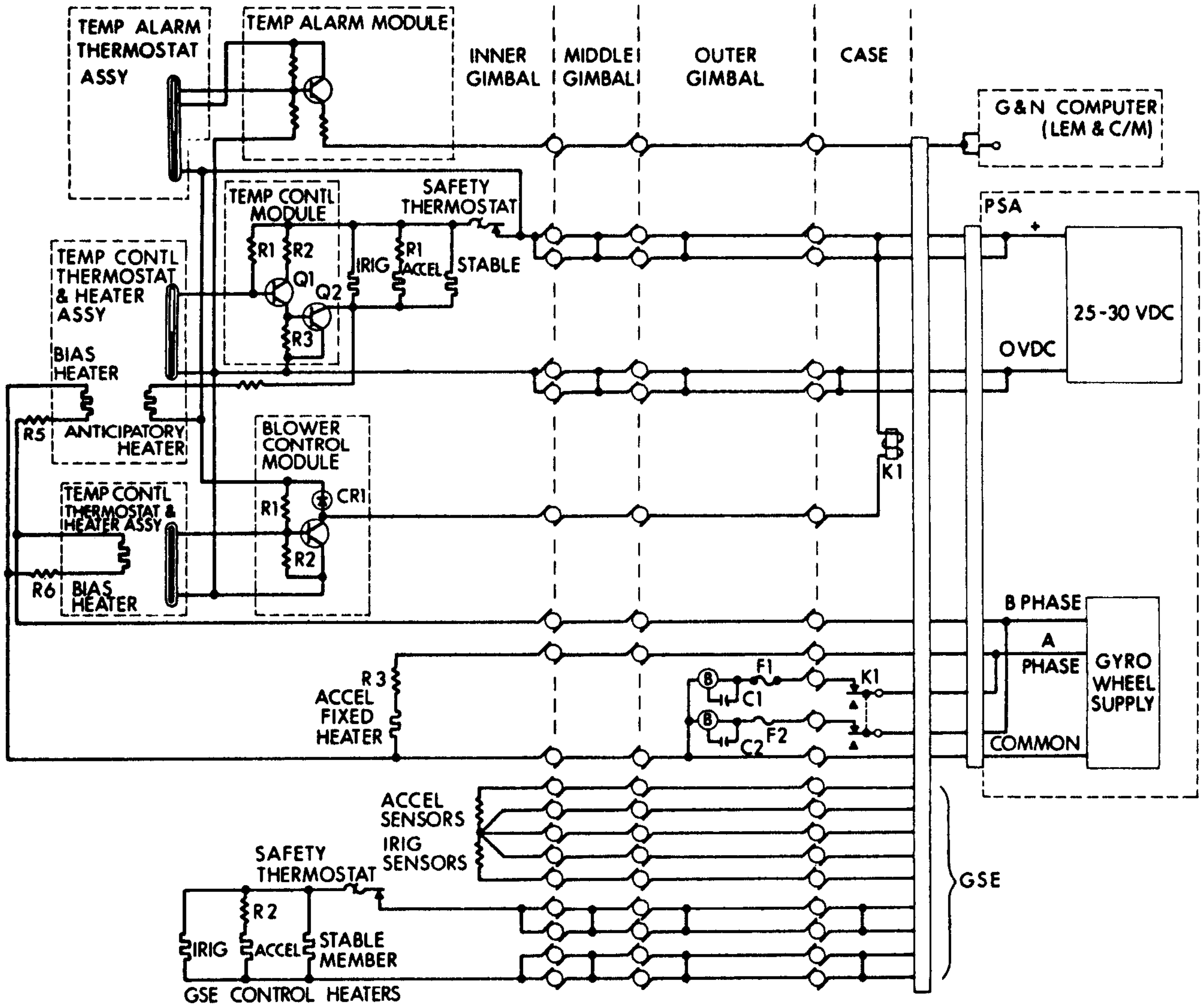


Fig. IV-3 Block II IMU Temperature Control System Schematic Diagram

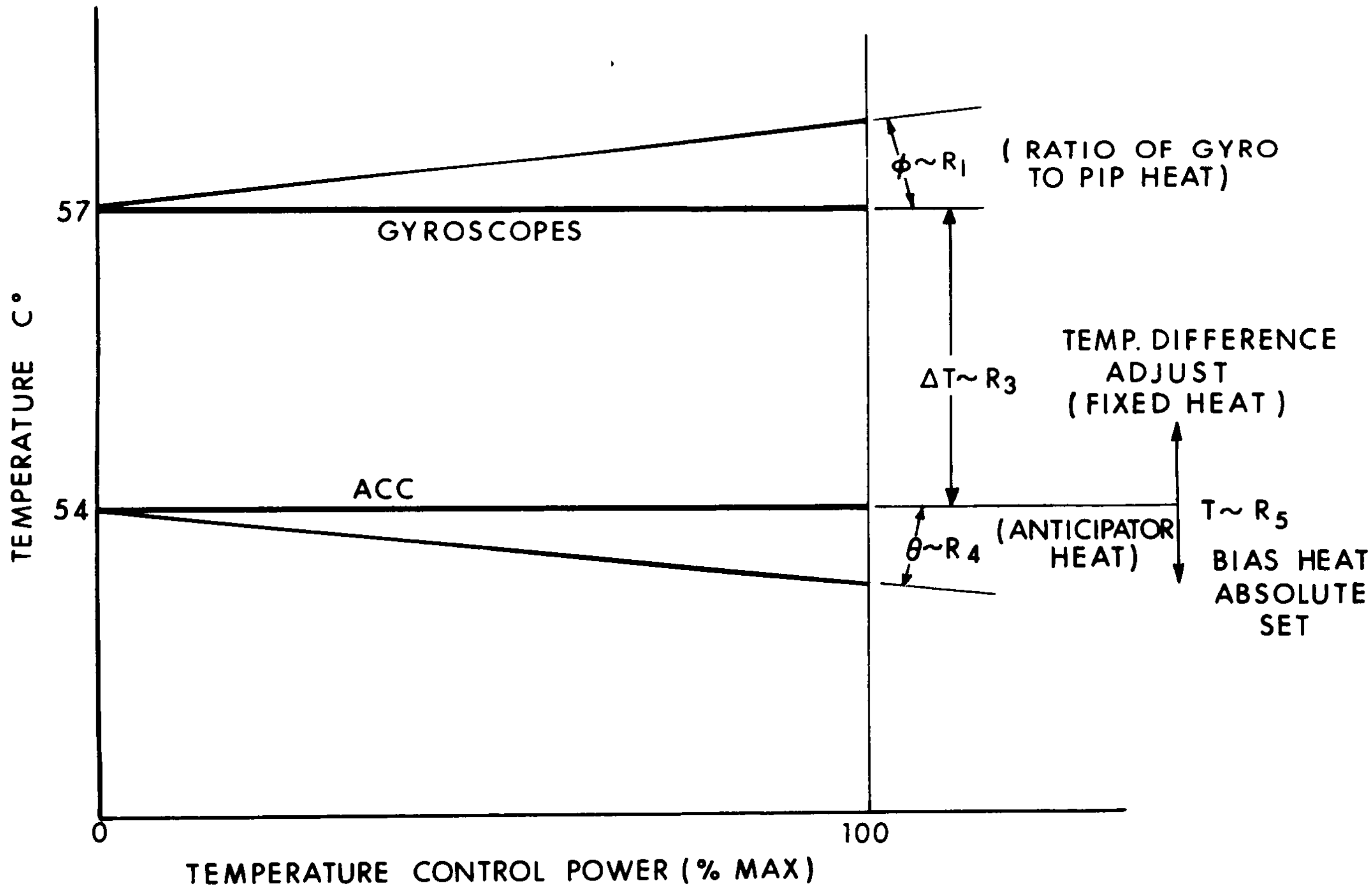


Fig. IV-4 IMU Temperature Control Trimming Resistor Functions done to reduce uncertainty torques. The thermostat dead zone is $0.17^{\circ}C$. A thermostat heater is used to prevent long delays and to control the limit cycle amplitudes of the inertial components. This heater is the anticipatory heater. Fixed heat is applied to the thermostat to set the absolute temperature of the inertial components.

Mounted on the outer gimbal are a pair of blowers, one on each end of a diameter. These also are thermostatically controlled, their purpose is to increase the dynamic range of the temperature control loop in the presence of environmental disturbances. Each inertial component is held to a constant temperature $\pm 1/3^{\circ}C$ under all environmental temperature and pressure design conditions. The blower control loop is required only under conditions of high environmental conditions or high heat dissipation on the stable member. About 18 watts are dissipated under normal operating conditions on the stable member. The principal portion of this power is in the gyro wheels, about 14 watts. A complete temperature control diagram is shown in Figure IV-3. The principal problems of design have been the reliable control of six inertial instruments with a single thermostat and the design of a precision temperature control system both under accelerating and free fall conditions with wide variations in the environmental disturbances. The temperature difference between IRIGs and PIPs is adjusted by properly proportioning the amount of power in each heater. This balance is obtained by adjustment of R_1. (Refer to Figure IV-3 and IV-4). As the power required to maintain the thermostat in its limit cycle operation is varied (because of environmental or power changes) the temperature error of the IRIGs and PIPs can change proportional to the power required. This proportionality can be adjusted to be either plus or minus, and is adjusted to zero. This results in a zero temperature error, or very nearly so, over the full power range possible. There is a fixed bias heat applied to the PIPs to bring them to their operating temperature under normal conditions. As the PIP temperature deviations are subject to gyro power dissipation, this fixed heat is from the same power supply as the gyro wheels.

The performance of the temperature control system may be seen in Figure IV-5. The thermostat cycle of operation at 50% duty cycle is $0.17^{\circ}C$ while at the same time the gyroscope limit cycle has been attenuated to 0.36 millidegrees C^{o} and that of the accelerometers to 5 millidegrees C^{o}. The stable member temperature continues to drop as the control power is increased as one would expect. But the inertial component temperature remains constant throughout the power range. The blower extends the dynamic range of operation, is used infrequently and is limit cycled by a thermostat. There is a separate sensor to detect temperature out of limits to caution the astronaut should this occur. In addition, high limit mechanical thermostats are used in every heater power line

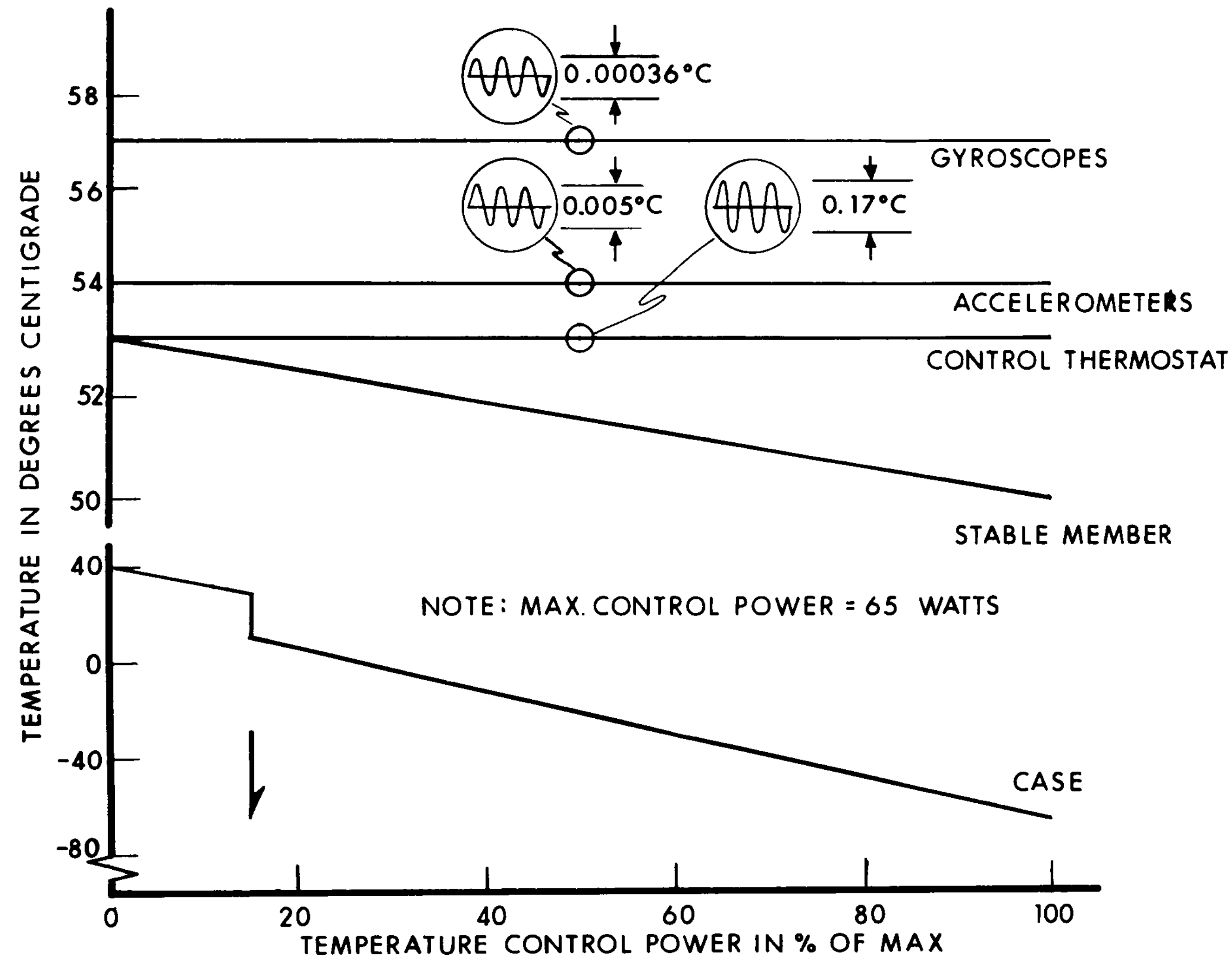


Fig. IV-5 IMU Temperature Profile vs Temperature Control Power to prevent an over heated condition. These are set to open the heater power at a temperature of about $5^{\circ}C$ over the normal control temperature. These have been rarely used but when necessary have prevented damage to valuable equipment.*

The GSE (Ground Support Equipment) heaters and sensors have been provided for ground handling. The IMU once assembled is maintained at temperature. This necessitates times and places of operation where normal power may not be available. Separate heaters and sensors are provided isolated from airborne control. Thus it is possible with the system in the spacecraft to be maintained at temperature with a portable temperature controller with normal power off or on.

Vibration - Before proceeding into the vibration environment it is well to discuss some of the mechanical construction features. The stable member itself is a sintered block of beryllium with drilled holes for each inertial component. These inertial components are pre-aligned in a test stand and assembled into the stable member in an aligned condition. This saves considerable adjustment of alignment at later stages of testing. A pre-wired harness with connectors is laid on to the stable member and interconnection is made. At each end of the stable member are intergimbal assemblies. These house the slip rings to electrically interconnect rotating members. The slip rings are 40 or 50 circuit elements depending upon axis location, the larger number required at the middle and outer axis. Much effort has gone into the design of reliable rotary slip rings which will give unlimited angular freedom of motion. There is a double pair of brushes for each circuit spaced 0.64 mm apart. Each circuit with two feet of #30 AWG has less than 1/3 of an ohm resistance. Electrical coupling does take place between adjacent circuits. Coupling is approximately inversely proportional to the spacing between circuits. With proper selection of circuit placement on the slip ring cross coupling effects can be reduced to negligible amounts. Each contact is rated for 2 amps of current. Only heater power utilizes this much current while everything else is less than 1 amp. High current carrying load slip rings are paralleled for reliability reasons. Early design considerations allow for about 10% spares. As the design progresses toward maturity these spares are usually reduced and finally all utilized for redundancy.

The gimbal construction is spherical. Gimbals are interconnected by means of an intergimbal assembly which contains the resolvers, dc gimbal torque motors, slip rings, stub shaft and a duplex pair of preloaded bearings. Surrounding the stable member is the middle gimbal, and surrounding the middle gimbal is the outer gimbal which is next to the case. The intergimbal assemblies contain either a resolver or a dc torque motor. Each axis utilizes one dc torque motor and a single resolver which is both a single speed resolver and a sixteen speed resolver utilizing the same excitation winding. In addition, on the inner axis is a

*This work done by Edward S. Hickey, MIT Instrumentation Laboratory

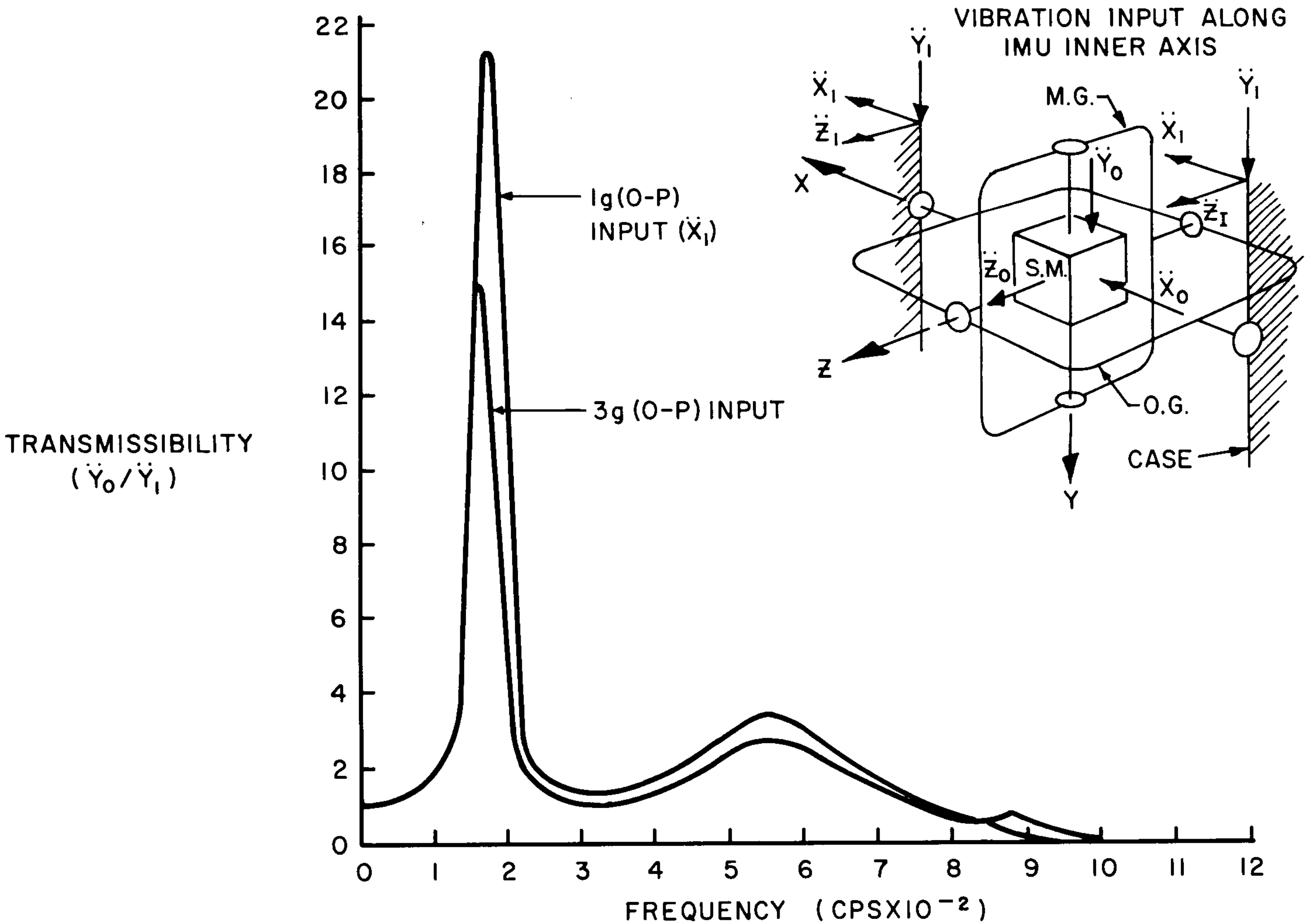


Fig. IV-6 Apollo IMU-VM (12.5) Vibration Test Results
Stable Member Response Curves

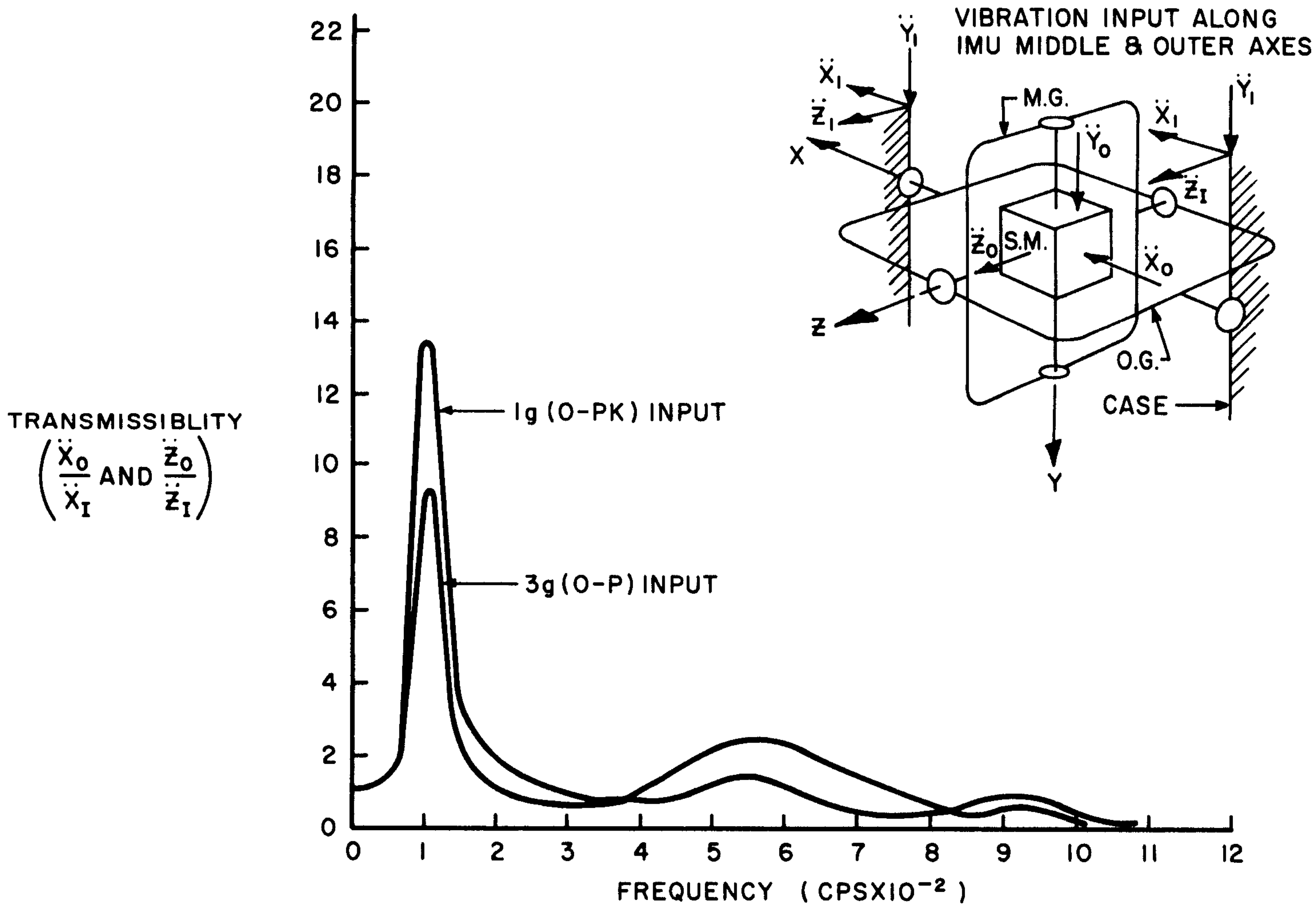


Fig. IV-7 Apollo IMU-VM (12.5) Vibration Test Results
Stable Member Response Curves

gyro error resolver to resolve X and Z gyro error voltages by the angle of the inner gimbal. Each axis has a duplex pair of preloaded bearings. These inter-gimbal assemblies are again made of beryllium for the housing and stub shaft in order to decrease weight.

The IMU, as a mechanical element, forms a system of masses, springs, and dampers, with non-linear behavior. As a complex mechanical system it will resonate at certain frequencies. The lowest frequency resonant vibration ratio of the amplitude of the stable member to that of the input to the case is the greatest. This ratio, called transmissibility, is non-linear in that the ratio decreases with increasing input amplitudes. The elements must be structurally sound enough to withstand the flight environment. Shown in Figures IV-6 and IV-7 is the transmissibility of the IMU. The expected flight environment is less 0.044 g^2/cps. Each IMU is given a workmanship vibration test from 20 to 2,000 cps at least 1 g rms input along each axis.

The resonant peaks may be reduced by adding vibration dampers along each axis which provide a Coulomb friction force for motion along the axis but which do not change the friction torque about the axis.

Gyroscopes and Stabilization - The gyroscopes used are single-degree-of-freedom integrating gyroscopes of a type previously described.[1] The angular momentum is $\frac{10^6}{2}$ gm cm^2/sec developed by a wheel support by a pair of preloaded ball bearings with a 13 n preload. This wheel has a hysteresis ring on it with a motor stator surrounding it on the inside of a spherical float. The float is a beryllium member again designed to yield the highest ratio of wheel to float weight. Flexible power leads carry power from the case to float. These power leads create less than 1/2 dyne cm of torque on the float. The fluid used for buoyant support is a brominated fluorocarbon with a density of 2.385 gms/CC at 58°C. The fluid is fractionally distilled to yield polymers of approximately the same length and nearly the same viscosity. This prevents fluid stratification under operating and storage conditions. The damping about the output axis is $4.6 \cdot 10^5$ dyne cm/radian/sec. In addition, to the fluid support there is an electromagnetic suspension system to provide both axial and radial suspension of the float with respect to the case. Such suspension systems have been previously described.[2]

The axis definition for the instrument has previously been described. Shown in Figure IV-9 is the signal generator which develops a voltage modulated by the angle of the float with respect to the case, A_{c-f}. This is the rotation of the stable member about the gyro non-rotating input axis. There is with each gyroscope a set of prealignment hardware. This provides integrally with each prealigned instrument the following items:

Fig. IV-8 Apollo II IRIG

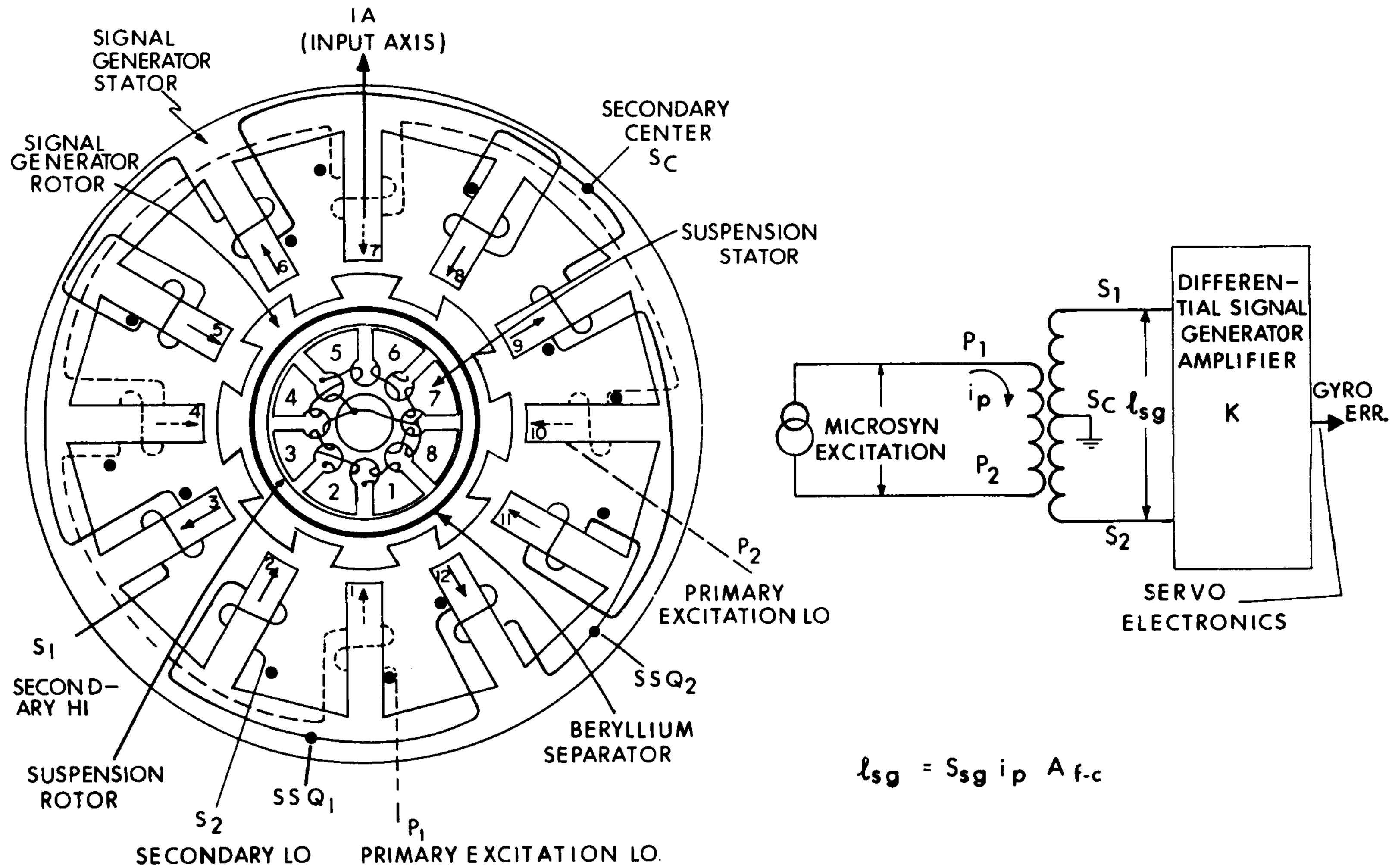


Fig. IV-9 Apollo II IRIG Signal and Suspension Microsyn

A. Suspension Capacitors for Microsyn Suspension

B. Temperature Sensor Normalization

C. End Mount Heater Pre-aligned to Gyro IA.

D. Torque Generator Normalization

E. Signal Generator Preamplifier with normalized gain

The gyro is prealigned on a test stand with the input axis aligned about the output axis relative to a slot in the mounting ring. The alignment is carried over to the stable member where a pin is precisely located to pick up the slot.

The use of prealigned components has made assembly techniques simpler and has brought about excellent correlation between component and system performance of the gyroscope. Deviations in performance between component and system are less than 10 meru (milliearth rate unit). 1 meru = 0.015 degrees per hour.

Each gimbal axis contains a dc torque motor used to null the deviations of the gyro input axis with respect to the case. These direct drive motors have been designed with sensitivities of 1.65 nm/amp. The gyro error signals are fed to the gimbal servo amplifier which with its proper dynamic compensation are used to stabilize the inner member. The performance of the gimbal servos is best described by their ability to attenuate angular deviations of the stable member in the presence of torque disturbances. Shown in Figure IV-10 is the performance of each axis of the IMU gimbal servo. Among the design problems are the mode switching from coarse align (resolver control) to fine align (gyro control), synchronization from turn-on or switch over conditions, performance under power supply changes, gain changes from geometrical variations, and torque disturbances.

Gyro torquing is required for precision alignment of the stable member at prelaunch and for in-flight alignment. The fine alignment procedure and mechanization has previously been described. The gyroscope has a torque generator used in a digital torque command loop. This loop accepts commands from the computer by a pulse train with equal pulse width spacing. The torque commanded is a constant magnitude and modulated only by the time interval. The computer could thus command one equivalent pulse incremental angle ($2\pi/2^{21}$ radians/pulse). One precision current switch is used for all three gyroscopes and the commands are multiplexed in one selector network. The torque level control will be described later. (See Chapter IV-2) The torque generator is normalized in gain to command a fixed angular velocity for a fixed current. This digital angle command from the computer to the IMU permits the stable member to be oriented within the uncertainty of the angle read system ($\Delta\theta$ = 40 arc sec). It furthermore, has the advantage of no current applied to the gyro during periods other than alignment. The gyroscopes are oriented on the stable member with Y and Z gyro output axis along

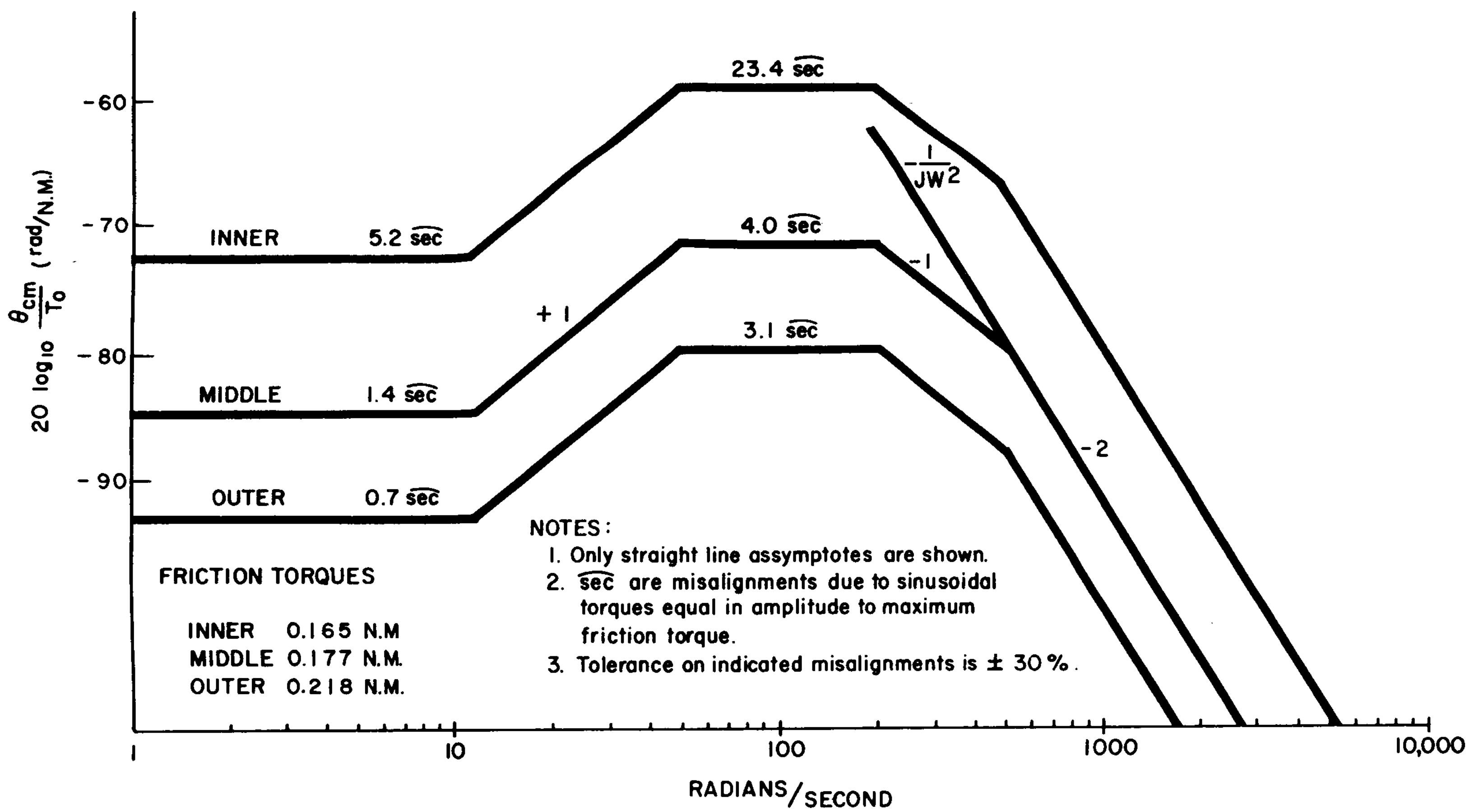


Fig. IV-10 Inertial Platform Misalignment Due to Friction Torques vs Frequency

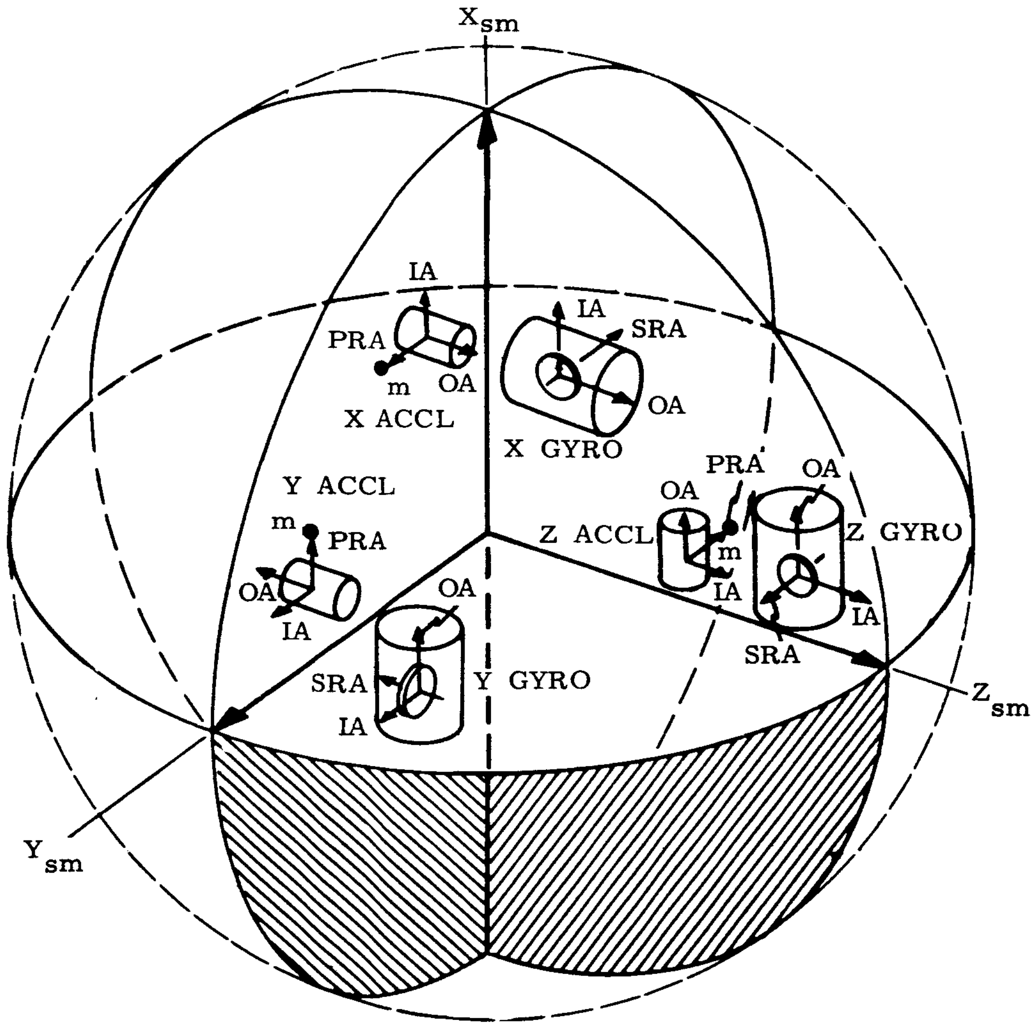
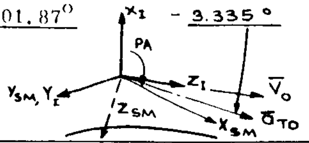


Fig. IV-11 Stable Member Geometry

$\Delta V = 3106.880$ m/sec MAX ACCELERATION $= 12.160$ m/sec^2

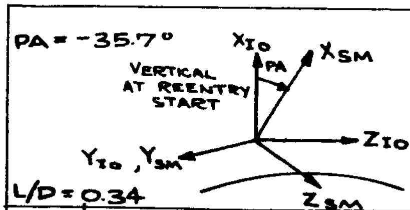
PA = -101.87° -3.335°



			Error Coefficient	Final Position Error in Target Axes in meters			Final Velocity Error in Target Axes in m/sec		
				Alt.	Track	Range	Alt.	Track	Range
STABLE Member	Initial S. M. Alignment Errors	$\Lambda_{(SM)XI}$	0.2 mr		30.10			0.25	
		$\Lambda_{(SM)YI}$	0.2 mr	84.21		-23.41	0.64		-0.14
		$\Lambda_{(SM)ZI}$	0.2 mr		-79.34			-0.56	
ACCELEROMETER	Accel. IA Non-orthogonality	X to Y	0.10 mr						
		X to Z	0.10 mr	-0.30		-1.82	-0.00006		0.0002
		Y to Z	0.10 mr						
	Bias Error	ACBX	0.2 cm/sec^2	-21.28		-92.72	-0.14		-0.59
		ACBY	0.2 cm/sec^2		-95.15			-0.60	
		ACBZ	0.2 cm/sec^2	96.06		-21.89	0.63		-0.14
	Scale Factor Error	SFEX	100 PPM	-9.42		-41.34	-0.07		-0.30
		SFEY	100 PPM		0			0	
		SFEZ	100 PPM	-2.13		0.60	-0.0012		0.0012
	Accel. Sq. Sensitive Indication Error	NCXX	10μg/g^2	-0.91		-3.95	-0.009		-0.03
		NCYY	10μg/g^2		0			0	
		NCZZ	10μg/g^2	0.02		-0.006	0.0001		-0.00003
GYRO	Bias Drift Drift Time before Traj start: 15 min (Init Mlm = 0.33 mr for 5 meru drift)	BDX Direct effect Eff on Init Mlm Combined eff	5 meru		0.12 49.58 49.70			-0.009 0.410 0.401	
		BDY Direct effect Eff on Init Mlm Combined eff	5 meru	17.94 138.02 155.96		-4.26 -38.30 -42.56	0.20 1.04 1.24		-0.04 -0.23 -0.27
		BDZ Direct effect Eff on Init Mlm Combined eff	5 meru		17.94 -130.11 -112.17			0.20 -0.93 -0.73	
	Acceleration Sensitive Drift	ADIAX	15 meru/g^2		0.91			-0.02	
		ADSRAY	10 meru/g^2	2.13		-0.61	0.01		0.005
		ADIAZ	15 meru/g^2		-3.34			-0.02	
	Acceleration Squared Sensitive Drift	$A^2D_{(IA)(IA)X}$	1 meru/g^2		-0.09			-0.001	
		$A^2D_{(SRA)(SRA)Y}$	1 meru/g^2	0.02		-0.006	0.0002		-0.00003
		$A^2D_{(IA)(IA)Z}$	1 meru/g^2		0.02			0.0001	

Fig. IV-12 Translunar Injection Trajectory Cutoff Coefficients

(V_0 = 11,000 m/sec, $_0$ = -6.0 deg., Max. G = 5.45, Platf. Angle = -35.7°



				Error Coefficient	Final Position Error in Target Axes in meters: Alt.	Track	Range
STABLE Member	Initial S.M. Alignment Errors (Uncorrelated)	$A_{(SM)XI}$		0.2 mr	39.52	287.58	8.51
		$A_{(SM)YI}$		0.2 mr	-717.44	-1.86	114.30
		$A_{(SM)ZI}$		0.2 mr	-13.38	545.98	30.70
ACCELEROMETER	Accel. IA Non-ortho-gonalty	X to Y		0.1 mr			
		X to Z		0.1 mr	-356.29	-0.91	6.99
		Y to Z		0.1 mr			
	Bias Error	ACBX		0.2 cm/sec^2	-484.27	-1.22	9.73
		ACBY		0.2 cm/sec^2	-1.22	-413.74	0
		ACBZ		0.2 cm/sec^2	10.64	0	-416.18
	Scale Factor Error	SFEX		100 PPM	65.36	0.30	-2.43
		SFEY		100 PPM	0	-15.81	0
		SFEZ		100 PPM	-7.90	0	306.13
	Accel. Sq. Sensitive Indication Error	NCXX		10μg/g^2	-14.90	0	0.30
		NCYY		10μg/g^2	0	-6.08	0
		NCZZ		10μg/g^2	2.13	0	-85.42
GYRO	Bias Drift Drift Time before Traj start: 45 min (Init Mlm = 0.98 mr for 5 meru drift)	BDX	Direct effect Eff on Init Mlm Combined eff	5 meru	0.30 194.86 195.16	259.01 1415.12 1674.13	2.13 42.26 44.39
		BDY	Direct effect Eff on Init Mlm Combined eff	5 meru	-284.84 -3531.87 -3816.71	-0.30 -8.51 -8.81	-2.74 562.40 559.66
		BDZ	Direct effect Eff on Init Mlm Combined eff	5 meru	-3.34 -65.97 -69.31	5.78 2687.36 2693.14	0.30 151.70 152.00
	Acceleration Sensitive Drift	ADIAX		15 meru/g	-0.30	-267.21	3.04
		ADSRAY		10 meru/g	-890.93	-1.22	2.74
		ADIAZ		15 meru/g	5.17	-10.03	-0.91
	Acceleration Squared Sensitive Drift	$A^2_{D(IA)(IA)X}$		1 meru/g^2	0	34.35	0
		$A^2_{D(SRA)(SRA)Y}$		1 meru/g^2	-245.63	-0.30	3.95
		$A^2_{D(IA)(IA)Z}$		1 meru/g^2	-2.43	-0.91	0.30

Fig. IV-13 3700 Kilometers Earth Reentry Trajectory Errors

X_{sm}. This axis is usually aligned along the thrust axis. This inertial component orientation reduces the acceleration sensitive torques in the gyros. The Y_{sm} axis is usually aligned such that it is normal to the plane of maneuver. For this reason the X gyro SRA is along Y_{sm} as can be seen in Figure IV-11. For the same reason the Z SRA axis is along Y_{sm}.

The largest error contributors per unit of error following any maneuver are the premaneuver errors. That is to say that the effects of gyro drift during thrusting create less error uncertainty in position and velocity at the end of thrust period than those misalignments due to drift between alignment and start of thrust. The error in free fall from the time of alignment to the start of any acceleration maneuver is time dependent upon the gyro non-acceleration sensitive drift (bias drift). As an example the ratio of error created by 1 meru of drift for 15 minutes prior to the start of translunar injection to that error created by the same drift during thrusting is approximately 400:1. A table of error coefficients for two thrusting maneuvers are shown in Figure IV-12, and Figure IV-13. This same fact is readily apparent. Every effort has been made to reduce gyro bias drift uncertainty and magnitude.

The torque generator has been designed to be dc torqued with low residual magnetism uncertainty torques. It is a twelve pole stator with an eight pole rotor creating 4500 dyne cm of torque with a current excitation of 0.1 amps. The beryllium separator prevents the use of the microsyn excitation as a reset coil or degausser for the rotor. Dipole storage in the rotor and stator would create residual torque in the gyro and this torque magnitude and sign would be a function of its past history. It would be possible to return the state of the magnetic material to its original condition in most cases by appropriate commands from the computer. However, the use of a reset coil with ac excitation from the microsyn excitation voltage supply serves to keep the state of the magnetic material constant following any pulse torque commands from the computer. There is a winding around each set of three poles which acts as a reset winding for both rotor and stator. Further use of this reset winding may be made if one recognizes that if the flux through one stator pole were unbalanced there would be a torque created on the float. The bias compensation winding on pole nine is used to create a torque equal and opposite to all the non-gravity sensitive torques and thus reduce the bias drift to zero. The torques creating bias drift come from two sources: flex leads, and microsyns. The microsyn torques are proportional to the microsyn excitation voltage squared. The bias compensation winding torque changes due to excitation voltage changes are exactly equal and of negative polarity relative to the microsyn torque changes from the same source. Only the flex lead torque magnitude may be compensated by the bias winding.

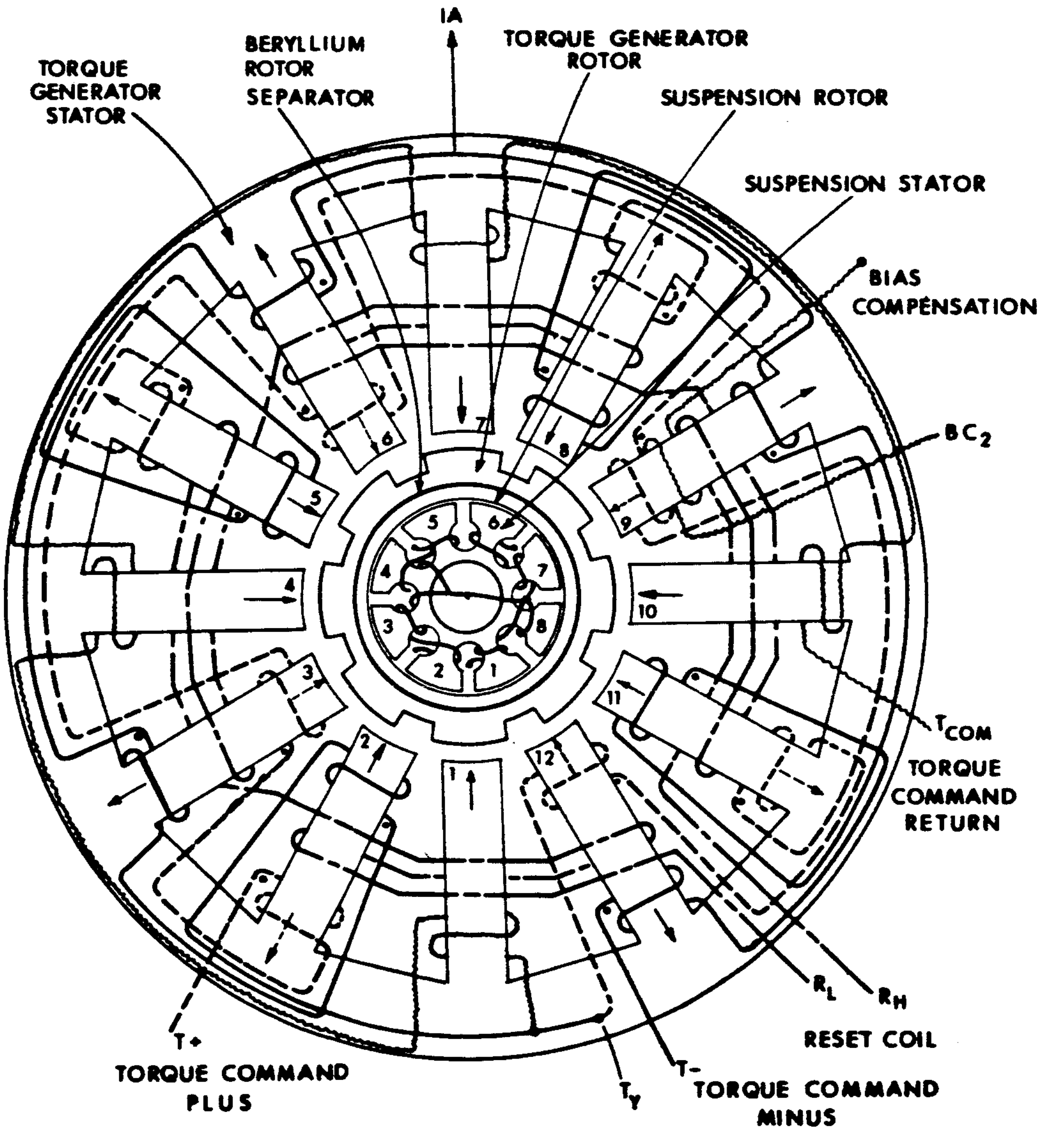


Fig. IV-14 Apollo II Torque Generator and Suspension

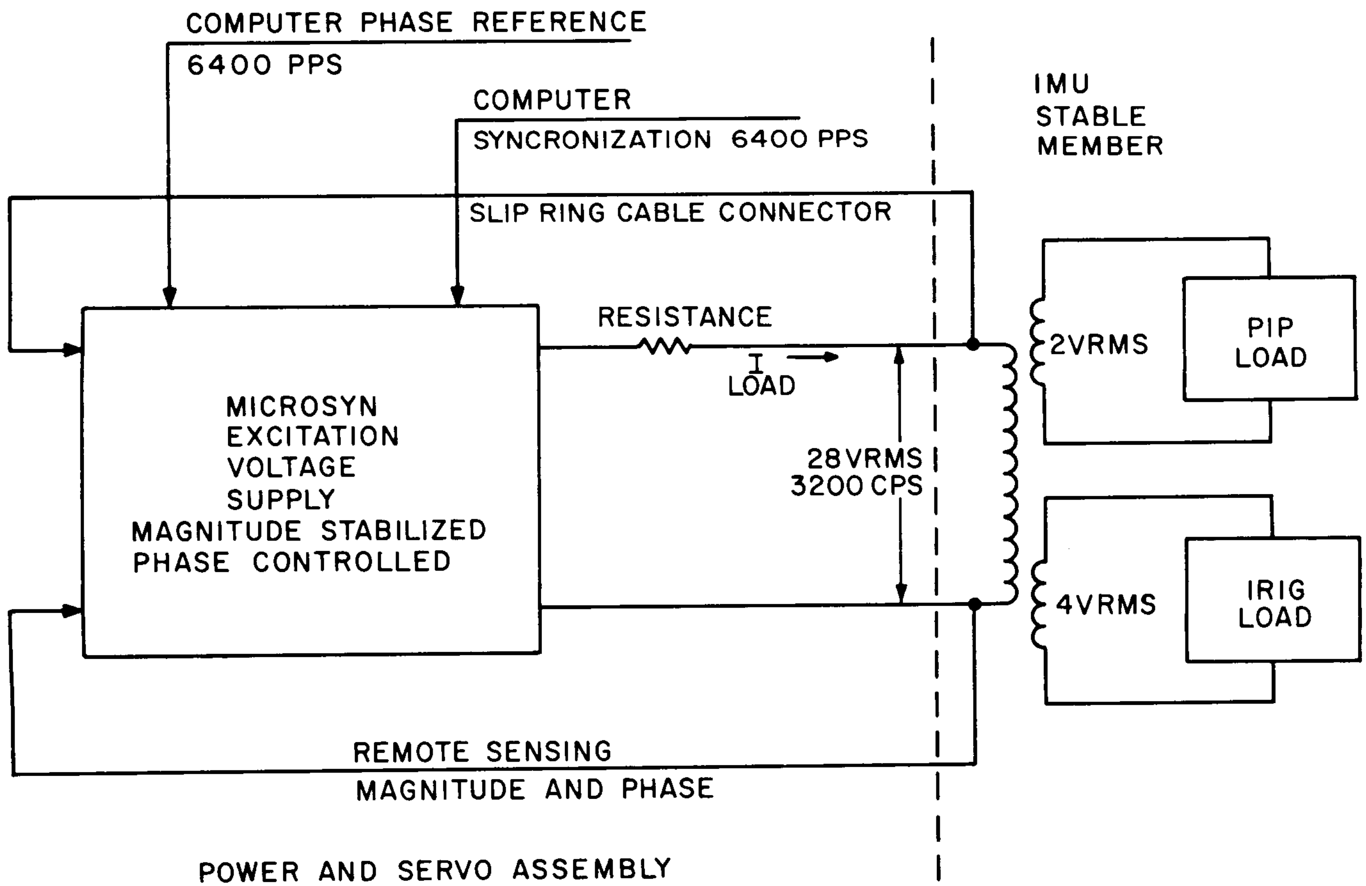


Fig. IV-15 Microsyn Excitation Voltage Supply

The microsyn excitation for both gyros and accelerometers is from a single source located in the power and servo assembly, PSA. The voltage level on the IMU stable member is stable to within 1% of the value required (2 volts for the PIPs and 4 volts for the IRIGs) under the design disturbance conditions. The magnitude and phase of the voltage on the stable member is controlled. All ac power supplies are synchronized to the guidance computer clock. Also dc to dc converter or dc supplies using multivibrators as ac sources for rectification are also synchronized to the computer. The method of synchronization is to use a multivibrator which will free run at a lower frequency without the computer pulses. This will assure operation of the IMU power supplier in the event of a computer failure. With the computer pulses these supplies will be synchronized. In addition, the microsyn excitation supply voltage is phase locked to the computer. The voltage stability and phase stability is required at the inertial component. For power transmission there is a step down transformer on the stable member. This reduces the slip ring current and the voltage drop effects due to slip ring, cable and connector resistance. The stable member voltage is sensed at the primary transformer side and compared to a voltage and phase reference. Phase is controlled to within $\pm$ 1/2° of the computer reference. On the stable member each PIP has the same length of wire from the transformer to the PIP input terminals on the instrument.

CHAPTER IV-2

THE PULSED INTEGRATING PENDULOUS ACCELEROMETER (PIPA)

Basic Operation - The PIPA block diagram is shown in Figure IV-16. The inertial sensor is a single-degree-of-freedom pendulum with a pendulosity of 1/4 gm cm. The pendulum is mounted on the stable member and is non-rotating with respect to inertial space. The signal generator is a variable reluctance device termed a signal generator microsyn. It is excited from a sinusoidal source which is synchronized and phase locked to the guidance computer. The signal generator output is voltage modulated by the angle of the pendulum float with respect to case (A_{c-f}) and the float angular velocity ($\dot{A}_{c-f}$). The voltage may be considered to contain only float angle information. This signal, amplified, is used in an interrogator. The interrogator is a peak detecting device used to determine sign and magnitude of the float angle at discrete times. These discrete times are the computer clock times (ΔT) synchronized with sinusoidal excitation. Only float angle sign is determined and the operation is binary. The interrogator output is a command to torque the float angle to null. The current switch directs a constant current source into either the odd poles or even poles of an eight pole torque microsyn. The torque generator creates a torque either positive or negative proportional to the square of the current in the windings. Current is controlled constant by comparison of a precision voltage reference with the voltage drop across a precision resistor in the current loop.

Dynamic Operation - The basic loop equation is neglecting uncertainty torques about the output axis and making the small angle assumption: Cos $A_{c-f} = 1$, Sin $A_{c-1} = A_{c-f}$.

$$J \ddot{A}_{I-f} + C \dot{A}_{c-f} = M_{tg} + ml\, a_{IA} \qquad \text{(IV-1)}$$

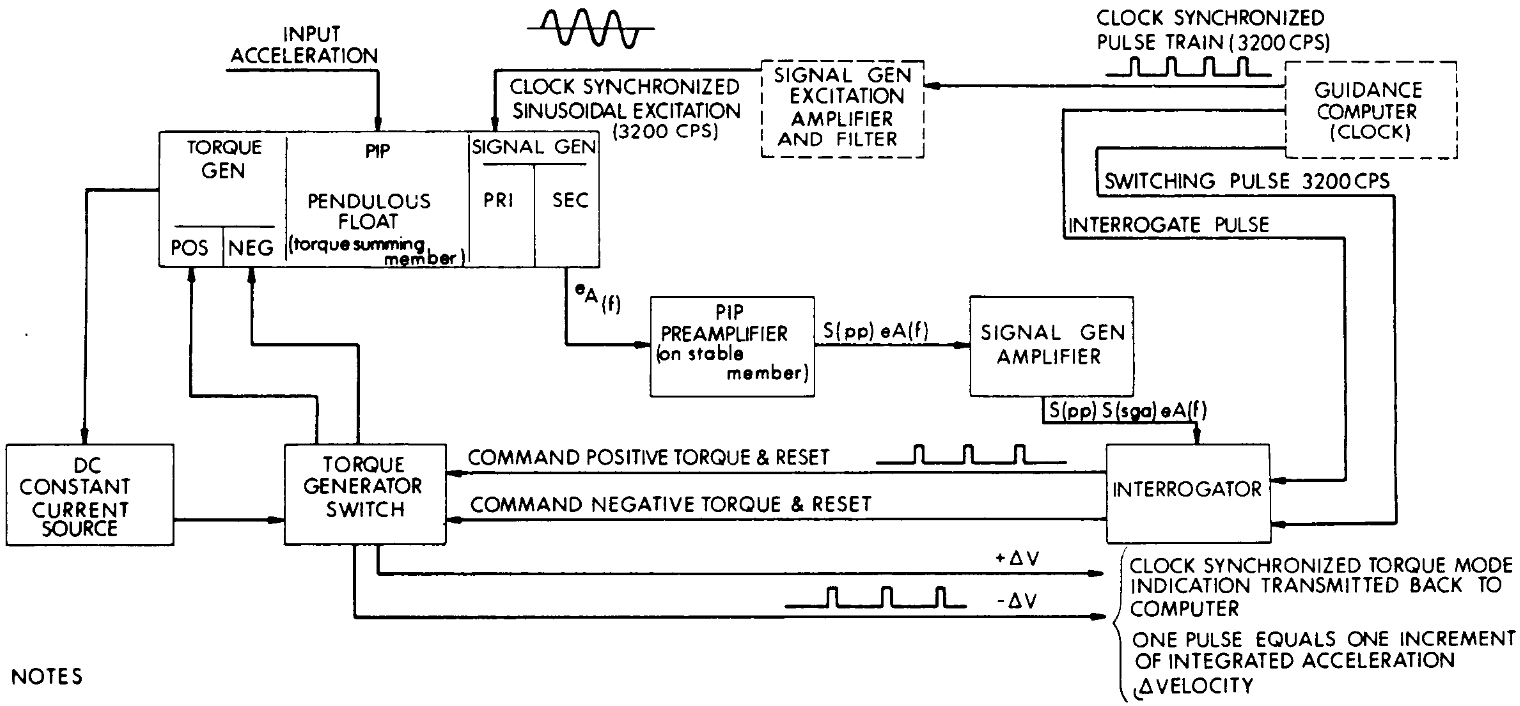
Where

J = Float Inertia

C = Damping about the output axis

A_{c-f} = Float angle with respect to the case (c-f)

A_{I-f} = Float angle with respect to Inertial Space



NOTES

1. PIP (PULSED INTEGRATING PENDULUM) 16 SIZE SINGLE DEGREE OF FREEDOM PENDULOUS UNIT
2. WAVE FORMS SHOW CONDITION FOR NEGATIVE INPUT ACCELERATION OF PIP CASE

ΔV COMMAND MODULE 5.85 CM/SEC/PULSE

ΔV LEM 1.0 CM/SEC/PULSE

Fig. IV-16 PIP Accelerometer Block Diagram

M_{tg} = Torque generator torque

a_{IA} = Acceleration along the input axis

$$V_{IA} = \int_o^t a_{IA}\, dt$$

Integrating yields

$$J\,\dot{A}_{I-f} + C\,A_{c-f} + \text{Initial conditions} = \int_o^t M_{tg}\,dt + ml\int_o^t a_{IA}\,dt$$

Considering the float storage and initial conditions for the accelerometer as an error (velocity stored, V_s)

$$V_s = \int_o^t M_{tg}\,dt + ml\,V_{IA}$$

$$ml\,V_{IA} - V_s = -\int_o^t M_{tg}\,dt \qquad \text{(IV-2)}$$

Since M_{tg} is controlled to be constant and permitted to change sign only at discrete times (ΔT) the integral may be replaced by summation:

$$ml\,V_{IA} - V_s = -\sum_o^{n+p} M_{tg}\,\Delta T = \Delta T\,M_{tg}\,(n-p) \qquad \text{(IV-3)}$$

where n = number of negative torque commands

p = number of positive torque commands

Neglecting for the moment V_s

$$V_{IA} = \frac{M_{tg}}{ml}\,\Delta T\,(n-p) \qquad \text{(IV-4)}$$

The accelerometer scale factor (SF) is then $\frac{M_{tg}\,\Delta T}{ml}$ per pulse and the integrated acceleration or velocity increment per pulse is the scale factor. The net velocity is the net number of minus and plus torque commands, (n - p) sign considered, times the scale factor.

The torque is constantly applied and only switched in direction. The system then behaves as a relay system with a delay due to sampling but no hysteresis. The system limit cycles and the limit cycle behavior is of importance near zero input acceleration as the stored velocity is a function of the magnitude of the limit cycle. The dynamic behavior may be described by collecting together all linear elements (the pendulum, amplifiers, signal generators, etc) as G (s) and the switching of torque as a relay either at + 1 or - 1 together with the sampler as N (s). (See Figure IV-19)

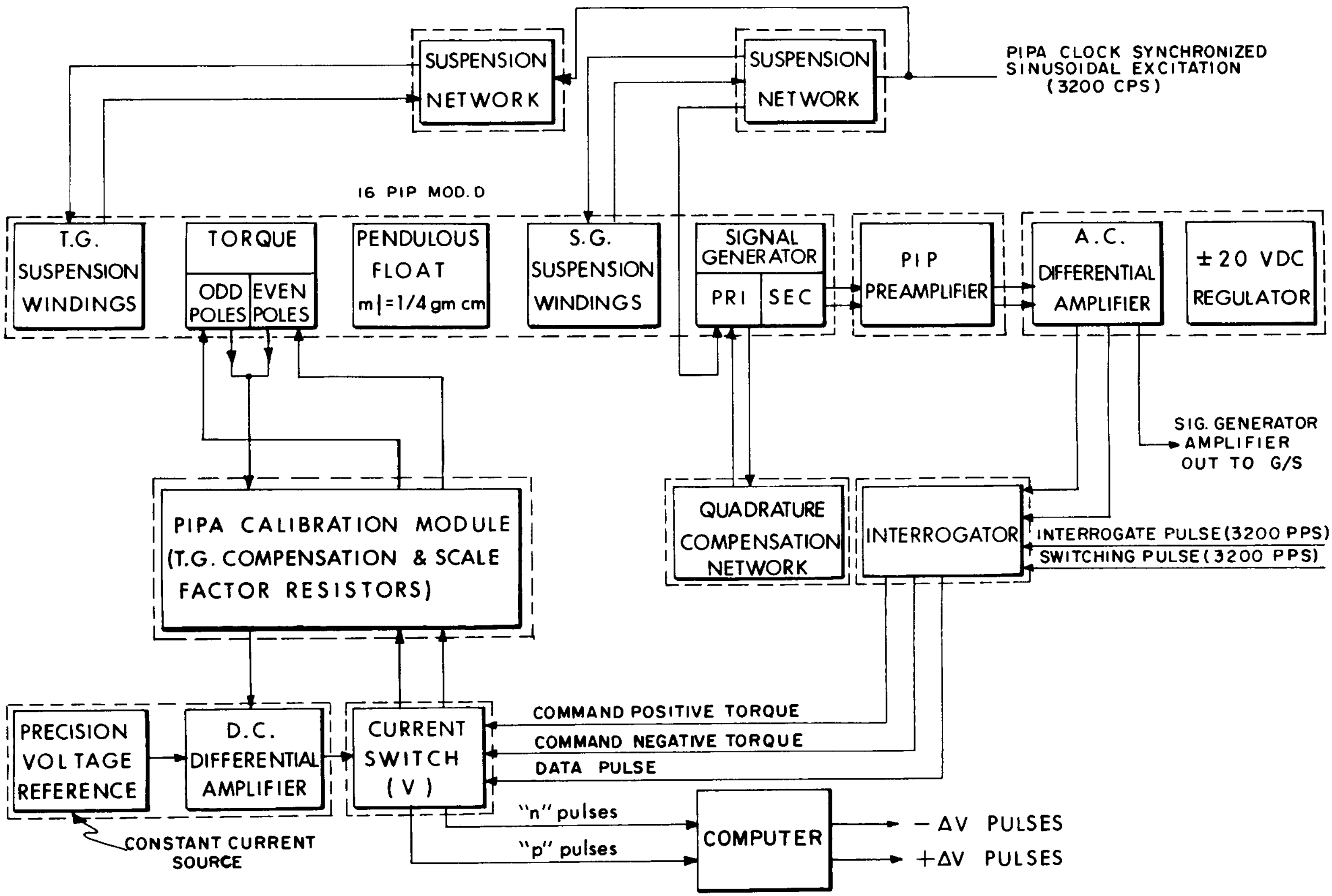


Fig. IV-17 PIP Accelerometer Module Block Diagram

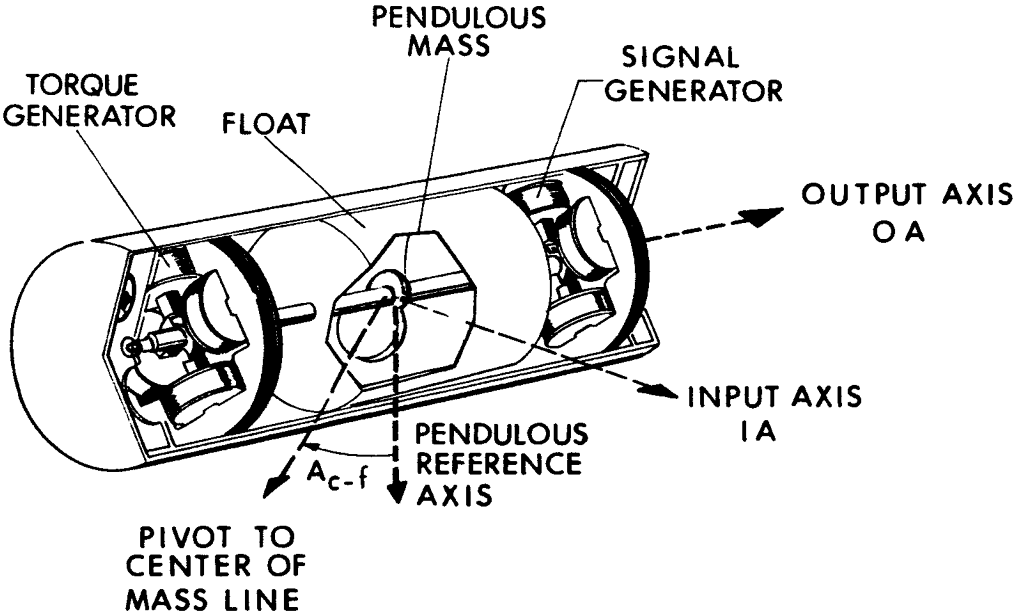


Fig. IV-18 Pulsed Integrating Pendulum Schematic

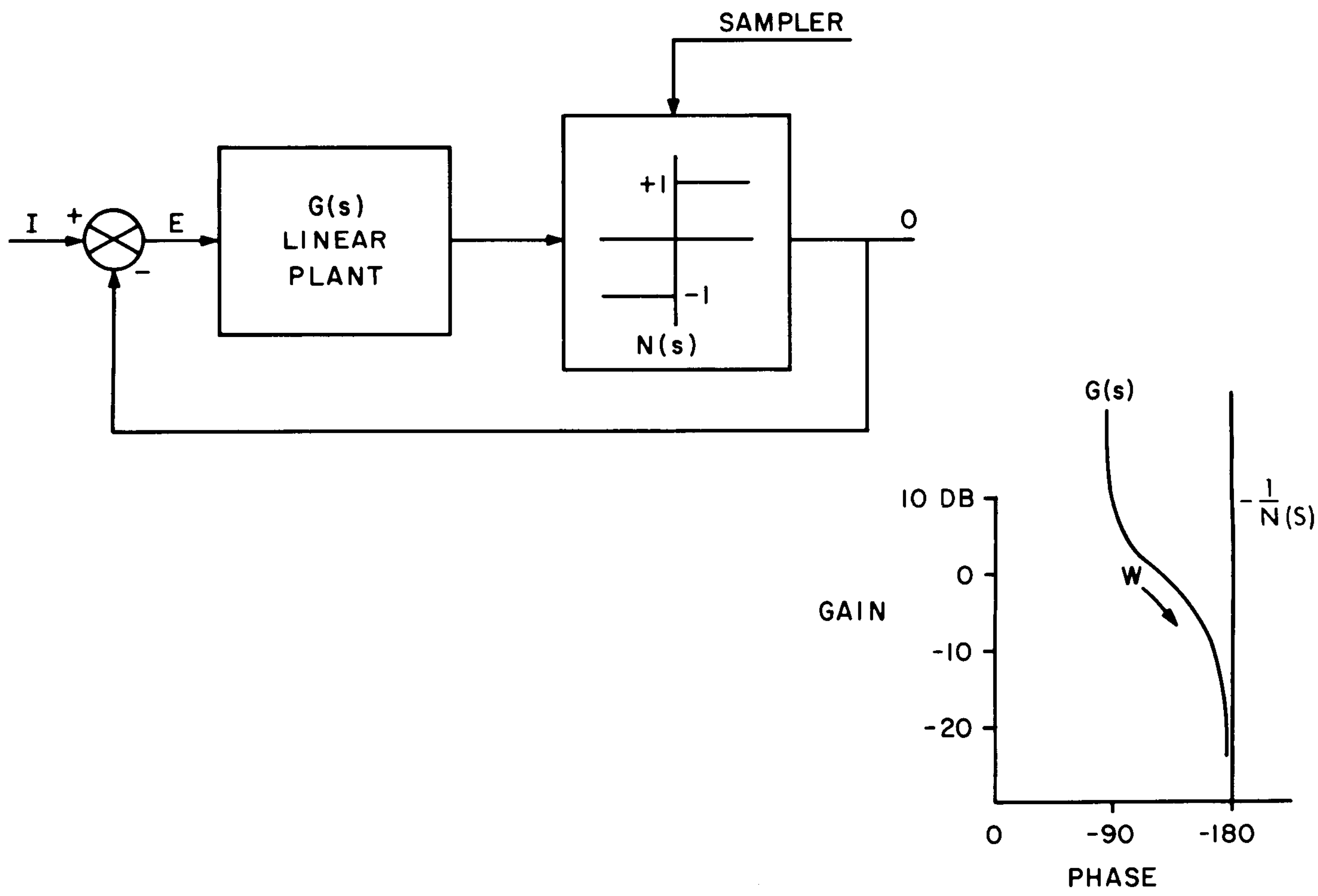


Fig. IV-19 Relay Control System

$$\frac{O}{I} = \frac{G(s)\ N(s)}{1 + G(s)\ N(s)}$$

This loop will oscillate when

$$1 + G(s)\ N(s) = O$$

or

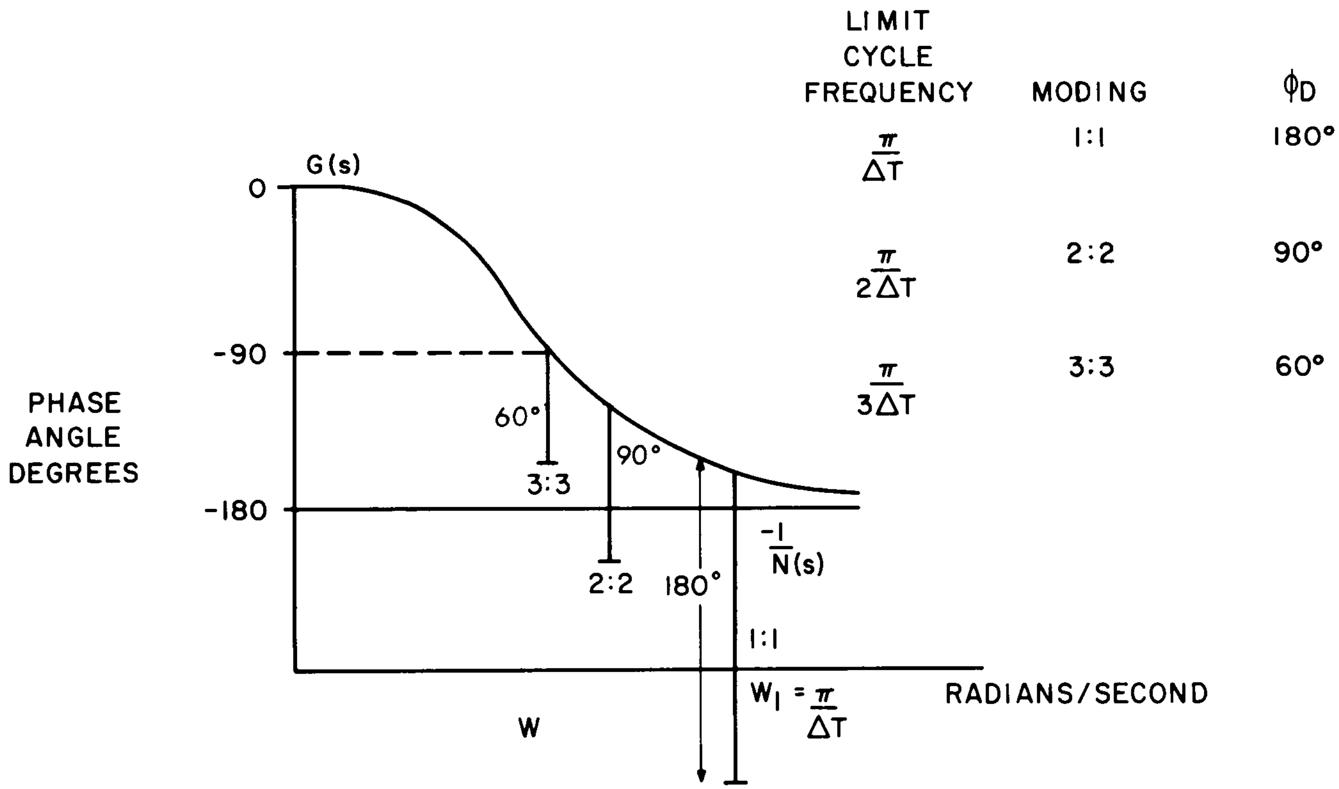
$$G(s) = -\frac{1}{N(s)}$$

Consider for an example only the dynamics of the inertial component as the linear plant.

$$G(s) = \frac{K}{s\left(1 + \frac{Js}{G}\right)} \qquad \text{(IV-6)}$$

The non-linear element, the switch, may be treated using the describing function technique.[3] The gain-phase plot of these two functions $G(s)$ and $\frac{-1}{N(s)}$ are shown in Figure IV-20. The intersection of $G(s)$ and $\frac{-1}{N(s)}$ would indicate that oscillations would exist at infinite frequency of zero amplitude. This however, neglects the phase delay due to sampling. While the phase delay due to sampling may be introduced on the gain phase plot it is equally valid to introduce it on the phase plot alone. The limit cycle oscillation is constrained by the interrogator to have a period of integral multiples of the clock time, ΔT. The stored velocity, V_s, is proportional to the magnitude of the limit cycle and it is therefore, desirous to have it small. The phase delay due to the sampler is $\frac{2\pi}{n}$ where n is the number of ΔT's per cycle in the limit cycle. The addition of the phase delay to G(s) shows intersections with $\frac{-1}{N(s)}$ and thus satisfies the conditions for $G(s) = \frac{-1}{N(s)}$. The moding or limit cycle may be described as 1:1 where there is one sampling pulse per half cycle. A n:n mode is where there are n sampling pulses per half cycle. The phase delays are shown as lines of phase shift at frequency $\omega = \frac{\pi}{n\Delta T}$. The illustrated graph shows as possible limit cycles either the 1:1 or 2:2 mode, and it is impossible to have a 3:3 or higher mode limit cycle operation.

It is possible here to decrease the moding by decreasing the ratio of J/C. For a physical instrument the value of J is fairly firmly established and not much alteration is possible. However, a fairly wide viscosity range of damping fluid is possible. This would seem to indicate that increasing the damping would reduce the limit cycle to its lowest mode. Such is usually the case; however, one must now consider the signal-to-noise ratio of the instrument. As the damping increases the float motion decreases. It becomes then a question of the smallest detectable float angle. That is to say what is the minimum angle from null at which you can positively identify the float angle as positive or negative.



SAMPLER PLACE DELAY = $\frac{2\pi}{n}$

n = NUMBER OF CLOCK INTERVALS (ΔT) PER CYCLE OF OSCILLATION

Fig. IV-20 PIPA Limit Cycle

Fig. IV-21 16 Pulsed Integrating Pendulum, Mod D

From previously, $V_s = J A_{c-f} + C A_{c-f}$ + Initial Conditions. Assuming initial conditions are zero and near null the minimum angle of detection is $A_{c-f} = A_m$. At this point the damping torque is much greater than the inertia reaction torque therefore.

$$\dot{A}_{c-f} = \frac{M_{tg}}{C} \tag{IV-6}$$

$$V_s = \frac{J}{C} M_{tg} + C \ A_m \tag{IV-7}$$

The damping for minimum stored velocity is

$$C = \sqrt{\frac{J \ M_{tg}}{A_m}} \tag{IV-8}$$

This relationship is heuristically correct in that a smaller detectable angle (higher signal-to-noise ratio) will permit increased damping. While decreasing J will reduce the float time constant. Increasing torque, M_{tg}, will in the same time interval increase float angle equivalent to reducing A_m.

PIP - The physical inertial component, the pendulum (PIP), is a single-degree-of-freedom pendulum. The pendulum is floated both to reduce friction uncertainty and to provide damping about all three axes. The output axis damping is to provide good dynamic behavior while the damping about the other axes are to provide geometrical stabilization of the float with respect to the case. This is necessary to assure stability of performance. There is an electromagnetic suspension system called ducosyn to provide both axial and radial centering forces to additionally stabilize the float with respect to the case. The float is a hollow cylindrical beryllium piece with ferrite rotors cemented on each end. There is a pivot shaft used for axial and radial stops of the suspension system. The pendulosity, 1/4 gm cm, is adjusted by a pin and screw which also serve as the output axis stop limiting rotation of A_{c-f} to less than $\pm 1^{\circ}$. On a diameter perpendicular to the pendulous axis are two screws used to achieve flotation. Surrounding the float is the damping block. The damping fluid and the clearances between the float and the case set the damping to about 120,000 dyne cm/radian/second. Located in the damping block are four bellows for volume compensation. On each end is the end housing which has the electromagnetic suspension system on the inner rotor and the signal generator or torque generator on the outer rotor. The rotor itself is a one piece device. The torque generator is an eight pole device with windings on the even poles for negative torque and on the odd poles for positive torque. The torque is proportional by the sensitivity of the torque generator, S_{tg}, and to the square of the current. The damping fluid is the only mechanical connection between the float and the case. This has made a pendulum with a friction uncertainty level of less then 0.02 dyne centimeters when only the suspensions and signal generator are excited. The float inertia about the pivot as output axis is approximately 14 gm/cm^2. The float time constant, J/C, is about 100 microseconds. The pendulum time constant for the float with the input reference axis and output axis perpendicular to gravity is eight minutes.

The signal generator is an eight pole device with limitations identical to those of the torque generator. It uses a primary excitation and a differential center tapped secondary. The signal generator is resistive and capacitive tuned for a second order system damping ratio of nearly unity and natural frequency slightly greater than the carrier frequency. This prevents excessive phase delays of the signal generator from affecting the moding and yet reduces high frequency noise generated and picked up in the signal generator.

The effects of an elastic restraint, K, may be seen by considering the basic equation:

$$J\,\ddot{A}_{I-f} + C\,\dot{A}_{c-f} + K\,A_{c-f} = M_{tg} + ml\,a_{IA} \qquad \text{(IV-9)}$$

K = elastic restraint coefficient

for the average values of the angle $\bar{A}_{c-f}$. This eliminates the limit cycle when considering small inputs.

$$J\,\ddot{\bar{A}}_{c-f} + C\,\dot{\bar{A}}_{c-f} + K\,\bar{A}_{c-f} = ml\,a_{IA} \qquad \text{(IV-10)}$$

Considering the inertia reaction torque small with respect to the damping torque and elastic restraint torque

$$C\,\dot{\bar{A}}_{c-f} + K\,\bar{A}_{c-f} = ml\,a_{IA}$$

$$\bar{A}_{c-f} = \frac{ml\,a_{IA}}{K}\left[1 - e^{-\frac{Kt}{C}}\right] \qquad \text{(IV-11)}$$

This is to say a finite positive K reduces the average angle through which the float will rotate. Any elastic restraint will alter the time at which an output, ΔV, will be indicated. A positive K will increase the apparent integration time and a negative K will decrease the apparent integration time. The limiting positive K case shows an interesting concept. Consider that there corresponding to a ΔV an equivalent average float angle ΔA.

$$\Delta\dot{A} = A\,\Delta T = \frac{M_{tg}}{C}\,\Delta t \qquad \text{(IV-12)}$$

For a steady state case as $t \longrightarrow \infty$

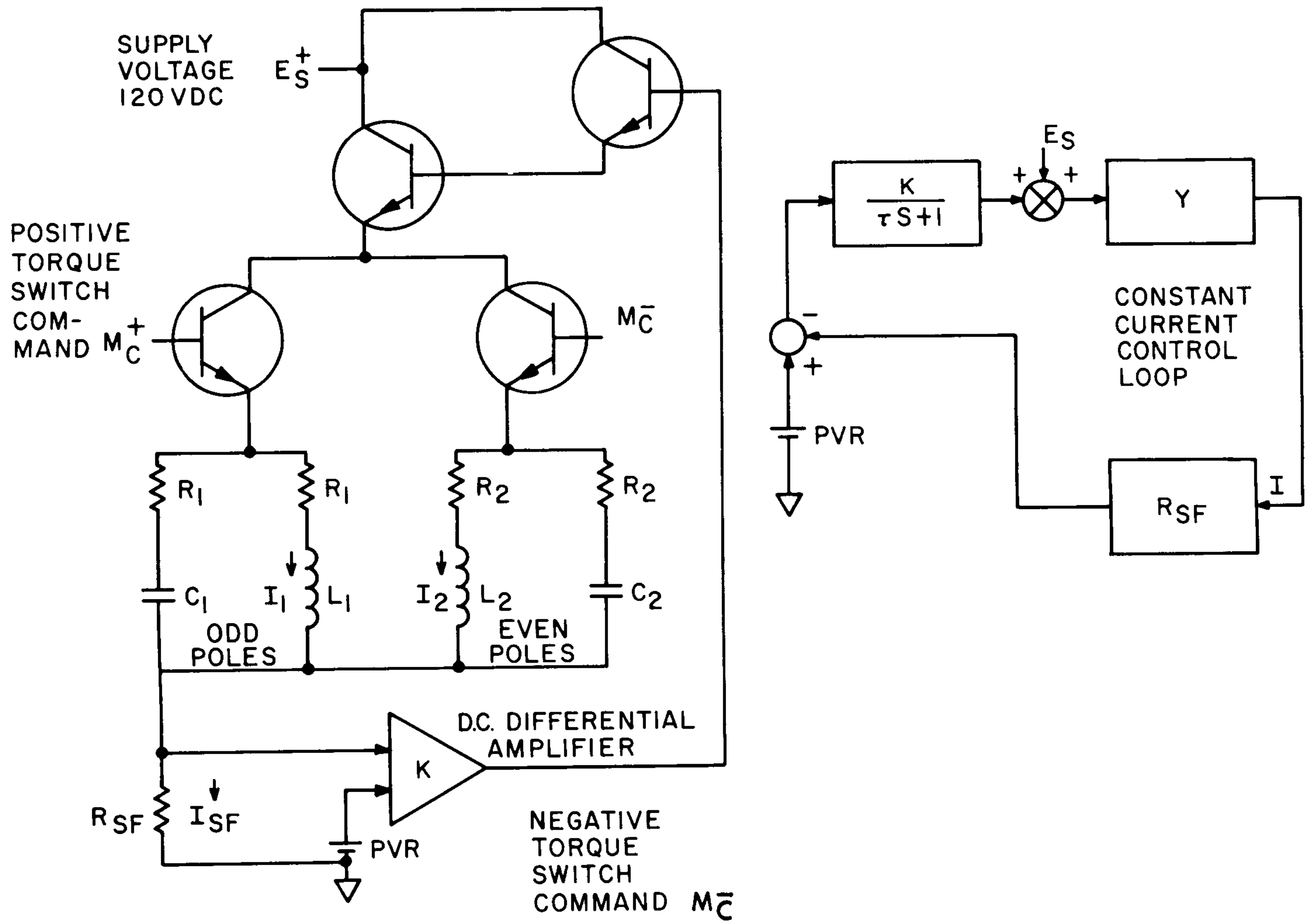


Fig. IV-22 Current Control and Current Switch

A_{c-f} = ΔA as a necessary condition for integrating an input acceleration

$$\frac{ml\ a_{IA}}{K} = \frac{M_{tg}}{C}\ \Delta T$$

or

$$K = \frac{ml\ a_{IA}}{M\ \Delta T} \qquad C = \frac{C\ a_{IA}}{\Delta V} \tag{IV-13}$$

This shows the minimum detectable input acceleration under conditions of a finite elastic restraint. Furthermore, this shows that increasing C reduces the effect of elastic restraint. We again return to the criteria of the minimum detectable angle A_m. For the optimum damping

$$K \leq \frac{a_{IA}\ ml}{\Delta T}\sqrt{\frac{J}{A_m\ M_{tg}}} = \frac{a_{IA}}{\Delta V}\sqrt{\frac{J\ M_{tg}}{A_m}} \tag{IV-14}$$

Refer to Fig. IV-22. The torque generator flats to create a pseudo-salient pole are not just a single planar surface. There a multiple planar surfaces near the cylindrical rotor portion designed to reduce the elastic restraint to a negligibly small value. The accelerometer itself is the best means of measuring the elastic restraint coefficient. The minimum detectable acceleration is below 0.05 cm/sec^2.

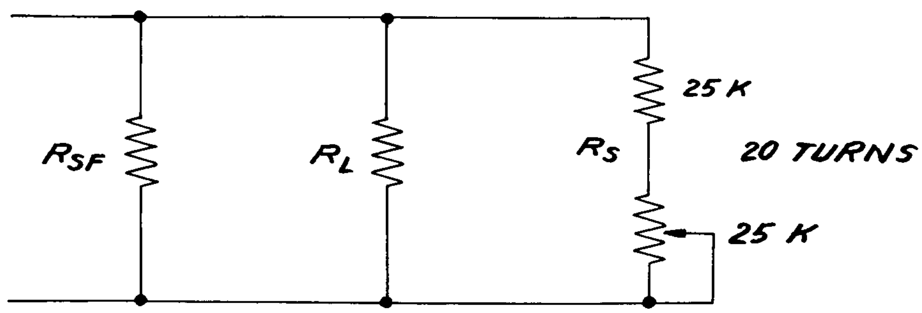
Scale Factor and Bias - The constant current loop with the reference and switching is shown in Fig. IV-23. A d.c. differential amplifier compares the voltage drop across a scale factor resistor with that of a precision voltage reference. The torque generator of the PIP is a V connected microsyn torque generator. Torque is developed proportional to the difference of the squares of the current in odd poles and even poles. Switching of current is controlled by appropriate drive voltages to the basis of the switch transistors for M_c^+ or M_c^-. The network of resistances and capacitance associated with each torque are chosen to make the impedance of the switched network resistive. That is to say $\frac{L_1}{R_1} = R_1 C_1$.
This for the amplifier has the output impedance always resistive. As will be shown later it is desirable to have the voltage across the compensated windings as high as possible to reduce the current rise time. However, the upper limit is controlled by the breakdown voltage of transistors used in this switching application. The amplifier gain is chosen to reduce the effects of supply voltage disturbances to acceptable limits. The dynamic response of the amplifier is chosen to reduce the current squared effects due to supply voltage changes to an acceptable vaue.

$$\frac{I}{E} = \frac{\frac{Y}{Y\ R_{SF}\ K}}{\tau\ s + 1} + 1 \tag{IV-15}$$

47

For a change of E_S of 10% or 12 volts the gain, K = 200,000, is determined such that the current changes are less than 10 ppm*. Dynamic response of the control loop, principally controlled by the amplifier, is set to minimize current transients. Loop compensation is required for stability and also to control current transients due to switching. The current rise time for the current in the torque generator (I_1, I_2) is desired to be a small as possible. The amplifier, drive impedance is tuned to be resistive to minimize current transients. Thus the torque generator impedance is tuned to be resistive with the addition of a resistor and capacitor. The current rise time, L_1/R_1 is then controlled by R_1 since L_1 is controlled by the torque generator. The current requirement, I_{SF}, is determined from the scale factor requirement and the torque generator sensitivity. This leaves only the supply voltage, E_S, as the variable for R_1 determination. The desire to increase R_1 is then determined as a compromise between the desire to increase R_1 and the limitation of E_S for voltage breakdown reasons, V_{CE}, of the current control transistors. Based upon the current state of the art for reliable producible transistors 120 volts was chosen as a suitable compromise. In the steady state case this leaves about 80 volts across C_1 or C_2.

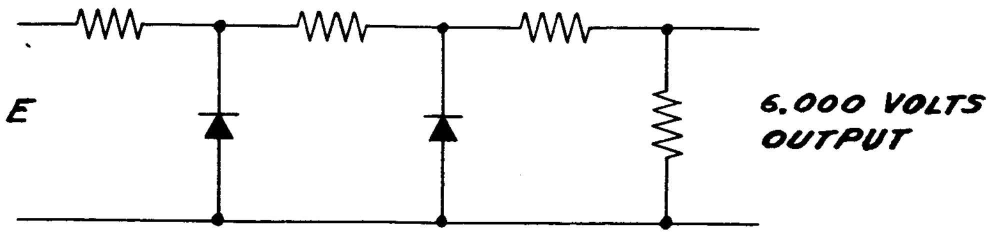
The stability of the current loop is determined by two other important items, the scale factor resistor and the voltage reference, call Precision Voltage Reference, PVR. Precision resistors with low temperature coefficients are commerically available. The method used for setting the desired conductance is to normalize all PVR's to a particular value, 6.000 volts and then adjust two fixed values R_L and R_{SF} with the ratio $\frac{R_L}{R_{SF}}$ between 100 and 1300:



With this method precision adjustment is possible by the use of the potentiometer. Resistance stability is generally better than 3 ppm/year with temperature coefficients of less than 3 ppm/C° deviation. The Precision Voltage Reference

* This work done by James L. Sitomer, Instrumentation Laboratory.

is a cascaded zener diode voltage source.



The second stage is chosen for a voltage where the temperature coefficient of the diode is nearly zero. This is slightly greater than 6 volts which is the point at which the temperature coefficient changes sign. The resistor divider network on the output is the normalization of the PVR to 6.000 volts. Voltage deviations of the PVR are less than 10 ppm/year with changes due to temperature less than 3 ppm/C$^{\circ}$. This then forms the major portion of the constant current loop.

As in any accelerometer the stability of parameters in one of the most important considerations. From previously

$$SF = \frac{S_{tg}\, i^2\, \Delta T}{ml}$$

$$dSF = \frac{\partial SF}{\partial S_{tg}}\, dS_{tg} + \frac{\partial SF}{\partial i}\, di + \frac{\partial SF}{\partial\, \Delta T} + \frac{\partial SF}{\partial\, ml}\, dml \qquad \text{(IV-16)}$$

$$\frac{dSF}{SF} = \frac{dS_{tg}}{S_{tg}} + \frac{2\, di}{i} + \frac{d\Delta T}{\Delta T} - \frac{d\,ml}{ml}$$

This is the fundamental relation describing stability of the scale factor. Mechanization is required to minimize these changes. The pendulosity is very stable. The current control has been previously described to minimize $\frac{di}{i}$. A single computer pulse train is used for switching the torque. As a precaution against any modulation of the switching pulse amplitude and the resulting timing changes a second pulse train, interrogration pulse, which precedes by a short time the switching pulse is used to set the state of the gates to allow switching with the switch pulse. This eliminates any apparent elastic restraint effects which might result from summing the float angle information with the switching pulse.

The primary scale factor stability parameter is the torque generator sensitivity.

The torque generator sensitivity is proportional to the permeability and inversely proportional to the gap. The gap is controlled by the magnetic suspension

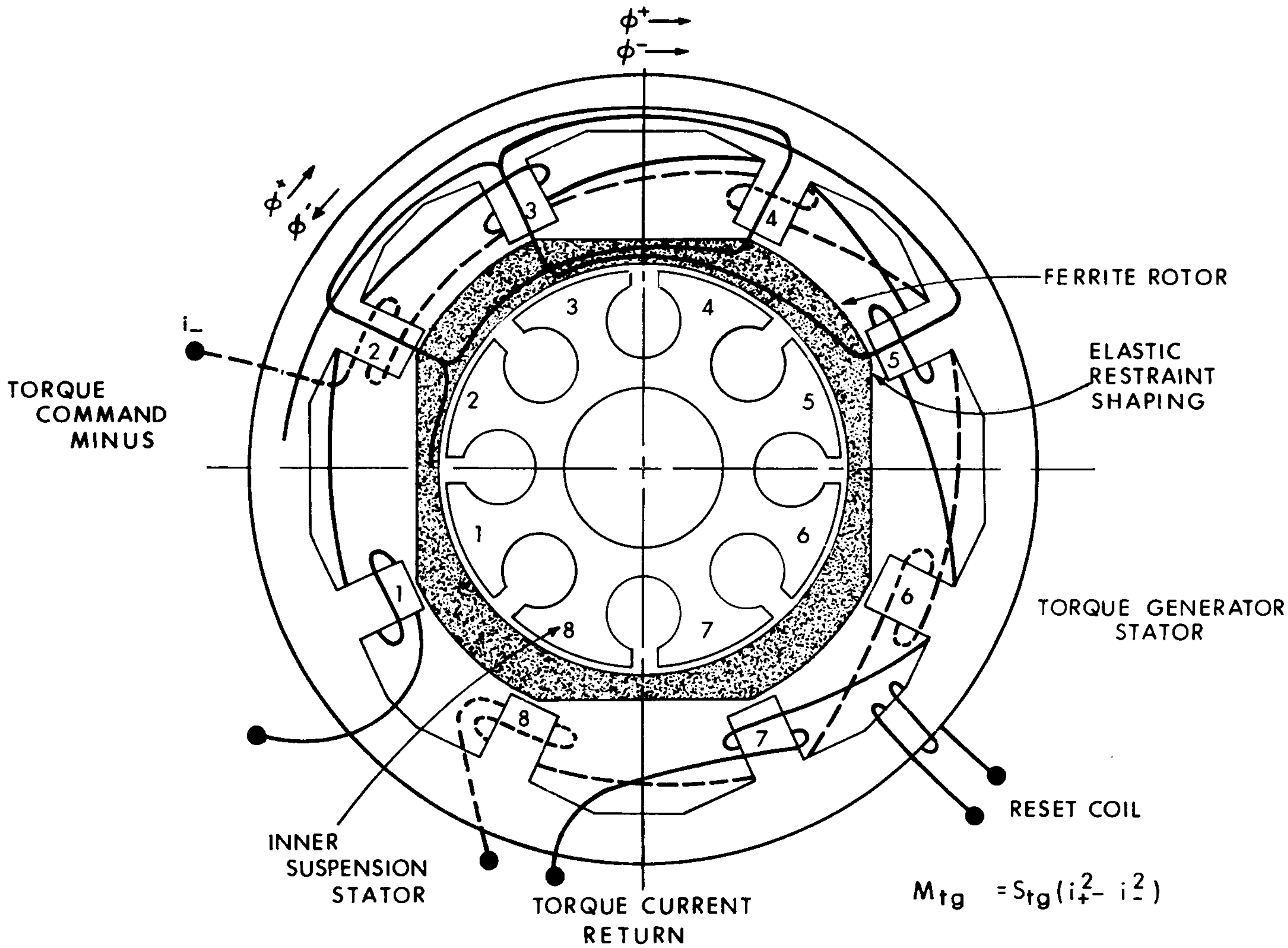


Fig. IV-23 Apollo 16 PIP Mod D Torque Generator Schematic Diagram system with radial force restraints of 11.6 n/mm and about 1/20 this stiffness for the axial restraint.

The stabilization of the permeability requires a number of things. First a review of the torque generator will show that at the stator locations between poles there are alternate locations of d.c. and a.c. flux. These d.c. flux regions are locations where memory can be maintained. This memory may be erased by the reset coil which has an a.c. flux in the back iron of the stator. The rotor memory is erased by the suspension system through the one piece rotor.

The quality of any accelerometer is determined by its performance stability. The bias, a_b, of the accelerometer (output with no input) and its stability has another relationship. The torque generator is compensated to the resistive. Upon switching, the current in one winding decays exponentially, while it rises exponentially in the other winding. The torque is proportional to the difference of the square of the current.

$$M_{tg} = S_{tg} \left(i_1^2 - i_2^2 \right)$$

If M_{tg} is integrated over one limit cycle period and averaged the bias due to the torque generator is obtained.

$$M_{avg} = \frac{1}{2n \, \Delta T} \int_0^{2n \, \Delta T} S_{tg} \left(i_1^2 - i_2^2 \right) dt = ml \; a_b$$

Thus bias acceleration, a_b, is proportional to difference to the current rise time.

$$a_b = \frac{\Delta V}{\Delta T} \; \frac{\tau_2 - \tau_1}{2n \; \Delta T} \qquad \text{(IV-17)}$$

It is important to have these two times identical or the bias would be a function of the period $2n \, \Delta T$. The accelerometer itself is the best instrument to use to adjust these two rise times to be identical. By introducing a phase delay (decreasing the damping for example) the moding may be altered. Then the change in bias is proportional to the current time constant differences.

The other parameter of primary significance in an accelerometer is the alignment and stability of the input axis. The mounting flange can be seen on the view of the pendulum. On that flange is a slot. The surface of the flange and the slot then form the reference alignment directions. The stable member contains the mounting surface and pin with the proper tolerances. The input axis is aligned about the output axis by rotating the PIP with respect to the slot. The mounting flange contains two rings which have mating spherical surfaces. Alignment of the input axis about the pendulous reference axis is obtained by sliding between these surfaces. The two rotations are uncoupled for small rotations. The pendulum is thus prealigned prior to its assembly into the IMU. Like the gyro it also contains a module of the necessary suspension capacitors and other normalizing components. For alignment reasons this module fastens into the stable member and not the pendulum.

CHAPTER IV-3

THE COUPLING DATA UNIT (CDU)

The Coupling Data Unit, provides the central angle junction box between the IMU, Optics, Computer and certain portions of the spacecraft analog electrical interfaces. There are three basic portions of the CDU. They are the angle read system or analog to digital conversion process, the digital to analog conversion process, and a portion of the moding controls for the guidance system. The analog to digital system will be covered in some detail and other portions only mentioned.

Analog to Digital Conversion - Angle information is stored in a two speed resolver system of a control member, for example a gimbal axis or an optical axis (see Figure IV-24.) The output of the resolver is proportional to $\sin\theta$, $\cos\theta$, $\sin n\theta$, $\cos n\theta$, where n is a binary number. It is a common technique to have both the single speed and multiple speed resolver use the same iron and utilize a single excitation winding. The second excitation winding space phased 90° with respect to the primary excitation is used for electrical zero adjustment. The elements of the angle read system are an analog multiplication of the resolver output, an analog summation, a sampler and quantizer, a storage counter to control the analog multiplication, and a. c. switches controlled by the counter to gate inputs to the analog multipliers.

The equation mechanized is:

$$\sin\theta \cos\psi - \cos\theta \sin\psi = \sin(\theta - \psi) \qquad \text{(IV-18)}$$

$$\cos\theta \cos\psi + \sin\theta \sin\psi = \cos(\theta - \psi) \qquad \text{(IV-19)}$$

Where ψ is a quantized angle in increments of 11 1/4 electrical degrees, and
θ is the angle of the control member

As shown in Figure IV-25 this output is summed with a quantized linear interpolation of difference between $\sin(\theta - \psi)$ and θ. The selection of the quantized angle ψ and the quantized linear interpolation angle ϕ, is based upon the contents of an angle counter register. The angle counter inputs are gated by the phase of the summed voltages. Thus when the contents of the counter are equal to the control member angle:

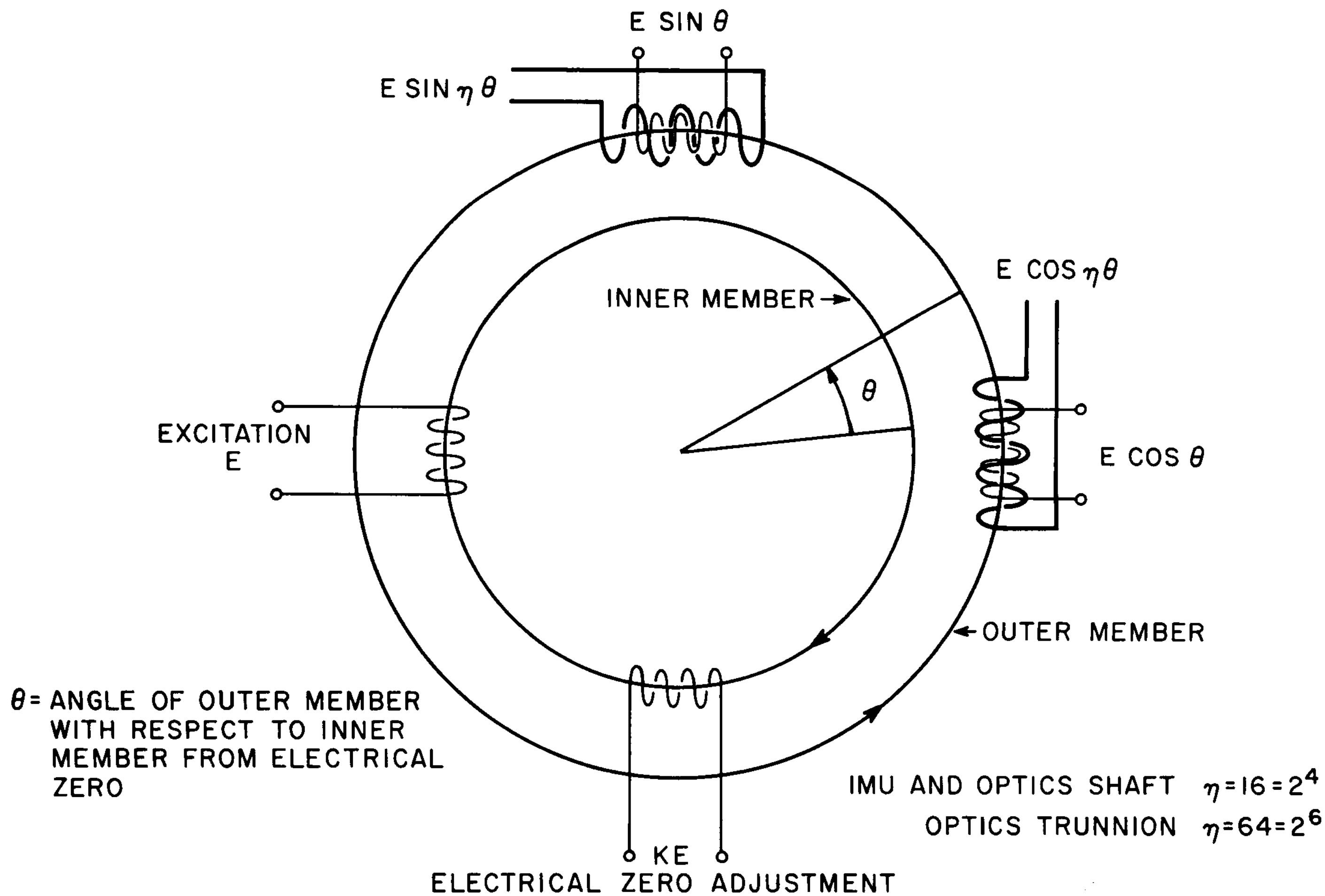
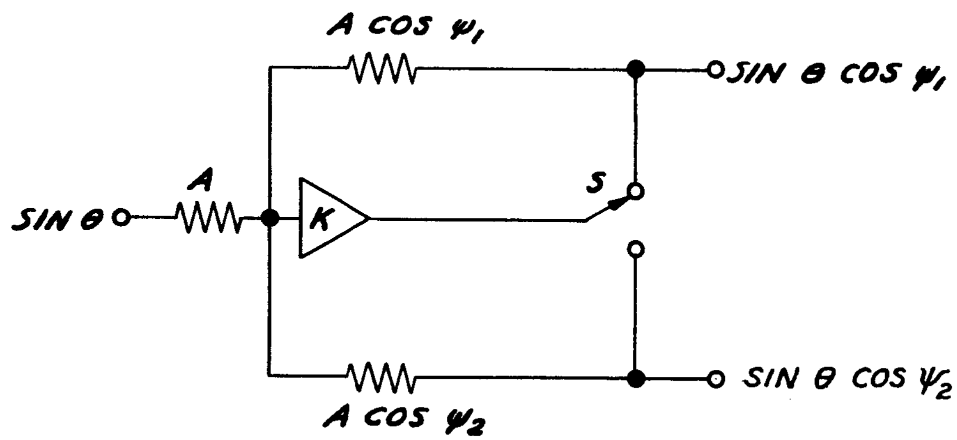


Fig. IV-24 Resolver Schematic

$$\sin(\theta - \psi) + K \cos(\theta - \psi) = 0 \qquad \text{(IV-20)}$$

The inputs to the counter are angle increments of the control member and these are parallel fed to the computer where the same control member angle information is stored.

Analog multiplication of the resolver $\sin\theta$ and $\cos\theta$ voltage is



accomplished by the use of an ac operational amplifier with the ratio of the feedback resistor to the input resistor equal to the cosine of the angle ψ_1. S is an ac transistorized switch gated closed or open by contents of the angle counter resistor. The switch is in series with the feedback resistor to take advantage of the high impedance open condition and low impedance closed condition which makes the impedance equivalent to a portion of the amplifier gain, K. Since the switch is effectively a single pole double throw switch the output of the open side differs from zero only by the input signal divided by the amplifier gain. Using the technique of transistorized ac switches and operational amplifiers the read system is mechanized. The read counter contains 16 bits. The lowest order bit is used to eliminate transmitting any limit cycle operation to the computer and thus creating unnecessary activity. The four highest order bits are used for quadrant selection and multiplication of the single speed resolver (ψ_1). In addition, the bits 2^9 - 2^{12} are used as an approximate linear interpolation of the single speed resolver to within 2.81° of the actual angle. The multiple speed resolver (sixteen speed) is the precision angle transmitting device. The zero-to-peak errors of this resolver are less than 20 seconds of arc. There is crossover between the sixteen speed and the single speed resolvers to assure synchronization of the reading of the sixteen speed resolver within the proper cycle. The lowest order bits are a linear interpolation of error using the voltage of the $\cos(\theta - \psi)$ as a source. This voltage has the same phase relation as the $\sin(\theta - \psi)$ of the sixteen speed resolver and is

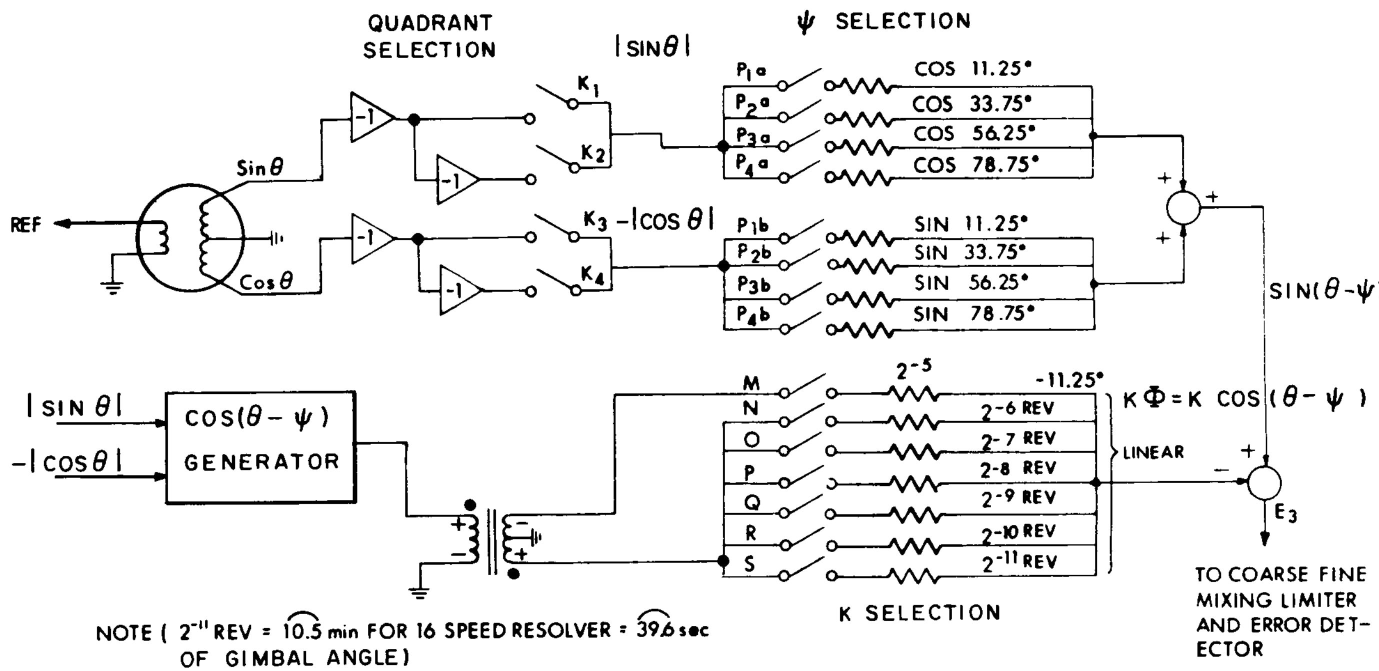
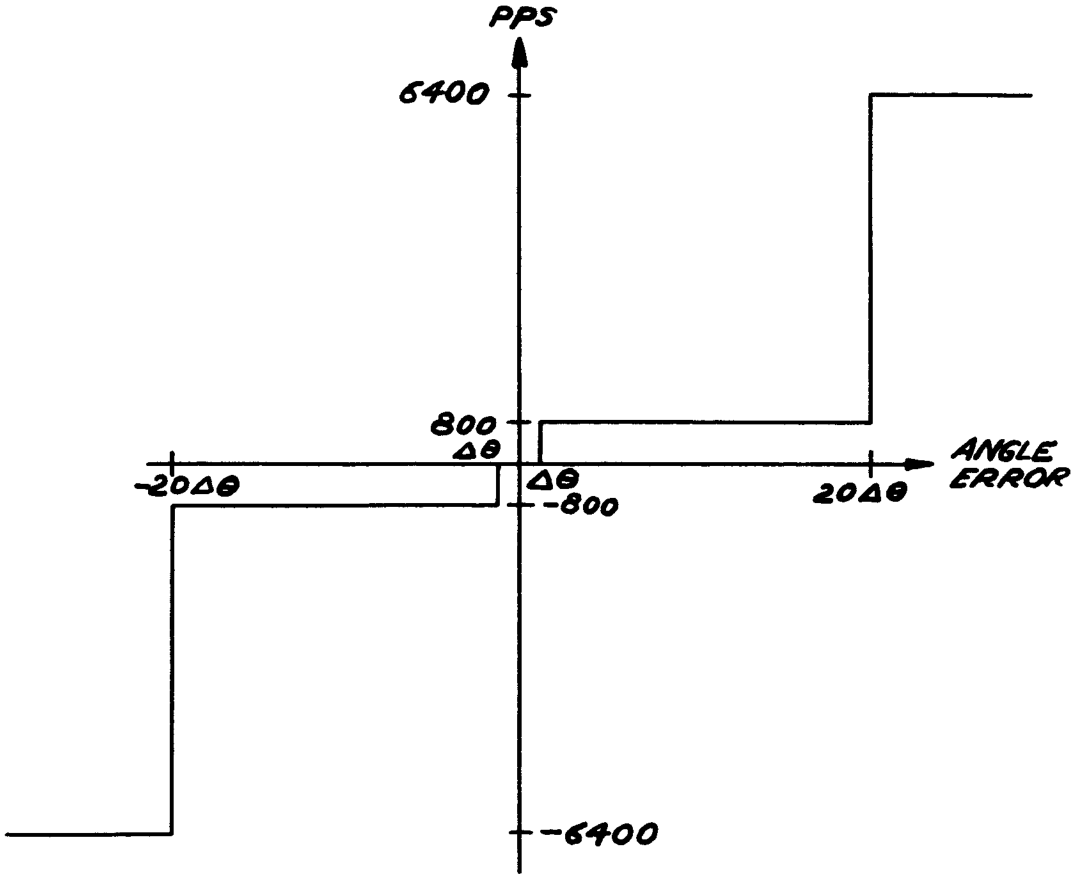


Fig. IV-25 Selection Logic - 16 Speed Resolver Digitizing Loop scaled correctly by the resolver attenuation.

Referring to Fig. IV-26 the input to the error detector is the sum of the single speed multiplication, the sixteen speed multiplication and the linear interpolation. There is a coarse-fine mixing network to assure synchronization and angle measurement using the precision resolver. The error detector contains an active feedback quadrature rejector network which for large error signals will not introduce dynamic errors for reading the angle, but for small errors will yield the proper precision. The output of the error detector is fed to both the rate selection logic and up-down counter logic. The contents of the counter are used to control the a. c. switches for the multiplication of the resolver voltages and the linear interpolator.

The error detector has three-state or ternary logic. The lowest order pulse rate command to the counter is 800 pulses per second. Using this as the lowest order assures switching of equal multiples of the resolver carrier frequency, 800 cps. This prevents rectification of the switched signal and altering of the dynamic operation of the read system.



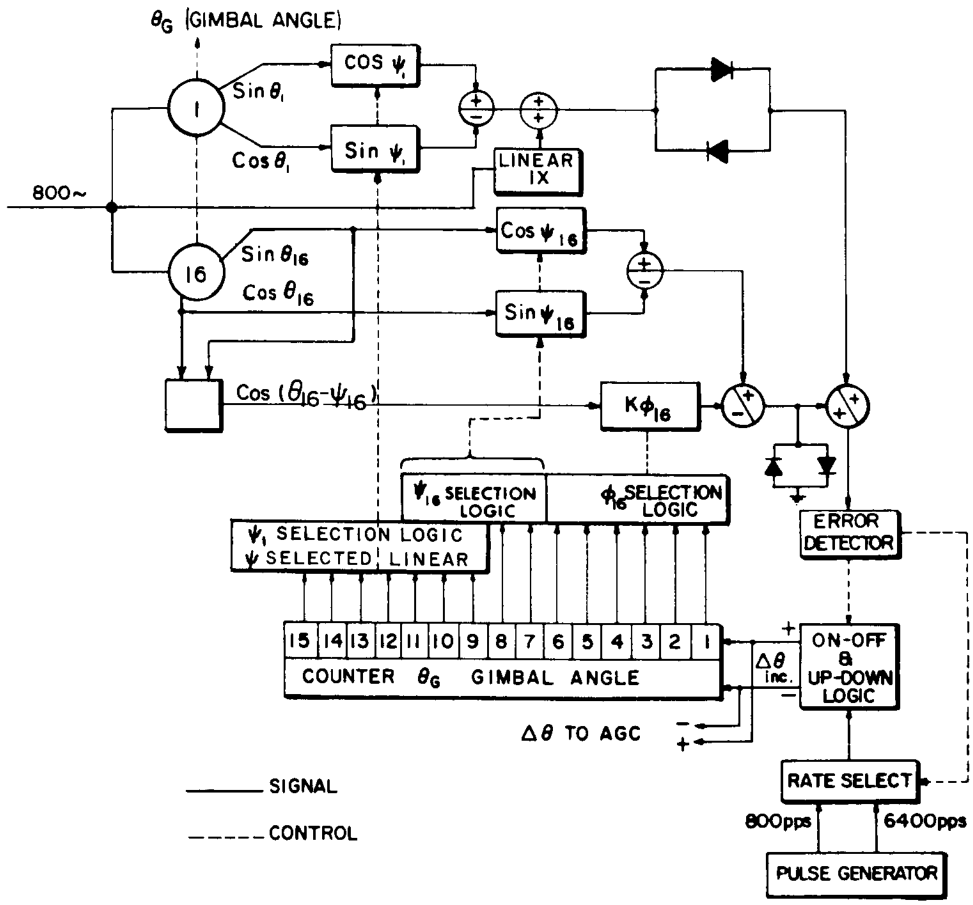


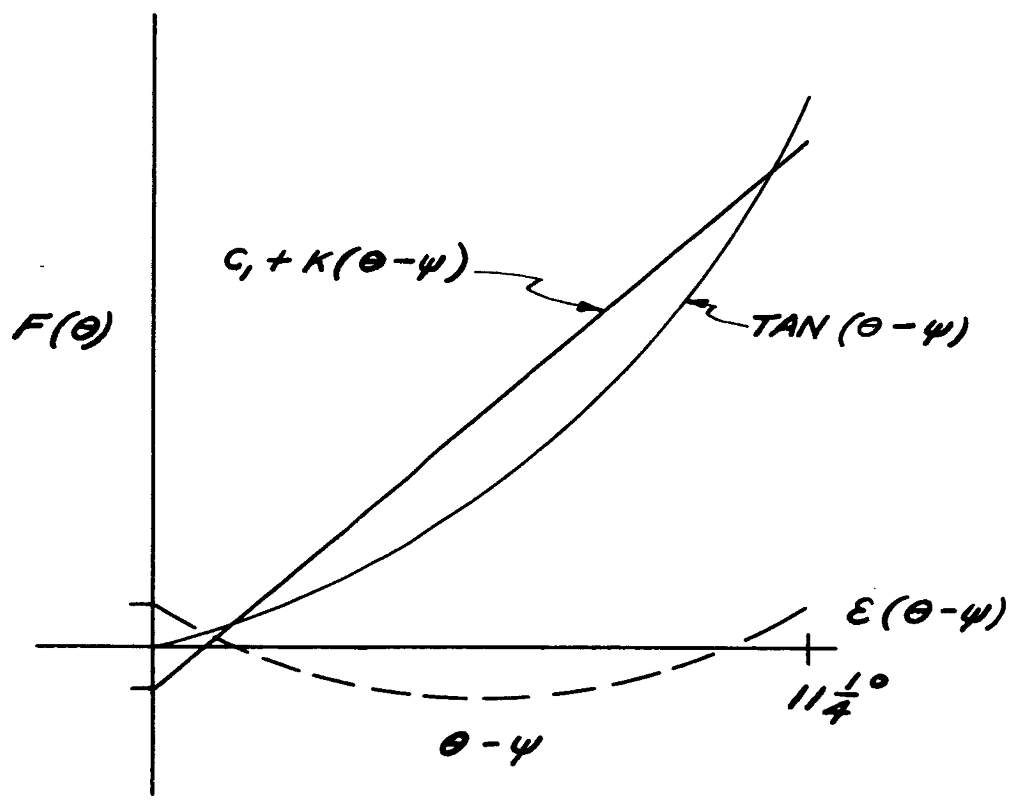
Fig IV-26 Coarse-Fine Mixing (1 and 16 Speed Resolver)

The high speed rate following command reduces dynamic error for high angular velocity inputs and the low speed command rate reduces the limit cycle error.

The linear interpolation constant K is adjusted to minimize the peak error

$$E = \sin(\theta - \psi) - K \cos(\theta - \psi) \qquad \text{(IV-21)}$$

By suitably choosing K the error for the speed resolver system can be reduced to less than 10 sec arc. For a 64 speed system as used on the optics trunnion this error is reduced to less than 3 sec arc.



In addition, a bias is added to further reduce the errors over the entire range of linear interpolation. This results in a system whose errors are within the predicted errors.

All digital functions including the memory are mechanized with the three input NOR gate, a silicon semiconductor micro-integrated circuit. This is the same element used in the computer. Direct coupled transistorized logic is used throughout. A multiple phase clock system, generated within the CDU and synchronized

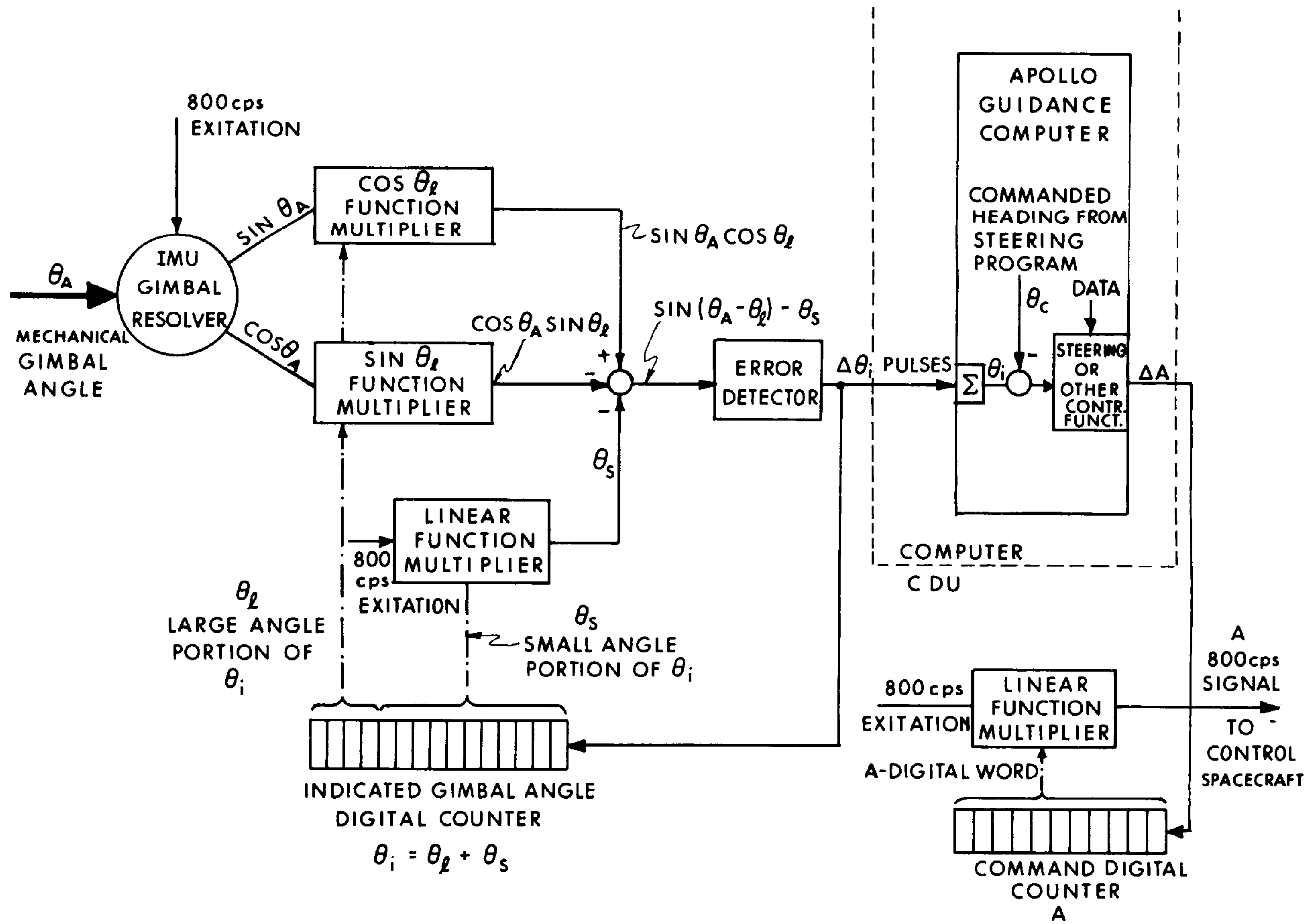


Fig. IV-27 Electronic Coupling Data Unit Schematic to the guidance computer, is used to control all functions.*

Moding and Digital to Analog Conversion - The guidance computer serves as the important link between the spacecraft and the sensing device. All angle transformations are made by the guidance computer based upon IMU gimbal angles or optics angles. For any steering function utilizing the IMU, the computer through the use of knowledge of the gimbal angles provides steering and attitude commands to the spacecraft via the digital to analog converter. The angle information is always stored in the resolvers and no mechanical rezeroing of the IMU is necessary to establish within the counter the IMU gimbal angle within the uncertainty of the resolver, the error of the read system and the bit size of the analog to digital converter.

The digital to analog converter system is required to accept digital commands from the guidance computer and generate analog voltages (a. c. and d. c.) proportional to these commands. The error angle counter is an eight bit up-down counter. Computer commands are stored in this counter. The counter is logically controlled to prevent it from being reset to zero in the event it is commanded to more than 2^8 increments. The analog error signals are developed by an 800 cps source gated by the counter contents through a resistance ladder to the input of an operational amplifier. For polarity reversal by switching is used to command the 0 phase or π phase 800 cps input to the ladder. All analog voltages used as steering commands to the LEM, Spacecraft or other portions of the Apollo launch vehicle are d. c. isolated from the guidance system either by transformers or an isolated demodulator.

The moding of the system is almost entirely controlled by the computer. Provision is made for two modes to be controlled by the astronaut. One mode is the cage function. In the event of a spacecraft tumble and loss of attitude, provision is made for the astronaut to cage the IMU gimbals with respect to the spacecraft after he has first stabilized the craft. This then gives him a means of obtaining a new reference in a very short time. The other manual mode is similar in that with the computer not operating he can use the same switch to cage the IMU with respect to the spacecraft, release it, and again use the IMU as an attitude reference.

All other moding is controlled by the computer as can be seen from Fig. IV-28 for the IMU. Optics moding has more manual control and is shown in Fig. IV-29. This moding will be described elsewhere. There are a number of interesting modes possible because of the flexibility of the CDU. The Coarse Align Mode, that of commanding the IMU gimbal angles by the resolvers, is rate controlled to limit angular velocity input to the gyroscopes. The input to the error detector is sum-

---

*This work is by Robert Crisp and Glenn Cushman, MIT Instrumentation Laboratory

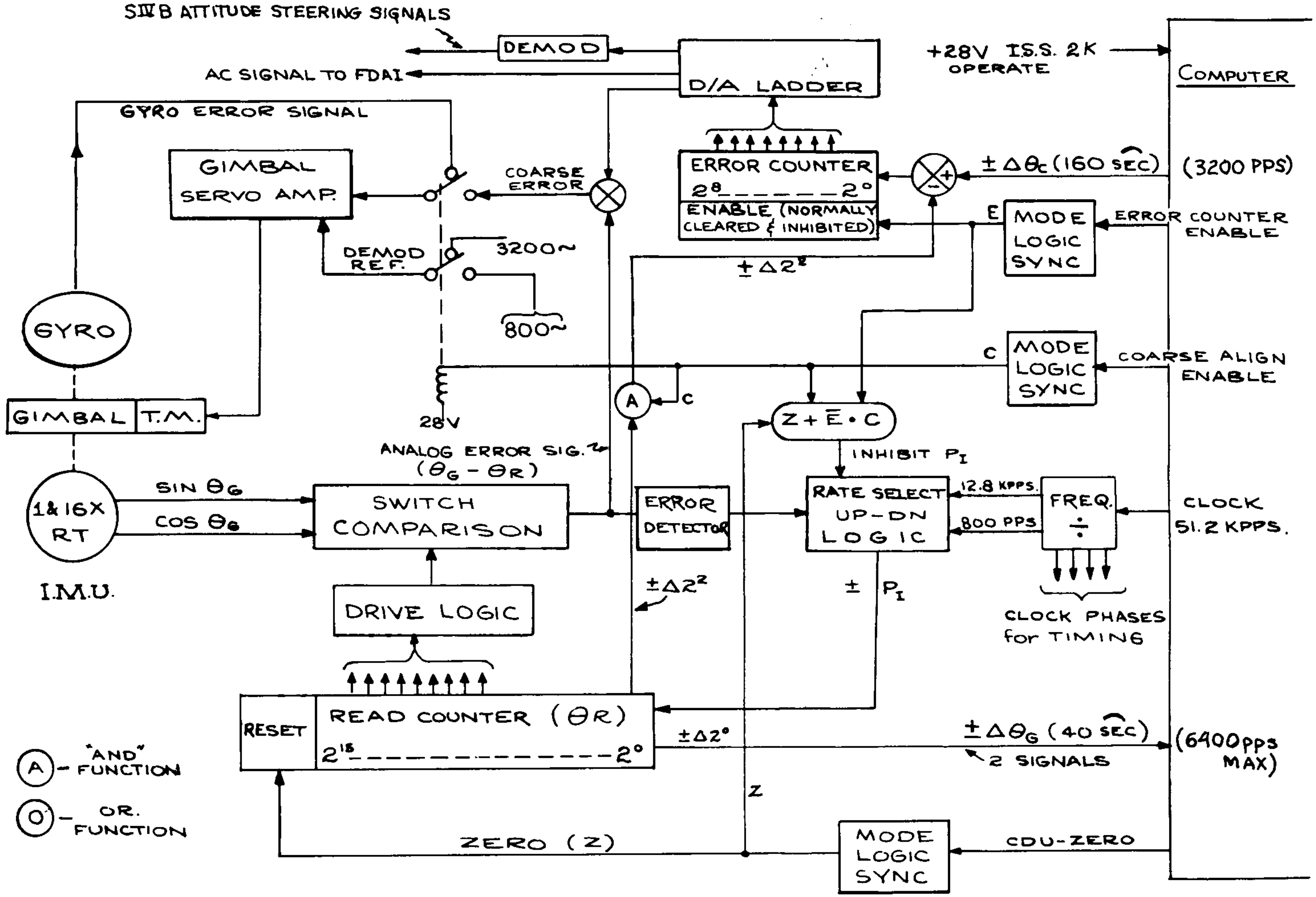


Fig IV-28 IMU-CDU Moding

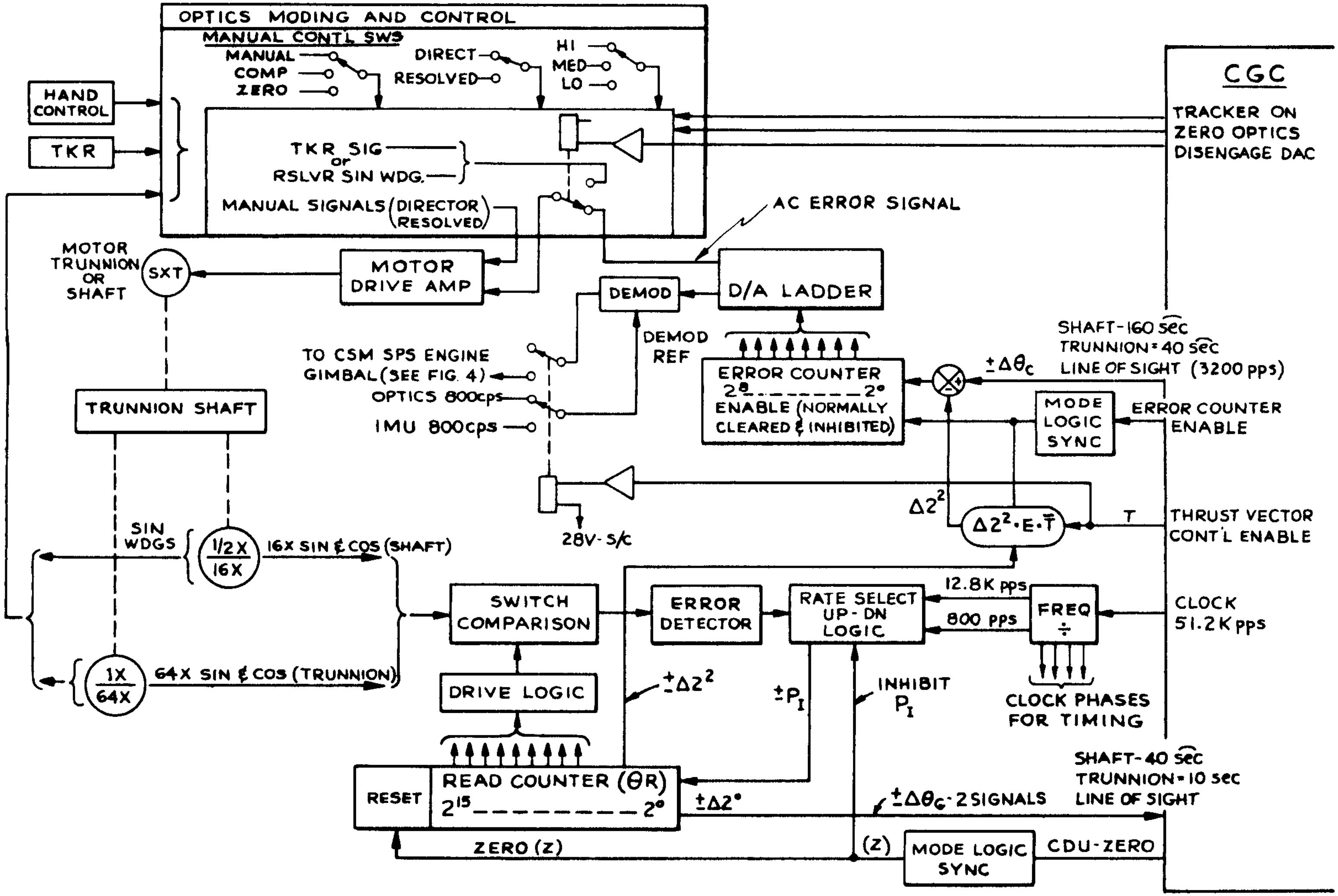


Fig. IV-29 Optics-CDU Moding (CSM only) med with the analog command from the computer to provide stable operation. The gimbal angle pulse increments from the read system are used as feedback pulse to the error angle counter.

BIBLIOGRAPHY

1. C. S. Draper, W. Wrigley, and L. R. Grohe, "The Floating Integrating Gyro and Its Application to Geometrical Stabilization Problems on Moving Bases," Institute of the Aeronautical Sciences Preprint 503, presented January 25, 1955.

2. P. J. Gilinson Jr., W. G. Denhard, and R. H. Frasier, "A Magnetic Support for Floated Inertial Instruments." Report No. R-277 Instrumentation Laboratory, MIT, Cambridge, Massachusetts, April 1960.

3. Thomas Freud Wiener, "Theoretical Analysis of Gimballess Inertial Reference Equipment Using Delta-Modulated Instruments, MIT, Cambridge, Massachusetts, Sc D Thesis.

PART V

OPTICAL MEASUREMENTS AND NAVIGATION PHENOMENA

by

D. Alexander Koso

D. ALEXANDER KOSO

Assistant Director, Instrumentation Laboratory
Massachusetts Institute of Technology

D. Alexander Koso, Assistant Director of Instrumentation Laboratory, Massachusetts Institute of Technology, heads the Laboratory group responsible for development of the optical subsystem — space sextant, scanning telescope and the associated electronics — used in the guidance system the Laboratory is developing for the Project Apollo spacecraft.

Mr. Koso was born in Bratislava, Czechoslovakia, March 13, 1935, and came to the United States in 1949. He was graduated from University High School, Minneapolis, Minn., in 1952. He received both the B.S. and M.S. degrees from M.I.T. in electrical engineering in 1957 and the degree of Electrical Engineer from M.I.T. in 1959.

While an undergraduate at M.I.T., Mr. Koso studied under the Institute's co-operative plan and was employed by the Philco Corporation. As a graduate student studying for the E.E. degree, he was a research assistant at the M.I.T. Electronic Systems Laboratory.

Mr. Koso joined the Instrumentation Laboratory in 1959 and worked a few years on navigation studies for manned boost-glide space vehicles. He was appointed an Assistant Director in 1963.

Part V

OPTICAL MEASUREMENTS AND NAVIGATION PHENOMENA

INTRODUCTION

During the orbital and midcourse phases of a space mission, inertial components (due to a lack of force) can no longer provide information about the position of the vehicle. The gyroscope can be used to provide an artificial set of fixed stars usable as a basic reference for measurements. However, external sensors have to be used to update the position information within the vehicle.

During the orbital phases of a mission, it is possible to treat the navigational problem with relative ease because one can write a set of linear constant coefficient equations which describe the propagation of errors with time. Each measurement then provides a linear equation between certain of these errors. In this chapter on-board measurements are considered which can be used to determine the orbit of a satellite.

CHAPTER V-1

NAVIGATION IN ORBIT

ERROR PROPAGATION IN ORBIT

The equations of motion in this analysis are written in polar coordinates to make linear approximation of these equations easier. Thus, the equations presented are closer to the equations of a local-vertical system; however, the analysis applies irrespective of the coordinate system used in the actual computation.

The equations of motion in polar coordinates are:

$$\ddot{R} = F_R + R\Omega^2 - E/R^2 \qquad \text{(V-1a)}$$

$$R\frac{d^2\theta}{dt^2} = R\dot{\Omega} = F_\theta - 2\dot{R}\Omega \qquad \text{(V-1b)}$$

where F_R and F_θ are the forces (per unit mass of the vehicle) in the radial and range direction respectively as illustrated in Fig. V-1.

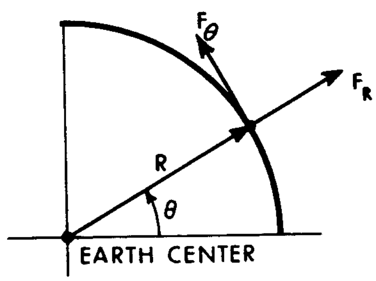


Figure V-1

For the purpose of analysis, set:

$$R = r_o + r \qquad \text{(V-2a)}$$

$$\Omega = \omega_o + \omega \qquad \text{(V-2b)}$$

where r_o and ω_o are constants.
Thus,

$$\dot{R} = \dot{r};\ \ddot{R} = \ddot{r};\ \dot{\Omega} = \dot{\omega};\ \theta - \omega_o t = \int_o^t \omega(x)dx$$

One can also assume without loss of generality that:

$$\omega_o^2 = E/r_o^3 \qquad \text{(V-3)}$$

If r_o is assumed to be about 300 kilometers above mean earth radius, then for orbits ranging from 150 to 450 km altitude:

$$r/r_o < .02 \qquad \text{(V-4)}$$

Similarly:

$$\omega/\omega_o < .05 \qquad \text{(V-5)}$$

Thus, the equations of motion can be linearized for satellites confined to nearly circular orbits.

Under the assumption that:

$$E/R^2 = E/r_o^2 - [2E/r_o^3]r \qquad \text{(V-6)}$$

Equations V-la and V-lb can be written as:

$$\ddot{r} = F_R + (r_o + r)(\omega_o^2 + 2\omega_o\omega + \omega^2) - E/r_o^2 + 2E/r_o^3\, r \qquad \text{(V-7a)}$$

$$(r_o + r)\dot{\omega} = F_\theta - 2\dot{r}(\omega_o + \omega) \qquad \text{(V-7b)}$$

Substitution of Eq. V-3 into V-7a and V-7b gives:

$$\ddot{r} = F_R + 3\omega_o^2 r + 2\,\omega_o\omega(r_o + r) + (r_o + r)\omega^2 \qquad \text{(V-8a)}$$

$$(r_o + r)\dot{\omega} = F_\theta - 2(\omega_o + \omega)r \qquad \text{(V-8b)}$$

Using the approximations of Eqs. V-4 and V-5,

$$r << r_o; \; \omega << \omega_o$$

Eqs. V-8a and V-8b become:

$$\ddot{r} = F_R + 3\omega_o^2 r + 2\omega_o r_o \omega \qquad \text{(V-9a)}$$

$$r_o\dot{\omega} = F_\theta - 2\omega_o\dot{r} \qquad \text{(V-9b)}$$

Equations V-9a and V-9b represent very closely the errors in the position computation of an orbiting vehicle. For relatively short flights, the two force inputs F_R and F_θ are also negligible, so that the major contribution of error is due to the uncertainty in initial conditions at the end of injection into orbit.

There is another error which can exist in an orbit determination. This one is due to a lack of knowledge of the exact direction of the ω_o vector (see Fig. V-2).

If the angular displacement is small, the error in the out of plane position is given by:

$$\Phi = E_\Phi \cos(\omega_o + \omega)t + \frac{E_{\dot{\Phi}}}{\omega_o} \sin(\omega_o + \omega)t \qquad \text{(V-10)}$$

For relatively small range errors, the ω term can be neglected. From a control viewpoint, the error in the position computation can be represented very closely by the two linear constant-coefficient systems shown in Fig. V-3. One set of equations is fourth order, the other second order. E_* is a step input representing the particular initial condition error.

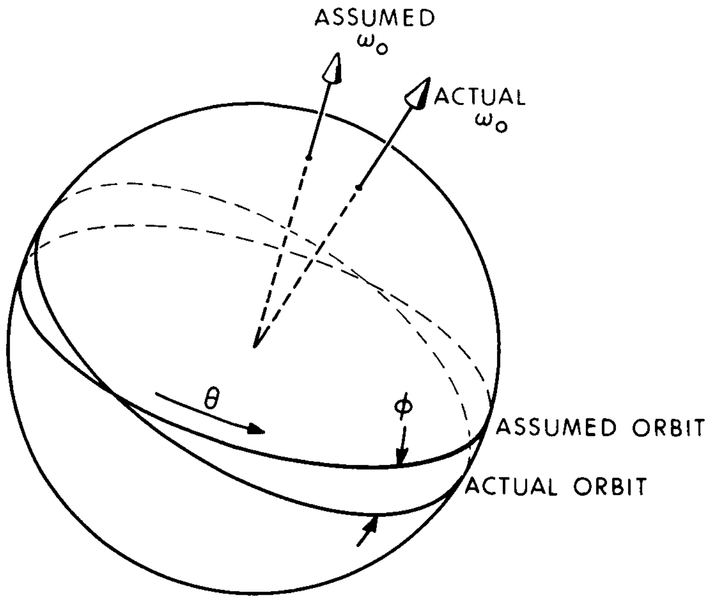


Fig. V-2 Orbital Geometry

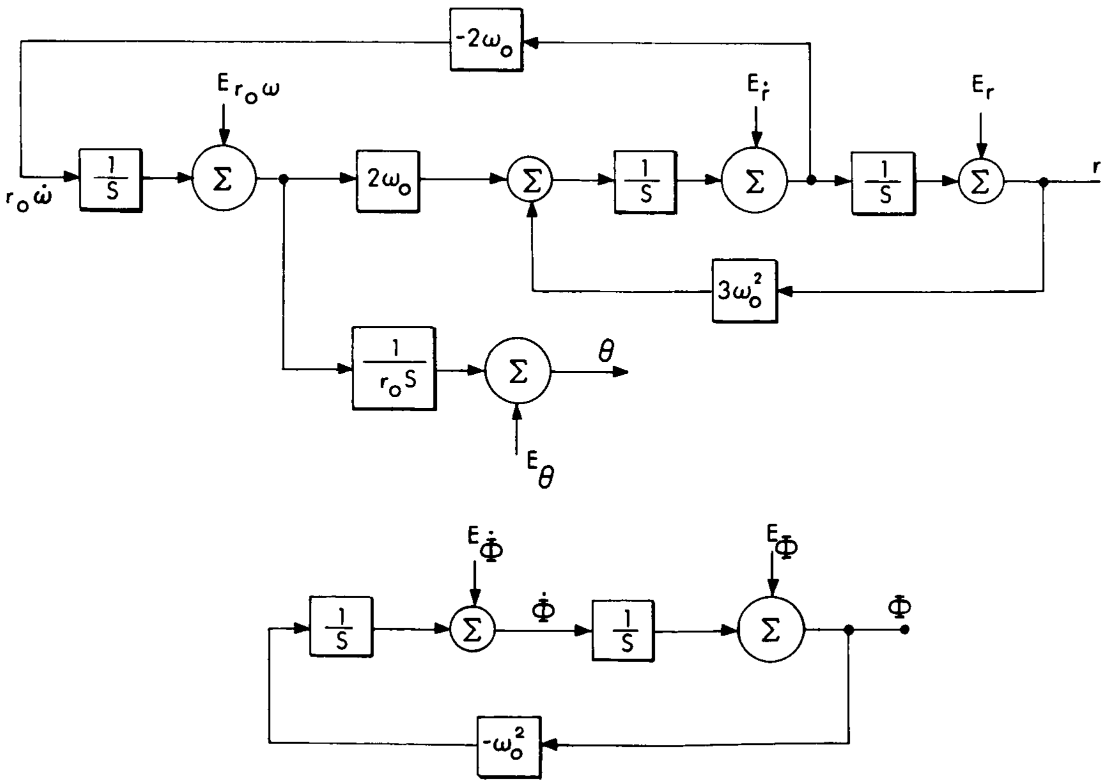


Fig. V-3 Block Diagram of Error Propagation

Table V-1 Summary of Errors due to Initial Conditions at Some Time, τ, after Insertion into Orbit

Error Source	Range	Range Rate	Altitude	Altitude Rage	Track	Track Rate
E_r	$\frac{6}{r_o}(\sin\omega_o t - \omega_o t)$	$6\omega_o(\cos\omega_o t - 1)$	$3(1-\cos\omega_o t)+1$	$3\omega_o \sin\omega_o t$	0	0
$E_{\dot{r}}$	$\frac{2}{r_o\omega_o}(\cos\omega_o t - 1)$	$-2\sin\omega_o t$	$\frac{1}{\omega_o}\sin\omega_o t$	$\cos\omega_o t$	0	0
E_θ	1	0	0	0	0	0
$E_{r_\theta \omega}$	$\frac{4}{r_o\omega_o}(\sin\omega_o t - \omega_o t) + \frac{1}{r_o}t$	$4(\cos\omega_o t - 1)+1$	$\frac{2}{\omega_o}(1-\cos\omega_o t)$	$2\sin\omega_o t$	0	0
E_ϕ	0	0	0	0	$\cos\omega_o t$	$-\omega_o \sin\omega_o t$
$E_{\dot{\phi}}$	0	0	0	0	$\frac{1}{\omega_o}\sin\omega_o t$	$\cos\omega_o t$

Two conclusions can be drawn from this Table:

Only range error grows with time

There is no error coupling between range and track errors

Generally, depending on the particular problem on hand, the navigational system has to be able to determine the position and velocity errors at some time, T, after injection into orbit. However, since this time, T, depends on a specific mission phase, only the problem of initial condition determination will be treated here. The actual errors at the specified time, T, can then easily be obtained from the initial condition errors and the expressions of Table V-1.

POSSIBLE MEASUREMENTS

As can be seen from the block diagram, Fig. V-3, there are six possible sources of error which have to be considered. The measurements which are taken have to provide some information about the Φ and θ errors. Thus, while a good altimeter can be used to determine $E_{r\omega_o}$, $E_{\dot{r}}$ and E_r, it cannot provide information about E_θ, E_Φ, and $E_{\dot{\Phi}}$. Similarly, a velocity meter, even if it could provide information about E_Φ and $E_{\dot{\Phi}}$ in addition to the variables provided for by the altimeter, would provide no information about E_θ.

While most of these considerations are of academic interest on the earth where radar tracking coverage is provided, they do become important during lunar operations, where - especially on the back side of the moon - ground based information is not available. Thus, the following measurements have been considered:

1. Bearing measurements to known landmarks
2. Bearing measurements to unknown landmarks
3. Star occultation measurements
4. Star horizon measurements
5. Star known landmark measurements

Measurements of two or three horizons simultaneously were rejected because the lines of sight have to be able to see almost a full hemisphere. This requires a sensor too close to the skin of the vehicle to make this type of a measurement feasible.

In midcourse, where the subtended angle of the planet is smaller, and this measurement becomes feasible from an equipment standpoint, the error sensitivities are so low that the accuracy requirement was deemed not feasible.

While this is not an exhaustive list of possible measurements, it does cover a large variety of applications and it can be instrumented with a relatively simple optical system.

KNOWN LANDMARK BEARING MEASUREMENT

Consider a known point on the earth' s surface and an assumed vehicle position as shown in Fig. V-4. The assumed position of the vehicle at time, T, is at a range θ_o (T) and ϕ_o (T) and an altitude H_o(T) from the known landmark. Now consider a projection of the landmark into the assumed orbital plane as shown in Fig. V-5. The angle α is the angle between the local horizontal at the landmark, and the projection of the vehicle into the assumed orbital plane.

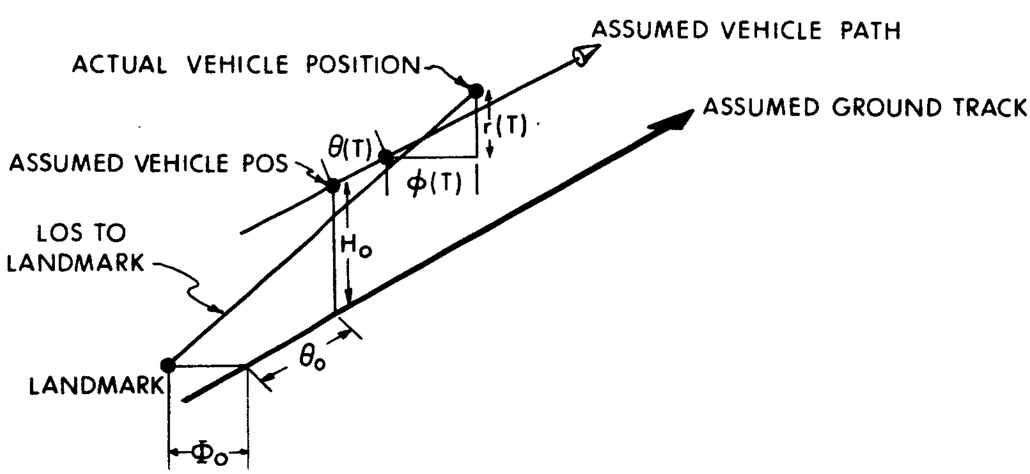


Fig. V-4 Known Landmark Measurement Geometry

77

Figure V-5 provides the first equation of orbit determination from a bearing measurement to a known landmark:

$$\frac{\theta_o(T) + \theta(T)}{H_o(T) + r(T)} = \cot \alpha \qquad \text{(V-11)}$$

Now consider a projection of Fig. V-4 into a plane which is orthogonal to the assumed orbital plane and which also contains the landmark local vertical as shown in Fig. V-6.

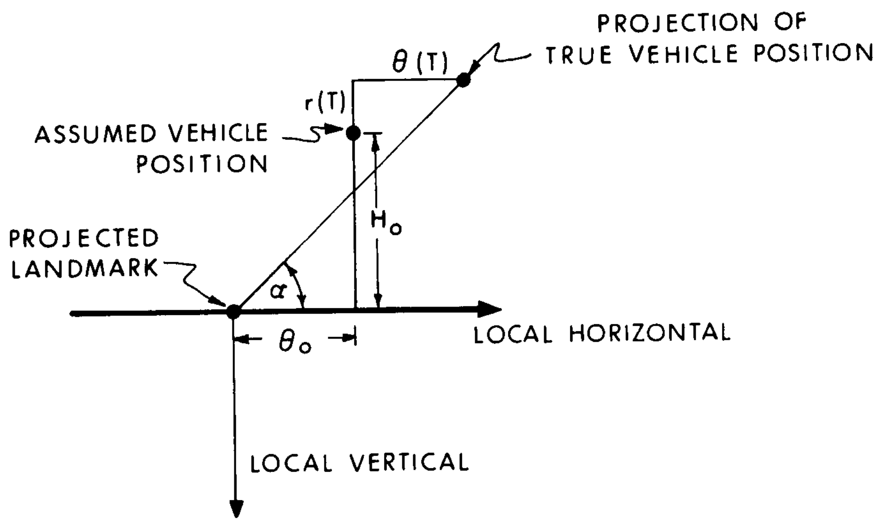


Fig. V-5 In Plane Measurement Geometry

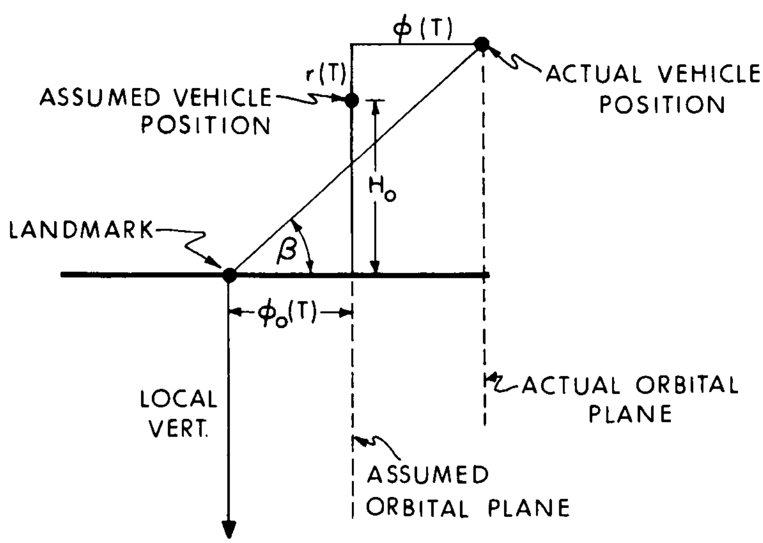


Fig. V-6 Out of Plane Measurement Geometry

This measurement provides the second equation of orbit determination:

$$\frac{\Phi_o(T) + \Phi(T)}{H_o(T) + r(T)} = \cot \beta \tag{V-12}$$

The cotangent has been chosen because $H_o(T) + r(T) > 0$ at all times and the equations are well behaved.

Let us assume that a landmark can be tracked when the angle between the local vertical at the landmark and the line of sight is less than 45^o. For a 150 kilometer orbit, this means that the landmark can be tracked for a period of less than 40 seconds. The exact amount depends on the distance between the landmark and the ground track of the vehicle.

Consider the time, T_1, when the landmark is first acquired. It is immediately possible to write two equations, V-11 and V-12, in the three unknowns $r(T_1)$, $\theta(T_1)$, and $\phi(T_1)$. Consider now the determination of other variables at time, T_1, by further bearing measurements to the same landmark. At the time, $t > T_1$, the position errors of the vehicle will have changed. The new variables are:

$$\theta(t) = \theta(T_1) + \omega(T_1)(t-T_1) \tag{V-13a}$$

$$r(t) = r(T_1) + \dot{r}(T_1)(t-T_1) + r_o\omega_o\theta(T_1) \tag{V-13b}$$

$$\phi(t) = \phi(T_1) + \dot{\phi}(T_1)(t-T_1) \tag{V-13c}$$

Substituting Eqs. V-13 into V-11 and V-12 provides a complete solution to the orbital navigational problem:

$$\begin{aligned}\theta(T_1)+(t-T_1)\omega(T_1)-\cot\alpha(t)r(T_1)-\cot\alpha(t)(t-T_1)[\dot{r}(T_1)+ r_o\omega_o\theta(T_1)]\\ = H_o(t) \cot\alpha(t) -\theta_o(t)\end{aligned} \tag{V-14a}$$

$$\begin{aligned}\phi(T_1) +(t-T_1)\dot{\phi}(T_1)-\cot\alpha(t)r(T_1)-\cot\alpha(t)(t-T_1)[\dot{r}(T_1)+r_o\omega_o\theta(T_1)]\\ = H_o(t) \cot\beta (t) - \Phi_o(t)\end{aligned} \tag{V-14b}$$

However, four bearing measurements to a single landmark are required before Eqs. V-14a and V-14b can be solved. Three measurements do provide six equations in the six unknowns; but they are not independent.

UNKNOWN LANDMARK BEARING MEASUREMENT

When the point on the earth's surface has unknown coordinates, the navigational system has to rely on the changes in the tracking angle as the vehicle passes over the landmark.

Let us assume that the astronaut can lock-on to an identifiable but otherwise unknown point on the earth's surface when the angle between the computed (or assumed) velocity vector and the line of sight to the point is approximately 45°. The exact number will depend on the skill of the astronaut and on the range error which the navigational system has accumulated by the time of the measurement.

Consider the measurement geometry of Fig. V-7 and the projections of this geometry into the orbital plane as shown in Fig. V-8.

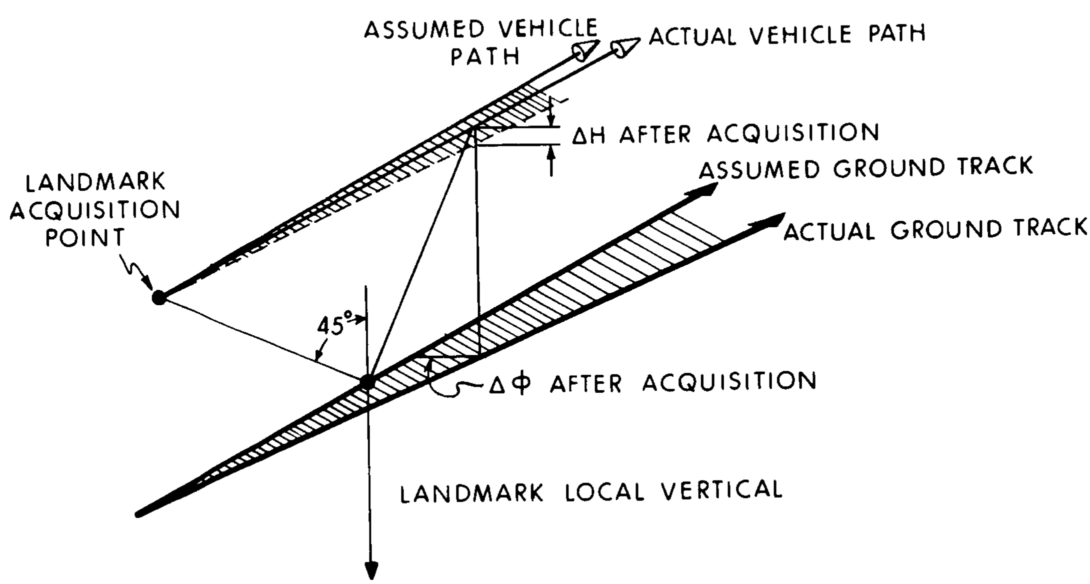


Figure V-7
Measured Geometry for an Unknown Landmark

As can be seen from Fig. V-9, it is impossible to determine the difference between an altitude error in the vehicle path and a lack of knowledge in the altitude of the landmark. For a general time, τ, after T_1, the actual bearing measurement given

$$\cot \alpha_a = \frac{S_o - r_o \omega_o \tau}{H_o + r(T_1) - \Delta H} + \frac{\cot \alpha(0)(r(T_1) - \Delta H)}{H_o + r(T_1) - \Delta H} \tag{V-16}$$

where ΔH is the altitude uncertainty of the landmark.

If the vehicle has an error $\omega(T_1)$ when tracking begins, then the measurement geometry is as shown in Fig. V-10.

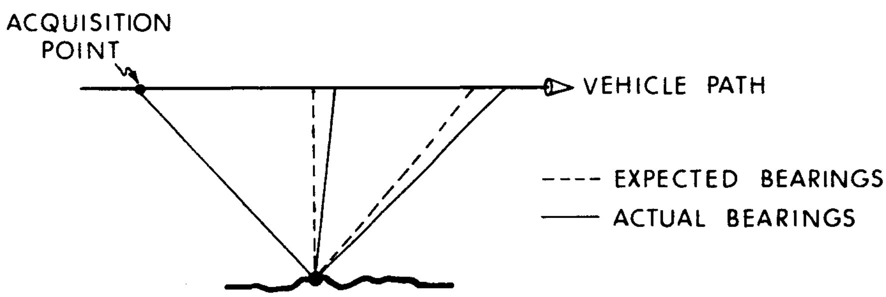


Fig. V-10 Tracking with an Erroneous Velocity Along Range

The actual bearing measurements in this case are:

$$\cot \alpha_a = \frac{S_o - (\omega_o + \omega(T_1)) r_o \tau}{H_o} \tag{V-17}$$

Again τ is the time after landmark acquisition.

Tracking with an initial vertical velocity error provides the geometry of Fig. V-11. In this case the bearing measurements are:

$$\cot \alpha_a = \frac{S_o - r_o \omega_o \tau}{H_o + \dot{r}(T_1)\tau} \tag{V-18}$$

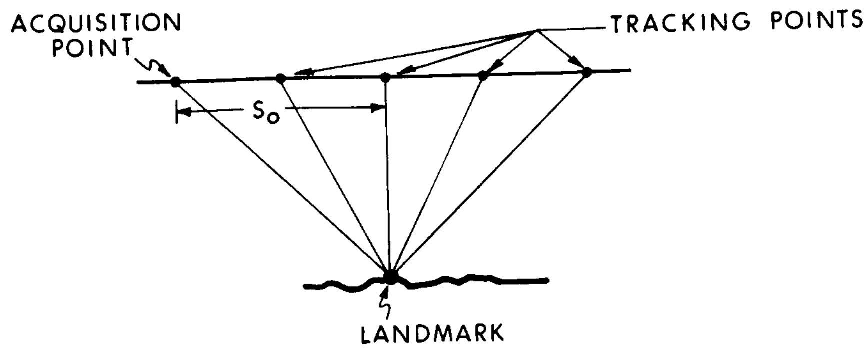


Fig. V-8 Nominal Tracking Sequence

The problem is the determination of the errors: $r(T_1)$, $\dot{r}(T_1)$, $\theta(T_1)$, $r_o\omega(T_1)$, $\phi(T_1)$, $\dot{\phi}(T_1)$ at the time of unknown landmark acquisition. As will be seen, it is not possible to determine all of the errors after the tracking of a single landmark; at least four landmarks have to be tracked before all of the orbital parameters of a vehicle can be determined.

Let us assume that the distance between the intersection of the landmark local vertical and assumed velocity vector and the point where tracking begins is S_o (Fig. V-8). For a general time, τ, after T_1(T_1 is the time when tracking begins), the computed bearing measurement is given by:

$$\cot \alpha_c = \frac{S_o - r_o \omega_o \tau}{H_o} \tag{V-15}$$

This variation in the bearing angle has to be compared to the angle which actually gets measured.

Consider first an erroneous altitude of the vehicle $r(T_1)$.

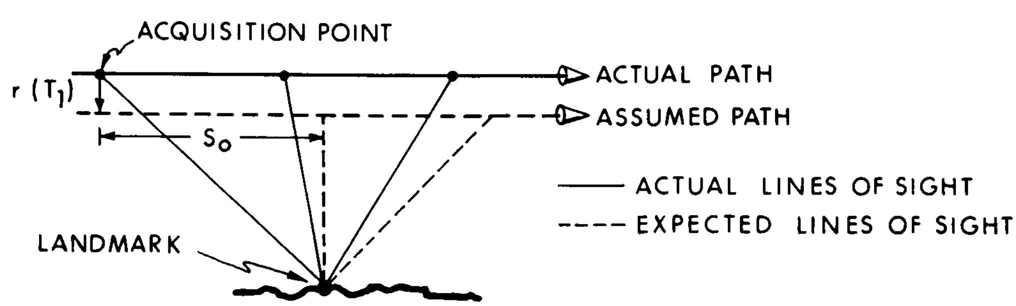


Fig. V-9 Tracking with an Erroneous Vehicle or Landmark Altitude

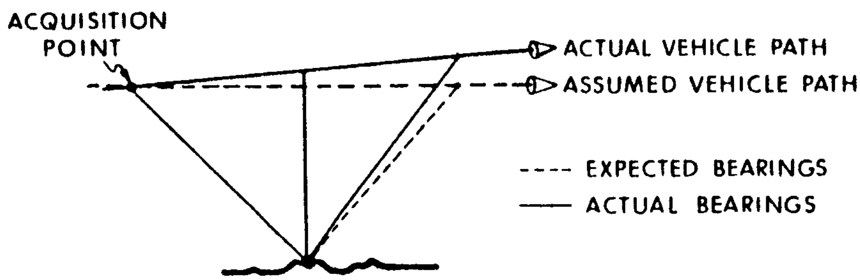


Fig. V-11 Tracking with an Erroneous Vertical Velocity

An initial range ($\theta(T_1)$) error appears identical to a vertical velocity error because at the time, T_1, the computed and actual velocity vectors differ by the angle $\theta(T_1)$ as shown in Fig. V-12. Thus it is impossible to distinguish between a range error at time, T_1, and a vertical velocity error at time, T_1.

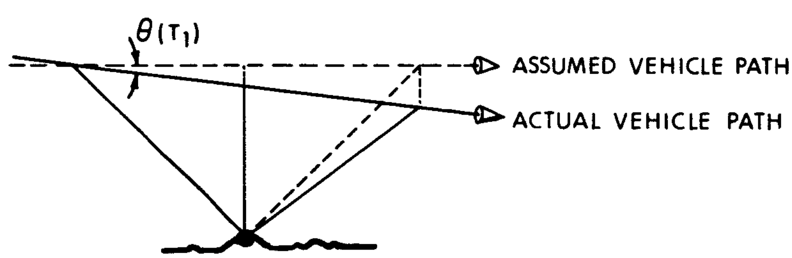


Figure V-12 Tracking with an Erroneous Range Error

The bearing measurement made along the actual trajectory is:

$$\cot \alpha_a = \frac{S_o - r_o \omega_o \tau}{H_o - \theta(T_1) r_o \omega_o \tau} \qquad \text{(V-19)}$$

Since the error terms in Eqs. V-16 to V-19 are generally small, further linear approximations of these equations can be made. Using the approximation,

$$\frac{1}{1+\delta} = 1 - \delta$$

Equation V-16 can be approximated by:

$$\cot \alpha_a = \frac{S_o - r_o \omega_o \tau}{H_o} + \frac{r_o \omega_o t(r(T_1) - \Delta H)}{H_o^2} \qquad \text{(V-20)}$$

(since $S_o = H_o \cot \alpha(0)$)

Equation V-17 can be approximated by:

$$\cot \alpha_a = \frac{S_o - r_o \omega_o \tau}{H_o} - \frac{\omega(T_1) r_o}{H_o} \tau \qquad \text{(V-21)}$$

Equation V-18 can be approximated by:

$$\cot \alpha_a = \frac{S_o - r_o \omega_o \tau}{H_o} - \frac{S_o \dot{r}(T_1)}{H_o^2} \tau + \frac{r_o \omega_o \dot{r}(T_1)}{H_o^2} \tau^2 \qquad \text{(V-22)}$$

Equation V-19 can be approximated by:

$$\cot \alpha_a = \frac{S_o - r_o \omega_o \tau}{H_o} + \frac{S_o r_o \omega_o \theta(T_1)}{H_o^2} \tau - \frac{r_o^2 \omega_o^2 \theta(T_1)}{H_o^2} \tau^2 \qquad \text{(V-23)}$$

The first term on the right hand side of Eqs. V-20 through V-23 is $\cot \alpha_c$. Thus, one can write in general:

$$\cot \alpha_a - \cot \alpha_c = \frac{1}{H_o^2} \left[r_o \omega_o \left[r(T_1) - \Delta H \right] \frac{r_o}{H_o} \omega(T_1) - S_o \left[r(T_1) - r_o \omega_o \theta(T_1) \right] \right] \tau \qquad \text{(V-24)}$$

$$+ \frac{r_o \omega_o}{H_o^2} \left[r(T_1) - r_o \omega_o \theta(T_1) \right] \tau^2$$

When one considers the magnitude of the terms in Eq. V-24, it becomes apparent that the $r_o \omega (T_1)(H_o)$ term can also be neglected. This leaves the relatively simple expression:

$$\cot \alpha_a - \cot \alpha_c = \frac{1}{H_o^2} \left[r_o \omega_o \left[r(T_1) - \Delta H \right] + S_o \left[r_o \omega_o \theta(T_1) - r(T_1) \right] \right] \tau$$

$$+ \frac{r_o \omega_o}{H_o^2} \left[\dot{r}(T_1) - r_o \omega_o \theta(T_1) \right] \tau^2 \qquad \text{(V-25)}$$

To determine the track errors in the vehicle trajectory, consider the projection of Fig. V-7 into a plane normal to the assumed velocity vector. The unknown landmark should also be contained in this plane. This measurement provides no information about the position error ϕ as shown in Fig. V-13.

Fig. V-13 Tracking with Erroneous Track Position

However, the measurement does provide information about the track velocity error ϕ as shown in Fig. V-14.

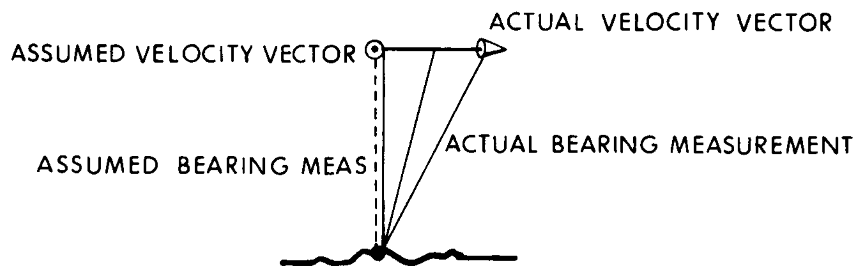


Fig. V-14 Tracking with Erroneous Track Velocity

To determine ϕ, two variables from the in plane component of this measurement have to be known:

The distance between the landmark and the vehicle velocity vector (just the sum - not the individual components)

The sum of the vertical velocity and range error (again only the sum).

These variables, though, are available from the in plane computation (Eq. V-25). The measurement geometry is as shown in Fig. V-15

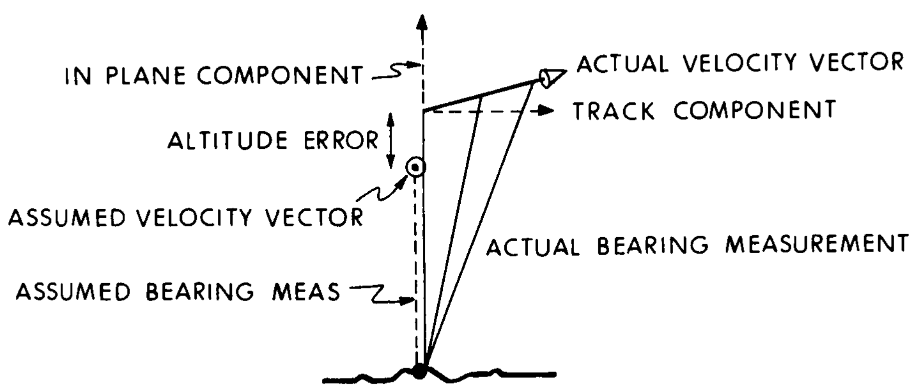


Fig. V-15 Tracking wi h Erroneous Track Velocity Altitude, Error, and Vertical Velocity Error

Consider the angle, β, between the assumed bearing measurement and the actual bearing measurement.

$$\sin \beta = \frac{\phi r_o \tau}{H_o + r(T_1) - \Delta H + (r(T_1) - r_o \omega_o \theta(T_1))\tau} \qquad \text{(V-26)}$$

Depending on the particular orbit, the denominator terms with exception of H_o or possibly $H_o + r(T_1) - \Delta H$ will be negligible. For small values of β $\sin \beta \approx \beta$ and in general:

$$\phi = \frac{\beta(\tau)\left[H_o + r(T_1) - \Delta H\right]}{r_o t} \qquad \text{(V-27)}$$

Equations V-25 and V-27 form the two basic equations which have to be solved to determine the orbital parameters when unknown landmarks are used.

Thus, as contrasted with the known landmark measurement, the unknown landmark measurement provides information about the velocity vector of the vehicle. Since each measurement to an unknown landmark provides only one equation for the determination of the four in plane orbital parameter unknowns, it is necessary to track four unknown landmarks to determine the initial condition errors. The first two of these measurements completely determine the track errors. The remaining two track computations - if carried out - serve only for redundancy.

The requirement for four landmarks instead of three appears here again, similarly to the requirements for four bearing measurements to a known landmark.

NAVIGATION USING STAR HORIZON MEASUREMENTS

Let us assume for the moment that it is possible to define an earth horizon by some means. A possible method is discussed in a later section of this paper, where a horizon is defined at approximately 30 KM above the earth's surface. The astronaut then maneuvers the vehicle so that the plane formed by the two lines of sight (star line of sight and horizon line of sight) also contains the center of the earth and measures the angle between the star and horizon.

This places the vehicle on a surface which touches the earth at the earth's artificial horizon as shown in Fig. V-16.

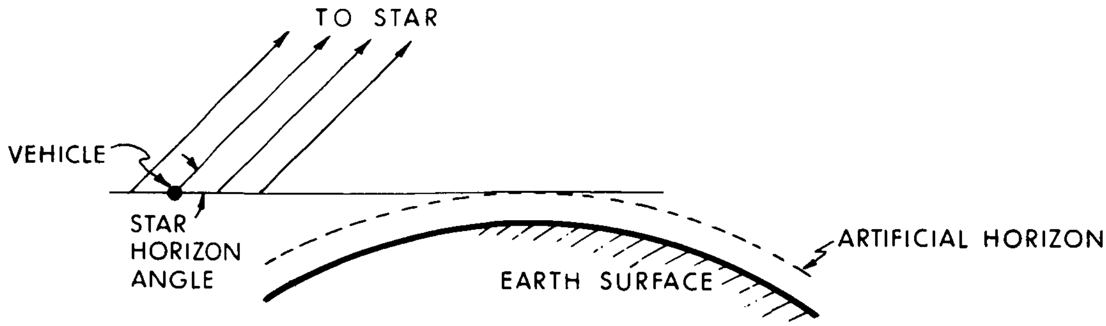


Fig. V-16 Star Horizon Measurement

Since the artificial horizon is at an altitude of approximately 30 kilometers, the angle between the local horizontal and the line to the horizon is about 0.2 radians.

Again it is possible to make a flat earth approximation for this measurement. Figure V-17 illustrates the star horizon measurement when there is no error in the vehicle position. The angle, γ, is the angle between the assumed orbital plane, and a plane containing the vehicle, earth center, and a vector from the vehicle to the star. Thus, when γ is zero, the measurement is insensitive to ϕ errors: when γ is 90^{o}, the measurement is insensitive to θ errors. The angle, $90^{o} - \alpha$ is the angle between the velocity vector at time, T_1, and the normal to the plane established by this measurement.

If the measurement is taken at a time, t, when the star horizon angle is equal to the expected star horizon angle at time, T_1, the vehicle is located on a plane (taking the position of the vehicle at time, T_1, as the origin) given by Eq. V-28. One equation relating the errors

$$r(t)\cos\alpha + \gamma(t)\sin\alpha\ \cos\gamma + \phi(t)\sin\alpha\ \sin\gamma = 0 \qquad \text{(V-28)}$$

at time, T_1, can be obtained by substitution of Eq. V-13a, b, and c into Eq. V-28. Thus, only one equation between the orbital parameter errors can be obtained from this type of measurement since the measurement involves only one angle and time.

The angle, α, depends only on the orbital altitude. The angle, γ, depends on star direction. It can be chosen to favor either the range error (by making γ small) or the track error (by choosing γ close to 90^{o}). The star horizon angle only determines the time, T_1, when the vehicle should pass through the plane given by Eq. V-28.

The measurements described in this chapter require different configurations of equipment. They also have certain constraints imposed on their usage as described below:

Known Landmark Bearing Measurement:

Advantages: Good error sensitivity
Moderate equipment accuracy requirement with man made beacons usable on both the light and dark side of the earth

Disadvantages: Landmarks have to be chosen for ease of recognition and unambiguity. Cloud cover can make landmarks not useable. Large map requirements. Surveying of some landmarks required. Barely usable on the moon due to poor maps. Requires a reference (e. g. inertial system)

Unknown Landmark Bearing Measurement:

Advantages: Reasonable error sensitivity. Usable anywhere on sunlit side and where ever there are lights on dark side of earth. No maps required. Usable even on the back side of the moon.

Disadvantages: Relatively high equipment accuracy required.
Requires a reference (e. g. inertial system)

Star Horizon Measurement:

Advantages: Reasonable error sensitivity
No reference required

Disadvantages: High equipment accuracy required useable only against sunlit earth or moon, requires automatic star tracker and photometer

Star Occultation Measurement:

Advantages: Good error sensitivity
No equipment required

Disadvantages: Usable on moon only

Since each measurement provides only one equation in the six unknown orbital errors, it is necessary to take at least six star horizon measurements to determine the orbital parameters.

STAR OCCULTATION MEASUREMENT

As a vehicle moves in an orbit around the earth or moon, stars will rise and set. Whenever a star sets, the disk of the moon or earth occults this star and the vehicle is at that moment crossing a cylinder with its axis in the direction of the

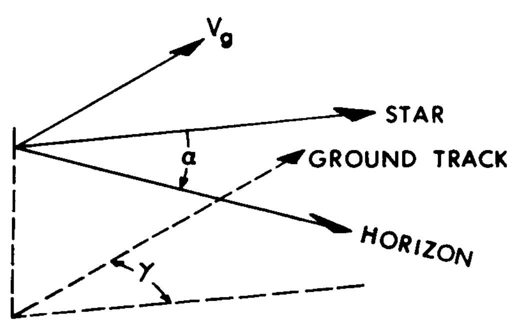


Fig. V-17 Star Horizon Measurement Geometry

star. The diameter of this cylinder is equal to the diameter of the earth or moon.

The measurement is very useful when the vehicle is in a lunar orbit. However, in earth orbit - due to the atmosphere - it is very difficult to determine when a star is occulted since attenuation of the starlight in the atmosphere will take place due to differential refraction, and due to attenuation.

This measurement can be considered as a special case of the star horizon measurement with a zero star horizon angle.

The position of the vehicle is given by the plane defined by Eq. V-28 and the error at time, T_1, by substitution of Eq. 13a, b, and c into Eq. V-28.

Again, six star occultation measurements have to be performed to completely determine the orbital parameters of the vehicle.

CHAPTER V-2

MIDCOURSE NAVIGATION

After injection into translunar or transearth orbit, it is no longer possible to use the simplifying assumptions of Eq. V-4 and V-5. Thus, one generally has to resort to computer solutions to solve the navigational problem. Navigation from ground based stations is again possible and has been used quite successfully. However, in this part only on-board measurements will be considered.

POSSIBLE MEASUREMENTS

Again in midcourse, as in the orbital case, use is made of optical measurements. However, the accuracy requirements in midcourse are so much greater than the accuracy requirements in orbit that bearing measurements with the inertial system used as a reference are no longer possible. As can be seen from Fig. V-4 and V-5 a measurement uncertainty of one milliradian produces position uncertainties of .15 to .2 km in earth orbit. A measurement with similar accuracy, however, is not too useful at a distance of 100,000 km.

At distances from the earth, which are comparable to the lunar distance, it is possible to use features on the earth or on the moon, or to use planets within our celestial system. Among these choices, the earth and moon provide the best accuracy because of their proximity to the trajectory.

Another factor is of great importance in midcourse. The bearing angles to landmarks or to the horizon change very slowly. In orbit these rates are (see again Fig. V-4 or V-7) in the order of degrees/second. In midcourse the rates reach low values of arc seconds per second. Thus, considerably more time can be taken by the operator to complete each measurement.

Measurements which have been considered are:

Star Landmark Angle Measurements

Star Horizon Angle Measurements

Star Occultation Measurements

The first two of these measurements are sextant measurements. Only one angle has to be read with great accuracy to complete the navigational measurement. The third measurement does not require any instrumentation when a star is occulted by the lunar disc. However, occultation measurements against the earth's disc require relatively precise knowledge of starlight attenuation through the atmosphere to determine a point of occultation. A photometer to make this measurement has not been included in the optical unit used in Project Apollo.

STAR LANDMARK MEASUREMENT

One precision angular measurement between a star and a landmark places a vehicle on a cone as shown in Fig. V-18. The apex of the cone is located at the landmark. The direction of the center line of the cone is parallel to the direction to the star. The cone half angle is equal to the measured star landmark angle.

Consider the plane formed by the vector from the vehicle to the landmark and the vehicle to the star. If another navigational measurement is made using the same landmark, and another star within this plane, then another cone of position is established as shown in Fig. V-19. The two conical surfaces touch at the line containing the vehicle and landmark. Locally, in the vicinity of the vehicle (assuming a distance to the landmark greater than 20,000 km and a star landmark angle greater than five degrees) this conical surface can be approximated by a plane; the only difference is that for a larger star landmark angle the flat plane approximation is better. No other information is provided by this measurement.

Thus, to obtain information for a second degree of freedom it is necessary to choose another star so that the plane containing the vectors from the vehicle to the second star and vehicle to the first star is approximately orthogonal to the plane shown in Fig. V-18.

The resultant locus of position of the vehicle is a line as shown in Fig. V-20. Any other measurement using the chosen landmark only provides redundant information.

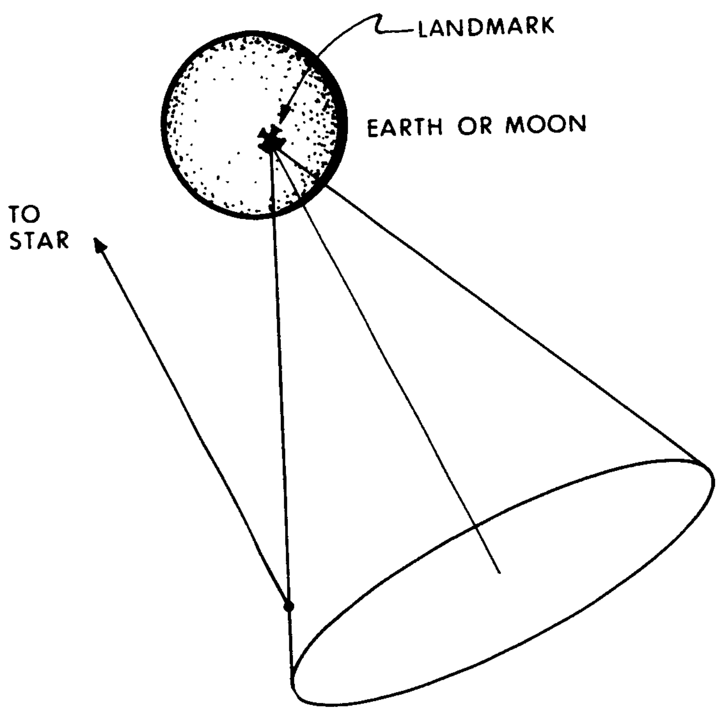


Fig. V-18 Star Landmark Measurement Geometry

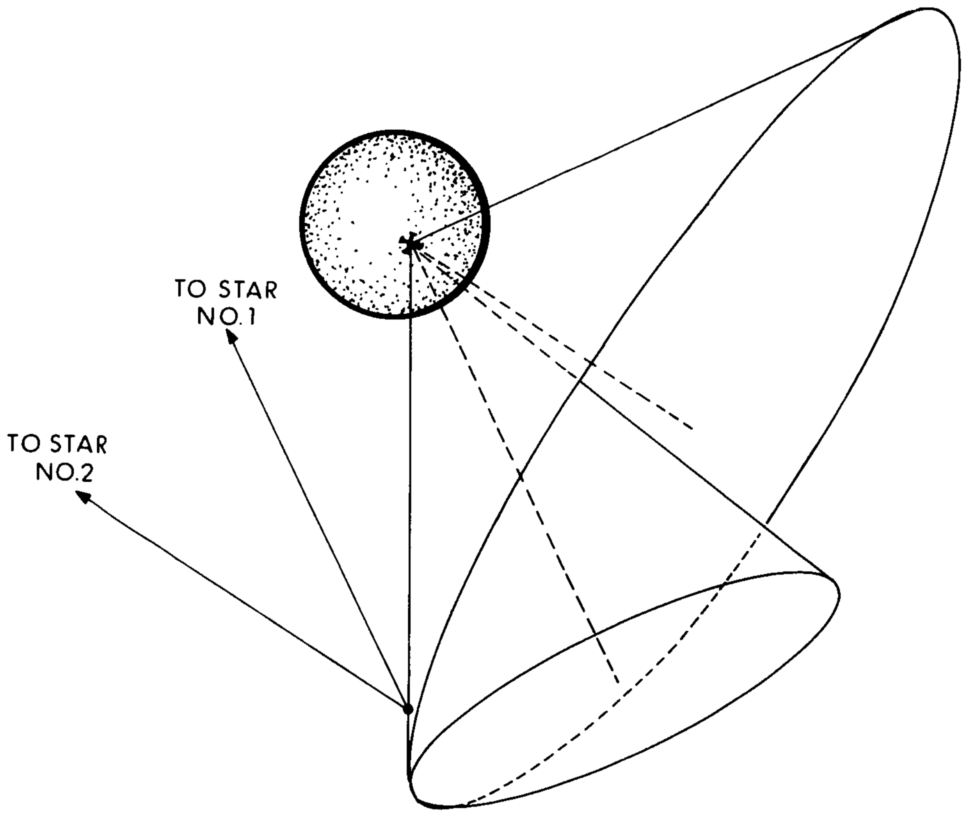


Figure V-19

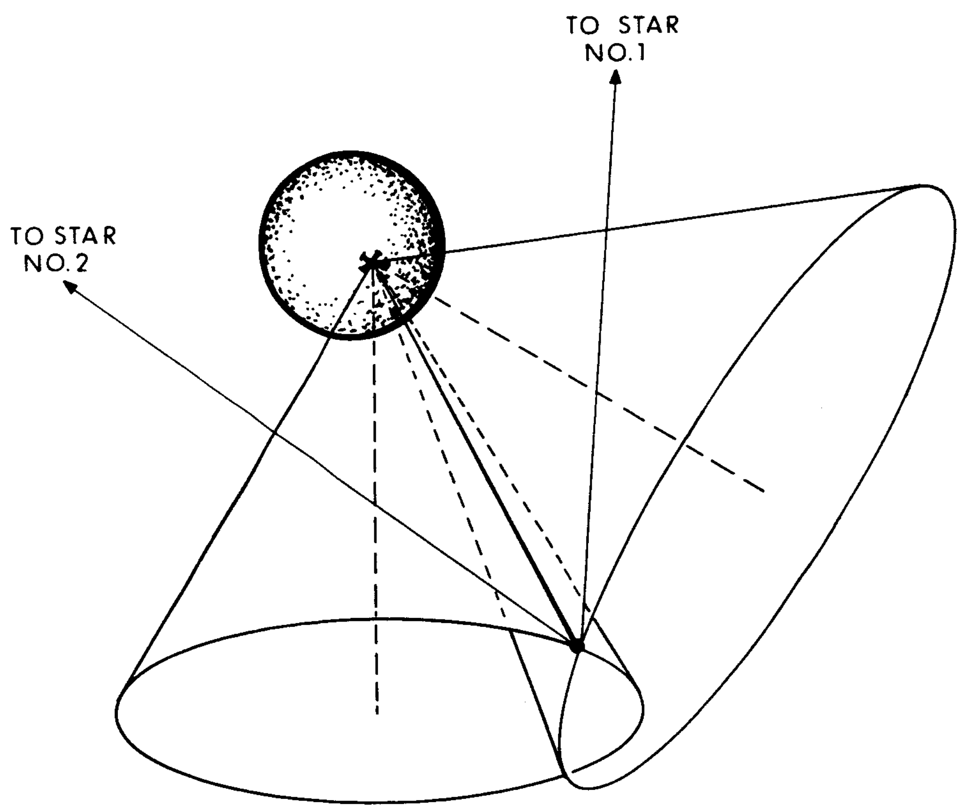


Fig. V-20 Measurement Geometry Using One Landmark and Two Stars

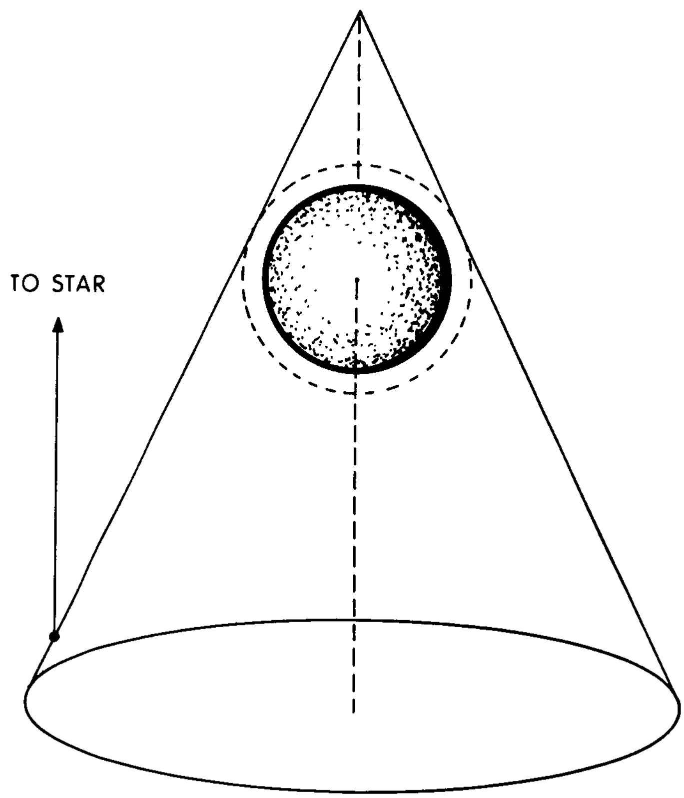


Fig. V-21 Star Horizon Measurement Geometry

The uncertainty of the vehicle along this line can only be reduced if another landmark is used.

Let us assume that a landmark can be used as long as the angle between the line-of-sight and the local vertical at the landmark is less than 45^{o}, then at a 20,000 km distance from the earth, the error along the line established in Fig. V-20 is about .4 km per arc second of error in the measurement. At 100,000 km from the earth star landmark measurements to two earth landmarks provide essentially no additional information when compared to measurements made with only one landmark.

Conversely, it can be said that the choice of earth (or moon) landmarks does not matter for star landmark measurement in midcourse. Any landmark, as long as its position is known and clouds are not present, can be used as well as any other landmark.

STAR HORIZON MEASUREMENT

Let us assume again that a suitable earth horizon as seen from space can be defined. The lunar horizon can be defined by the lunar disc. The astronaut maneuvers the vehicle so that the horizon scan takes place in a plane normal to the horizon. When the measurement between star and horizon is made, the vehicle is located on a cone as shown in Fig. V-21. The center line of the cone is parallel to the starlight. The half angle of the cone is equal to the star horizon measurement angle. The location of the apex of the cone can be obtained from Fig. V-22, where R is the radius of the earth or moon respectively.

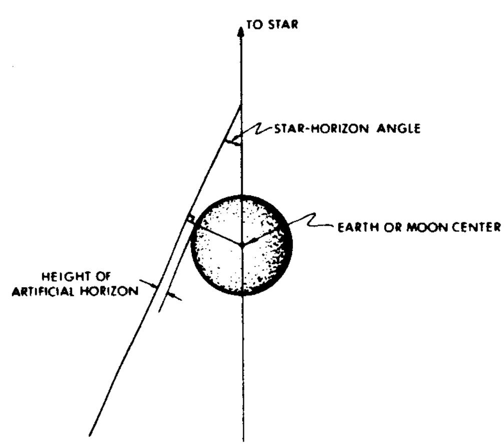


Fig. V-22 Star Horizon Cone Apex Location

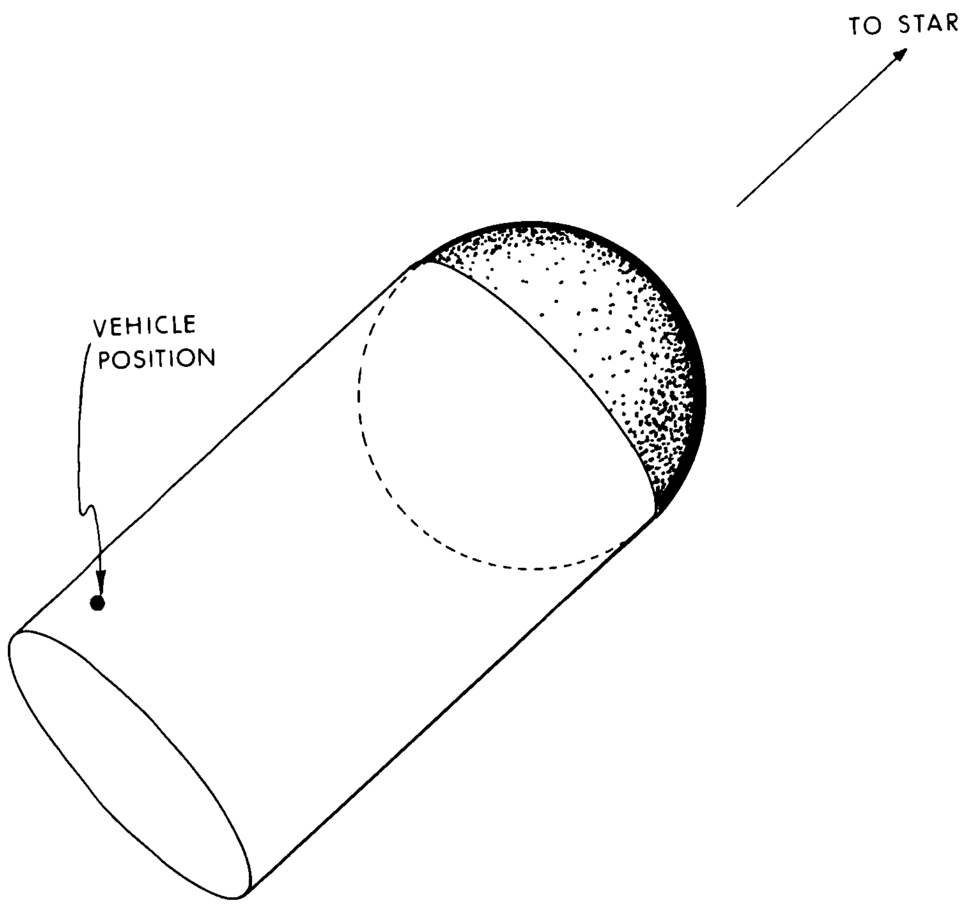


Fig. V-23 Star Occultation Measurement Geometry

Thus, as can be seen, the geometry of a star horizon measurement and a star landmark measurement is almost identical. The only difference between the two measurements is the location of the apex of the cone which defines the vehicle position.

STAR OCCULTATION MEASUREMENT

When a star is occulted by the earth's or moon's disc, the vehicle is located on a cylinder as shown in Fig. V-23. The center of the cylinder passes through the center of the earth. The diameter of the cylinder is equal to the earth's or moon's disc.

Since star occultations depend on the trajectory of the vehicle, it is not possible to choose the measurement geometry to any extent. However, the measurement does have the advantage that no equipment is required to perform it. The only requirement is timing of the star occultation.

ERROR SENSITIVITIES NEAR THE EARTH AND MOON

When the distance between the earth or moon and the vehicle is less than about 20,000 km, the angle between the two horizons is sufficiently large so that an advantage can be gained from star horizon measurements using first one horizon and then the other horizon. Since errors at injection are costliest (from a correction fuel requirement standpoint) when they are made in the plane of the trajectory, it is important to determine their magnitudes early during the midcourse phase.

Let us assume that at the end of injection the vehicle has an error in the magnitude of the velocity vector. The direction of the velocity vector and position of the vehicle are assumed to be correct. Figure V-24 shows in exaggerated fashion how the path of the venicle deviates from the nominal (or expected) path. The differences between the actual angles and expected angles are shown in Figs. V-25 to V-29. Let us assume that the vehicle injected with an altitude error of one km. The difference between the expected and actual star horizon angle as defined in Fig. V-24 is shown in Fig. V-25. The error sensitivity for readings taken against either horizon is almost identical when the vehicle reaches a distance of about 30,000 km from earth center (100 minutes after injection). Similar conclusions can be reached from Fig. V-26 (error sensitivity due to wrong injection velocity) and Fig. V-27 (error sensitivity due to an error in the direction of the velocity vector at burnout).

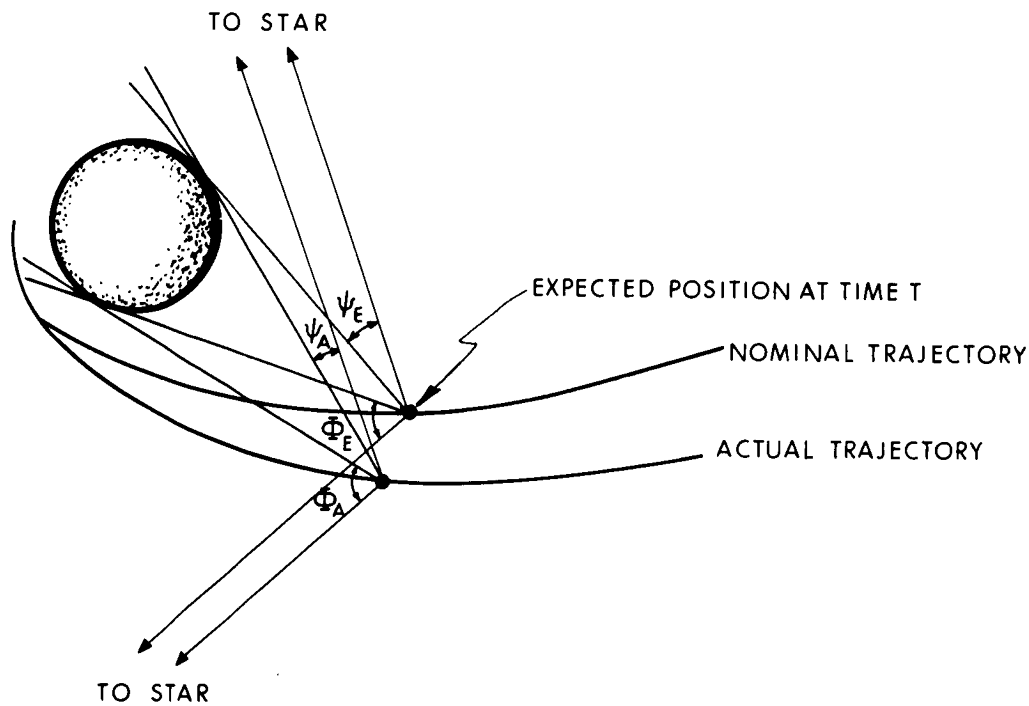


Fig. V-24 Measurements to Both Horizons

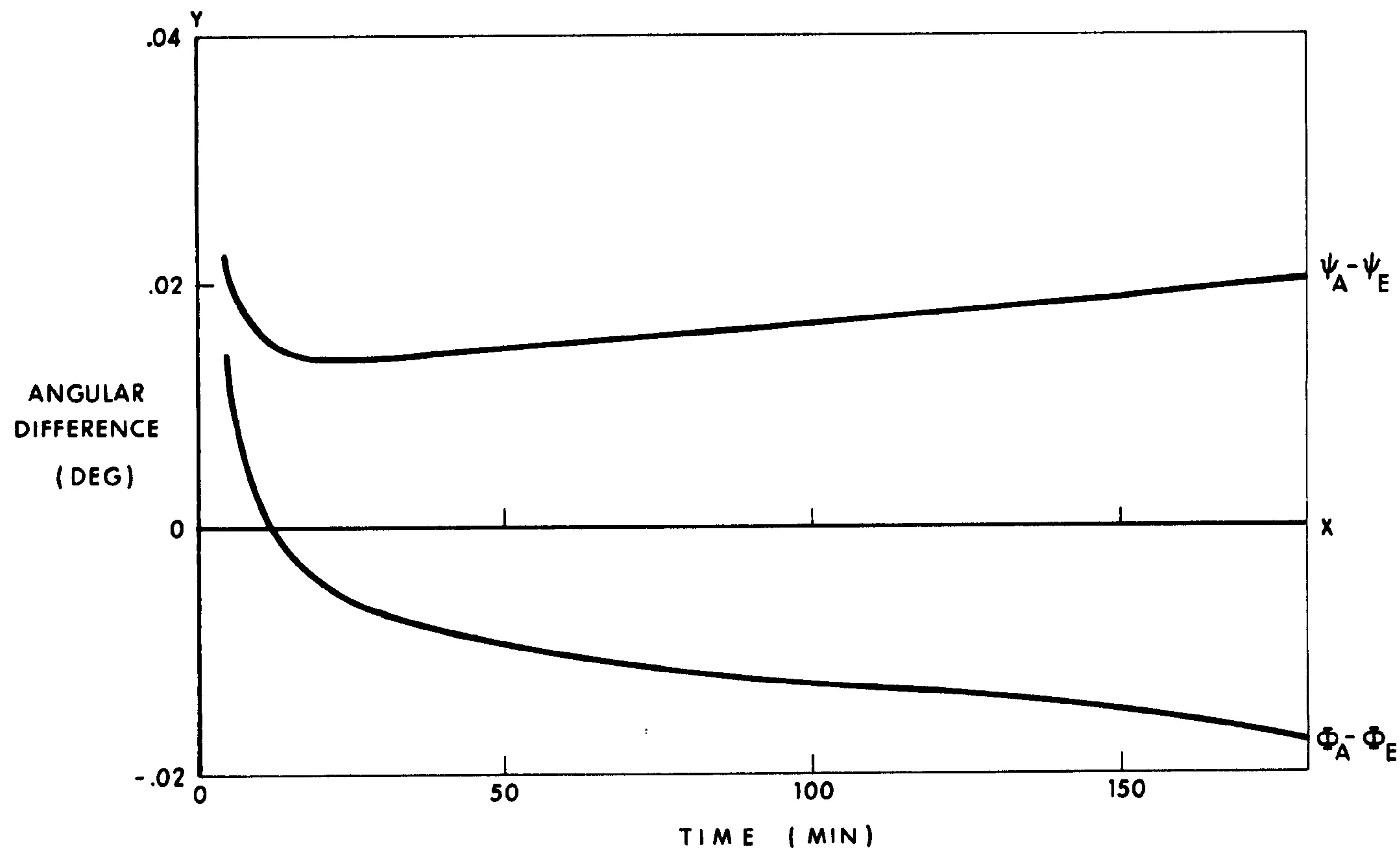


Fig. V-25 Change in Angle Measurement due to Initial Altitude Error

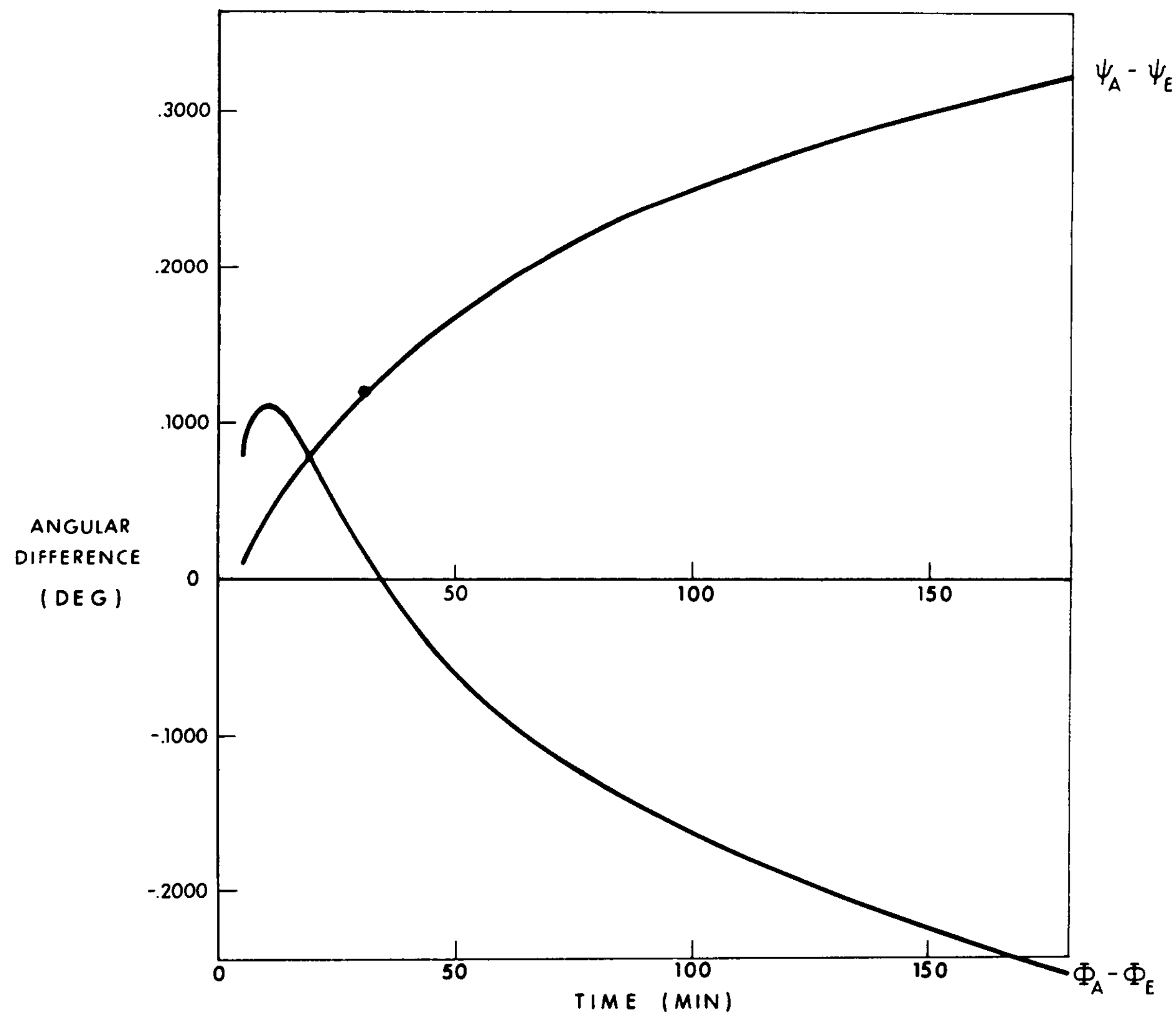


Fig. V-26 Change in Angle Measurement due to Initial Velocity Error of 1 Km/H2

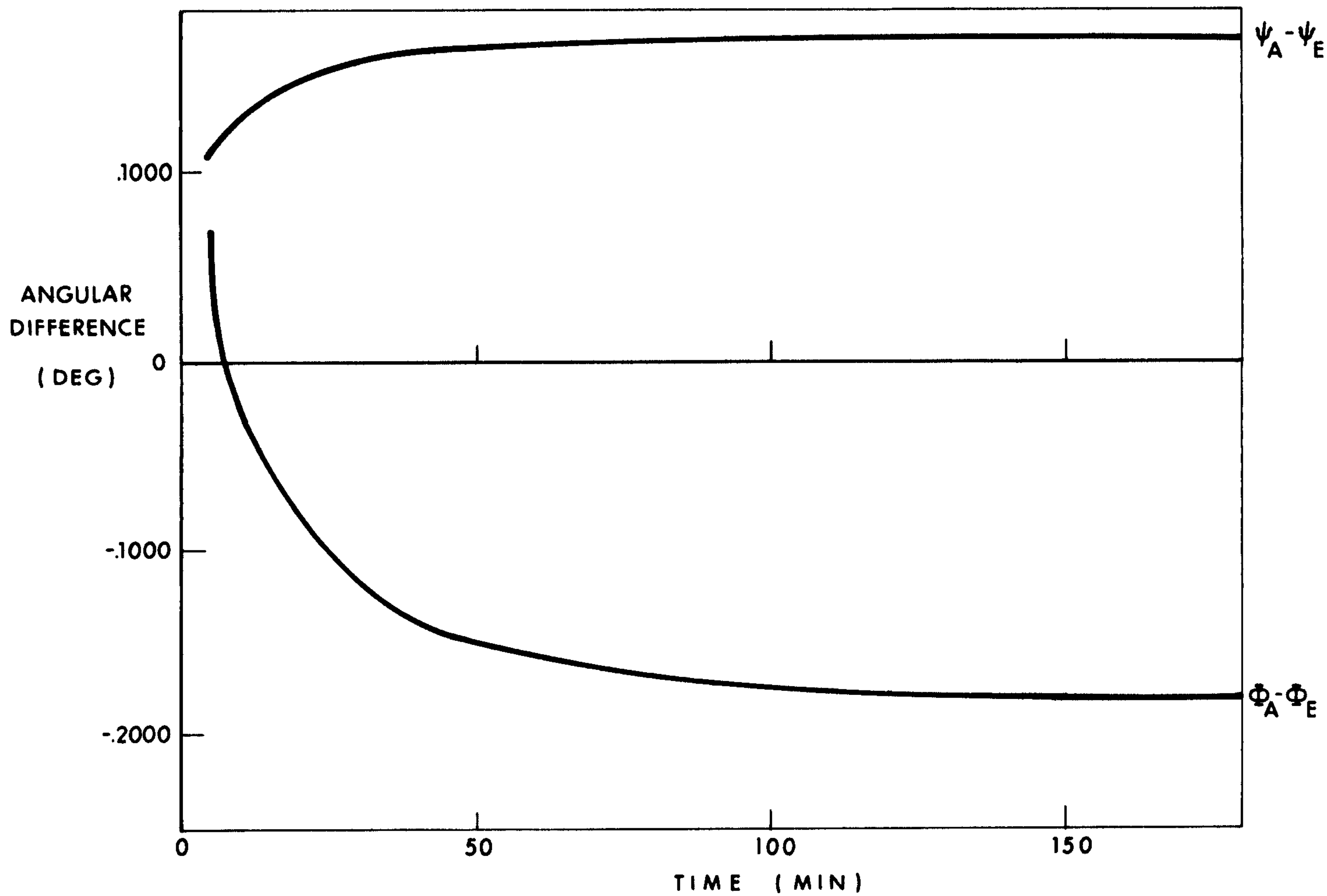


Fig. V-27 Change in Angle Measurement due to Initial Velocity Direction Error of 1 Milliradian

103

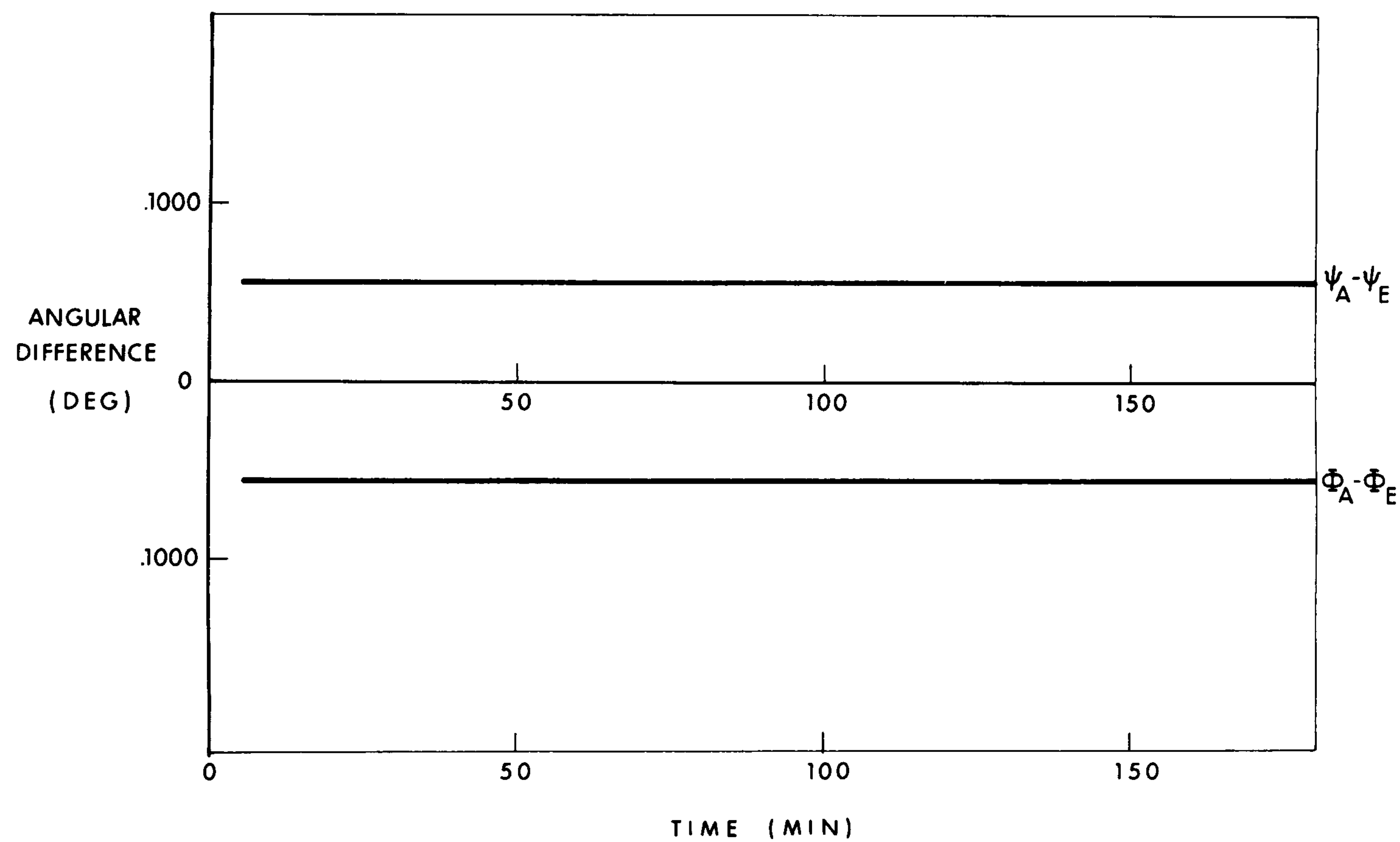


Fig V-28 Change in Angle Measurement due to Initial Range Error of 1 Milliradian

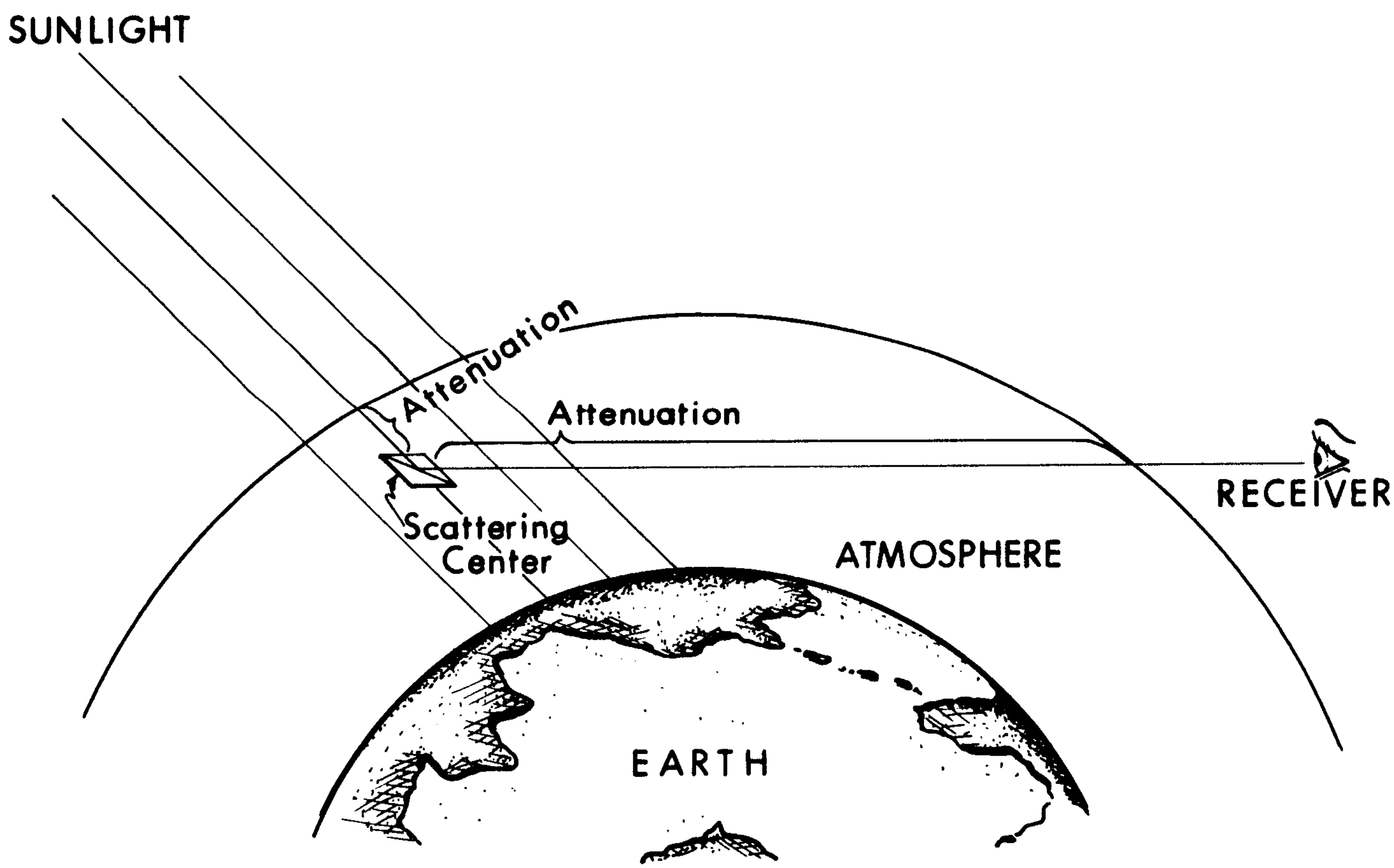


Fig. V-29 Location of Horizon Phenomena

A range error (θ in Chapter I) simply rotates the transfer ellipse by the angular range error θ. Thus, the error sensitivity is constant during the first part of the mission.

If navigational sightings cannot be taken during the first one or one-half hours, either due to crew tasks or constraints imposed by radiation belts, the astronaut or computer has to chose only between an earth feature or a lunar feature. Choosing between specific landmarks or horizon on either the earth or the moon does not provide any major improvement in the knowledge of the vehicle position.

AN ARTIFICIAL EARTH HORIZON

Clouds and atmosphere on the earth, makes it generally impossible to see the earth's surface near the horizon; thus it is necessary to define a horizon which is sufficiently high above the clouds to be visible during most of the mission time.

To date, most of the horizon sensing has been accomplished in the infrared region of the spectrum. However, sunlight scattered within the earth's atmosphere also provides a possible horizon. Since the upper atmosphere scatters most of the blue and near ultraviolet light, a horizon model based on scattered sunlight is being used. Figure V-29 illustrates the measurement geometry used. A comparison is made between the maximum intensity and 1/2 of the maximum. The 1/2 max. establishes a horizon located about 30 km above the earth's surface.

CHAPTER V-3

THE APOLLO OPTICAL UNIT

The varied navigational measurements which are made during the Apollo mission are used as a basis for the design of the optical unit shown in Fig. V-30 The unit consists of a sextant, an automatic star tracker and photometer, and a scanning telescope.

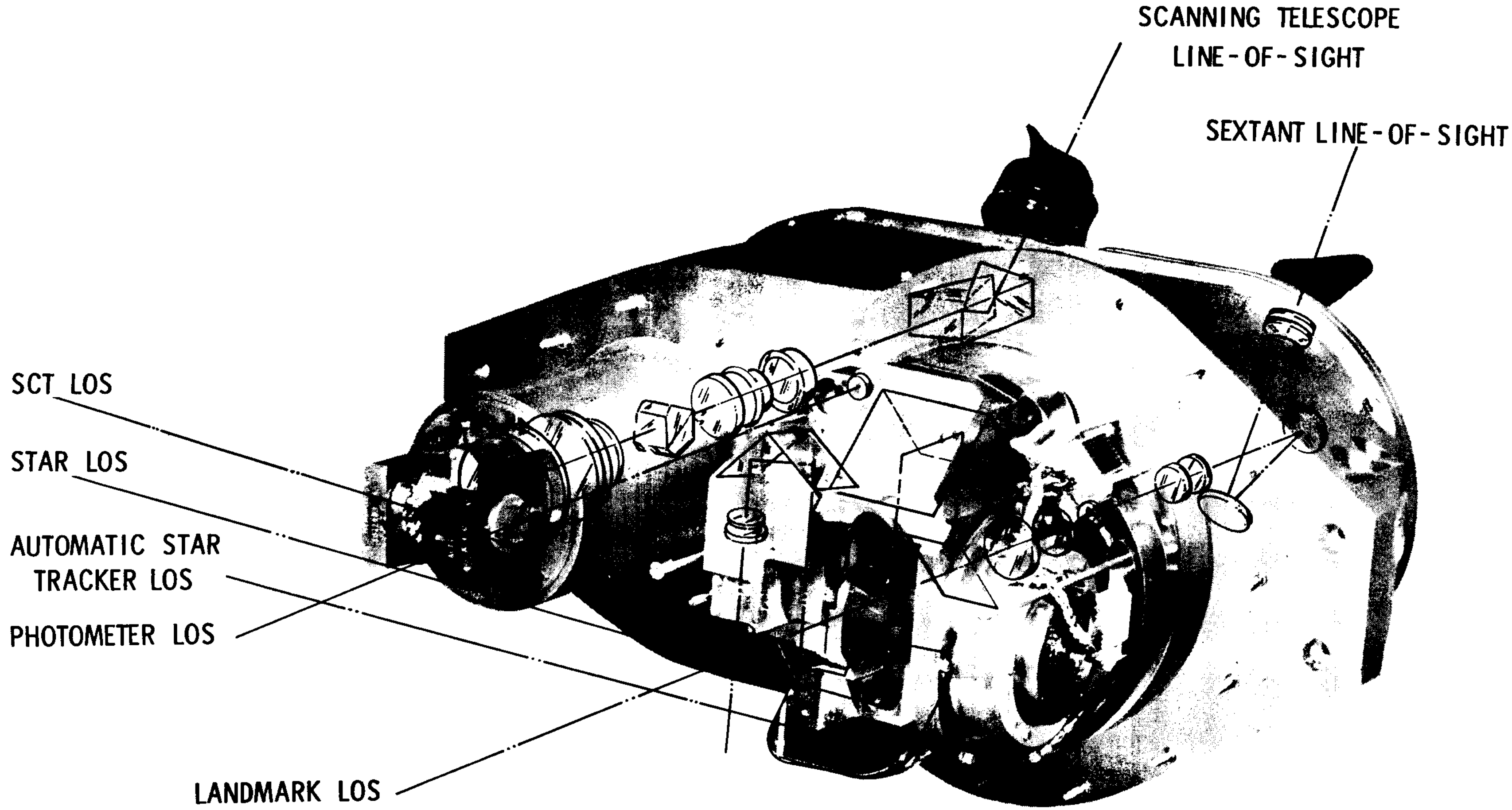
During midcourse when it is necessary to make accurate star landmark and star horizon measurements, the sextant is used. The scanning telescope is used only as an instrument for target acquisition.

Both instruments are used during the orbital portion of the mission. The scanning telescope is used for known landmark bearing measurements. Its large field of view is essential for landmark identification and acquisition. The movable line-of-sight of the sextant is used for unknown landmark tracking. The automatic features are used for star-horizon measurements. Star occultation measurements need no particular bearing read out since only the time of occultation is required. Either instrument or a window can be used for this particular measurement.

SPACE SEXTANT

The sextant consists of a 28 power, 1.8 degree field of view telescope located within the optical base (see Fig. V-30). Before light enters the telescope, it passes through the sextant head, where the two lines-of-sight of the instrument are combined.

The landmark line-of-sight passes through the beamsplitter and from there directly into the telescope. To move this line-of-sight, it is necessary to move the vehicle. The beamsplitter passes about 20 percent of the light to the telescope. Also contained within the beamsplitter is a filter, which attenuates most of the blue and green light entering the landmark line-of-sight. This reduces the blue haze which is generally visible on the earth from high altitudes without appreciably degrading the characteristics of lunar landmarks.



V-30 Apollo Optical Unit

The star line-of-sight is directed to the telescope by a movable mirror, two fixed mirrors, and the beamsplitter. About 80 percent of the starlight passes through the beamsplitter into the telescope.

The movable mirror is positioned by a conventional A.C. servo system. The read out of the mirror position (this is the most accurate angle read out within the guidance system) depends on a 64 speed (128 pole) resolver. The angles of this resolver are read by a coupling data unit as described in Part IV.

Angles ranging from zero to 50 degrees between the two lines-of-sight can be read with this instrument. The mirror can be moved further to 90 degrees, where the image of the reticle is reflected back upon itself. The astronaut can use this point as a check of the alignment stability of the instrument. Viewing of a bright star or planet with the line-of-sight angle set to zero degrees provides another point where the sextant angular read out can be easily checked. These self-checking features have been incorporated to make sure that a check of the instrument accuracy can be made before and after each midcourse navigational sighting.

The head assembly and telescope rotate within the optical base. This motion provides the second degree of freedom of motion for the star line-of-sight. Motion in this direction is limited to $\pm$ 270 degrees from a predetermined zero. The power to the components located on the sextant head is carried through a flex cable. Limitation of the freedom of motion of the sextant thus makes use of slip rings not required.

STAR TRACKER-PHOTOMETER

The star tracker is included within the optical unit to enable the astronaut to make measurements between the earth's horizon and a star. Both instruments including their detectors and preamplifiers are completely located on the sextant head.

The photometer measures the intensity of the near ultraviolet sunlight scattered by the earth's atmosphere. Light collected by the photometer aperture is filtered and modulated by a vibrating reed before it reaches the photodetector. The signal is further amplified and the level of the signal is detected with circuitry within the power and servo assembly of the guidance system.

The star tracker is a narrow, field-of-view (1/2 deg. X 1/2 deg.) instrument.

Star light is directed into the aperture of the telescope by the star line-of-sight movable mirror and two redirection mirrors. Crossed tuning forks are used to provide information about the star position within the tracker' s field of view. The electronics required to process the star tracker signal are also located in the power and servo assembly.

The two degrees of freedom required to position the star tracker are identical to the degrees of freedom provided for the star line-of-sight. Thus, when a star is tracked the astronaut can compare the star position in the sextant field to the center of the reticle. This provides a visual check on the operation of the star tracker.

SCANNING TELESCOPE

The scanning telescope is a single line-of-sight, unit power, 60 degree field of view instrument used for making known landmark bearing measurements in earth and lunar orbit and also for general use, such as acquisition of landmarks and stars for sextant sightings.

The complete telescope assembly rotates within the optical base, similar to the sextant. The two shafts are continuously tied together with a conventional A. C. servo system.

The light enters the telescope through a double dove prism. This prism can be positioned so that the telescope line-of-sight is parallel to the landmark line-of-sight (for midcourse landmark acquisition) or parallel to the star line-of-sight for star acquisition. A third position, offset from the landmark line-of-sight by 25 degrees, is also provided. In this position the scanning telescope covers the complete sextant field of view. This position is used when the astronaut tries to hold one image in one of the sextant lines-of-sight while searching for the second image.

To provide backup in case of an electrical failure, the scanning telescope also contains two mechanical counters connected to the movable parts of the telescope. These counters can be read and the angles manually entered into the computer in case of a malfunction within the angle encoding loop.

Two knobs which can be used for manual positioning of the telescope line-of-sight are also provided. These provide a backup for the A. C. servo components which are normally used to position the sextant and scanning telescope.

BIBLIOGRAPHY

Chandrasekhar, S., Radiactive Transfer, Dover Publications 1960

Peterson, R., Final Report on Special Techniques for Space Navigation, Control Data Corporation, July 1964

Seward, H.H., The Blue-White Boundary Horizon Sensor, MIT Instrumentation Laboratory Report E-1206, September 1962

Wolff, M., The Profile of an Exponential Atmosphere Viewed from Outer Space and Consequences for Space Navigation, MIT Instrumentation Laboratory Report E-1634, September 1964

Zucherbraun, J.,"High-Reliability Scanners for Stellar Navigation", Electronics, May 11, 1962

V-46 BLANK

N66 28762

PART VI

GUIDANCE COMPUTER DESIGN

by

Dr. Albert Hopkins

DR. ALBERT L. HOPKINS, JR.

Assistant Director, Instrumentation Laboratory
Massachusetts Institute of Technology

Dr. Albert L. Hopkins, Jr., Assistant Director of Instrumentation Laboratory, Massachusetts Institute of Technology, is a specialist in the development of computers for inertial guidance and navigation systems.

In his present assignment in the Laboratory's Digital Development Group, Dr. Hopkins is responsible for design of memory circuits and logic employed in the guidance system computer for the Project Apollo spacecraft.

Dr. Hopkins was born in Chicago, Ill., May 6, 1931, and was graduated from Exeter Academy, Exeter, N.H., in 1949. He attended Harvard University where he received the A.B. degree in engineering sciences in 1953, the A.M. degree in applied mathematics in 1954 and the Ph.D. degree in applied mathematics in 1957. He was Instructor in Control Systems Engineering in the Harvard Division of Engineering and Applied Physics from 1957-60 under Professor Howard Aiken.

Dr. Hopkins joined Instrumentation's Digital Development Group in 1960.

Part VI

GUIDANCE COMPUTER DESIGN

INTRODUCTION

Some 22 years have elapsed since the first digital computer was completed. For the past several years, digital guidance computers have flown in airplanes, missiles, and rockets. Some of these vehicles are scarcely bigger than an early computer, whose performance is surpassed by the guidance computers they carry.

The first general-purpose computer and the first high speed electronic number processor were each put into use in about 1943. Numbers were stored in relay controlled counter wheels and in vacuum tube ring counters. The next few years brought the high speed general-purpose computer, the acoustic delay line, and the electrostatic storage tube. In the early 1950's higher densities were achieved in logic by the use of semiconductor diodes, and in memory by the introduction of magnetic cores, drums, and tapes. The transistor was developed in 1948, and was employed in some experimental small computers intended as prototype airborne computers in the early 1950's. Transistors began to be employed in large-scale computers in the late 1950's, beginning with computers for military applications. Only now are semiconductor components beginning to match the enormous density which was achieved in magnetic memories ten years ago. In today's guidance computers, we are realizing an overall density thousands of times greater than in computers of fifteen years ago. Part of this difference is due to advances in mechanical design which have been made purely because of the extreme importance of weight and volume in airborne applications. In cases where size is not an important factor, densities are lower by an order of magnitude.

Performance has been increased over the years by advances in logical design as well as in component size and speed. Early contributions of the number bus, binary arithmetic, and common storage have been followed by such improvements as methods of fast arithmetic, indexing and other address modification schemes, multiprogramming, and program interrupt.

Progress in mathematical areas has been a significant factor in our ability to employ digital computers in airborne guidance at this stage of their development. Computer programming developments have given us automatic programming and internal "software" routines such as executive control and interpretive programs. Recent efforts have produced numerical methods of celestial mechanics which do not overtax the limited resources of today's guidance computers; and sampled-data theory has given rise to methods of stable control of unstable vehicles using digital techniques. Vast amounts of programming and analysis are needed before today's missions become tractable for existing computer performance. It is the purpose of the remainder of this essay to discuss the present state of the art in logical design, hardware, and software, particularly as it is applied in the Apollo Guidance Computer.

The reasons for having an airborne computer are so numerous that the computer engineer is apt to take the parochial viewpoint that the guidance computer is the principal part of the guidance system, and that the other parts, such as inertial, optical, radar, and radio elements are ancillary units for sensing and communication. This picture of a guidance system is somewhat distorted, but not entirely so. In 1962, J. F. Shea, then Deputy Director of Manned Space Flight at NASA, and presently Apollo Project Manager, said, "Although engineers in each discipline tend to regard their particular developments as the most critical, once the propulsion capability has been provided the key to reliable execution of a wide range of complex, long-duration missions is the computational capacity provided aboard the spacecraft."[1]

The Apollo Guidance Computer will be incorporated into the Guidance and Navigation Systems in the Command and Lunar Excursion Modules. Its functions are to aid in operating the inertial and optical subsystems, to provide steering signals where human reaction is too slow, to perform spacecraft attitude control with minimum fuel expenditure, to maintain timing references, to communicate with the Astronauts via display lights and keyboard, to communicate with ground tracking stations via digital data links, and to perform the calculations necessary to deduce position and velocity relative to the earth and the moon from the input data available during all flight phases from boost through lunar landing and rendezvous to final entry and landing.

Chapter VI-1

CHARACTERISTICS OF GUIDANCE COMPUTERS

GENERAL

Guidance computers are designed to meet a number of severe constraints. The extent to which the constraints are met depends upon the ingenuity of the designers; but to judge from comparisons among existing computers, it depends far more strongly on the limitations of available hardware. The more significant differences in performance among computers can be traced to the degree to which their designers are willing to commit themselves to advanced technology, and processes previously untried. Today, the complex microcircuit and the multi-layer etched board are areas in which some computer makers are attempting to attain significant advantages, while others wait and watch to see the inevitable problems arise and become solved before venturing.

Requirements for guidance computers are extreme reliability, low weight and power consumption, high performance in terms of mathematical answers per second and inputs and outputs serviced, and flexibility to grow with the scope of the mission. Although immense achievements have been made within the last few years, the present generation of guidance computers rapidly becomes obsolete; for missions already in the planning stages call for much greater achievements in design, programming, production, test, and repair than have so far been realized.

Some feeling for the range of guidance computer characteristics may be obtained from the tabulation in Table VI-1 of published data [1, 2, 3] on computers designed within the past few years. Comparisons are often misleading, especially if one is trying to prove superiority of one computer over another. It is reasonable and often necessary to choose among computers for a specific application, but it is not easy, for subtle differences can be of great importance. It is less difficult and more valid to draw conclusions about the similarities of various computers from a chart of comparative characteristics. Size and weight data have been omitted here because they tend to be particularly misleading in the absence of knowledge of the particular input and output configuration of the computers. This, in turn, is hard to present because of its detailed nature in some cases and its scanty description in others. Sizes range from 0.2 to 2 cubic feet (0.005 to 0.05 cubic meters), and densities are close to that of water.

	Serial or Parallel	Negative Numbers	Word Length Bits	Number of Operations	Addition Time	Multi-plication Time	Power Consumption	Bits of Memory
Univac type 1824	Parallel	2's complement	24	41	8 μsec	64 μsec	110 watts	2×10^5
AC Spark Plug "Magic"	Serial	2's complement	24	16 (?)	70	258	90	1×10^5
Burroughs D-210	Parallel	2's complement	24	32	30	570	1-100	7×10^4
Arma Micro Computer	Serial	2's complement	22	19	27	135	50	5×10^4
IBM Saturn V Computer	Serial	2's complement	28	18	82	328	131	5×10^5
Autonetics D26C	Parallel	2's complement	30	100	6	18	192	3×10^5
MIT/IL Apollo Computer	Parallel	1's complement	16	34	24	48	90	6×10^5

Table VI-1 Selected Characteristics of Several Guidance Computers

LOGICAL DESIGN[4, 5]

Word Length - It is desirable to minimize word length in a guidance computer. Memory sense amplifiers, being high-gain class A amplifiers, are considerably harder to operate with wide margins (of temperature, voltages, input signal) than, for example, circuits made of NOR gates. Memory digit drivers are also critical circuits whose number is equal to the number of bits in a word. Similarly, the time required for carry propagation in a parallel adder or for circulation in a serial machine is proportional to word length, and, moreover, the very size of a computer is dependent on word length.

Factors which discourage the minimization of word length are the numbers of bits required for data words, input and output variables, and instruction words. These numbers are functions of mission requirements and details of logical design. Most guidance computers have word lengths of around 24 bits. The Apollo Guidance Computer is unique among those listed in having 16 bits (of which one is a parity check bit). As explained later, the difference is due largely to a decision to use multiple-precision arithmetic for variables concerned with guidance and navigation. Even the longest word in the list (30 bits) is short by comparison to the large scale computer installations, where size is not of as great concern as are speed and programming ease.

Instruction Repertoire - The implicit requirements for any Von Neumann-type computer demand that facilities exist for:

A. Fetching from memory
B. Storing in memory
C. Negating (complementing)
D. Combining two operands (e.g., addition)
E. Address modification
F. Normal sequencing (specifying the location of the next instruction)
G. Conditional sequence changing.

A single instruction can provide several of these facilities, so that a very limited repertoire is possible,[6] although a large burden is thereby placed on program storage, and speed is limited. For a relatively small additional cost in complexity, a more comfortable repertoire is obtained. An operation set of eight instructions can provide flexibility without sacrificing simplicity. All of the computers listed go beyond this, however, and in general it is done to obtain speed at a cost in hardware. In some instances, the taking of square roots and the conversion of numbers between binary and decimal appear as single instructions. More commonly, the instruction sets contain convenient data handling, branching, and arithmetic operations with from about 2^4 to 2^5 codes.

Speed - It is well known that the overall speed of a computer can be enhanced by its logical design, usually at an equipment cost. This may take the form of having separate adders in a parallel machine for indexing and arithmetic; or it may consist of providing circuits to speed up multiplication by processing several multiplier bits at a time. Alternatively, speed can be obtained by providing single instructions which perform complex jobs such as the two mentioned in the preceding paragraph. Speed is important in guidance computers, and logical complexities are employed in order to gain speed in virtually every guidance computer design; but size and reliability restrictions are of sufficient importance to limit the number and extent of such complexities. In data processing computers, where size is less important and where speed is a competitive issue, logic circuits are employed somewhat into the area of diminishing returns. Guidance computers, as a result, are generally slower than their ground based relatives.

Input and Output - Guidance computers, and control computers in general, differ from data processors most significantly in the area of input and output. Modern data processors generally communicate with peripheral equipment which is complex and sophisticated enough to send and receive data over parallel channels without the computer having to spend much time overseeing the process. In some cases the computer sends data to a remote buffer register upon receipt of an indication that the remote unit is ready. In other cases the remote unit interrogates the computer memory as often as necessary, thus eliminating the buffer. In guidance computers, however, the input and output are not generally exchanged with such sophisticated machines. Owing partly to the non-digital nature of such electromechanical machines as inertial measurement units and rocket steering servos and partly to the strong desire to keep interface circuits and cables as small as possible, the guidance computer spends a substantial part of its time (or equipment) budget on maintaining communication with these units.

Another interesting contrast exists between data processors and guidance computers. The former are designed to spread a work load out over a period of time to achieve a good balance between internal computing and input-output activity. One figure of merit of a data processing installation is the degree to which it can keep its various facilities busy by time-sharing them among various independent users. If a large demand occurs for time on a printer, for example, the results to be printed will be buffered on a magnetic tape to be printed later when the facility is available. In a guidance computer, the central processor is time-shared among numerous jobs, but the allowable delays in reacting to a large demand for input-output service are measured in milliseconds rather than minutes; and the logical design of computers and systems must reflect this fact.

Fault Diagnosis - Another area of interesting contrast between data processors and guidance computers is in the area of self-checking, or fault diagnosis. Since time on a large computer is valued at hundreds of dollars per hour, it is economically necessary to locate and correct faults very rapidly. For this reason, modern computers are equipped with circuits whose function is to make fault location nearly automatic.

Guidance computers cannot afford to carry extra hardware for this purpose. It is important, however, to be able to detect that an error has occurred in flight so that the proper course of action may be taken. This action might be to switch to a back-up computer or other means of control, or possibly it may mean that a missile must be destroyed in order that it not stray far off course. The most common means of fault detection is by a programmed self-check which is run at all times when the computer is not otherwise occupied. More refined checking may be done by a limited amount of circuitry. For example, some guidance computers employ a parity test on the contents of memory. Still other types of alarms are included in the Apollo Guidance Computer, including tests for prolonged or insufficient interrupt activity and various sorts of program freezes. The sum total of these checks and alarms reduces to a small value the probability that a malfunction shall go undetected.

HARDWARE

Memory Devices - The ferrite coincident current core memory is the cornerstone of computer technology, providing fast random access at a few cents per bit in the megabit range. Thin film memories have had a large research investment and have surpassed core memories in speed and bit density by little or none at all. Their cost is relatively high and their capacity is more limited than core. Plated wire promises to be a substantial improvement, but is not yet advanced enough to be producible or reliable in data processing or guidance applications.

Core memory is clearly ahead of thin film in data processing applications. The matter is controversial with respect to guidance computers, where the higher cost and the capacity limitations of film are less important. Film offers somewhat higher speed, where core offers the economy of coincident selection plus a large output signal. Density and reliability are unresolved issues between the two.

High capacity electromechanical memories such as drums and discs are disappearing from guidance computer use. This is so for three reasons: a substantial increase in packaging density of core and film memories, the serial access nature of discs and drums, and the limited time that discs and drums can operate without maintenance. The high bit densities and large capacities attainable with electromechanical memories make them virtually indispensable to large scale data processing installations, however.

Fixed memory is not used in data processing to any large degree except as a means of implementing internal machine logic such as in a program sequence generator. Its broader use in guidance computers stems from its potential for indestructibility and high density. Indestructibility is a two-edged sword. It requires that program and data be determined well in advance of use; moreover it places a limitation on changes in mission plan such as may be required periodically in ballistic missile applications. Wherever these limitations are not overly constraining, a fixed memory offers assurance that the computer program is identical through all phases of testing and in flight. It moreover permits recovery from temporary malfunctions which would alter the contents of an erasable memory.

Some types of memories compromise between reliability and unchangability by having the ability to be electrically alterable. Modifications of film and core memories have this property, although coincident selection is less apt to be possible in the core versions, most of which are relatives of transfluxors. The bit densities of such memories have been well below those which are available in permanent memories.

Logic Devices - In the past few years integrated circuits, or microcircuits, have been adopted nearly universally by guidance computer designers for at least the logic portion of their computer designs. Prior to the advent of microcircuits, magnetic cores were strong contenders as logic elements against all-transistor circuitry. Core circuits were no larger, and in addition were capable of operating on substantially lower power than all-transistor circuits. Although special applications may yet exist which favor the magnetic core, the very small size of microcircuits and their high speed and proven reliability make them preferred in nearly all instances over cores. With each passing year, moreover, the power consumption of new microcircuit logic devices has been substantially reduced. One can now expect to consume less power with microcircuits than with cores at full speed. The latter elements still retain the advantage of reduced power consumption at low speed operation.

Data processing machines are only now beginning to use microcircuit techniques, because of numerous problems which have attended the large scale production of microcircuits. If these problems are solved, we may expect to see the same increase in the ratio of performance to size in the logic area of data processors which was seen in guidance computers a few years ago.

The primary area in which advances need to be made is in interconnection of logic units.[7] A poor job of mechanical design results in unreliable connections or low component density or both; yet it has so far proven quite difficult to arrive at a structure which is satisfactory in all respects. Some of the essential and important requirements are reliability, ease of manufacture, thermal conductance, mechanical soundness, convenient shape, means of inspection and repair, and high density. Some of the methods which achieve high density are seriously lacking in some of the other attributes listed. The multi-layer etched board appears to be a means whereby the technology of interconnection can be advanced, but there exists some disagreement as to its qualifications in its present state of development. A highly satisfactory, though somewhat less dense method, is the welded wire matrix, which is also a multi-layer device, but not made in an integral unit and not so small as the etched board.

Word Length	15 Bits + 1 Parity
Number System	One's Complement
Memory Cycle Time	11.7 μsec
Fixed Memory Registers	36,864 Words
Erasable Memory Registers	2,048 Words
Number of Normal Instructions	34
Number of Involuntary Instructions (Interrupt, Increment, etc.)	10
Interrupt Options	10
Addition Time	23.4 μsec
Multiplication Time	46.8 μsec
Double Precision Addition Time	35.1 μsec
Double Precision Multiplication Subroutine Time	575 μsec
Increment Time	11.7 μsec
Number of Counters	29
Power Consumption	100 Watts (AGC + DSKY's)
Weight	58 Pounds (Computer Only)
Size	1.0 Cubic Foot (Computer Only)

Table VI-2 AGC Characteristics

Chapter VI-2

CHARACTERISTICS OF THE APOLLO GUIDANCE COMPUTER

LOGICAL DESIGN

General - The AGC has three principal sections. The first is a memory, the fixed (read only) portion of which has 36,864 words, and the erasable portion of which has 2048 words. The next section may be called the central section; it includes an adder, an instruction decoder,(SQ), a memory address decoder, (S), and a number of addressable registers with either special features or special use. The third section is the sequence generator which includes a portion for generating various microprograms and a portion for processing various interrupting requests.

The backbone of the AGC is the set of 16 write buses; these are the means for transferring information between the various registers shown in Fig. VI-1. The arrowheads to and from the various registers show the possible directions of information flow. In Fig. VI-1, the data paths are shown as solid lines; the control paths are shown as broken lines.

The Fixed Memory is made of wired-in "ropes" which are compact and reliable devices. The number of bits so wired is in excess of 5×10^5. The cycle time is 12 μsec.

The erasable memory is a coincident current ferrite core system with the same cycle time as the fixed memory. Instructions can address registers in either memory, and can be stored in either memory. The only logical difference between the two memories is the inability to change the contents of the fixed part by program steps.

Each word in memory is 16 bits long (15 data bits and an odd parity bit). Data words are stored as signed 14 bit words using a one's complement convention. Instruction words consist of 3 order code bits and 12 address code bits.

The contents of the address register S do not always determine uniquely the address of the memory word. For example, the 2048 erasable registers are accessed via a 1024 word address field. This is done with a three-bit auxiliary address contained in the "Erasable Bank" register, which is under program control. Part of the address field is one-to-one: addresses between 0 and 1377 (octal, or base eight) always refer to the same registers. Addresses 1400 - 1777 (octal) are ambiguous, and refer to one of 5 sets of 256 words according to the number stored in the Erasable Bank register.

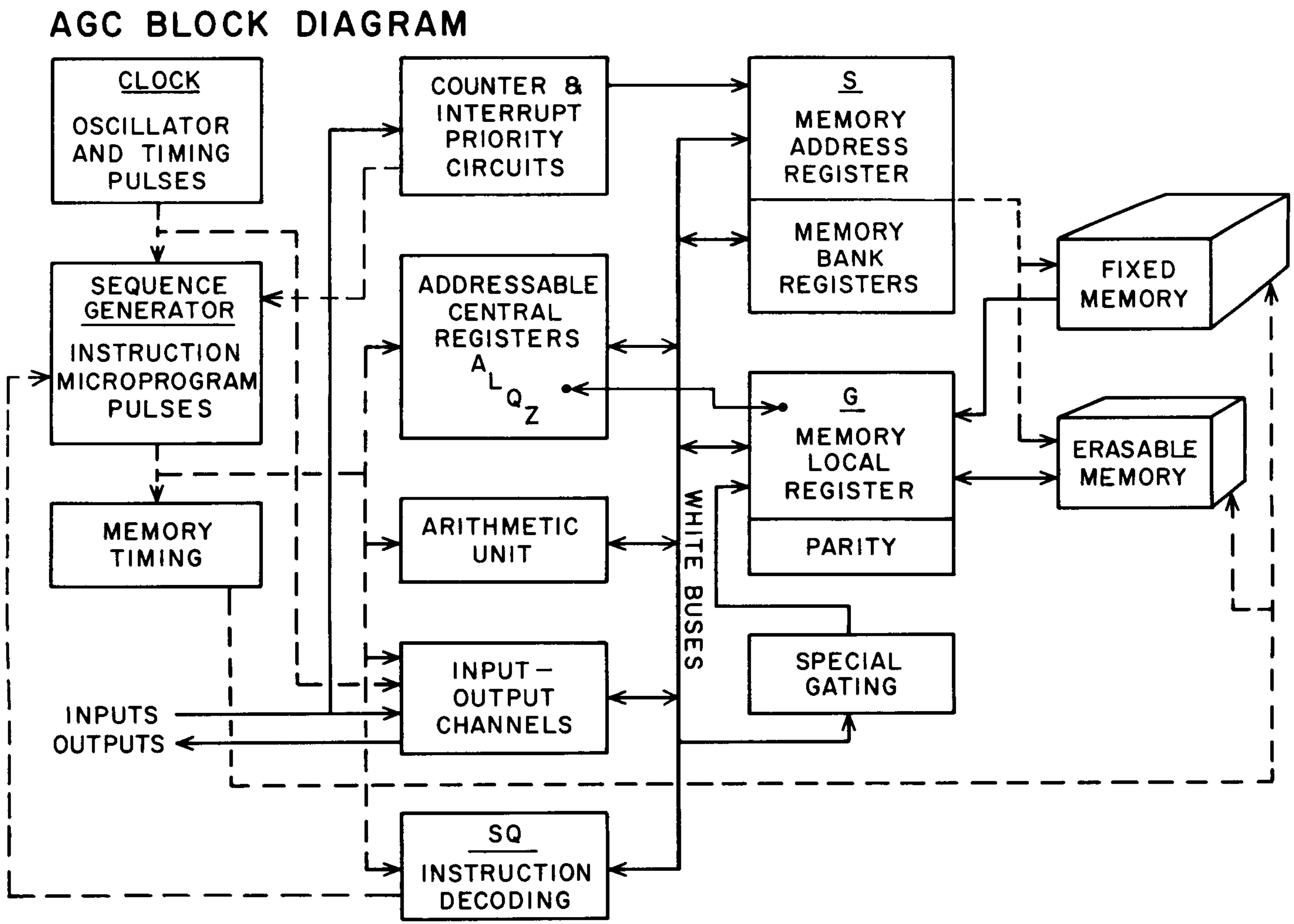


Fig. VI-1 AGC Block Diagram

The 3072 word fixed-memory address field encompasses 36,864 words by means of a 5-bit "Fixed Bank" register and a one-bit "Fixed Extension" channel. Addresses between 2000 and 3777 (octal) are ambiguous, and refer to one of 34 banks of 1024 words each according to the number in the Fixed Bank register. If this number exceeds 30 (octal) then the bank selection further depends on the Fixed Extension bit. The Bank registers and the Extension channel are addressable, and are all in the non-ambiguous portions of the erasable memory and channel fields.

Transfers in and out of memory are made by way of a memory local register G. For certain specific addresses, the word being transferred into G is not sent directly, but is modified by a special gating network. The transformations on the word sent to G are right shift, right cycle, left cycle, and 7-position right shift for editing interpretive instruction words.

The middle part of Fig. VI-1 shows the central section in block form. It contains the address register S and the memory bank registers which were mentioned above. There is also a block of addressable registers called "central and special registers," which will be discussed later, an arithmetic unit, and an instruction decoder register, SQ. The arithmetic unit is an adder with shifting gates and control logic. The SQ register bears the same relation to instructions as the S register bears to memory locations; neither S nor SQ are explicitly addressable. The central and special registers are A, L, Q, Z, and a set of input and output channels. Their properties are shown in Table VI-3.

The sequence generator provides the basic memory timing and the sequences of control pulses (microprograms) which constitute instructions. It also contains the priority interrupt circuitry and a scaling network which provides various pulse frequencies used by the computer and the rest of the navigation system.

Instructions are arranged so as to last an integral number of memory cycles. The list of instructions is treated in detail later. In addition to these there are a number of "involuntary" sequences, not under normal program control, which may break into the normal sequence of instructions. These are triggered either by external events, or by certain overflows within the AGC, and may be divided into two categories: counter incrementing and program interruption.

Counter incrementing may take place between any two instructions. External requests for incrementing a counter are stored in a counter priority circuit. At the end of every instruction a test is made to see if any incrementing requests exist. If not, the next instruction is executed directly. If a request is present, an incrementing memory cycle is executed. Each "counter" is a specific location in erasable memory. The incrementing cycle consists of reading out the word stored in the counter register, incrementing it (positively or negatively) or shifting it, and storing the results back in

Register	Octal Address	Purpose
A	0000	Central accumulator. Most instructions refer to A.
L	0001	Lower accumulator. Used in multiply, divide, and all double-precision operations.
Q	0002	Return address register. If a transfer control (TC) operation occurred at line L, (Q) = L + 1.
EB	0003	Erasable bank register, bits 9, 10, 11.
FB	0004	Fixed bank register, bits 11, 12, 13, 14, 15.
Z	0005	Program counter. Contains L + 1, where L is the address of the instruction presently being executed.
BB	0006	Both bank registers: Erasable, bits 1, 2, 3. Fixed, bits 11, 12, 13, 14, 15.
--	0007	Contains zero.

Table VI-3 Addressable Special and Central Registers the register of origin. All outstanding counter incrementing requests are processed before proceeding to the next instruction. This type of interrupt provides for asynchronous incremental or serial entry of information into the working erasable memory. The program steps may refer directly to a counter register to obtain the desired information and do not have to refer to input buffers. Overflows from one counter may be used as inputs to another. A further property of this system is that the time available for normal program steps is reduced linearly by the amount of counter activity present at any given time.

Program interruption also occurs between program steps. An interruption consists of storing the contents of the program counter and transferring control to a fixed location. Each interrupt option has a different location associated with it. Interrupting programs may not be interrupted, but interrupt requests are not lost, and are processed as soon as the earlier interrupted program is resumed.

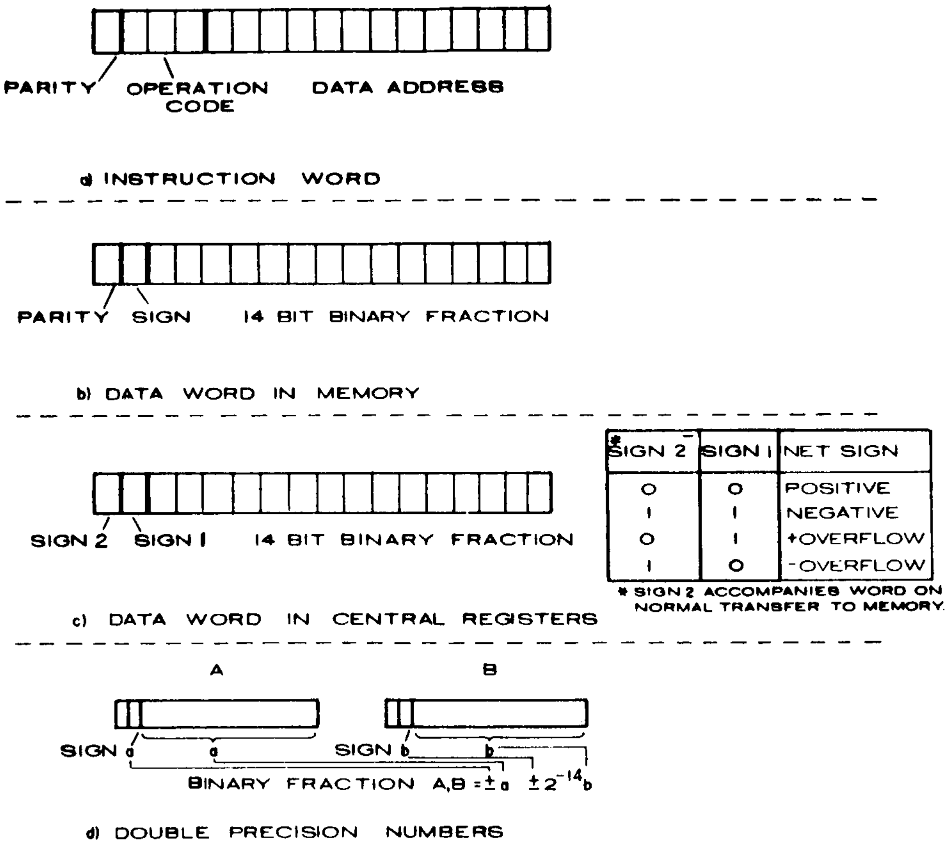
Word Length - The AGC is a "common storage" machine, which means that instructions may be executed from erasable memory as well as from fixed memory, and that data (obviously constants, in the case of fixed memory) may be stored in either memory. The word sizes of both types of memory must be compatible in some sense; the easiest solution is to have equal word lengths. The AGC is somewhat unique in its very short word length, and the reasons for it are of some interest. The principal factors in the choice of word length are:

A. Precision desired in the representation of navigational variables;
B. Range of the input variables which are entered serially or incrementally;
C. Instruction word format. Division of instruction words into two fields, one for operation code and one for address.

As a start, the word length (15 bits) for two previous machines in this series[4] was kept in mind as a satisfactory word length from the point of view of mechanization; i.e., the number of sense amplifiers and inhibit drivers, and the carry propagation time, etc., were all considered satisfactory. The influence of these principal factors will be taken up in turn.

The data words used in the AGC may be divided roughly into two classes: data words used in elaborate navigational computations, and data words used in the control of various appliances in the system. Initial estimates of the precision required by the first class ranged from 27 to 32 bits $0(10^{8 \pm 1})$. The second class of variables could almost always be represented with 15 bits. The fact that navigational variables require about twice the desired 15-bit word length means that there is not much advantage to word sizes between 15 and 28 bits, as far as precision of representation of variables is concerned, because double-precision numbers must be used in any event. Because of the doubly signed number representation for double-precision words, the equivalent word length is 29 bits (including sign), rather than 30, for a basic word length of 15 bits.

129



*SIGN 2	SIGN 1	NET SIGN
0	0	POSITIVE
1	1	NEGATIVE
0	1	+OVERFLOW
1	0	-OVERFLOW

* SIGN 2 ACCOMPANIES WORD ON NORMAL TRANSFER TO MEMORY.

Fig. VI-2 Word Formats

The initial estimates for the proportion of 15-bit vs. 29-bit quantities to be stored in both fixed and erasable memories indicated the overwhelming preponderance of the former. It was also estimated that a significant portion of the computing had to do with control, telemetry and display activities, all of which can be handled more economically with short words. A short word allows faster and more efficient use of erasable storage because it reduces fractional word operations, such as packing and editing; it also means a more efficient encoding of small integers.

As a control computer, the AGC must make analog-to-digital conversions, many of which are of angles. Two principal forms of conversion exist: one renders a whole number, the other produces a train of pulses which must be counted to yield the desired number. The latter type of conversion is employed by the AGC, using the counter incrementing feature. When the number of bits of precision required is greater than the computer's word length, the effective length of the counter must be extended into a second register, either by programmed scanning of the counter register, or by using a second counter register to receive the overflows of the first. Whether programmed scanning is feasible depends largely on how frequently this scanning must be done. The cost of using an extra counter register is directly measured in terms of the priority circuit associated with it. In the AGC, the equipment saved by reducing the word length below 15 bits would probably not match the additional expense incurred in double-precision extension of many input variables. The question is academic, however, since a lower bound on the word length is effectively placed by the format of the instruction word.

An initial decision was made that instructions would consist of an operation code and a single address. The straightforward choices of packing one or two such instructions per word were the only ones seriously considered, although other schemes, such as packing one and a half instructions per word, are possible.[1] The two previous computers had a 3-bit field for operation codes and a 12-bit field for addresses, to accommodate their 8 instruction order codes and 4096 words of memory. In the initial core-transistor version of the AGC, the 8 instruction order codes were in reality augmented by the various special registers provided, such as shift right, cycle left, edit, so that a transfer in and out of one of these registers would accomplish actions normally specified by the order code. These registers were considered to be more economical than the corresponding instruction decoding and control pulse sequence generation. Hence the 3 bits assigned to the order code were considered adequate, albeit not generous. Furthermore, as will be seen, it is possible to expand the number of order codes.

The address field of 12 bits presented a different problem. At the time of the design of the previous computers, it was estimated that 4000 words would satisfy the the storage requirements. By the time of redesign it was clear that the requirement was for 10^4 words, or more, and the question then became whether the proposed extension of the address field by a bank register was more economical than the addition of bits to the word length. For reasons of modularity of equipment, adding more bits to the word length would result in adding more bits to all the central and special registers, which amounts to increasing the size of the non-memory portion of the AGC.

In summary, the 15-bit word length seemed practical enough so that the additional cost of extra bits in terms of size, weight, and reliability did not seem warranted. A 14-bit word length was thought impractical because of the problems with certain input variables, and it would further restrict the already cramped instruction word format. Word lengths of 17 or 18 bits would result in certain conceptual simplicities in the decoding of instructions and addresses, but would not help in the representation of navigational variables. These require 28 bits, so they must be represented in double precision in any event.

Number Representation - In the absence of the need to represent numbers of both signs, the discussion of number representation would not extend beyond the fact that numbers in the AGC are expressed to base two. But the accommodation of both positive and negative numbers requires that the logical designer choose among at least 3 possible forms of binary arithmetic. These 3 principal alternatives are: one's complement, two's complement, and sign and magnitude.

In one's complement arithmetic, the sign of a number is reversed by complementing every digit; and "end around carry" is required in addition of two numbers. In two's complement arithmetic, sign reversal is effected by complementing each bit and adding a low order one, or some equivalent operation. Sign and magnitude representation is typically used where direct human interrogation of memory is desired, as in "postmortem" memory dumps, for example. The addition of numbers of opposite sign requires either one's or two's complementation or comparison of magnitude, and sometimes may use both. The one's complement notation has the advantage of having easy sign reversal, which is equivalent to Boolean complementation; hence a single machine instruction performs both functions. Zero is ambiguously represented by all zero's and by all one's, so that the number of numerical states in an n-bit word is $2^n - 1$. Two's complement arithmetic is advantageous where end around carry is difficult to mechanize, as is particularly true in serial computers. An n-bit word has 2^n states, which is desirable for input conversions from such devices as pattern generators, geared encoders, or binary scalers. Sign reversal is awkward, however, since a full addition is required in the process.

In a standard one's complement adder, overflow is detected by examining carries into and out of the sign position. These overflow indications must be "caught on the fly" and stored separately if they are to be acted upon later. The number system

		STANDARD					MODIFIED					
		S_1	4	3	2	1	S_2	S_1	4	3	2	1
EXAMPLE 1:	Both operands positive; Sum positive, no overflow. Identical results in both systems.	0	0	0	0	1	0	0	0	0	0	1
		0	0	0	1	1	0	0	0	0	1	1
		0	0	1	0	0	0	0	0	1	0	0
EXAMPLE 2:	Both operands positive; positive overflow. Standard result is negative; Modified result is positive using S_2 as sign of the answer. Positive overflow indicated by $S_1 \cdot \bar{S}_2$.	0	1	0	0	1	0	0	1	0	0	1
		0	1	0	1	1	0	0	1	0	1	1
		1	0	1	0	0	0	1	0	1	0	0
EXAMPLE 3:	Both operands negative; Sum negative no overflow. End around carry occurs. Identical results in both systems using either S_1 or S_2 as the sign of the answer.	1	1	1	1	0	1	1	1	1	1	0
		1	1	1	0	0	1	1	1	1	0	0
		1	1	0	1	0	1	1	1	0	1	0
						1 carry						1 carry
		1	1	0	1	1	1	1	1	0	1	1
EXAMPLE 4:	Both operands negative; negative overflow. Standard result is positive; modified result is negative using S_2 as the sign of the answer. Negative overflow indicated by $\bar{S}_1 \cdot S_2$.	1	0	1	1	0	1	1	0	1	1	0
		1	0	1	0	0	1	1	0	1	0	0
		0	1	0	1	0	1	0	1	0	1	0
						1 carry						1 carry
		0	1	0	1	1	1	0	1	0	1	1
EXAMPLE 5:	Operands have opposite sign: Sum positive. Identical results in both systems.	1	1	1	1	0	1	1	1	1	1	0
		0	0	0	1	1	0	0	0	0	0	1
		0	0	0	0	1	0	0	0	0	0	1
						1 carry						1 carry
		0	0	0	1	0	0	0	0	0	1	0
EXAMPLE 6:	Operands have opposite sign; sum negative. Identical results in both systems.	1	1	1	0	0	1	1	1	1	0	0
		0	0	0	0	1	0	0	0	0	0	1
		1	1	1	0	1	1	1	1	1	0	1

Fig. VI-3 Illustrative example of properties of modified one's complement system.

adopted in the AGC has the advantage of being a one's complement system with the additional feature of having a static indication of overflow. The implementation of the method depends on the AGC's not using a parity bit in most central registers. Because of certain modular advantages, 16, rather than 15, columns are available in all of the central registers, including the adder. Where the parity bit is not required, the extra bit position is used as an extra column. The virtue of the 16-bit adder is that the overflow of a 15-bit sum is readily detectable upon examination of the two high order bits of the sum (see Fig. VI-3). If both of these bits are the same, there is no overflow. If they are different, overflow has occurred with the sign of the highest order bit.

The interface between the 16-bit adder and the 15-bit memory is arranged so that the sign bit of a word coming from memory enters both of the two high order adder columns. These are denoted S_2 and S_1 since they both have the significance of sign bits. When a word is transferred to memory, only one of these two signs can be stored. In the AGC the S_2 bit is stored, which is the standard one's complement sign except in the event of overflow, in which case it is the sign of the two operands. This preservation of sign on overflow is an important asset in dealing with carries between component words of multiple-precision numbers.

Multiple-Precision Arithmetic - A short word computer can be effective only if the multiple-precision routines are efficient corresponding to their share of the computer's work load. In the AGC's application there is enough use for multiple-precision arithmetic to warrant consideration in the choice of number system and in the organization of the instruction set. A variety of formats for multiple-precision representation are possible; probably the most common of these is the identical sign representation in which the sign bits of all component words agree. The method used in the AGC allows the signs of the components to be different.

Independent signs arise naturally in multiple-precision addition and subtraction, and the identical sign representation is costly because sign reconciliation is required after every operation. For example, (+6, +4) + (-4, -6) = (+2, -2), a mixed sign representation of (+1, +8). Since addition and subtraction are the most frequent operations, it is economical to store the result as it occurs and reconcile signs only when necessary. When overflow occurs in the addition of two components, a one with the sign of the overflow is carried to the addition of the next higher components. The sum that overflowed retains the sign of its operands. This overflow is termed an interflow to distinguish it from an overflow that arises when the maximum multiple-precision number is exceeded.

For triple and higher orders of precision, multiplication and division become excessively complex, unlike addition and subtraction where the complexity is only linear with the order of precision. Apollo programs do not require greater than double-precision multiplication and division, however. The algorithm for double-precision multiplication is directly applicable to numbers in the independent sign notation. The treatment of interflow is simplified by a double-precision add instruction. Double-precision division is exceptional in that the independent sign notation may not be used; both operands must be made positive in identical sign form, and the divisor normalized so that the left-most non-sign bit is <u>one</u>. A few triple-precision quantities are used in the AGC. These are added and subtracted using independent sign notation with interflow and overflow features the same as those used for double-precision arithmetic.

<u>Instruction Set</u> - The major goals in the AGC were efficient use of memory, reasonable speed of computing, potential for elegant programming, efficient multiple-precision arithmetic, efficient processing of input and output, and reasonable simplicity of the sequence generator. The constraints affecting the order code as a whole were the word length, one's complement notation, parallel data transfer, and the characteristics of the editing registers. The following rules governing the design of instructions arose from these goals and constraints: three bits of an instruction word are devoted to operation code, address modification must be convenient and efficient, there should be a multiply instruction yielding a double length product, facility for multiple precision must be available, and a Boolean combinatorial operation should be available. These rules are by no means complete, but give a good indication of what kind instruction set was desired.

The three bits reserved for instruction codes are capable of rendering a selection among eight operations with no further refinement. Two techniques are employed in the AGC to expand the number of operations 4-fold. These are called "extension" and "partial codes" respectively. Extension is like using a teletype shift code; when an Extend instruction occurs, it signifies that the next instruction code in sequence is to be interpreted otherwise than normally. By this means, the instruction set could be expanded almost indefinitely at a penalty in speed, for a memory cycle time is required for each extension. In the AGC the size of the instruction set is doubled by an Extend operation, which calls forth the less-often used instructions. For example, code 000 selects the Transfer Control instruction unless it is preceded by an Extend, in which case it selects an Input-Output instruction.

Partial codes are instruction codes which encroach upon the address field. This technique capitalizes upon the essential difference between fixed and erasable memory. More specifically, a wider variety of instructions are applicable to erasable than to fixed memory; for example, all instructions which modify the operand register are not fully applicable for fixed memory. Since the fixed memory address field in the AGC is 3 times as big as the erasable memory field, it is possible to pack 3 extra erasable memory instructions into that portion of the entire address field. Thus operation code 101 for addresses 0 through 1777 (octal) selects the Index instruction for the erasable memory, whose address field is also 0 through 1777 (octal). The same operation code for addresses 2000 - 3777 (octal) selects a Double Exchange instruction for erasable memory, whose addresses are obtained by reducing the address modulo 2000 (octal). In a similar way, the Transfer to Storage instruction is selected by the same code for addresses 4000 - 5777 (octal), and the Exchange A instruction for addresses 6000 - 7777 (octal), both for erasable memory. Alternatively, the entire fixed memory field may select a different instruction for fixed memory, or else the same instruction may be selected over the entire address field.

Table VI-4 lists the normal AGC instructions. These include facility for double-precision data handling and addition. Many of these instructions are similar to one another and share microprogram steps.

Input and output are handled to a large extent by special registers called channels, which are not accessible through the regular address field. In the version of the AGC prior to the present one, this was not true; the input and output registers were addressable for any instruction. Here, the channels are accessible by the input-output instructions alone. A slight extra degree of freedom is provided by making the Lower accumulator (L) and Return address (Q) registers accessible through channels 1 and 2 as well as through regular addresses 1 and 2. This is primarily to allow the programmer to take advantage of the <u>or</u> and <u>exclusive or</u> input-output instructions.

The remainder of the AGC instructions are involuntary or address dependent, and are listed in Table VI-5. The last four are not really instructions, but are rather editing operations on all words written into the specified four addresses. They are tabulated as instructions only because such operations have instruction status in most computers.

HARDWARE

<u>Logic</u> - The design of the Apollo Guidance Computer began at a time when microcircuits were first being produced. Microcircuits held great promise, but were not well enough proven for the design of this computer to be based on them; magnetic core and transistor logic had been used in its immediate ancestry and was scheduled to be used here. Nevertheless, during the first year of design, microcircuits were evaluated for possible use in the AGC. When it became clear that microcircuits could be reliably produced with rigid specifications, the decision was made to substitute them for the core-transistor logic. In the course of this change, the power consumption increased by a factor of three, but size and weight were reduced by half, and performance and speed were doubled. Moreover, though it could not be known at the time, the reliability of the logic hardware was greatly increased.

One of the important decisions made at that time was to confine the use of logic microcircuits to a single type to avoid having to develop successfully a number of

A. Sequence Changing

	1. Transfer control, set return address	1 MCT	All Memory
	2. Transfer control only	1 MCT	Fixed only
	3. Four way skip and diminish by one	2 MCT	Erasable
*	4. Branch on zero	1 or 2	Fixed only
*	5. Branch on zero or minus	1 or 2	Fixed only

B. Fetching and Storing

	1. Clear and add to Accumulator, A	2 MCT	All
	2. Clear and subtract from Accumulator, A	2 MCT	All
*	3. Double clear and add to A and Lower Accumulator, L	3 MCT	All
*	4. Double clear and subtract from A and L	3 MCT	All
	5. Transfer to storage	2 MCT	Erasable
	6. Exchange A with storage	2 MCT	Erasable
	7. Double exchange A and L with storage	3 MCT	Erasable
	8. Exchange L with storage	2 MCT	Erasable
*	9. Exchange Q with storage	2 MCT	Erasable

C. Instruction Modification

	1. Index (add to next instruction)	2 MCT	Erasable
*	2. Index and extend	2 MCT	All

* Requires Extend instruction
MCT = Memory Cycle Time

Table VI-4 Normal Instructions (Part 1)

D. Arithmetic and Logic

1.	Add to A	2 MCT	All
* 2.	Subtract from A	2 MCT	Erasable
3.	Add to Storage and A	2 MCT	Erasable
* 4.	Modular subtract from A (mixed number system)	2 MCT	Erasable
5.	Add 1 to storage (Increment)	2 MCT	Erasable
* 6.	Increase absolute value of storage by 1 (Augment)	2 MCT	Erasable
* 7.	Decrease absolute value of storage by 1 (Diminish)	2 MCT	Erasable
8.	Double add A and L to storage	3 MCT	Erasable
9.	Logical product to A	2 MCT	All
* 10.	Multiply; product to A and L	3 MCT	All
* 11.	Divide A and L by storage; quotient to A	6 MCT	Erasable

E. Input Output

* 1.	Transfer channel to A	2 MCT	Channels
* 2.	Transfer A to channel	2 MCT	Channels
* 3.	Logical product (of A and channel) to A	2 MCT	Channels
* 4.	Logical product to channel and A	2 MCT	Channels
* 5.	Logical sum to A	2 MCT	Channels
* 6.	Logical sum to channel and A	2 MCT	Channels
* 7.	Exclusive or to A	2 MCT	Channels

* Requires Extend instruction
MCT = Memory Cycle Time

Table VI-4 Normal Instructions (Part 2)

A. <u>Involuntary</u>

1. Transfer to interrupt program, store c(Z) and c(B)	3 MCT	Limited
2. Increment by 1	1 MCT	Counters
3. Increment by -1	1 MCT	Counters
4. Diminish absolute value by 1	1 MCT	Counters
5. Shift left	1 MCT	Counters
6. Shift left and add 1	1 MCT	Counters

B. <u>Address Dependent</u>

1. Resume interrupted program = Index 0017 (octal)	2 MCT
2. Extend = Transfer control 0006	1 MCT
3. Inhibit interrupt = Transfer control 0004	1 MCT
4. Permit interrupt = Transfer control 0003	1 MCT
5. Cycle right each access	Address 20 (octal)
6. Shift right each access	Address 21 (octal)
7. Cycle left each access	Address 22 (octal)
8. Shift right seven places each access	Address 23 (octal)

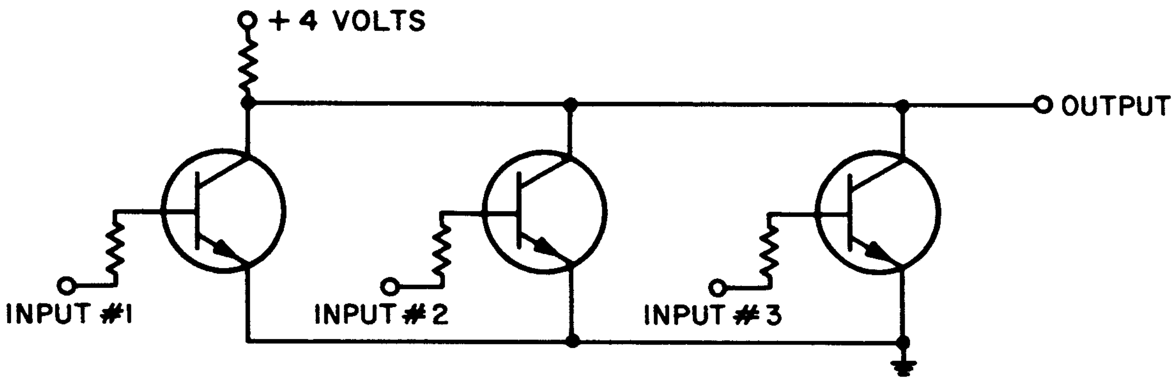
MCT = Memory Cycle Time

Table VI-5 Special Instructions

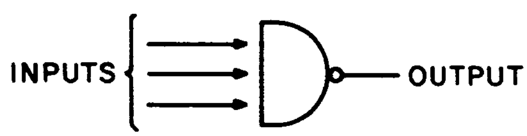
different devices. A logic circuit was required whose operational function was capable of synthesizing all switching functions, and which was simple enough to be controllable, testable, and producible. The circuit chosen was a NOR gate which employs a configuration known as modified direct-coupled transistor logic (DCTL). Three transistors in parallel, along with four resistors, form a three-input gate with a fan-out capability of approximately 5, and an average propagation delay of about 20 nanoseconds, while dissipating about twelve milliwatts of power. A recent modification of design has resulted in a new unit with approximately the same specifications except for a power dissipation of 5 milliwatts instead of 12. These gates are designed to operate over the temperature range from 0 to 70°C.

The importance of using a single circuit should not be underestimated. Thousands of logic gates are employed in each computer and barring the use of redundancy techniques, every one may be considered critical. Indeed, most redundancy techniques depend on randomness of failures; and in general new components and assembly methods introduce failure modes which make erroneous the basic assumptions on which the redundancy is based. High reliability is essential for every gate. It can best be attained by standardization, and can only be demonstrated by the evaluation of large samples.[8] Had a second type of logic microcircuit been employed in the AGC, the number of logic elements could have been reduced by about 20 percent; but it is clear that to have done so would have been false economy, for neither of the two circuits would have accumulated the high mean time to failure and high confidence level that the one NOR-circuit has.

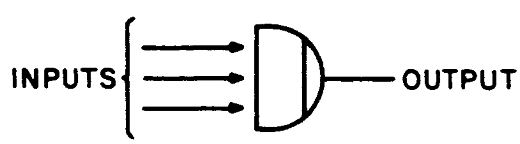
Logic equations expressed in the familiar AND, OR, NOT notation may readily be realized with NOR operators. A two-level and-or expression is realizable in a two-level NOR-circuit. The NOR function of three variables is as follows: $N(x, y, z) = \bar{x}\,\bar{y}\,\bar{z} = \overline{x + y + z}$. An AND function is $A(x, y, z) = x\,y\,z$, and an OR function is $0(x, y, z) = x + y + z$. By comparison, $N(x, y, z) = A(\bar{x}, \bar{y}, \bar{z}) = \overline{0(x, y, z)}$. The NOT operation, or complementation, is the NOR function of one variable; i.e., $\bar{x} = N(x)$. Complex Boolean expressions ordinarily arise only in connection with non-sequential, or combinational aspects of the computer logic. Sequential operations require storage; and the basic logic storage element is the flip-flop. Two NOR gates form a flip-flop if the output of each is an input to the other, and if all other inputs are normally zero. If one of these other inputs is momentarily made equal to <u>one</u>, the flip-flop is forced into one state, whereas if a free input on the opposite gate is made equal to <u>one</u>, the other state is obtained. Most frequently, the condition for setting a flip-flop to a particular state is that two or more other signals simultaneously take on prescribed values. Detection of such coincidence requires a NOR operation separate from the flip-flop plus any NOR operation required to invert (complement) the inputs.



a) EQUIVALENT CIRCUIT OF NOR GATE



b) DIAGRAM NOTATION FOR NOR GATE



c) DIAGRAM NOTATION FOR UNPOWERED NOR GATE (+4 VOLTS DISCONNECTED)

Fig. VI-4 The NOR Gate

It is frequently necessary to implement NOR functions of more than three variables, and also to be able to drive more than five inputs with a single output. For these reasons, NOR gates may be combined so as to increase either the input (fan-in) capacity, or the output (fan-out) capacity, or both. Fan-in is increased by connecting the outputs of unpowered gates to the output of a powered gate. This provides a fan-in of three times the total number of gates. Fan-out is increased by connecting the outputs of powered gates together. Both fan-in and fan-out are increased, but the fan-in is not available because it is necessary to have each input signal connected to as many inputs in common as there are powered gates connected together. This is done in order to be able to saturate the transistors, whose current gain is limited. By simultaneous application of these techniques however, it is possible to increase both fan-in and fan-out at the same time.

An illustrative example of the employment of NOR logic in the AGC is provided by the Central register flip-flops. Digits are transferred from one register to another via a common set of wires called write buses. The sending and receiving flip-flops are selected by read and write pulses, respectively, applied to gates which either set or interrogate the flip-flop of the corresponding register. Figure VI-5 shows a hypothetical set of three flip-flops similar to those in one column of the AGC central register section. Dashed lines imply the existence of other registers than the three shown. Diagonal lines, or slashes, after signal names denote inverse polarity. Thus WRITE BUS/ is normally in the one state, and changes to zero while transferring a one. Suppose REG 1 contains a one, i. e., the top gate of its flip-flop has an output of zero. At the time that the READ 1/ signal goes to zero from its normally one state, the output of the read gate, CONTENT 1, becomes a one. This propagates through a read bus fan-in and an inverter and fan-out amplifier to make WRITE BUS/ become zero. Suppose that WRITE 2/ is made zero concurrently with READ 1/. Then the coincidence of zero's at the write gate of REG 2 generates a one at the input to the upper gate of its flip-flop, thus setting the bit to one.

If REG 1 had contained a zero, the write bus would have remained at one, and no setting input would have appeared at the upper gate of REG 2. The CLEAR 2 pulse, which always occurs during the first half of WRITE 2, would have forced the flip-flop to the zero state, where it would remain; whereas when a one is transferred, the SET 2 signal persists after the CLEAR 2, and thus forces the register back to the one state. Thus the simultaneous occurrence of READ 1/, WRITE 2/, and the short CLEAR 2 pulses transfer the content of REG 1 to REG 2. Only the content of REG 2 may be altered in the process. REG 1 and REG 3 retain their original contents. An instance of gates being used to increase fan-in is shown where several CONTENT signals are mixed together to form the signal READ BUS/. An increase in fan-out is achieved by the two gates connected in parallel to form the signal WRITE BUS/.

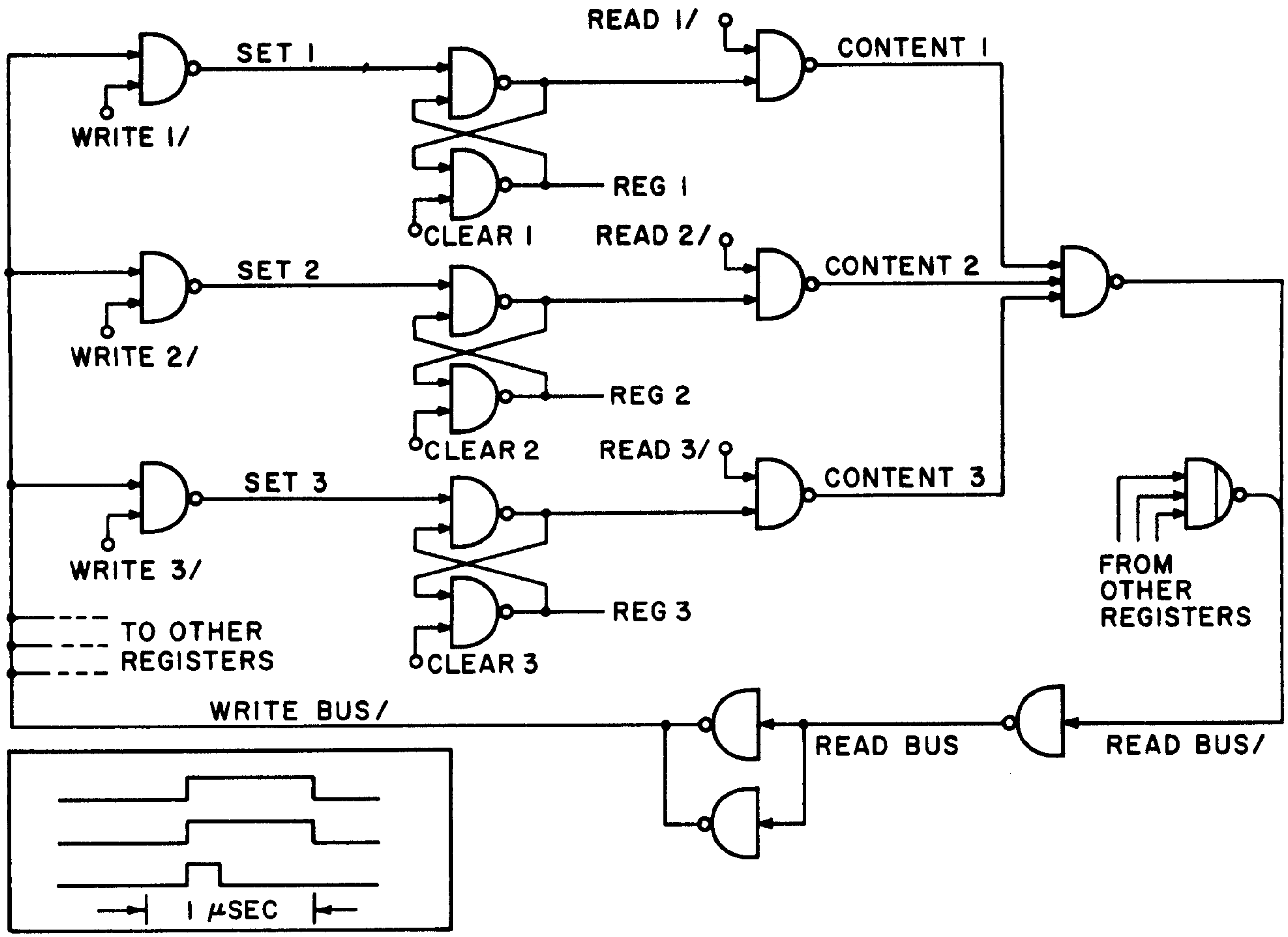


Fig. VI-5 NOR Gate Circuit Resembling One Column of AGC Central Registers

The main problem of mechanical design in guidance computer logic is the creation of signal interconnections; indeed, approximately 3/4 of the volume of the AGC is used for this purpose. Interconnections are primarily of two types: between modules by wrapped wire, and within a module by welded matrix.

The carrier into which all modules are inserted is called a "tray." The AGC comprises two trays: one for logic, power supply and interface modules, and the other for memory and ancillary circuit modules. The 15,000 jacks on the tray into which signal pins are inserted pass through the tray and extend out the other side in the form of posts with square cross sections. Interconnections between pins are made by wires whose ends are tightly wrapped around the posts without the use of any further contact mechanism such as solder or welds. This method has several advantages: it is executed by a machine, which requires only a few seconds per wire, it is controlled by a punched card input, it can easily be altered if a change is desired, wires can be run point to point if desired, and the reliability of the connection is extremely high since there is no single point where bending stress is applied. Moreover, it is compatible with hand wiring, which is required wherever wires are twisted together to protect low level signals, or where heavy gauge wire is needed in order to accommodate high currents.

In the AGC, one of the basic goals has been to make the electronic circuits in small pieces which are easily installed and removed, for the sake of producibility, testing, easy diagnosis and economical maintenance. This can only be realized insofar as it does not excessively degrade the overall packaging density of the computer; for volume is, of course, critically limited in the spacecraft. It was found expedient to make 24 modules each containing 120 microcircuit units, separated into two independent groups of sixty. The twelve-milliwatt gates are packaged one to a unit; sixty gates are connected together into a circuit with 72 pins to bring signals in and out. The more recent five-milliwatt gates are packaged two to a unit, because of their less severe heat transfer requirement. These are organized into sub-groups of thirty, such that sixty gates are again connected together; and 72 pins are again available to each sixty gates. The modules are the same size, so that the low-power units are packaged with double the density of the high power units. Accordingly, the density of pins and interconnections has been doubled along with that of the gates.

The main method of connection internal to the module is by matrix. Gates are disposed in a single row within each sixty-gate sub-group. An array of vertical wires (at right angles to the row of gates) access every connection to the gates. Horizontal conductors (parallel to the row of gates) carry signals from gate to gate and from gate to pin. Connections between horizontal and vertical matrix members as well as between matrix members and gate connection pins are made by a spot welding process. The process was developed in an earlier guidance computer project in order to eliminate the problems of cold solder joints.

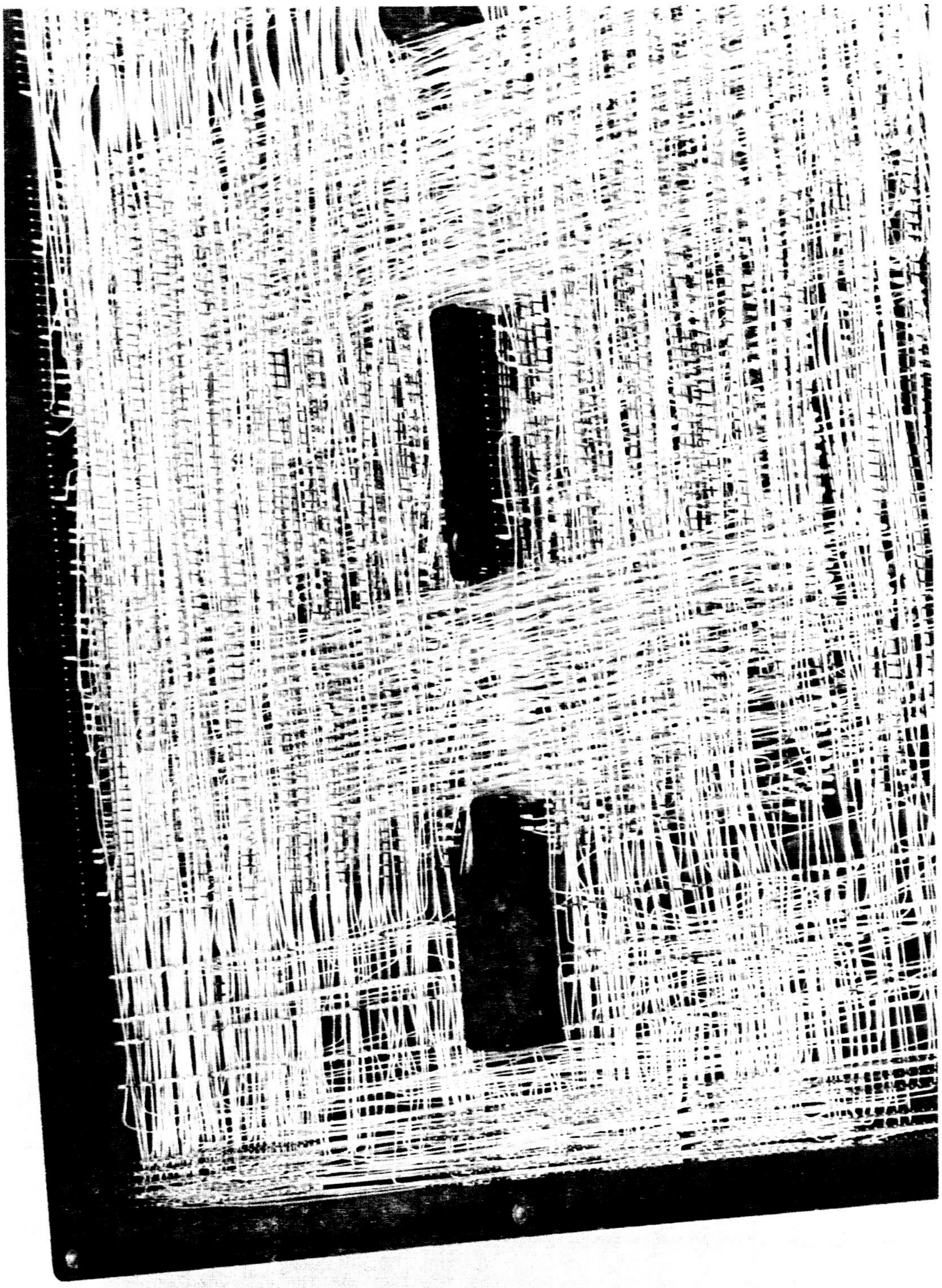


Fig. VI-6 Photograph of Wire-wrapped Tray Connections

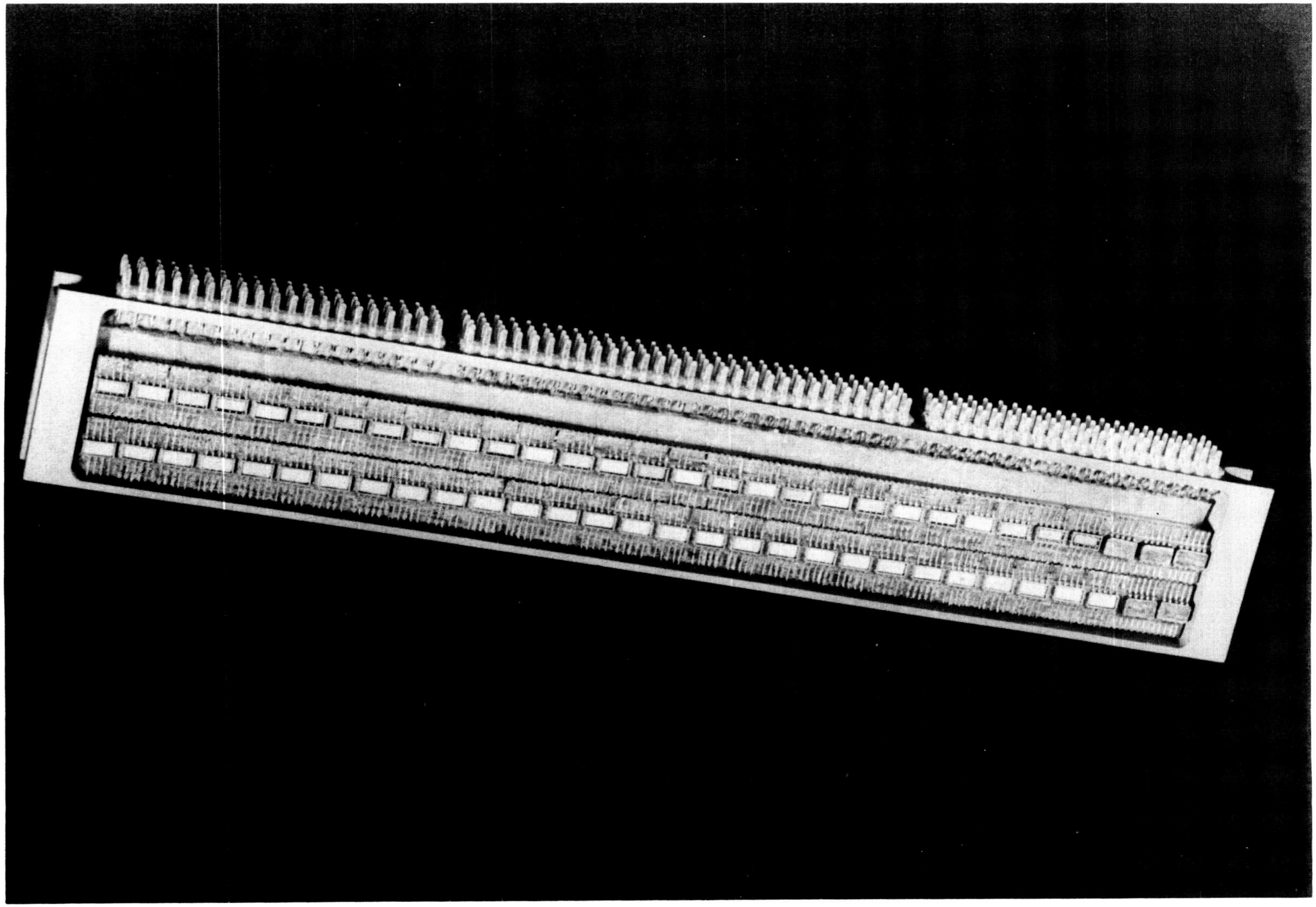


Fig. VI-7 Logic Module Containing 120 Microcircuit Units

Power distribution is a special problem in this computer. The current drawn by the gates is about 6 Amperes for low power gates and about 13 Amperes for high power. For the sake of efficiency, these large currents must be distributed from the power supply to the logic modules with very little d.c. voltage drop. Moreover, the current return, or zero-volt distribution must not sustain any a.c. voltages of such a frequency or amplitude as to turn on or turn off a gate inadvertently. This is all accomplished by building an interlaced gridwork of heavy conductors upon the terminal posts of the tray. Each group of sixty gates shares a ground plane in a module which is brought out at three equally spaced places to connect to this gridwork, which provides multiple paths for return current much the same as a ground plane. The other power line, the positive voltage, is distributed by a gridwork circuit to two points on the power bus shared in a module by each sixty-gate group.

The main tray structure of the AGC is an aluminum alloy frame into which the modules are affixed by jacking screws, providing a good thermal path between modules and tray. The tray in turn is screwed to a cold plate where heat is removed. A model of the AGC is shown in Figs. VI-8 and VI-9. The front is shown in the first figure, with the outlines of the six fixed memory modules just visible. One of the six is partially extracted. The second figure shows the rear and the connector deck with system and test connector covers. The ruler is calibrated in inches (one inch = 2.54 cm.).

Memory - The erasable memory of the AGC was inherited from its core-transistor logic ancestor.[9] It is a conventional coincident-current ferrite core array whose ferrite compound yields a combination of high squareness and a comparatively low sensitivity to temperature. Moreover, the silicon transistor circuits which drive this memory vary their outputs with temperature in such a way as to match the requirements of the cores over a wide range, from $0^{\circ}C$ to $70^{\circ}C$. Coincident current selection affords an economy in selection circuitry at the expense of speed in comparison with linear (word) selection. This is advantageous to the AGC, where the memory cycle time is already long, largely due to the fixed memory. The 2048 word array is wired in 32 x 64 planes with no splices in the wires for highest reliability. The planes are folded to fit into a 9 cubic inch module along with two diodes for each select line. Bi-directional currents are generated in each selection wire by a double-ended transistor switching network. The selection of one wire in 32 is made by twelve switch circuits in an 8 x 4 array; the selection of one wire in 64 is made in an 8 x 8 array. The operation of the switching network is illustrated in Fig. VI-10. The transistors are driven by magnetic cores, which offer two advantages over all-transistor circuits: small size, and storage of address for data regeneration. Again, this circuit economy is realized at the expense of speed. The timing of the currents which operate the switch cores is based on the duration of the write current in the memory array, which is

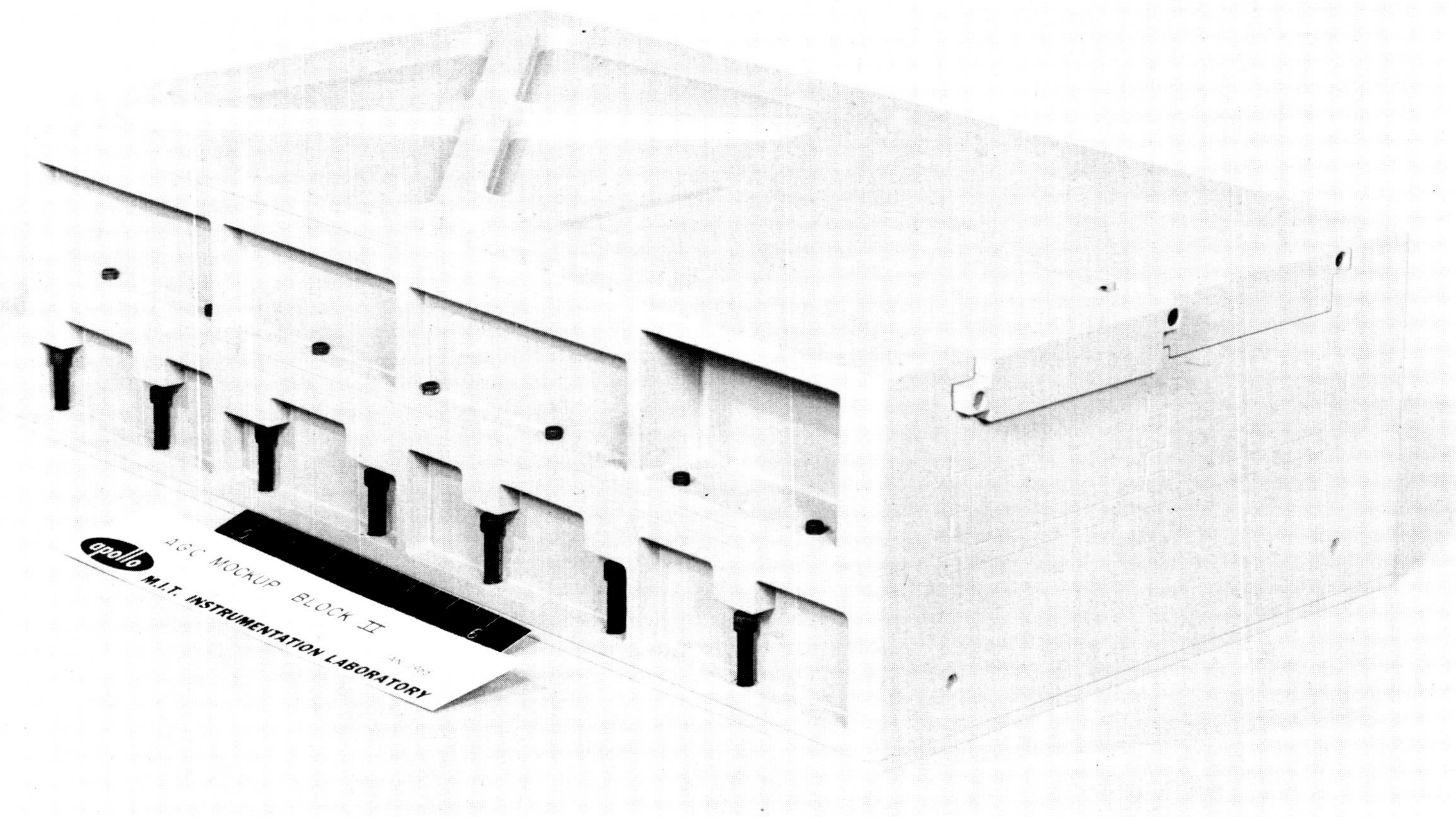


Fig. VI-8 AGC Mockup - Front View

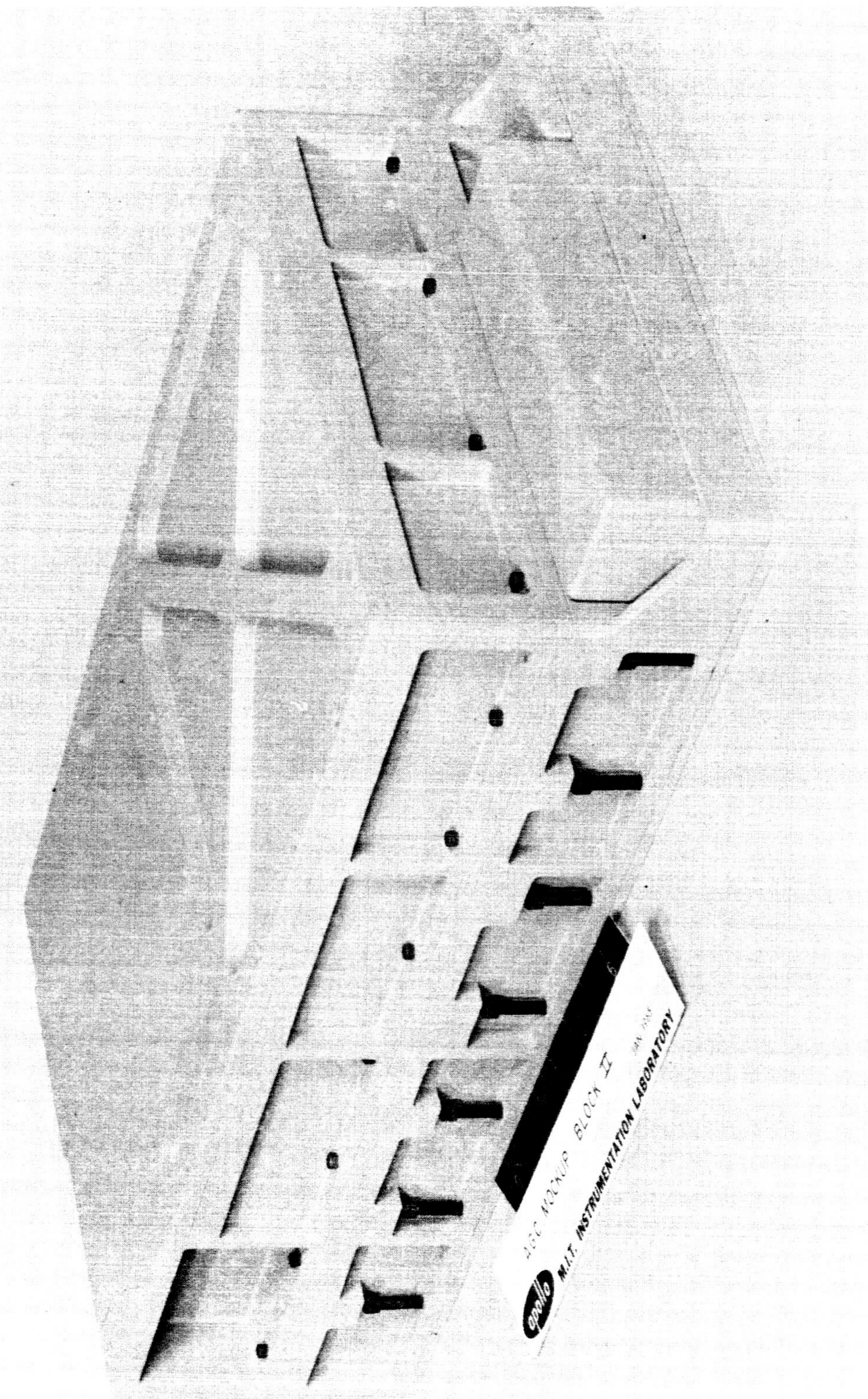


Fig. VI-9 AGC Mockup - Rear View

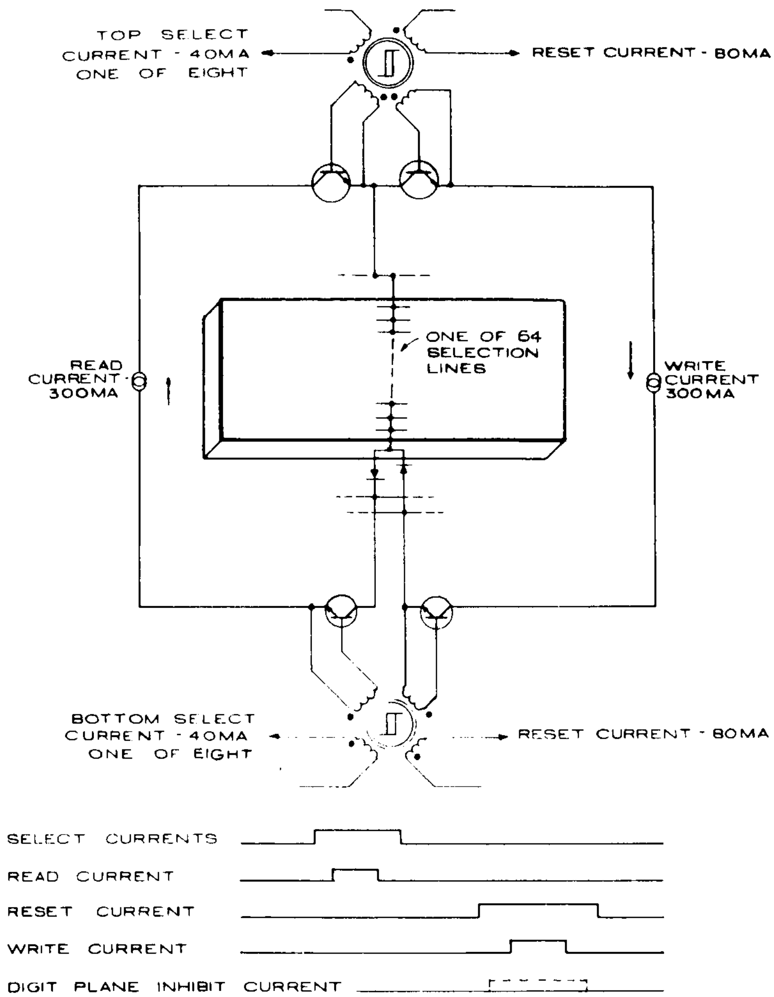


Fig. VI-10 Erasable Memory Current Switching

2 microseconds. Two current drivers with controlled rise times, one for reading and one for writing, are used on each of the two drive select networks. Sixteen more such drivers are used to drive the digit lines which control the writing phase of the memory cycle. Current amplitude is governed by the forward voltage drop across a silicon junction, so that temperature compensation is achieved without any circuit complications for changes in coercive force.

The output signal from the memory cores has an amplitude of about 50 millivolts. Transformer coupling to the sense amplifiers provides common mode noise rejection and a voltage gain of two. The sense amplifiers have a differential first stage operated in the linear, or class A mode. A second stage provides threshold discrimination, rectification, and gating, or strobing. Three reference voltages are generated for the sense amplifiers by a circuit whose temperature characteristics compensate for amplitude and noise changes in the memory.

The sense amplifiers are implemented as integrated circuits in order to realize a number of advantages inherent in single-chip semiconductor circuits. Differential amplifiers pose a special problem in component matching both internal to a single amplifier in achieving balance, and among a group of amplifiers in achieving uniform behavior for common reference voltages. In discrete-component amplifiers a great deal of time and effort go to specifying and selecting matching sets of circuit components. In an integrated circuit, however, balance is readily achieved owing to the extremely close match between transistors on the same silicon chip exposed to the same chemical processes. A similar situation holds for uniformity from one amplifier to another. Within the same batch, amplifiers tend to be very much alike, their differences being easily compensated by external trimming resistors. The efforts expended in specifying the integrated sense amplifier and in a program of reliability testing are little if any more costly than for discrete component amplifiers. Indeed, for a given degree of matching, the cost may be expected to be lower for the integrated circuit. Considerably more complex than the integrated NOR gate used in the AGC, the sense amplifier is used in far less number (32 per computer), so that is is feasible to test and screen each amplifier more comprehensively than a NOR gate, of which there are over 5,000 in each AGC. Performance of the sense amplifiers has been superior, with no spontaneous failures recorded in about a million operational device-hours as of this writing. Their small size is advantageous in obtaining temperature tracking, since it is not difficult to keep them at a temperature close to that of the memory cores. Where sense amplifiers have historically been the "weak link" of computer memory systems, the integrated sense amplifier has already been proven to be at least on a par with the rest of the memory electronics.

The AGC fixed memory is of the transformer type and was developed at MIT.[5, 10] It is designated a "core rope" memory owing to the physical resemblance

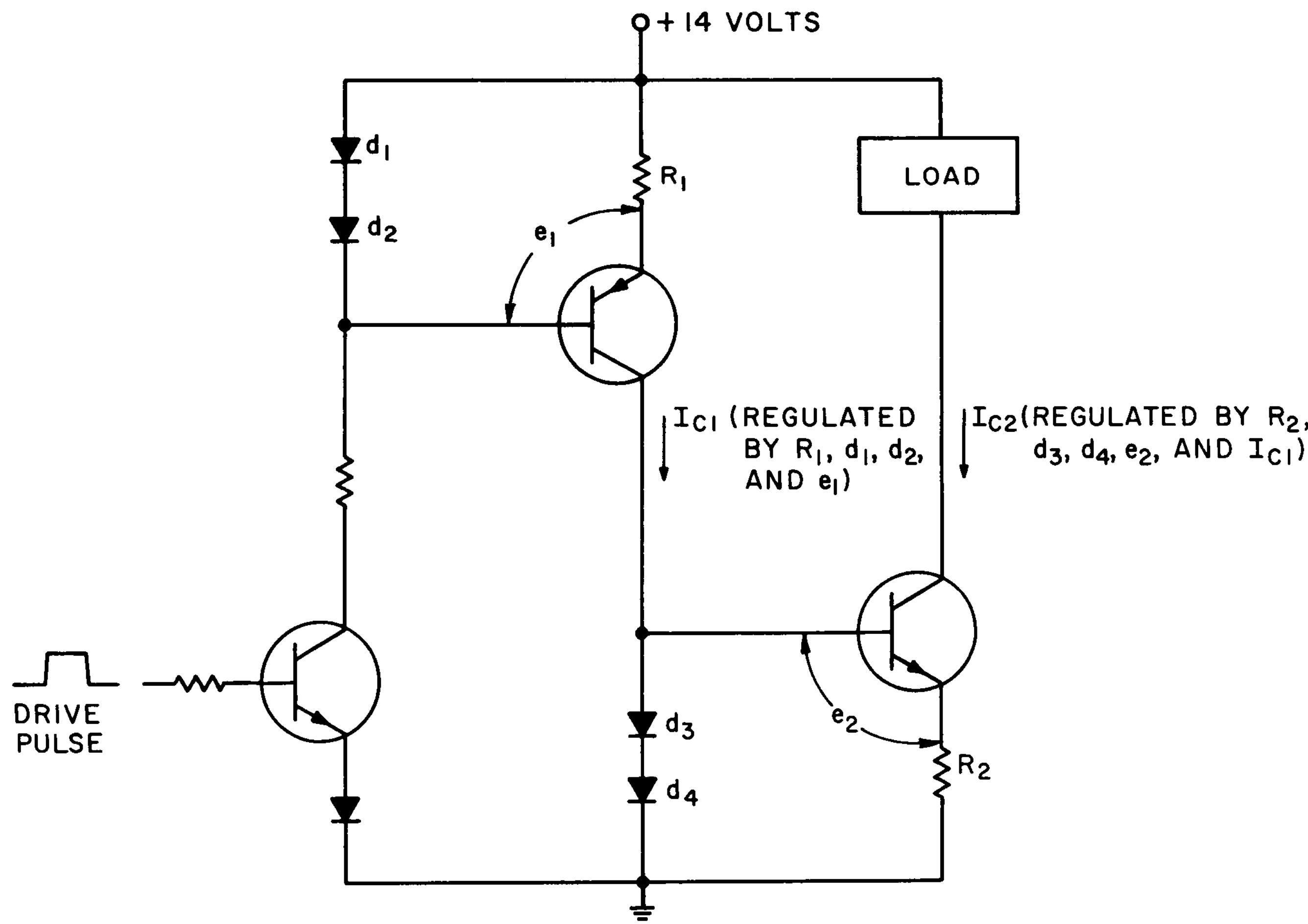


Fig. VI-11 Regulated Pulsed Current Driver Circuit

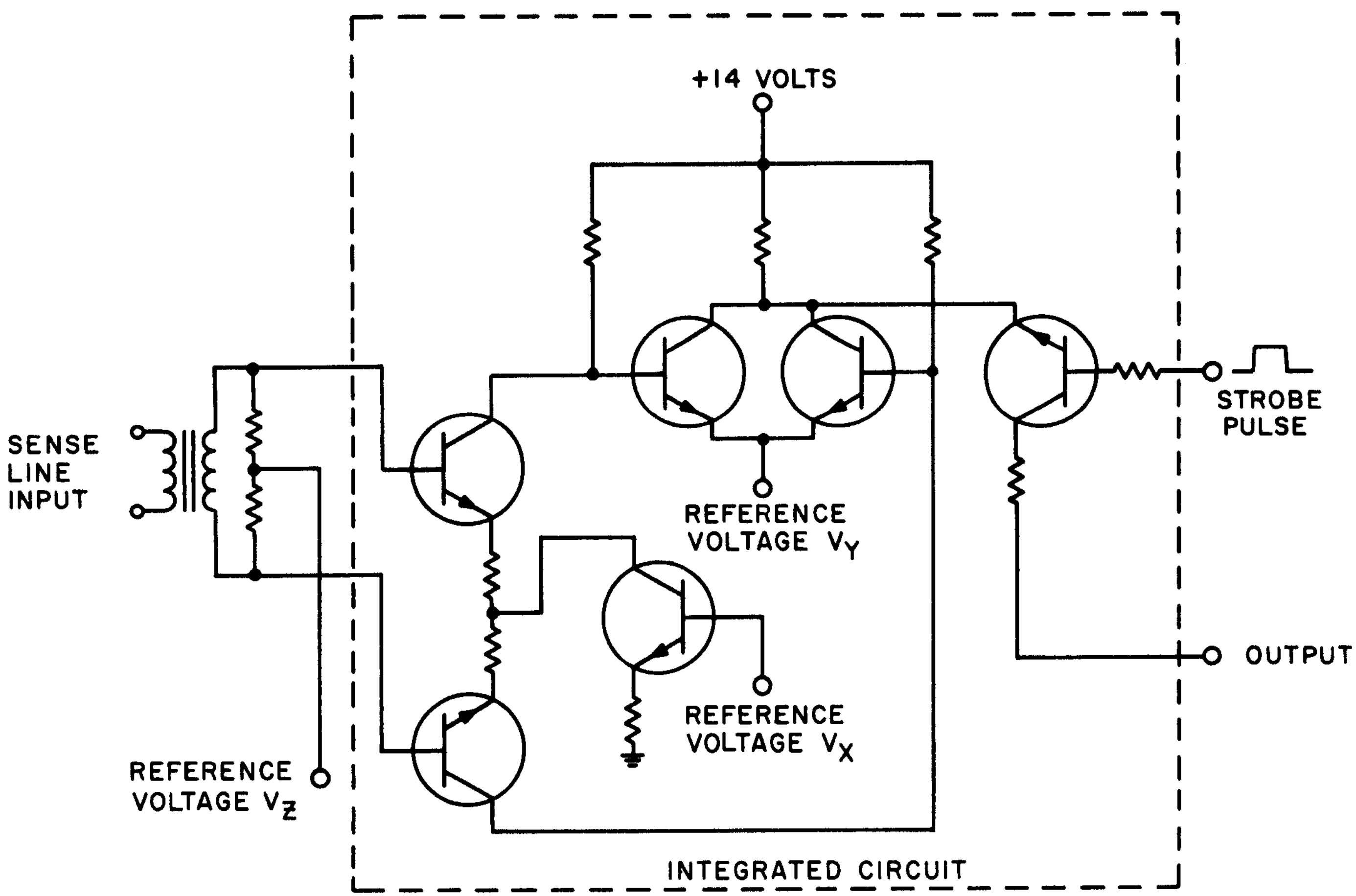


Fig. VI-12 Sense Amplifier Circuit

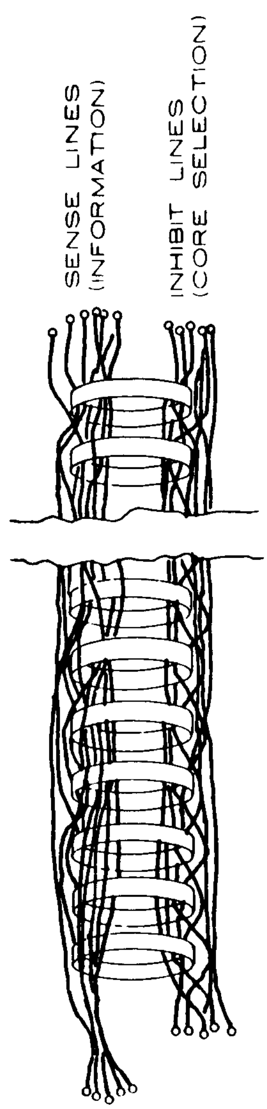


Fig VI-13 Inhibit and Sense Lines through a Rope - Conceptual Drawing

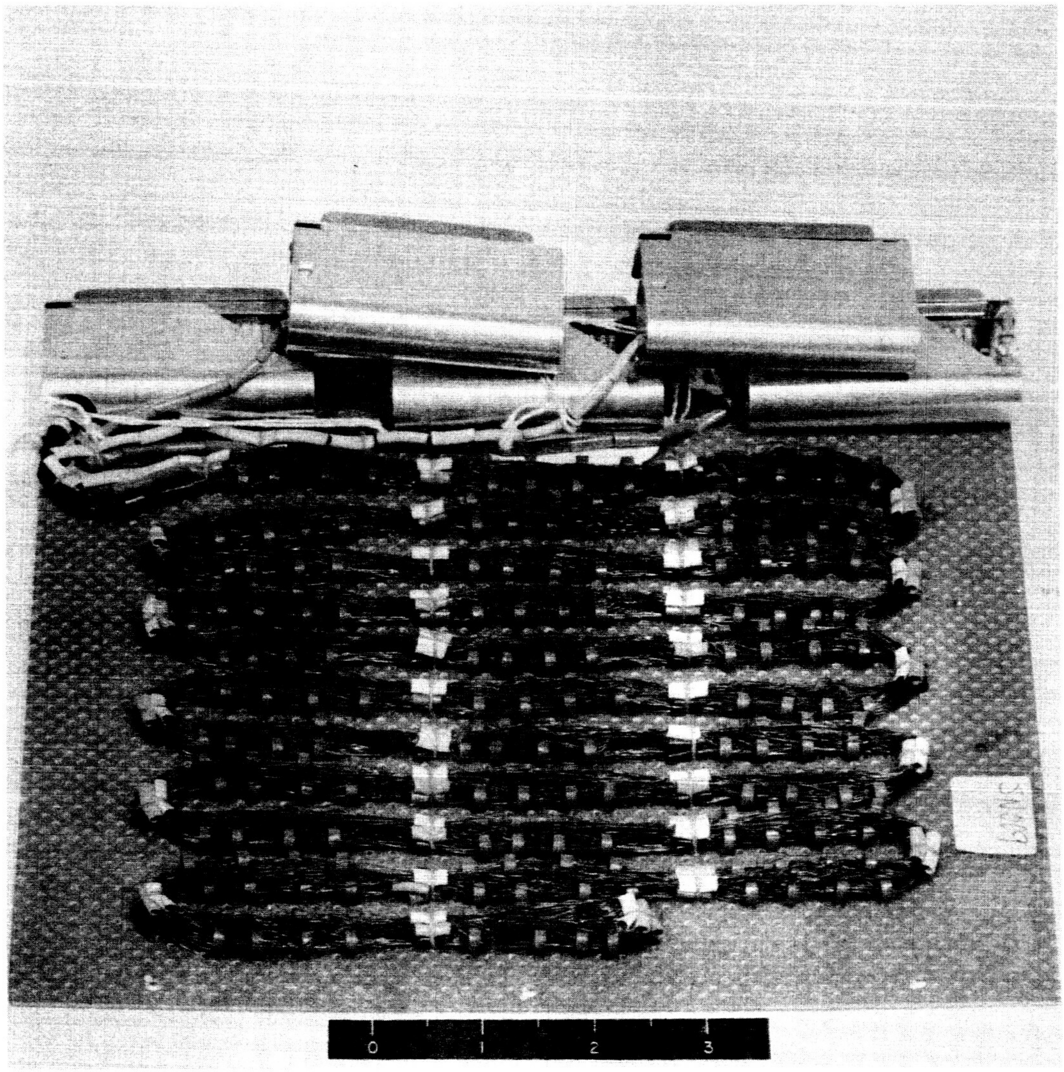


Fig. VI-14 Early Model of Core Rope

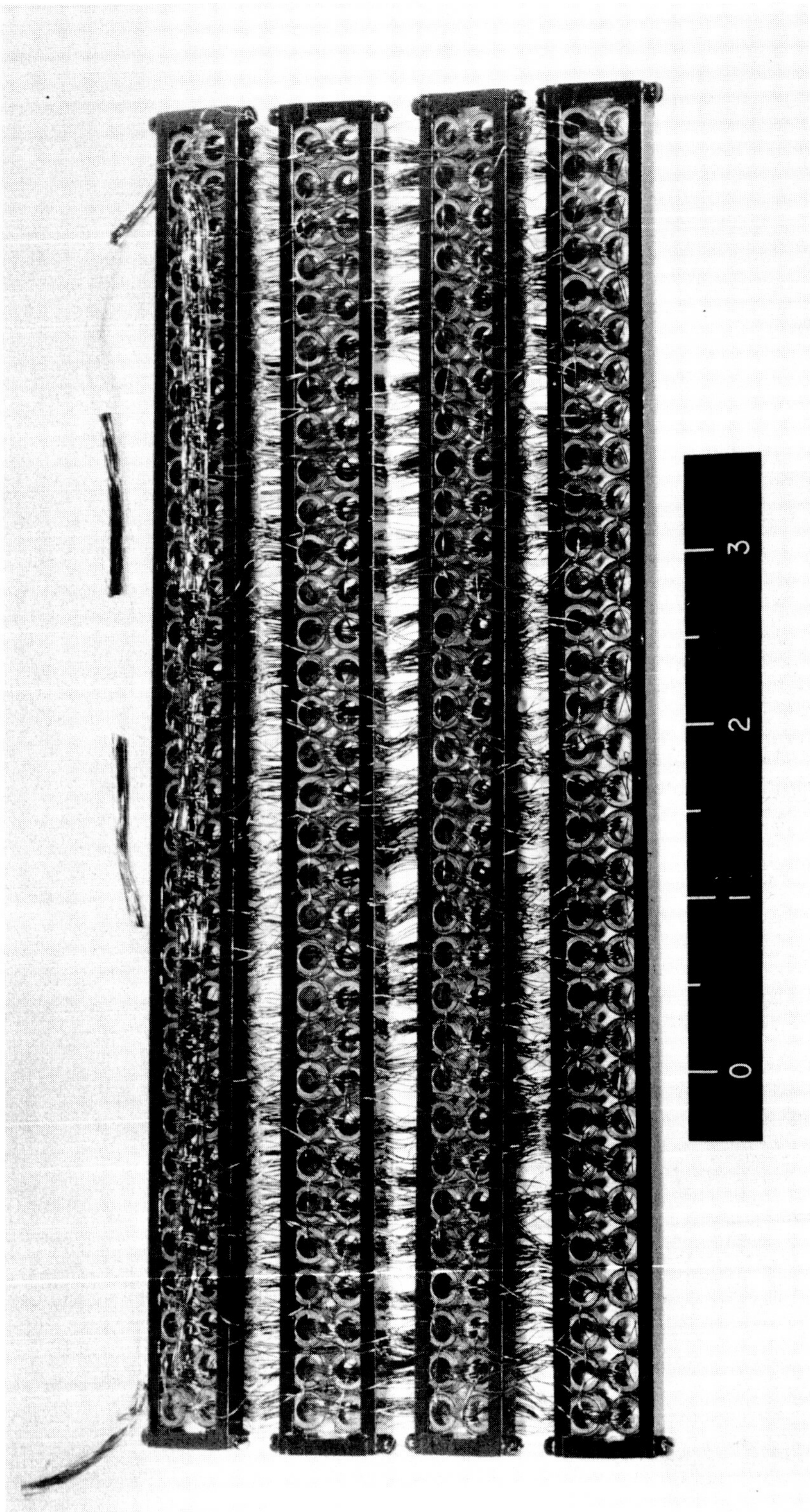


Fig. VI-15 Prototype Core Rope, 1963

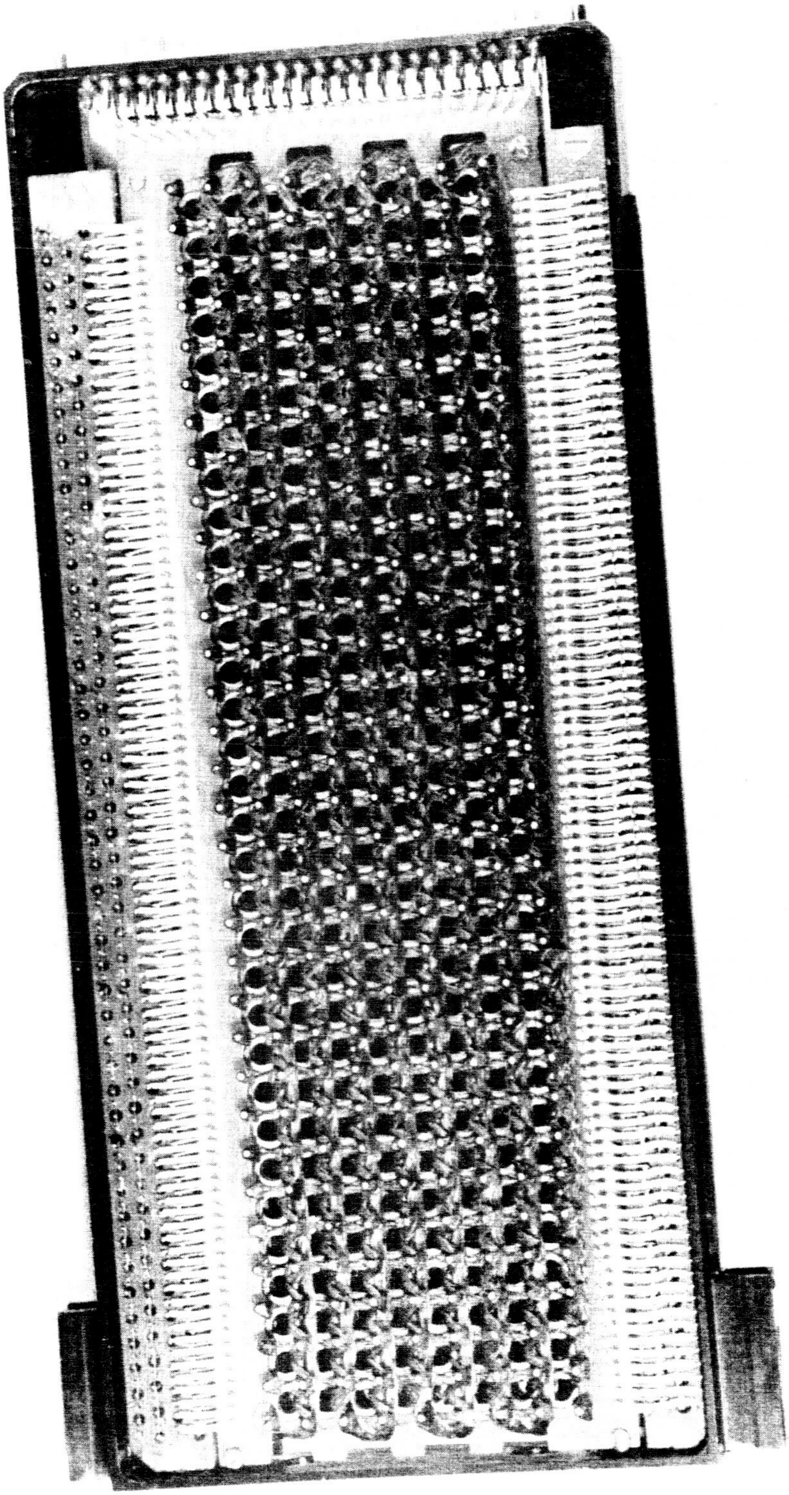


Fig. VI-16 AGC Core Rope Module

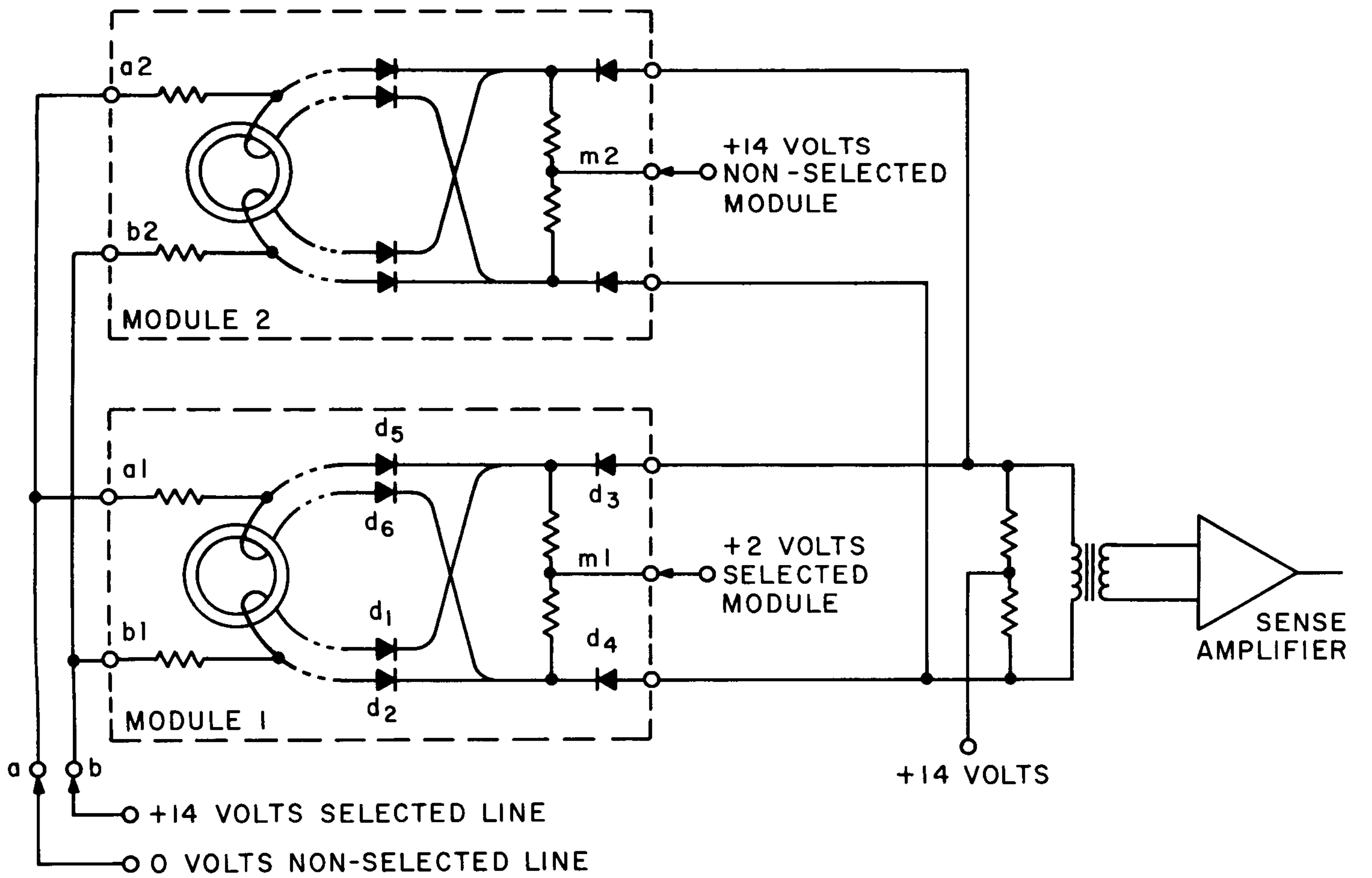


Fig. VI-17 Simplified Rope Sense Line Switching Diagram of early models to lengths of rope. Incorporated into its wiring structure is an address decoding property, because of which its cycle time is not as short as that of some other transformer memories whose address decoding is external. The resulting bit density is extremely high: approximately 1500 bits per cubic inch, (or about 100 bits per cubic centimeter) including all driving and sensing electronics, interconnections, and packaging hardware. This high density of storage is achieved by "storing" a large number of bits in each magnetic core. A stored bit is a one whenever a sense wire threads a core, and is a zero whenever it fails to thread a core. The total number of bits is the number of cores multiplied by the number of sense lines having a chance to thread the cores. The AGC's memory is composed of six modules. Each module contains 512 cores and 192 sense lines and hence contains 192 x 512 = 98,304 bits of information. This information is permanently wired; once the module has been manufactured, not a single bit can be changed, either intentionally or unintentionally, except by physical destruction or by failure of one or more of a number of semiconductor diodes whose functions are described below.

In the operation of the rope memory, a core is switched in a module, thus inducing a voltage drop in every sense line which threads the core. Only one word is read at a time, so that of the 192 sense lines, only 16 are connected to the sense amplifiers to detect which have voltage drops and hence store one's. Thus it is that each core stores 12 words; and within each module a switching network is included in order to transmit no more than one of the 12 to the module's output terminals. The principle of the switching network is illustrated in Fig. VI-17. It consists of diodes and resistors connected so as to block the sense line's output when sense line diodes are reverse-biased as in the case of d_5 and d_6, and to transmit it when the sense line diodes are forward biased, as in the case of d_1 and d_2. A second-stage switch composed of d_3 and d_4 is used to select one of six modules' outputs to be transmitted to the sense amplifiers. Only the selected line in the selected module is transmitted; all others are blocked by one or two sets of reverse-biased diodes. All of these selection diodes are physically located in the rope modules to minimize the number of terminals necessary for each module. The application of selection voltages to the line and module select terminals is a part of the address decoding which is external to the rope; the balance is internal. The means by which a single core in a module is caused to switch is as follows: First, a switching current is applied, which attempts to set 128 cores. Four such current lines serve a 512 core module. Second, an inhibit current is simultaneously applied to either the first half or the second half of each group of 128 cores. Two inhibit lines exist for this purpose. Third, another inhibit current is simultaneously applied to either the first or the second half of each half-group. Two more inhibit lines exist for this purpose. Fourth, six more pairs of inhibit lines exist for the purpose of reducing the uninhibited core groups successively by halves until only one core is left uninhibited. One member of each inhibit pair carries current at a time; there are 8 pairs in all to select among 2^7 cores, of which 7 pairs correspond to the 7 low order address bits. The eighth pair is logically redundant, being selected by the parity of the address. The redundancy is used to reduce the amount of current required in each inhibit line. After the selected core has switched, a reset current is passed through all cores. Only the core which was just set will change state, and the sense amplifiers may be gated on during either the set or the reset part of the cycle to read information out of the memory. The noise level during reset is lower than during set for a number of reasons, but the access time, which is the time it takes to read the memory after the address is available, is longer that way. Both ways have been used in the AGC; the newer design uses the longer access time and produces the address earlier to make up for it.

INTERFACE METHODS

General - Information transfer between the AGC and its environment occupies a substantial fraction of the computer's hardware and its time budget. An attempt has been made to minimize the number of different types of circuits involved. This minimizes engineering effort and makes the computer more easily produced and tested. The impact of this design philosophy on the system has been substantial, and it came about only by active cooperation between subsystem design and system integration groups. The nature of information handled through the interfaces is varied. In some cases computer words are transferred bodily into and out of the computer. Prelaunch and in-flight radio links are maintained between the computer and ground control. Owing to the great difference in data rates between up and down directions, the mechanizations differ considerably. The down link operates at a relatively high rate (50 AGC words or 800 bits per second) and is made so as to occupy a minimum of the computer's time budget. The circuit is relatively expensive. It serializes a word stored in parallel in a flip-flop register and, upon command, sends the bits in a burst to the central timing system of the spacecraft. The up link uses memory cycles to effect a serial to parallel conversion of data. Each bit received requires a memory cycle; a maximum of 160 bits per second are allowed. The cost in equipment is small. The same procedure used in the up link for serial to parallel conversion is used for whole word transfers out of the computer to digital spacecraft display units. It is also used to accept data from the radar measurement subsystem and can be used if desired to communicate between the two AGC's in the command module and the LEM.

Incremental information transfer is similar to serial information transfer in that a sequence of pulses is transmitted over a single channel. It differs in that each pulse represents the same value, or weight, as opposed to serial transfer, where two adjacent pulses differ in weight by a factor of two, and where the concept of positional

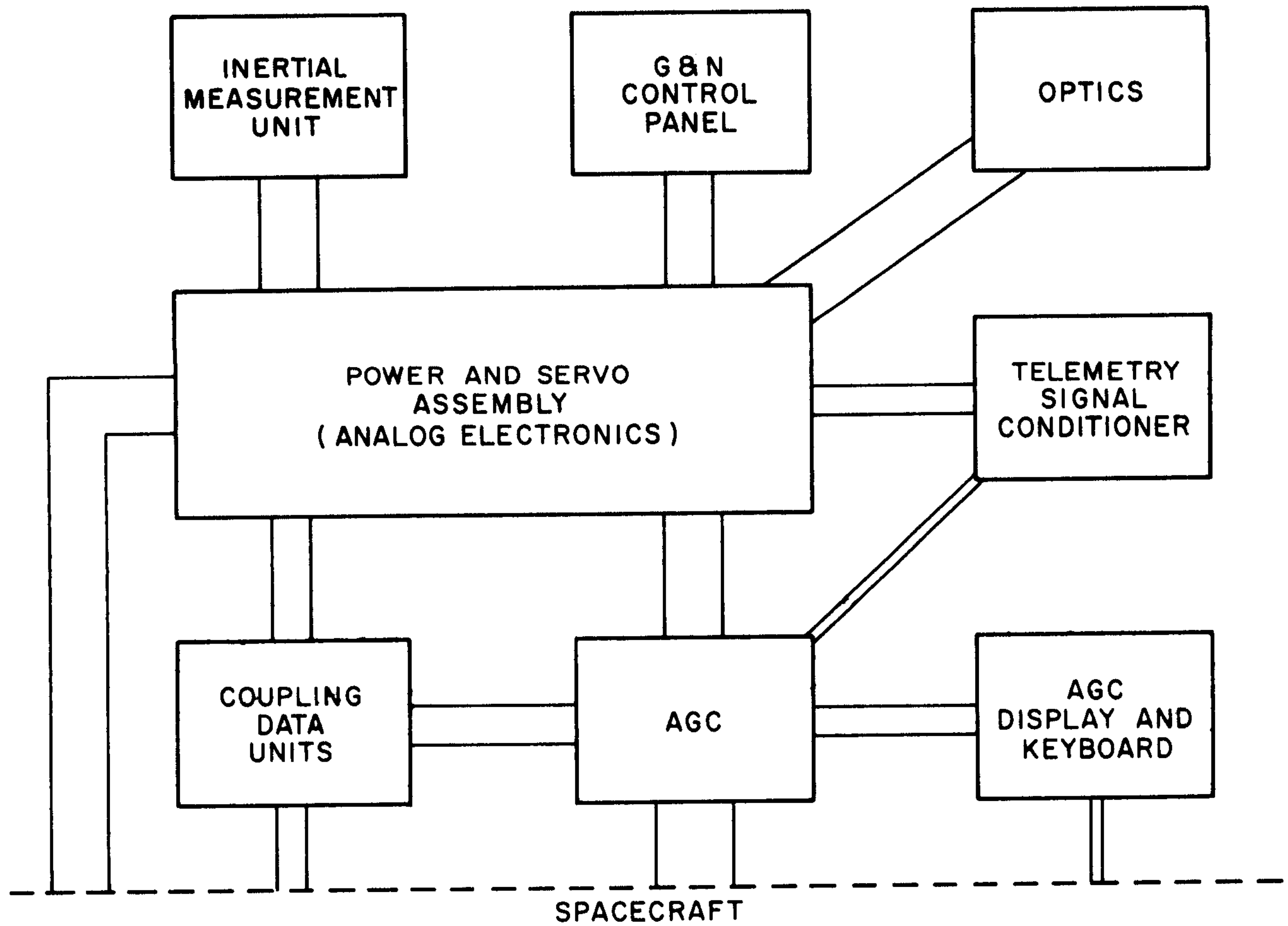


Fig. VI-18 Simplified G & N Interconnection Diagram notation is employed. An incremental receiver counts pulses to form a word, where a serial receiver shifts pulses to form a word. Incremental information transfer was adopted as a means of analog data transmission in order to maintain high precision and standardization. In the conversion of gimbal angles and optics angles, an intermediate transformation to incremental form is made in the Coupling Data Units (CDU's), the inputs to which are electrical resolvers. The alternative to having this extra conversion is to measure the time difference between zero crossings of resolver outputs, to do which may involve equally "expensive" hardware. The Apollo accelerometers are incremental by nature, producing a pulse output to the computer for each unit change in velocity. Incremental transfer is also used for angle commands from the computer to the gyros and the CDU's, and for thrust control and certain display functions in the spacecraft. Pulses are sent in groups or "bursts" at a fixed rate. Pulse rate multipliers would be required in order to send smooth, continuous pulse trains, and these are more costly in equipment.

Discrete signals are individual or small groups of binary digits which give commands or feedback for discrete actions, such as switch closures, mission phase changes, jet firings, display initiations and many other similar controlled events. The display portion of the computer communicates with the computer proper by discrete signals in groups which carry encoded information. Serial transmission might be suitable for this communication, but would be costlier owing to the small number of bits involved.

The computer is the primary source of timing signals for all spacecraft systems; and within the guidance and navigation system it furnishes in the neighborhood of twenty time pulse signals to various subsystems.

The development of the Apollo Guidance system has followed a number of principles which reflect experience gained in previous missile-borne control systems: electrical isolation and asynchronism. Electrical isolation is an important point which has both electrical and logical implications. The computer is connected to the power supply return at a single place, thus avoiding "ground loops." Isolation of interface signals is accomplished by transformer coupling, by switch closure (relays), or by a high resistance d.c. current signal. Input and output circuits are designed so that no damage can be caused by improper connections at the interface, such as short circuits to ground. The ability to accept asynchronous inputs, i.e., those not related to computer timing signals, is desirable because it affords a design with a minimum of reference to signals outside the computer, and reduces the number of signals across the interface. This principle is particularly important for signals whose functions are to interrupt normal program sequences.

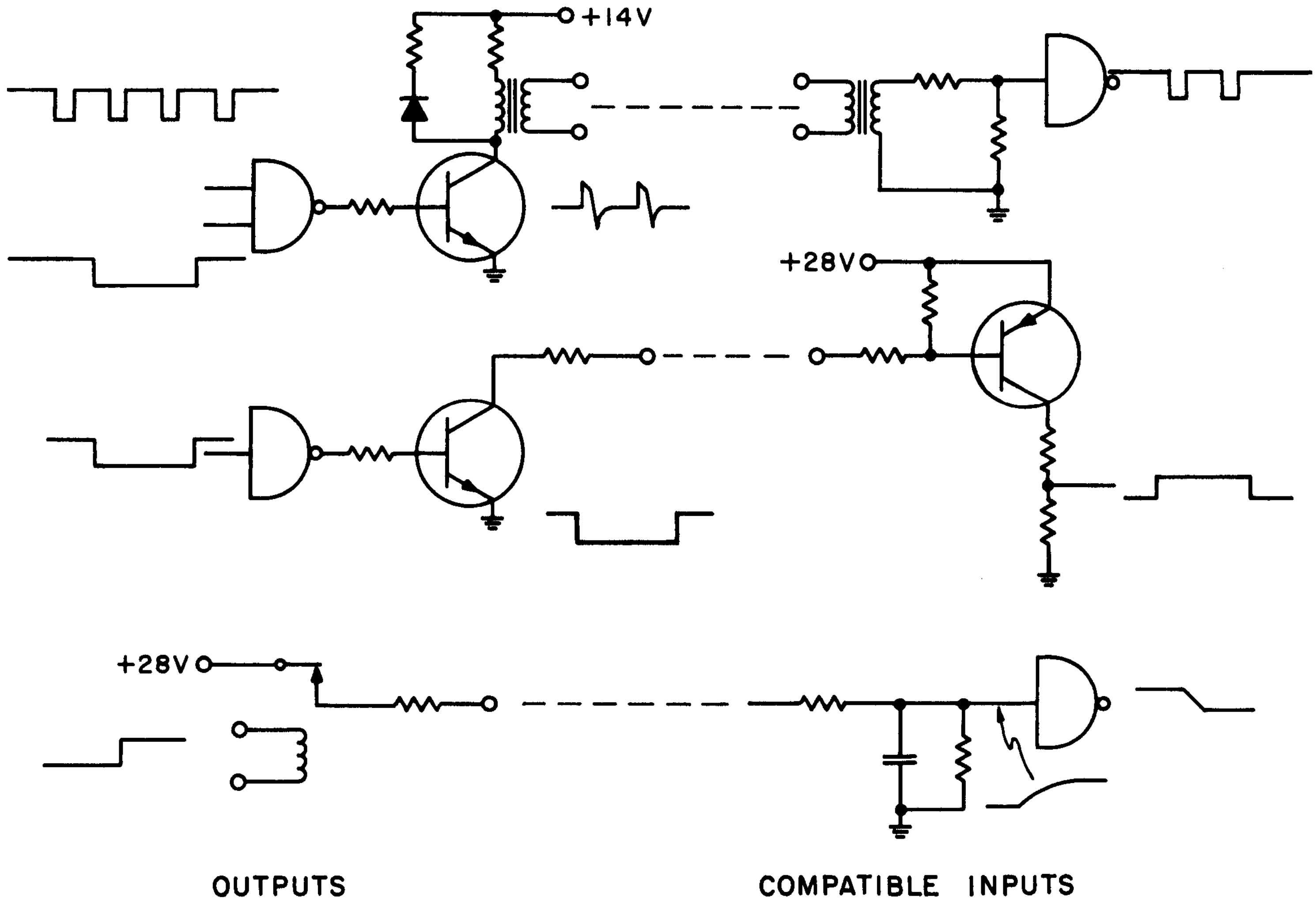


Fig. VI-19 Examples of AGC Interface Circuits

Number of Discrete Inputs	73
Number of Input Pulses (Serial and Incremental)	33
Number of DC Output Discretes	68
Number of Variable Pulsed Outputs (Serial, Incremental, and Discrete)	43
Number of Fixed Pulsed Outputs	10
Number of Connector Wires	365

Table VI-6 AGC Interface Summary

Inputs - Incremental and serial inputs are received in special erasable memory cells called "counter registers." They are made special by the fact that pulses received by the computer cause short interruptions of the program sequence during which one of these registers is interrogated and modified. Just which counter register is involved and what the modification to its contents is are determined by the particular input being responded to. Since there are 29 of these registers, some provision must be made to accommodate simultaneous requests for servicing several counter registers. The circuit which accomplishes this and in addition satisfies the principle of asynchronism is known as the "Counter Priority" circuit. This circuit stores incoming pulses until they can be processed. If more than one request is pending, preference is given to the one for the counter register whose address is lowest. When the "Counter Increment" cycle begins, the priority circuit delivers to the S register the address of the counter register having the outstanding request of highest priority. At about the same time, the choice is made as to how the counter's contents will be modified when they are obtained. This choice is based upon the source of the request. The counter word is shifted if it is one of the serial data receiving counters. If the "one" input received the request, a low order one is added after shifting; if the "zero" input received the request nothing is added. When the register is full, a program interrupt is requested. For a counter which is an incremental receiver, a low order plus one or minus one is added, depending upon whether the positive or negative requesting input was received. Since counter words are in the erasable memory, they are always readily accessible by any program. Each increment or shift requires a memory cycle for its execution, so the aggregate counting rate has to be limited in order to avoid an excessive use of the computer's time budget. In some instances, this requires having a logic circuit between the interface and the priority circuit which prevents the input pulse rate from exceeding a chosen level.

Two types of discrete inputs to the computer are distinguished -- interrupting and non-interrupting, the latter class being much larger than the former. Non-interrupting discrete inputs are signals which can be interrogated by input-output channel instructions. They are mechanized either as d.c. inputs through a filter to a logic gate, or as a.c. signals, transformer coupled to a flip-flop which is reset after interrogation. Interrupting inputs, in addition to appearing where they can be interrogated, announce their presence to the computer's sequence generator by initiating a program interrupt. A priority circuit similar to the Counter Priority circuit buffers the asynchronism of the inputs and resolves ambiguities caused by simultaneous interrupt requests. At the earliest permissible time, the Interrupt Priority circuit forces a transfer of control to a particular address, where the computer program interrogates the appropriate inputs and initiates any necessary action. The original program is then resumed at the point where it was interrupted. Interrupt programs never exceed a few milliseconds in running time.

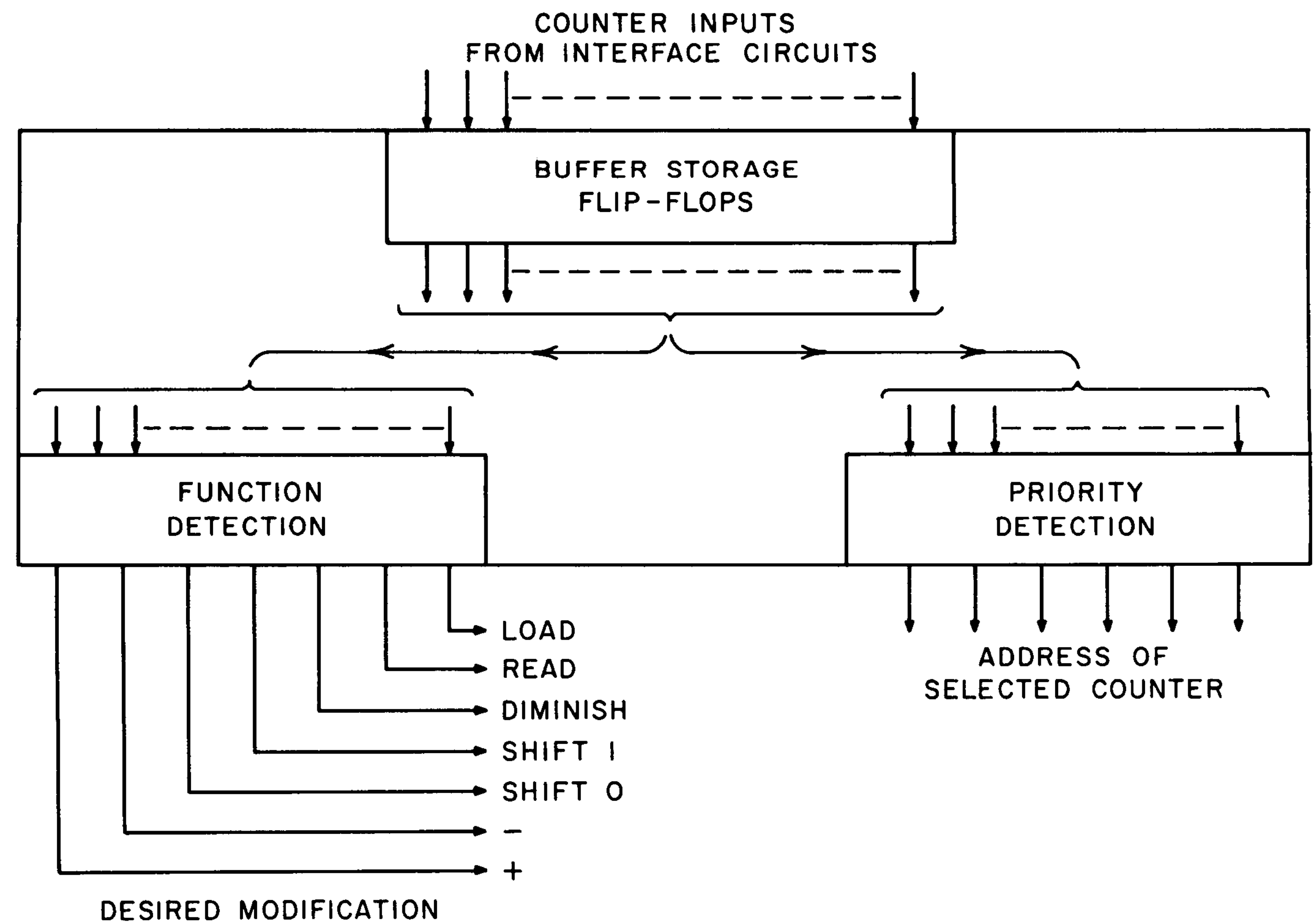


Fig. VI-20 Structure of Counter Priority Circuit

Outputs - With one major exception, the AGC uses its counter register processing facility to make all conversions from parallel words to serial and incremental pulse trains. The exception is the down data link circuit mentioned previously, where a rather costly circuit is used so that the high bit rate will not detract from the time budget. Serial outputs originate from counter registers which contain the word to be transmitted. Fifteen successive shifting requests are applied to the Counter Priority circuit; each time an overflow occurs during the shifting process, a one pulse is sent to a transformer output circuit. If no overflow occurs, a pulse representing a zero is sent to another output. This is a self-timing form of serial transmission, and is fully compatible with the serial input counter circuits.

Incremental transmission is made by placing a number in an output counter register and activating the Counter Priority circuit at a fixed rate of 3200 times per second. Each time the counter is processed, the number in the counter register is diminished by a low order one of such a sign that the register's contents approach zero. An output pulse is generated concurrently each time on one of two lines, depending on the sign of the number in the counter register. When the number has reached zero, the periodic activation of the Counter Priority circuit ceases, and the pulse burst terminates. Bursts of anywhere from one to 16,384 pulses can be generated this way. Two forms of digital-to-incremental-to-analog conversion are used in the Apollo Guidance Equipment. The simpler of the two is used for gyro torquing. During an output pulse burst, a precision current source is gated on, so that an amount of charge proportional to the desired angle change is forced through the torquing element. A single precision element is used for this form of conversion, but external storage is required for the result. In this case the storage is in the mechanical gyro angle. The second form uses a counting flip-flop register and a resistance ladder. The number in the register controls the switching of a set of precision resistors in an operational amplifier network such that the amplifier output is proportional to that number. These ladder networks are located physically in the Coupling Data Units. An incremental form of information transfer from the computer to the Coupling Data Units is used in order to minimize the interface. Several precision elements are needed for this kind of conversion, but no external storage is needed. Thus these analog signals are available as voltages for driving such equipment as attitude displays and steering gimbals for a rocket engine.

Discrete outputs are controlled either directly or indirectly by program. Typically, a discrete output is turned on by placing a one in the proper bit position of an output channel, which sets a flip-flop. If the output is transformer coupled, the flip-flop signal drives a transistor output circuit. For higher power levels, the transistor output circuit drives a relay located in the Display and Keyboard unit, and the relay's contacts are connected to the interface.

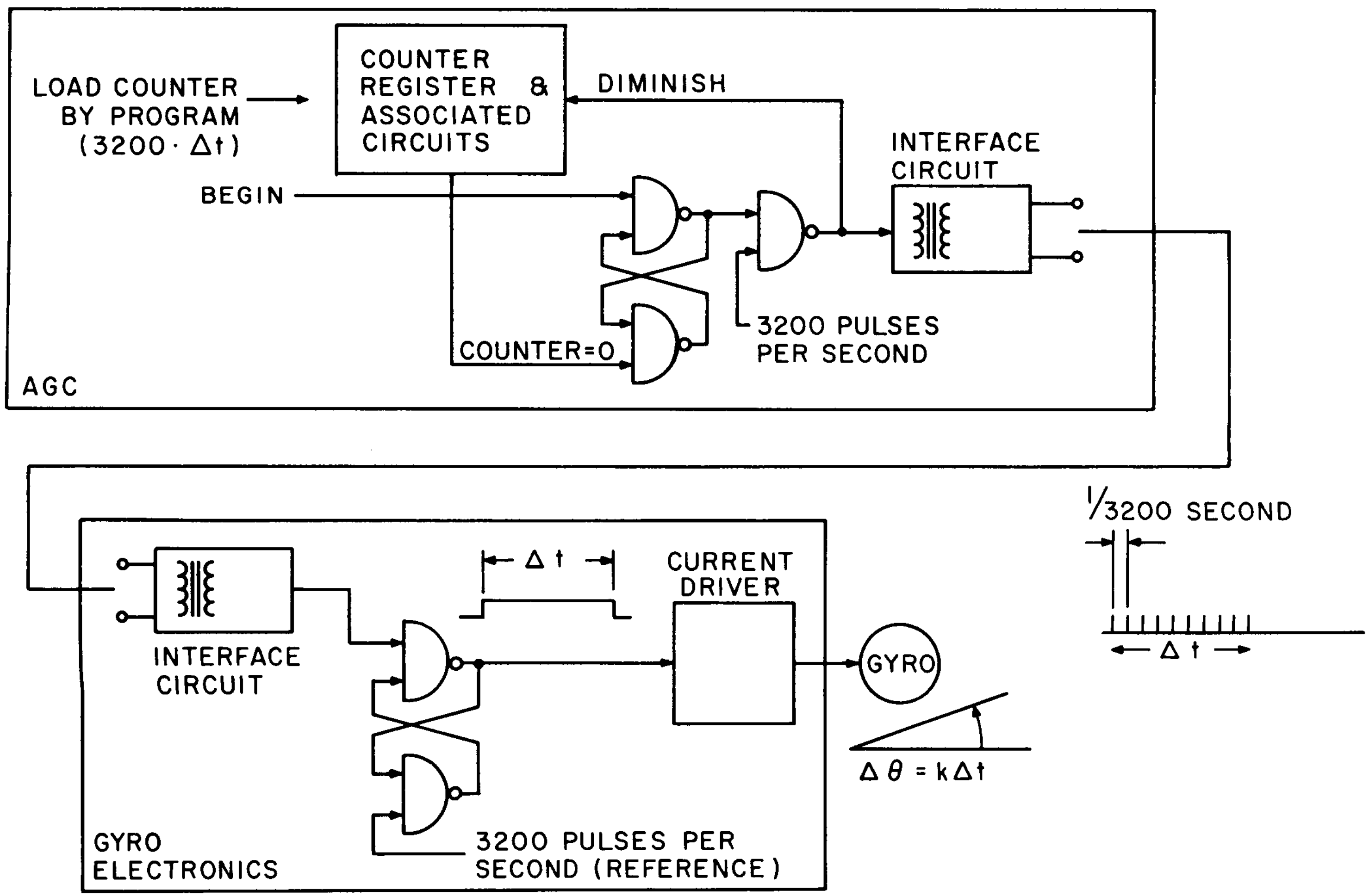


Fig. VI-21 Digital-to-Analog Conversion by TimeDuration of Precision Current

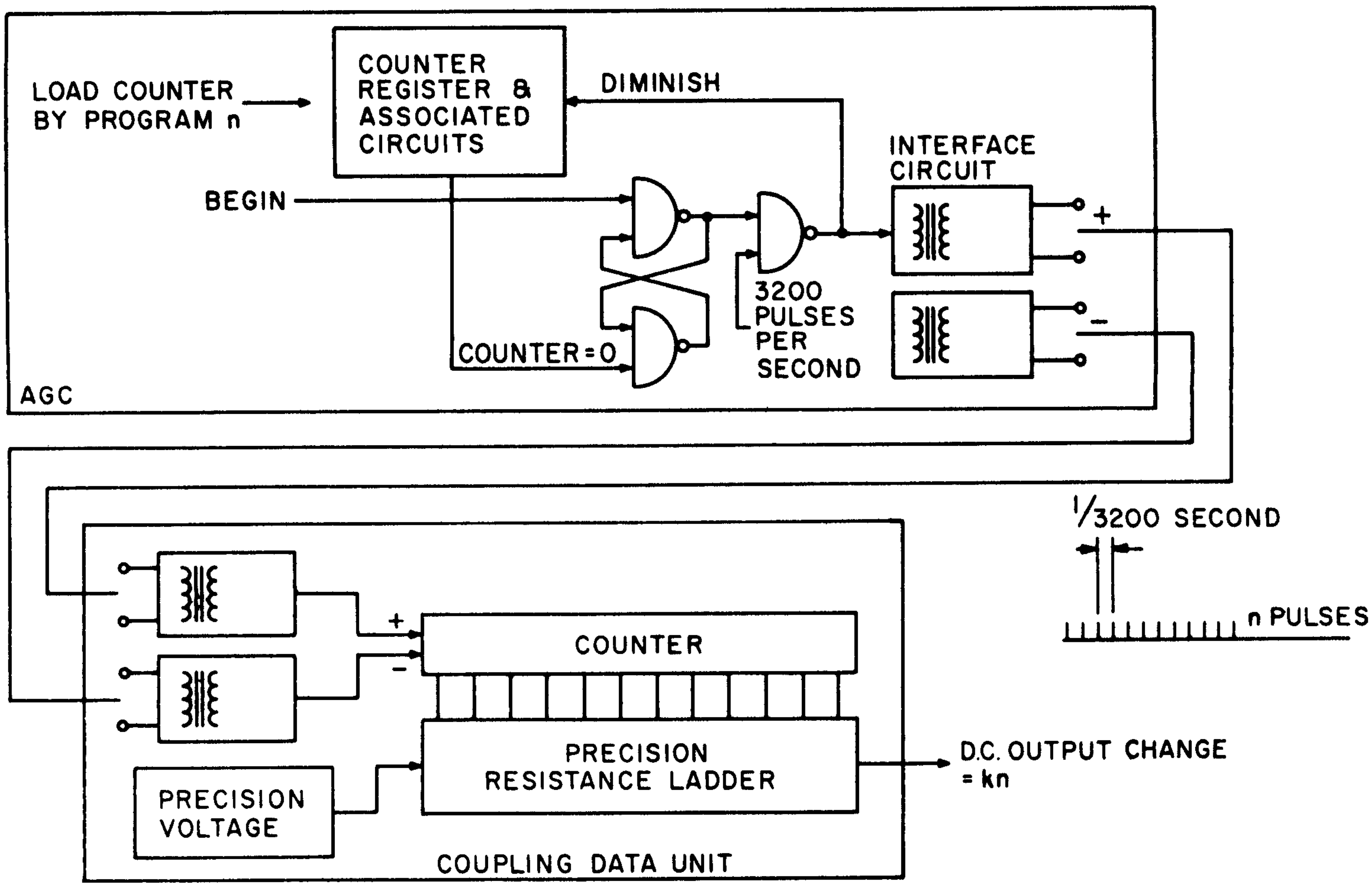


Fig. VI-22 Digital-to-Analog Conversion by Resistance Ladder

Purpose	Serial or Incremental	Nature of Signal	Maximum Rate	Implementation
Digital down data link	S	50 bursts/sec of 40 pulses each	51.2 KC	Special circuit
Digital up data link	S	10 or fewer bursts/sec of 16 pulses each	1.0 KC	Counter increment
Digital cross links	S	10 or fewer bursts/sec of 16 pulses each	3.2 KC	Counter increment
Altitude display	S	2 or fewer bursts/sec of 16 pulses each	3.2 KC	Counter increment
Radar data link	S	10 or fewer bursts/sec of 16 pulses each	3.2 KC	Counter increment
Gimbal and optics angles	I	All rates to maximum	6.4 KC	Counter increment
PIPA velocity increments	I	All rates 0 to maximum	3.2 KC	Counter increment
Gyro torquing	I	Occasional bursts of 0 to 2^{14} pulses	3.2 KC	Counter increment
Engine thrust control	I	Occasional bursts of 0 to 2^{11} pulses	3.2 KC	Counter increment
Engine steering	I	Occasional bursts of 0 to 2^{11} pulses	3.2 KC	Counter increment

Table VI-7 Partial List of Serial and Incremental Interface Characteristics

Fixed outputs are steady pulse trains which are used to synchronize other equipment. Nearly all of these are transformer coupled, and are generated simply by driving the transformer circuit with the appropriate pulse signal.

Display and Keyboard - The Display and Keyboard unit (abbreviated DSKY) is in some respects like an integral computer part, yet it is operated with the same interface circuits used for connection to other subsystems and systems. Since it serves as the channel for human communication with the computer, it needs a rather high peak data rate without either being very large in itself or having overly many wires between itself and the computer located a few meters distant along a cable.

The principal part of the display is the set of three light registers, each containing five decimal digits composed of electroluminescent segmented numerical lights. Five digits are used so that an AGC word of 15 bits can be displayed in one light register by five octal digits. No fewer than three registers are used because of the frequent need to display the three components of a vector. No more than three are used because the extra space and weight penalty cannot be justified. In addition to the numerical lights, a sign position is included in each light register. The convention is used that when a sign appears, the number is to be interpreted as decimal. Otherwise it is taken to be octal. Electroluminescent lights are small and easy to read, and require relatively little power. They are driven by an 800 cycle alternating voltage supply, switched by miniature latching relays. These have a substantial power advantage over the equivalent electronic circuitry. The contacts, well suited to the high a.c. voltage, are used for decoding, while the latching action provides a storage function. Both latching and non-latching relays are used for interfaces with other subsystems and systems, and are located in the DSKY's where they share driving circuits with the light register relays. A double-ended selection matrix is used for actuating the relays. This is organized so that one of fourteen groups of eleven relays is set at a time. Eleven signals are required from the computer to govern the configuration of the eleven relays in a group, and four more bits are used to select one group out of fourteen, making fifteen bits. The fact that this is the size of an AGC word is not entirely coincidental. This arrangement allows one word in the DSKY output channel to control enough relays to light two numbers in a light register and one stroke of a sign.

Digits are entered into the computer from a keyboard of nineteen push buttons including the ten decimal digits, plus and minus, and a number of auxiliary items. No more than one key is depressed at a time, so the nineteen key functions can be encoded into five signals. This is done by a diode matrix in the keyboard section of the DSKY. In order that each key depression can be quickly gathered and interpreted, the key code inputs to the computer are of the interrupting type. The key input channel is interrogated by the keyboard interrupt program, which also makes a request to the computer's executive program to process the character at the earliest opportunity. A

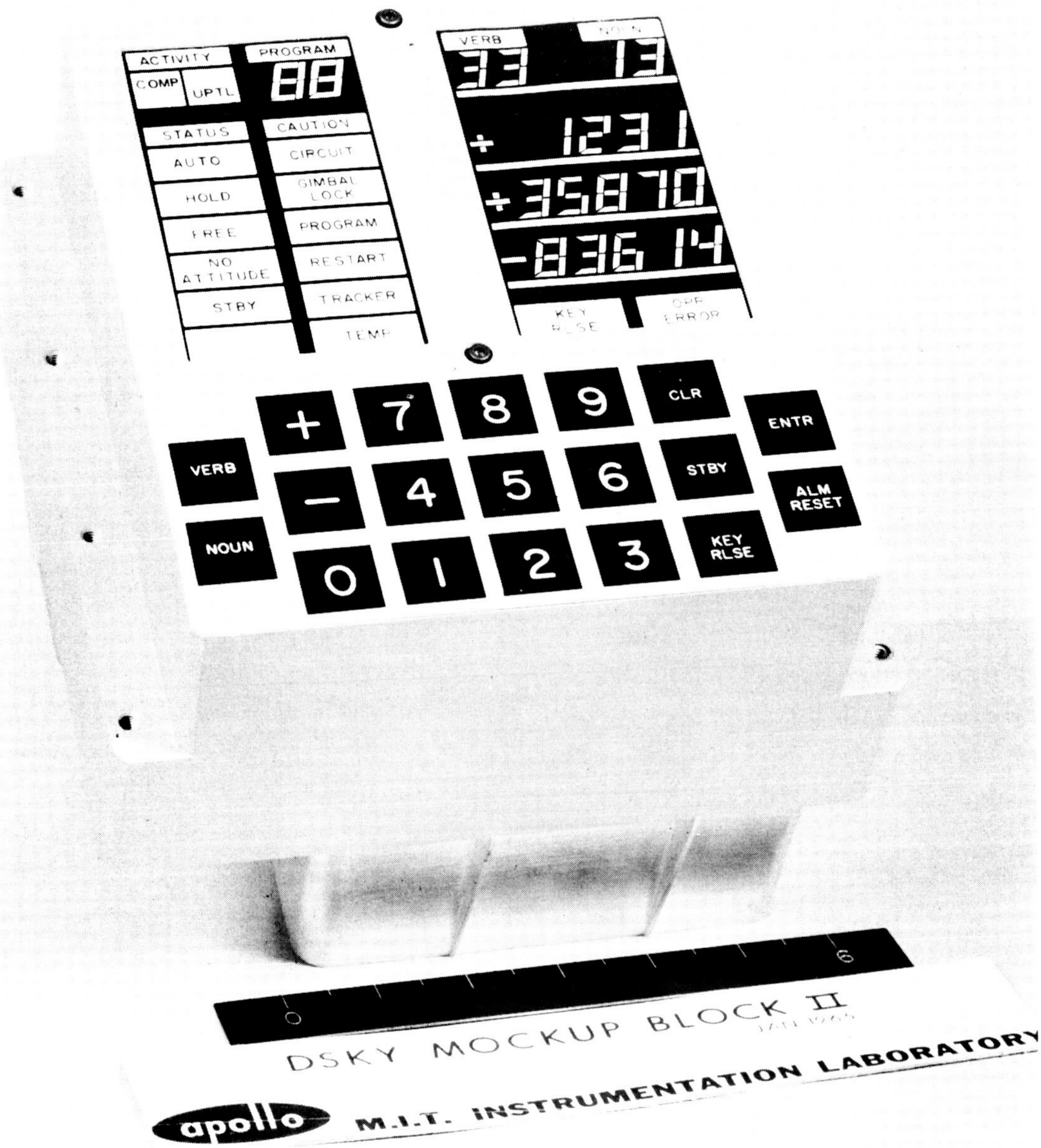


Fig. VI-23 AGC Display and Keyboard

"trap" circuit logically differentiates the leading edge of the key code signal so that no more than one interrupt is made for each depression of a key. This trap circuit is reset by a signal through all of the normally closed contacts of the keys; the reset signal is present only when all keys are released. Problems from key contact bounce are avoided by having flip-flops at the input channel to receive the key code signals.

In addition to the three light registers, the display has other digit displays labeled verb, noun, and program. The keyboard has keys labeled verb, noun, enter, and clear as well as three others. These are used to enable concise yet flexible communication between man and computer. Commands and requests are made in the form of sentences each with an object and an action, such as "display velocity" or "load desired angle." The first is typical of a command from man to machine; the second is typical of a request from machine to man. The DSKY is designed to transmit such simple commands and requests made up of a limited vocabulary of 63 actions, or "verbs" and 63 objects, or "nouns." These verbs and nouns are, of course, displayed by number rather than by written word; so it is necessary to learn them or else to have a reference document at hand. To command the computer, the operator depresses the verb key followed by two octal digit keys. This enters the desired verb into the computer, where it is stored and also sent back to the DSKY to be displayed in the verb lights. The operator next enters the desired noun in similar fashion using the noun key, and it is displayed in the noun lights. When the verb and noun are specified, the enter key is depressed, whereupon the computer begins to take action on the command.

When the computer requests action from the operator, a verb and a noun are displayed in the lights, and a relay is closed which causes the verb and noun lights to flash on and off so as to attract the operator's attention and inform him that the verb and noun are of computer origin. To illustrate the usefulness of the requesting mode, consider the procedure for loading a set of three desired angles for the IMU gimbals. The operator keys in the verb and noun numbers for "load 3 components, IMU gimbal angles." When the enter key is depressed, the computer requests that the first angle be keyed in by flashing and changing the verb and noun lights to read "load first component, IMU gimbal angles." Now the operator keys in the angle digits, and as he does so the digits appear in light register number one. When all five digits have been keyed, the enter key is depressed. The verb and noun lights continue to flash, but call for the second component; and when it has been keyed and entered, they call for the third component. When the third component has been entered, the flashing stops, indicating that all requests have been responded to. If a mistake in keying is observed, the clear key allows the operator to change any of the three angles previously keyed until the third angle has been entered.

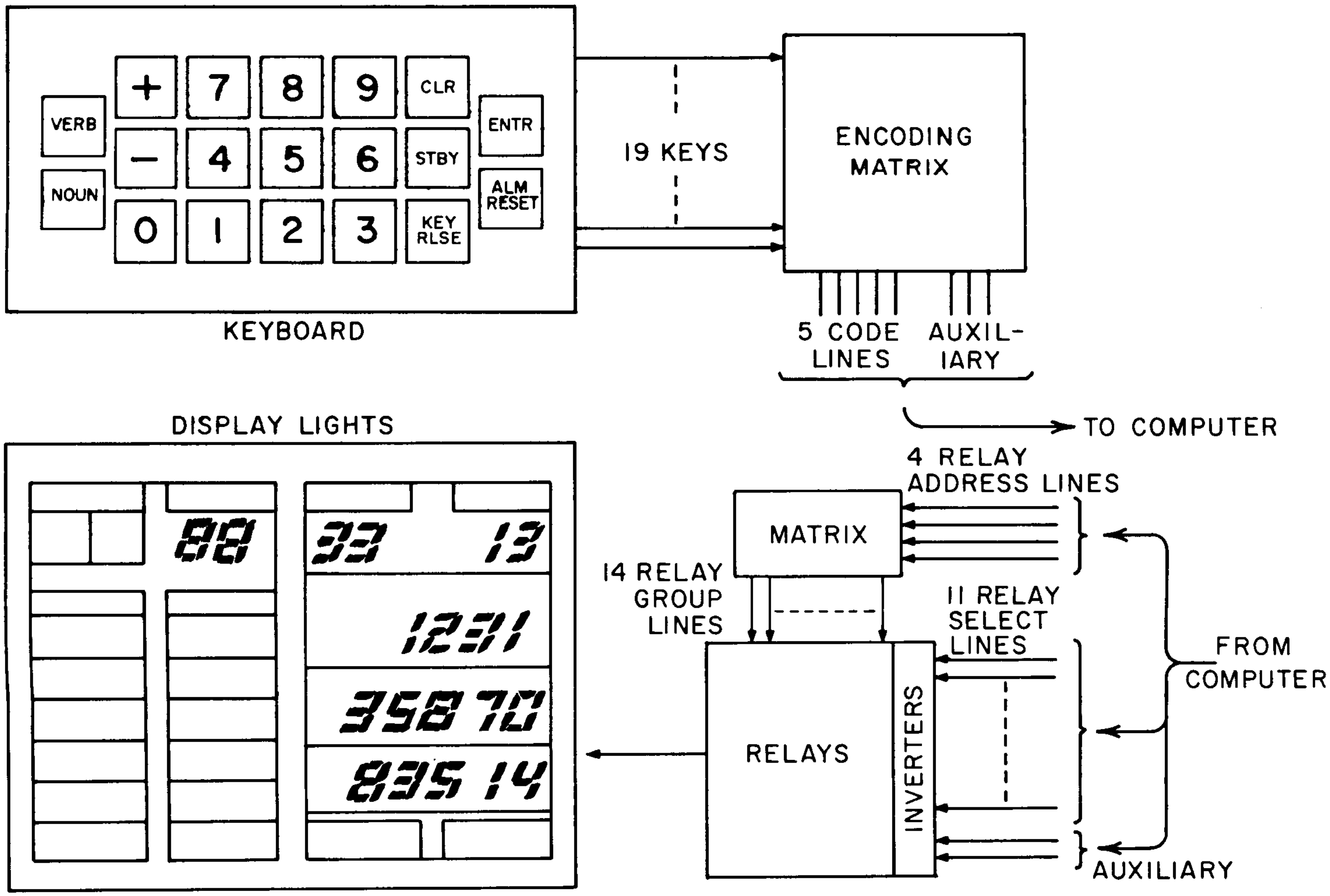


Fig. VI-24 Block Diagram of AGC Display and Keyboard Unit

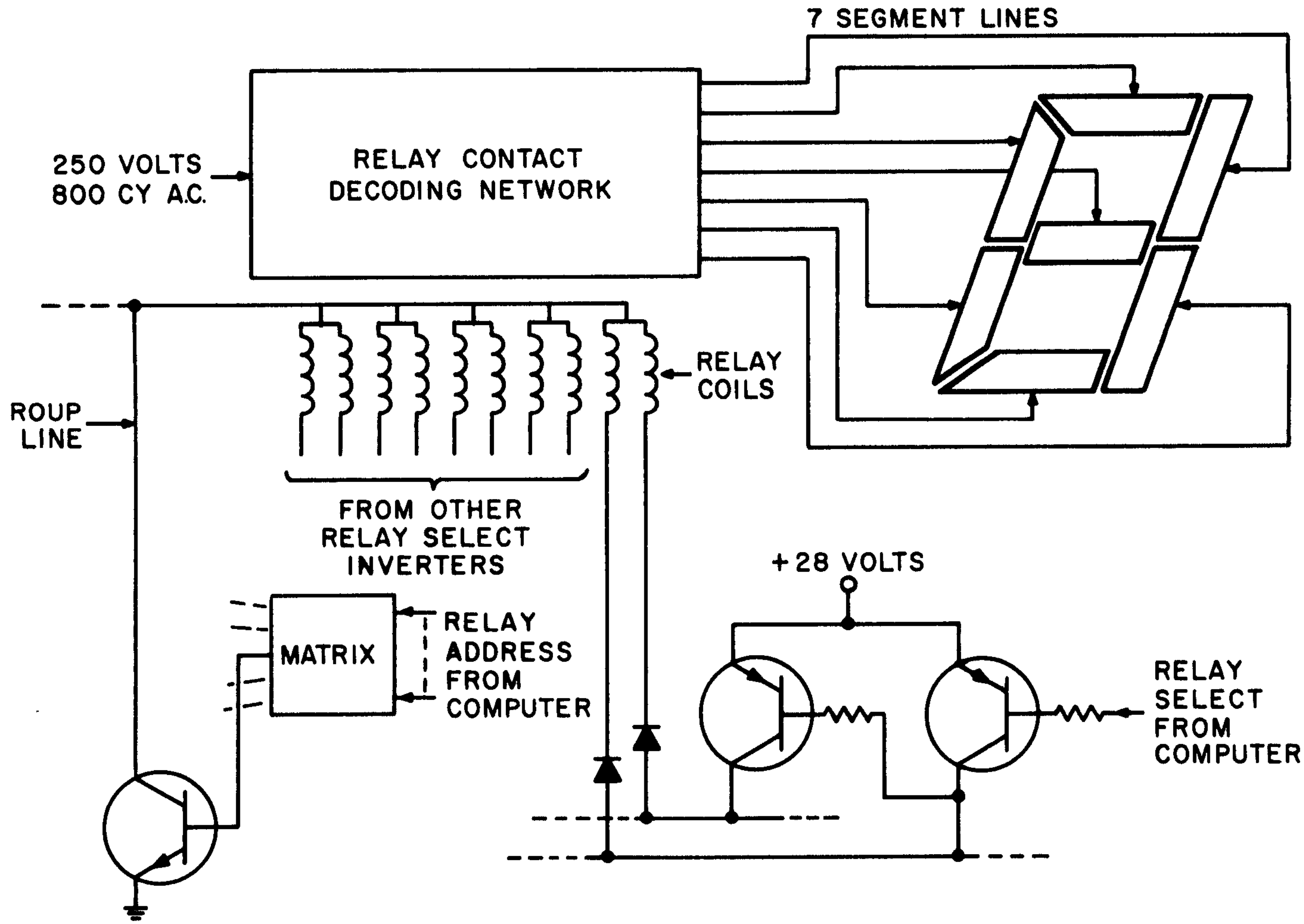


Fig. VI-25 Display Circuit Scheme

Program lights give the operator an indication showing what major programs the computer is running. Additional features of the DSKY are discrete alarm and condition lamps, a condition light reset key, and a key with which the operator relinquishes his use of the display lights to the computer. The last named function is useful because it is not always known a priori whether the operator's command has a higher priority than the computer's request. This is resolved by having the operator make the decision. If a keyboard entry sequence is in progress at a time when the computer program has a request or a result to display, a condition lamp is turned on to notify the operator of the fact. When he is ready to have the computer use the display, he need only depress this release key.

UTILITY PROGRAMS (11, 12, 13)

Interpreter - Most of the AGC programs relevant to guidance and navigation are written in a parenthesis-free pseudocode notation for economy of storage. In a short word computer, such a notation is especially valuable, for it permits up to 32,768 addresses to be accessible in a single word without sacrificing efficiency in program storage. This notation is encoded and stored in the AGC as a list of data words. An AGC program called the "interpreter," translates this list into a sequence of subroutine linkages which result in the execution of the pseudocode program. A pseudocode program consists of lists called "equations." Each equation consists of a string of operators followed by a string of addresses to be used by the operators. Two operators are stored in an AGC word, each one being 7 bits long. A partial list of operators appears in Table VI-8.

Use of the interpreter accomplishes a saving in instruction storage over programs generated in an automatic compiler, and it affords the programmer a rapid and concise form of program expression which liberates him from the time consuming job of programming in basic machine language. In so doing it expands the instruction set into a comprehensive mathematical language accommodating matrix and vector operations upon numbers of 28 bits and sign. This is made possible at the modest cost of a few hundred words of program storage and the cost of about an order of magnitude in execution time over comparable long word computers.

Executive - All AGC programs operate under control of the Executive routine except those which are executed in the interrupt mode. Executive controlled programs are called "jobs" as distinct from so-called "tasks," which are controlled by the Waitlist routine and completed during interrupt time. The functions of the Executive are to control priority of jobs, to permit time sharing of erasable storage, and to maintain a display discrete signal denoting "Computer Activity."

Jobs are usually initiated during interrupt by a task program or a keyboard program. The job is specified by its starting address and another number which gives

DP = double precision

Operator	Average Execution Time, Milliseconds
DP Add	0.66
DP Subtract	0.66
DP Multiply	1.1
DP Divide	2.5
DP Sine	5.6
DP Cosine	5.8
DP Arc Sine	9.3
DP Arc Cosine	9.1
DP Square Root	1.9
DP Square	0.76
DP Vector Add	0.92
DP Vector Subtract	0.92
DP Vector x Matrix	9.0
DP Matrix x Vector	9.0
DP Vector x Scaler	3.3
DP Vector Cross Product	5.0
DP Vector Dot Product	3.1

Table VI-8 Partial List of Interpretive Operators it a priority ranking. As the job runs, it periodically checks to see if another job of higher priority is waiting to be executed. If so, control is transferred away until the first job again becomes the one with highest priority. No more than 20 milliseconds may elapse between these periodic priority checks.

When a job is geared to the occurrence of certain external events and must wait a period of time until an event occurs it may be suspended or "put to sleep." The job's temporary storage is left intact through the period of inactivity. When the anticipated event occurs the job is "awakened" by transfer of control to an address which may be different from its starting address. If a job of higher priority is in progress, the "awakening" will be postponed until it ends.

When a job is finished it transfers control to a terminating sequence which releases its temporary storage to be used by another job. Approximately ten jobs may be scheduled for execution or in partial stages of completion at a time.

Waitlist - The function of the Waitlist routine is to provide timing control for other program sections. Waitlist tasks are run in the interrupt mode, and must be of short duration, 4 milliseconds or less. If an interrupt program were to run longer it could cause an excessive delay in other interrupts waiting to be serviced, since one interrupt program inhibits all others until it calls for resumption of the normal program.

The Waitlist program derives its timing from one of the counter registers in the AGC. The Counter Priority stage which controls this counter is driven by a periodic pulse train from the computer's clock and scaler such that it is incremented every 10 milliseconds. When the counter overflows, the interrupt occurs which calls the Waitlist program. Before the interrupting program resumes normal program it presets the counter so as to overflow after a desired number of 10 millisecond periods up to a limit of 12,000 for a maximum delay of 2 minutes.

If the Waitlist is to initiate a lengthy computation, then the task will initiate an Executive routine call so that the computation is performed as a job during non-interrupted time.

Display and Keyboard - The programs associated with operation of the two Display and Keyboard units are basic to the employment of the AGC in the Apollo Guidance and Navigation system. These programs are long, but their duty cycle is low, so that their use of the time budget is reasonably small.

Key depressions interrupt to a program which samples the key code, makes a job request to the Executive, and then resumes. When this job is initiated, it examines the code and makes numerous branches based on past and present codes to select the appropriate action. Nearly always, a modification of the light registers in the display is called for. A periodic interrupt program similar to Waitlist, but occurring at fixed time intervals, performs the required display interface manipulations after it has been initiated by the job. More complex situations occur as a result of lengthy processing of data and periodic re-activation of a display function. For example, it is possible to call for a periodic decimal display of a binary quantity, for which the Waitlist is required to awaken a sampling and display job every second. This job samples the desired register or registers and makes the conversion to decimal according to the appropriate scaling for the quantity, i.e., whether it is an angle, a fraction, an integer, etc., and where the decimal point is located.

The Keyboard and Display programs are highly sophisticated routines to which a certain amount of computer hardware is expressly devoted for the sake of efficiency. They also make full use of the Executive and Waitlist functions to furnish a highly responsive and flexible medium of communication.

Chapter VI-3

MECHANIZED AIDS TO DESIGN AND PRODUCTION

MANUFACTURING

A number of automated processes are employed in the production of the Apollo Guidance Computer hardware. This has been done largely for the sake of minimizing human error and thus minimizing the consequent problems of reworking parts which were improperly built.

The case of the computer consists of metal pieces processed on a numerically controlled milling machine.

Signal matrices which interconnect microcircuit logic packages within a 60-package module are made semiautomatically, using punched paper tape to control a punching die which forms a matrix layer from a thin sheet of metal. Layers are insulated and stacked by hand before being hand-welded to the logic packs.

Another tape-controlled semiautomatic process is used for threading sense wires in Core Rope memory modules. Information on paper tape is used to position a rope fixture for an operator to pass a wire bobbin through core holes in which <u>one</u>'s are to be stored, bypassing those where zeros belong. The bobbin is passed as many times as there are <u>one</u>'s on the particular sense wire; and following each pass the tape is advanced so as to cause the fixture to be properly positioned for the next pass. When all 192 wires have been fully threaded and terminated, the module is tested on a rope memory tester, which operates the module as it will be operated in the computer. A punched tape input to the tester is compared against the information content of the module.

Interconnections between modules are made by wire which is terminated by tightly wrapping it about a rectangular post. The wrapping process may be done manually or automatically. In the AGC most of the wiring is done automatically by a machine whose information input is in the form of punched cards. The machine positions the wire over the pins with as many as 2 right angle (90°) bends in it, cuts it, strips the insulation at the ends, and wraps both ends

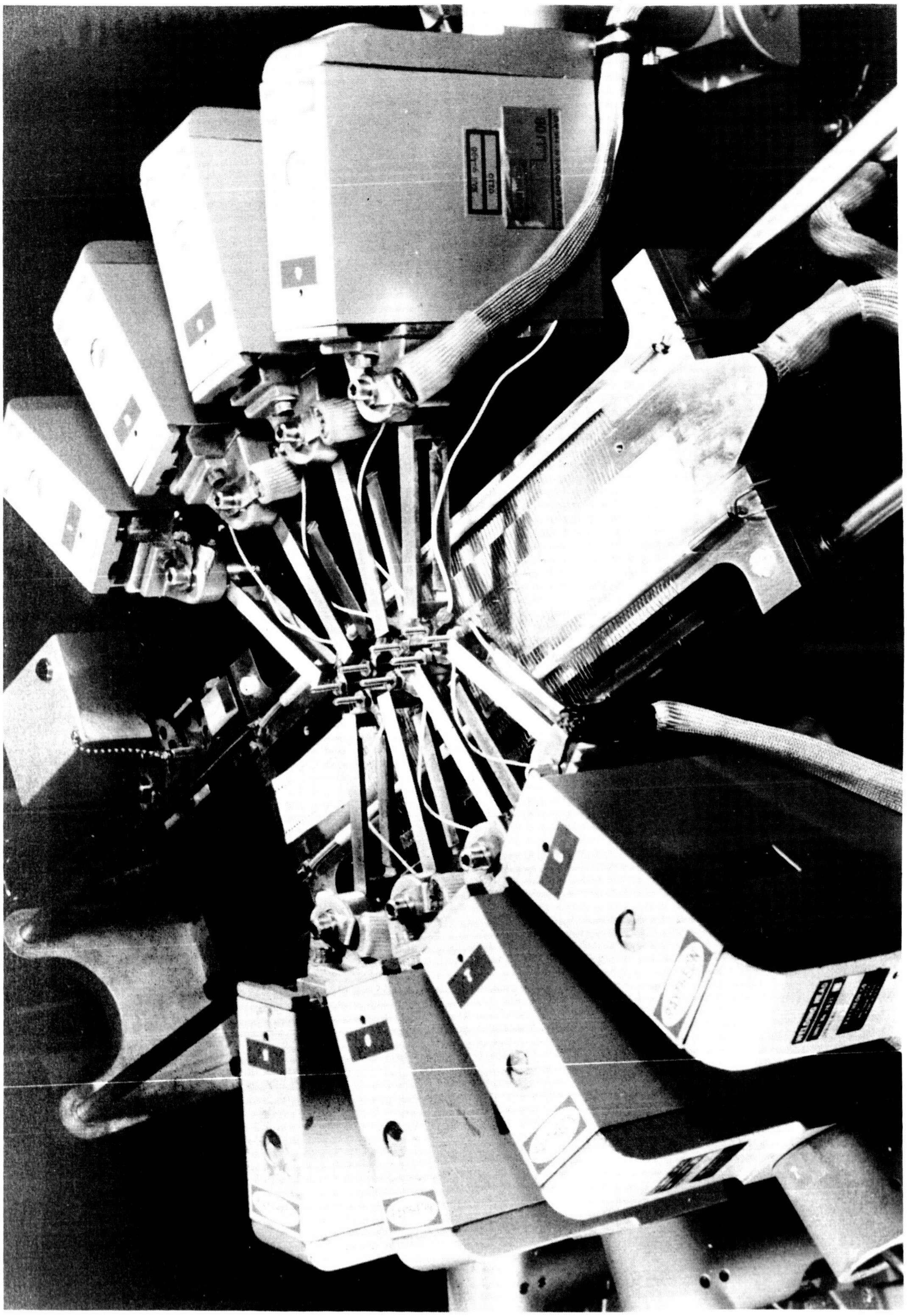


Fig. VI-26 Automatic Fabrication of Welded Matrix

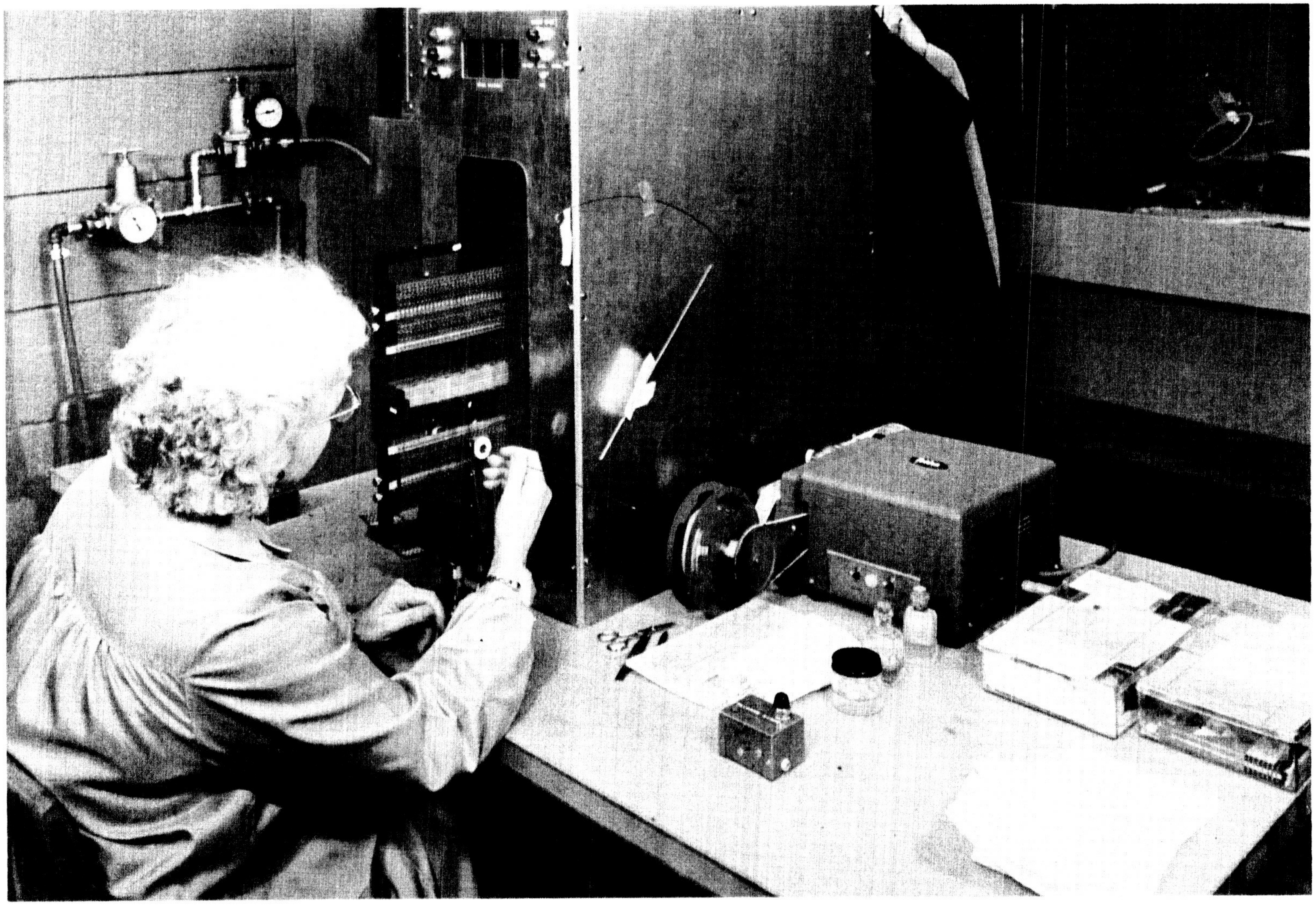


Fig. VI-27 Semiautomatic Core Rope Fabrication

INTERCONNECTION WIRING

The raw data for interconnection of modules is necessarily originated by hand. There have been several instances of computer makers using automated logical design procedures in which the manual input was in the form of Boolean expressions to be mechanized. Such procedures, attractive as they sound, are difficult to prepare and check out, and owing to their necessary inflexibility are not as efficient in hardware utilization as manual design methods. In the AGC, logic circuits are assigned to modules when they are drawn, and terminal assignments are made at the same time. A name is given each terminal signal using the rule that all terminals bearing the same signal name shall ultimately be connected together, and no connections shall be made between terminals bearing different names. A punched card is prepared for every used terminal of every module, whether it be a logic module or any other kind. These cards are accepted by a so-called "Wirelist" program, which sorts the inputs by signal name to show the terminal groups. On each card pertaining to a logic module, a number is added stating the loading or generating nature of the circuits connected to the terminal within the module. The Wirelist program processes these numbers showing for each signal name (group) whether the load exceeds the drive or vice versa, and by how much. This is a useful feature, for whereas it is not difficult to analyze loading of a signal entirely contained in a module and hence drawn on a single sheet of paper, it is perplexing to analyze loading on a signal which goes several places and appears on several different drawings.

The Wirelist program prepares a printed document showing the terminal groups listed alphabetically, and also showing the signal names for each module terminal listed in numerical order. The information file, obtained from the cards and stored on tape, constitutes the input to the program which prepares the card deck for the wire-wrap process.

The preparation of the card deck for wire wrapping would be a tedious task without the aid of automatic data processing. Starting from a magnetic tape file listing all of the interconnections in the AGC, a computer program assigns a path to each wire, punches the card deck, and prints a listing of what it has done. The program examines in order each group of terminals which are to be wired together. Since there is no pre-specified order in which the terminals in a group are to be joined to one another, the program tries to minimize the wire length by a sequential trial and error process. Not all possible configurations can be tried owing to the excessive time required; so an algorithm is used which can find an optimal or nearly optimal solution in a short time. The program thus arrives at a definition of the connections within a group, i.e., a list of terminal pairs.

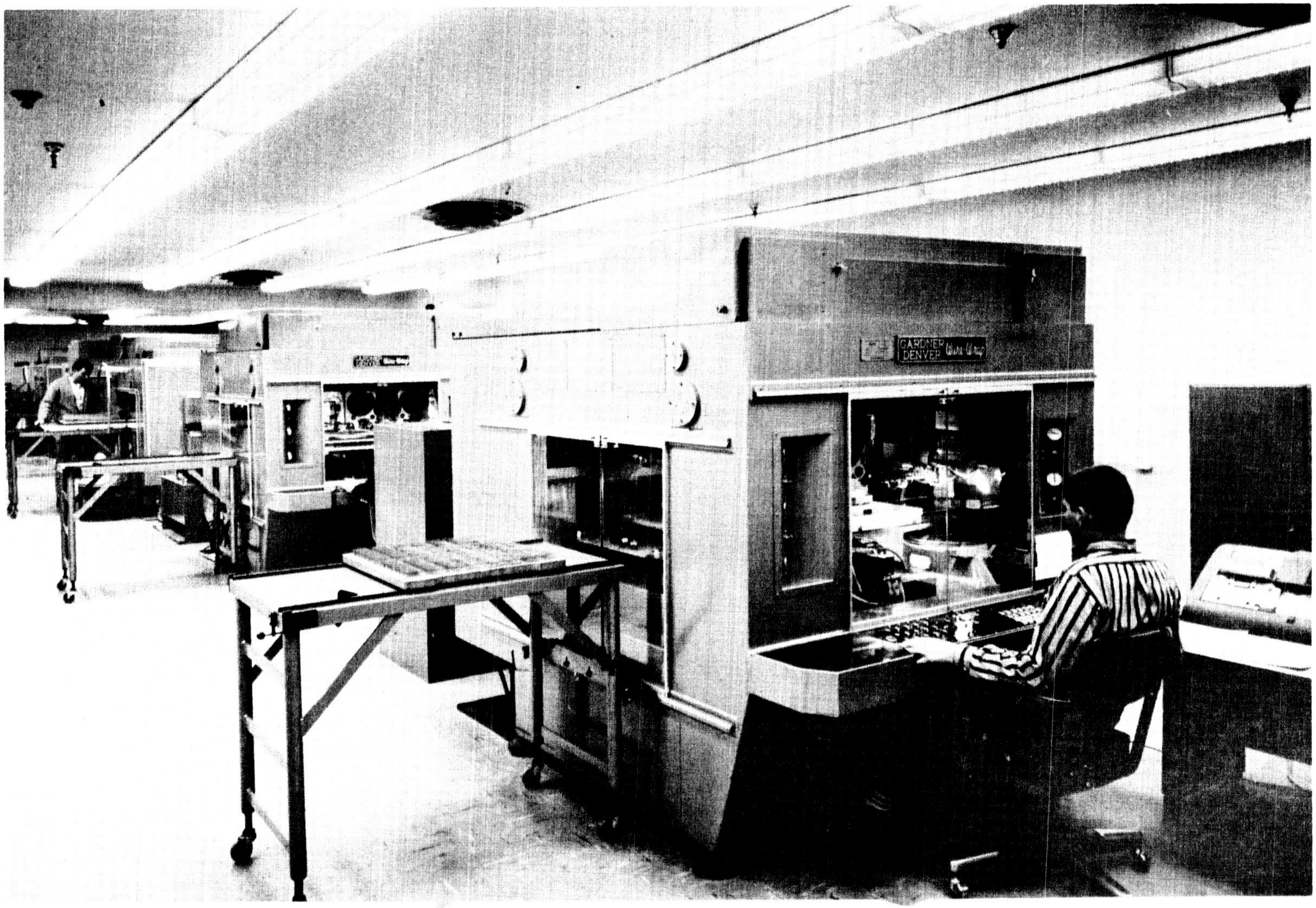


Fig. VI-28 Wire Wrapping Machine (Photo Courtesy Gardner Denver Co.)

It is next necessary to decide how a wire will proceed from one terminal to the other, a straight line is only possible if the terminals happen to lie in the same row or column. A rectangular layout pattern is used so as to avoid choking the gaps between terminals. The program first tries to assign either a straight wire or one with a single bend. If wires previously assigned have blocked the gaps through which this would have to pass, then another assignment is tried. If all possible assignments have been tried without success, no control card is punched for this wire. Rather an entry is made on the printed listing calling for a manual insertion of the wire after the automatic wrapping is complete.

PROGRAM PREPARATION

Assembler - The standard programming language for the AGC is an assembly language, in which each machine word of program is represented by a symbolic expression on an 80-column punched card. In fact, it can be said that there are two programming languages, basic and interpretive, and that any sizable program contains large amounts of both. As these languages are mutually exclusive, that is, no expression in one can be mistaken for any expression in the other, the assembler readily handles any mixture of the two. The punched-card format consists of three main fields.

A. Location field, which may be used to assign a symbolic name to the location of the machine word defined on the card;

B. Operation Code field, containing a symbolic code specifying the operation to be done, which determines whether the expression on the card is in basic or interpretive language; and

C. Address field, which in basic language usually contains a wholly or partially symbolic expression that specifies the address of the location to be operated upon.

There are many exceptions in detail to those definitions; only the most common use of each is given. Ample space is provided also for explanatory remarks, which are an integral part of the assembly-language file of a program. The printed listings made during assemblies, which for complete mission programs run to several hundred pages, thus constitute a medium of communication among the 50 or so engineers who do most of the programming.

The assembly program itself runs on the Instrumentation Laboratory's Honeywell 1800 data processing installation and performs in parallel its two major tasks, assembly and updating.

The assembly process translates programs in assembly language into absolute binary form for simulation and manufacturing, prints a listing in which the symbolic and absolute forms of each word of the program are displayed side by side, and prints diagnostic information about syntactical errors. The assembly process also allocates memory space to the program and to its variables and constants. The updating process maintains magnetic tape files of current programs in both assembly-language and absolute form, greatly reducing the need to handle large numbers of punched cards. For example, a program may be revised by presenting to the assembler just enough cards to specify the changes.

Assembly and updating take from less than 30 seconds to several minutes, depending on the size of the program being assembled and the amount of information on the file tape. The absolute binary files generated by assembly and maintained by updating are the input to the AGC simulator program and to program manufacturing activities that are described in later sections.

It may be instructive to trace briefly the history of part of a mission program -- re-entry guidance, for example -- from the conceptual stage to actual readiness for flight.

A. The mathematical ideas are blocked out roughly, tolerances guessed, variables and effects judged to be significant or negligible, and some such decisions are left to be settled by trial and error.

B. A procedure employing the concepts is worked out, using one mathematical model for the spacecraft with its guidance and navigation system, and another for the environment. This procedure is then employed in a data processor program for testing.

C. The program is compiled, tested, revised, and retested, until the mathematical properties of the procedure are satisfactory. Because these programs retain 2 to 4 more digits of precision than double-precision AGC programs, the variables at this stage may be considered free of truncation or round-off error. It is desirable to do as much pinning down of the procedure as possible in steps B and C, since these programs run a good deal faster than real time.

D. An AGC program is written by translating the relevant parts of the program into AGC assembly language; the more mathematical parts, such as position and velocity updating, into interpretive language, and the more logical, such as turning reaction control jets on and off, into basic.

E. In combination with the utility programs described previously, the AGC program is assembled. After two or three revisions, when the grammatical errors that can be detected by the assembler are eliminated, the program checkout is begun by use of the AGC simulator, another data processor program, which is described later. In advanced stages of checkout, this simulation incorporates the part of the program of steps B and C that models the environment. AGC simulation discovers the numerical properties of the procedure, since all effects of scaling, truncation, and roundoff are present. Steps D and E are repeated until the program fulfills the goals determined in step A. Severe problems may send the engineers back to step B, or even to step A.

F. Up to this point, everything has been done on the initiative of the 3- or 4-man "working group" whose specialty is the particular phase of the mission. Now, however, this AGC program must be integrated with the rest of the mission program. Using the updating facilities of the assembler, the working group transfers its own coding to the mission program and, in cooperation with the group in charge of program integration, checks out not only its operation but its ability to "get along with" the other parts of the mission program, e.g., staying within its part of the time budget. Here again, the AGC simulator is the primary tool.

G. At this point, or sometimes before step F, it is necessary to run an AGC attached to guidance and navigation and ground support equipment. This is the last procedure devoted entirely to the checkout of an AGC program.

H. When all of steps A through G for a whole mission program have been completed, the assembly listing of the program is given the status of an engineering drawing. Only now are rope memory modules wired; and further testing is for the benefit of other subsystems as much as of the program.

It should be explained that by "mission" is meant not only a flying mission but lesser responsibilities as well. Early mission programs, shipped with computers to other contractors to aid them in testing their systems, consisted mostly of utility programs.

Simulator - The assembler is designed to detect programmer errors of the nature of inconsistencies, but it is not capable of checking program validity in general. The hazards are numerous: faulty analysis of a problem, incorrect scaling, wrong use of instructions, interference with other programs, wrong timing, endless loops, and many other pitfalls familiar to programmers. Strictly speaking, the AGC program can never be fully tested before it is operated in a system in flight. Short of this, however, it is possible to prove out a program to a high degree of confidence by simulations of the program operating in an environment. Several possible approaches have been taken to the simulation study. In one, a computer and other parts of the guidance and navigation system are operated in conjunction with a real-time hybrid analog and digital simulation of the environment.

The other approach to simulation is an all digital simulation program which is run on the large scale data processing installation at the Instrumentation Laboratory. This is an important tool in the development of system oriented programs. Since it does not run in real time, it is possible to halt for the purpose of recording information relevant to program progress, such as periodic values of control constants or guidance and navigation variables, traces of interpretive instructions, environmental data and so forth. Initial conditions are easily set, and there is no limit, in principle, to the extent to which one can reproduce anticipated operational environments. The penalty is having running times from 2 to 40 times as long as real time (10 times is typical).

The Simulator comprises 3 major sections. The first simulates the AGC, and even operates an AGC Display and Keyboard (DSKY) connected on-line to the data processor. The second simulates the environment, and is constantly being added to and improved as operating experience is gained. These two sections run largely independent of one another, as their functions are basically incompatible. One is a function of elapsed time, and the other is a function of AGC program execution status. The third section exists for the purpose of communication between the other two. It is capable of representing the environment to the computer over short periods of time while the two operate in isolation. From time to time the simulation halts while the AGC and environment sections reconcile with one another. The communicator section is re-initialized; and the simulation proceeds, with the communicator extrapolating the recent past history of the environment.

Information which is recorded out is subsequently edited in such a way that system analysts can easily process it with their own programs in order to make error analyses, study correlation of events, or perform any other mathematical or editing operation.

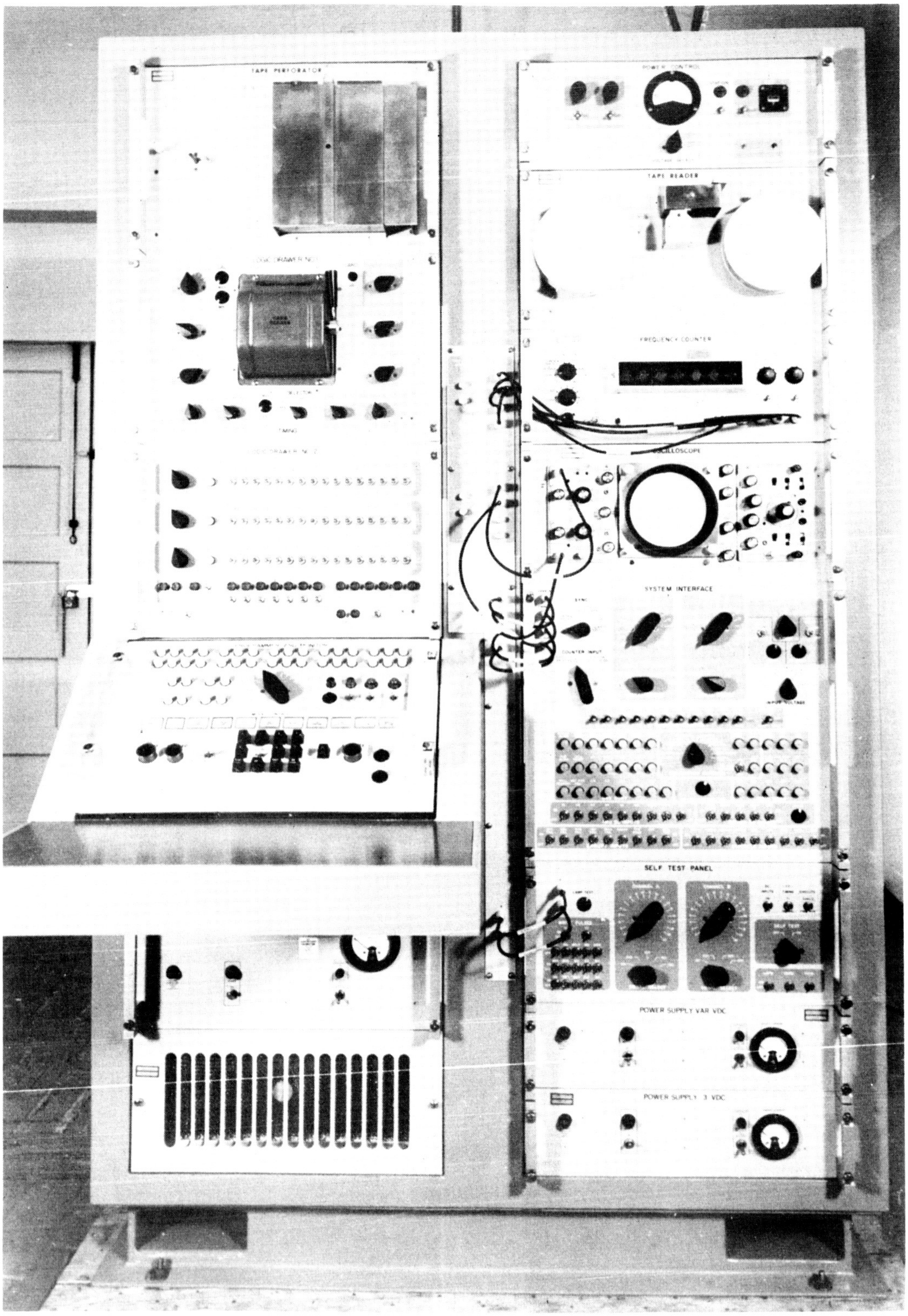


Fig. VI-29 Computer Test Set

Chapter VI-4

GROUND SUPPORT EQUIPMENT

TEST SET

Because of volume and weight restrictions, guidance computers are built without the extensive maintenance features found in data processing computers. During all stages of design, assembly, and field testing it is important to have facility for manual intervention into the computer's operation, as well as a means for certifying thall all circuits are properly operating. This facility is provided by a separate unit which connects to the conputer through what is usually a separate interface.

The unit which does this job with the AGC is called the AGC Test Set. It contains a number of flip-flop registers which can be made to serve various roles by selecting among numerous modes of operation. Over 100 signals are exchanged between the AGC and Test Set through an interface called the test connector. The Test Set has access to the AGC's write buses. It can sense them and also force signals upon them. This feature, together with the AGC's timing pulses and central register control pulses, permits the Test Set to follow the progress of AGC programs.

In its monitor mode the Test Set uses flip-flop registers to duplicate, or mimic, the contents of AGC central registers. Any of these registers may be displayed in lights on the control panel. Since in particular the S, or address, register and the G, or memory local, register are displayed, it is possible to see what is stored in memory. The successive contents of any erasable register can be displayed by commanding the Test Set to sample the G register and display its contents each time the address is the same as the address specified by hand set switches on the panel.

Facility is also available for causing the computer to halt when it interrogates a specified address, or when the content of a specified register reaches a specified value, or at the end of each instruction, memory cycle, or alarm. The computer may be made to proceed from where it is, or it may be started at any desired instruction. Any register may be read to the lights whether or not it is accessed by program, and any erasable location may be loaded from the Test Set with any desired value.

The Test Set contains circuits for exercising and testing all of the AGC's interface circuits. A separate cable connects the Test Set with the AGC's system connector and a switch panel causes an oscilloscope to be connected to an AGC output, or else a signal generator of appropriate characteristics to an AGC input. This feature is used to make a detailed test of output signals, including their rise time, amplitude, duration, and any other important characteristics. A faster and less complete check of the interfaces can be made by a special connector which exercises pulse inputs by connecting them to pulse outputs, and exercises DC inputs by connecting them to DC outputs and supplying a dummy load. The check can then be made by an AGC program read into the erasable memory.

AGC MONITOR

Before an AGC program is wired into fixed memory modules, it must be exercised on an AGC. For this purpose, there exists a form of ground support equipment incorporating some features of the Test Set together with an erasable type of memory whose contents are sent to the AGC when it interrogates a fixed memory address. This machine is the AGC Monitor, so called because it encompasses the monitor feature of the Test Set. The memory control circuits of the Monitor are arranged so that only a part of the AGC fixed memory is read from the Monitor, while the remainder is read from the AGC. This is done, moreover, without having to remove fixed memory modules from the AGC. Rope simulations are done in 1024 word segments. When the AGC Fixed Bank register and S register are observed to be set to access a simulated bank, the sensing of the AGC fixed memory is inhibited by a signal from the monitor while the simulated word is transferred into the G, or memory local, register of the AGC.

The memory used in the Monitor is a set of 9 erasable memory units, each containing a 4096 word coincident-current ferrite core stack with driving electronics similar to those in the AGC. The logic in the Monitor and in the Test Set is made from microcircuit NOR gates. One reason this is done is to put to use those gates which are not found qualified for flight hardware, but are still satisfactory for less stringent environments.

Chapter VI-5

CONCLUSION

Guidance computer engineering is a simultaneous effort in mechanical, electrical, and logical design disciplines.

In order to obtain high efficiency in terms of performance, volume, and power consumption, guidance computers are designed to work in a single specific system, unlike most commercial computers. They commonly have fewer and less flexible instructions, shorter word length, and less complex arithmetic units. Compactness and short term reliability predominate over considerations of programming ease, maintainability and manufacturing cost. Whereas the commercial computer designer strives to maximize answers per month, the guidance computer designer seeks the capability of handling high peak loads, and is concerned with answers per second.

The guidance computer receives data from various sources and delivers answers to various destinations over tens or hundreds of signal paths, each requiring appropriate conditioning circuitry at origin and at destination. One of the challenging design problems is to minimize the number of these interface signals, and moreover to minimize the number of different circuits used.

The AGC is quite representative of the state of the art in guidance computers as contrasted with the rest of computer technology. Most conspicuous of the attributes common in guidance computers are the extensive use of microcircuits and high density interconnections, a dense fixed memory of about half a million bits and a small erasable memory, and a short word length.

The computer was designed to employ certain utility programs. The Interpreter program allows efficient expression of double precision matrix programs for navigation, attitude control, and steering. The Executive program allots computer time among various jobs according to a priority schedule. The Waitlist program provides interrupted entry to other programs at specified intervals of real time.

Guidance computers are often supported by commercial computers for automatic programming and in various areas of mechanical and electrical design. A large scale computer is used in connection with the Apollo Guidance Computer in several respects. It assembles and makes simulation runs on programs; it generates the input card decks for automatic wire wrapping machinery used in computer manufacture; and it also prepares punched tape for use in fixed memory fabrication, and information input to ground support equipment.

The constraints on the guidance computer designer are severe. In addition to the requirements of size, performance, and reliability is the urgency for early delivery which stems from trying to get the best equipment possible without introducing unnecessary delay in flight schedules. For this reason, we may expect that guidance computer engineering will continue to be a highly productive, competive discipline for years to come.

ACKNOWLEDGEMENT

The Apollo Guidance Computer has been used here to exemplify the state of the art in broad scope, from hardware to software. The AGC owes its existence to the pioneering work of Ramon L. Alonso, Eldon C. Hall, and J. Halcombe Laning, Jr. Its development has encompassed the efforts of many workers, some of whose contributions are referenced below. Others, whose contributions are substantial but unwritten include D. J. Bowler, E. J. Duggan, J. S. Miller, R. F. Morse, and H. A. Thaler. To them and to the Raytheon Company, manufacturers of the Apollo Guidance Computer, the author is grateful for advice and assistance in preparing this description of their achievements.

BIBLIOGRAPHY

1. Proceedings of the Conference on Spaceborne Computer Engineering, I. R. E., P. G. E. C., Anaheim, California, October 30 - 31, 1962.

2. M. M. Dickinson, J. B. Jackson, and G. C. Randa, "Saturn V Launch Vehicle Digital Computer and Data Adapter," Proceedings 1964 Fall Joint Computer Conference, Spartan Books, Inc., Baltimore, Maryland, 1964.

3. H. Wyle, S. Z. Rubenstein, L. Miller, and S. E. Githens, "The Logical Design of the D26C: A Microminiature Real Time Processing System," National Electronics Conference, May, 1964.

4. R. L. Alonso, H. Blair-Smith, and A. L. Hopkins, "Some Aspects of the Logical Design of a Control Computer: A Case Study," IEEE Transactions on Electronic Computers, Vol. EC-12, No. 5, December, 1963.

5. R. L. Alonso and J. H. Laning, Jr., "Design Principles for a General Control Computer," Institute of Aeronautical Sciences, New York, New York, S. M. Fairchild Publication Fund Paper FF-29, April, 1960.

6. R. L. Alonso, A. I. Green, H. E. Maurer, and R. E. Oleksiak, "A Digital Control Computer; Developmental Model 1B," Report No. R-358, Instrumentation Laboratory, MIT, Cambridge, Massachusetts, April, 1962.

7. T. C. Taylor, "A Proposed Circuit Structure for Computer Logic Based on Semiconductor Flat Packs Interconnected by Multilayer Circuit Boards," Report No. E-1699, Instrumentation Laboratory, MIT, Cambridge, Massachusetts, December, 1964.

8. J. Partridge, L. D. Hanley, and E. C. Hall, "Progress Report on Attainable Reliability of Integrated Circuits for Systems Application," Report No. E-1679, Instrumentation Laboratory, MIT, Cambridge, Massachusetts, November, 1964.

9. D. Shansky, "Erasable Store Mod 3C," Report No. E-1158, Instrumentation Laboratory, MIT, Cambridge, Massachusetts, July, 1962.

10. R. L. Alonso and A. L. Hopkins, "The Apollo Guidance Computer," Report No. R-416, Instrumentation Laboratory, MIT, Cambridge, Massachusetts, August, 1963.

11. C. A. Muntz, "User's Guide to the Block II AGC/LGC Interpreter," Report No. R-489, Instrumentation Laboratory, MIT, Cambridge, Massachusetts, April, 1965.

12. A. I. Green and J. J. Rocchio, "Keyboard and Display System Program for AGC," Report No. E-1574, Instrumentation Laboratory, MIT, Cambridge, Massachusetts, August, 1964.

195

13. T. J. Lawton and C. A. Muntz, "Organization of Computation and Control in the Apollo Guidance Computer," Instrumentation Laboratory, MIT, Cambridge, Massachusetts, April, 1965.

PART VII

SPACE VEHICLE CONTROL SYSTEMS

by

Dr. Wallace E. Vander Velde

DR. WALLACE E. VANDER VELDE

Professor, Dept. of Aeronautics and Astronautics
Massachusetts Institute of Technology

Dr. Wallace E. Vander Velde is an Associate Professor of Aeronatuics and Astronautics and Consultant to the Instrumentation Laboratory at Massachusetts Institute of Technology. In his academic position he is an authority on automatic control systems for fixed wing aircraft, helicopters, missiles, satellites, and space vehicles. At the Instrumentation Laboratory he is working with the Apollo Spacecraft Guidance and Navigation Group.

Dr. Vander Velde was born June 4, 1929, at Jamestown, Michigan. He received the Bachelor of Science degree in Aeronautical Engineering from Purdue University in 1951. Following military service, in 1953, he entered M. I. T. as a graduate student and received the degree of Doctor of Science in Instrumentation in 1956..

From 1956 to 1957, Dr. Vander Velde was director of applications engineering for the GPS Instrument Co., Inc., Newton, Massachusetts. He returned to M. I. T. in 1957 as Assistant Professor of Aeronautical Engineering and was appointed Associate Professor of Aeronautics and Astronautics in 1961 and Professor of Aeronautics and Astronautics in June, 1965.

Part VII

SPACE VEHICLE FLIGHT CONTROL

INTRODUCTION

The development of the theory and practice of automatic control has to a large extent gone hand in hand with the development of aerospace flight control systems. Although the earliest application of feedback control as a deliberately-conceived and consciously-applied technique preceded the invention of the airplane by about 50 years, it was the exacting requirements of aircraft flight control which for years made the greatest demands on the developing theory of feedback control and which stimulated much of its growth. In recent years, the advance of "modern" control theory has been led in large part by workers responding to the need for more sophisticated control theories and techniques for application to aerospace flight control problems. The requirements for control efficiency, accomodation of changing controlled-member characteristics, and reliable control in complicated and demanding situations have resulted in a far broader application of optimal controls, adaptive controls, and digital-computer control systems to the problems of aerospace vehicle flight control than to any other area of application.

Aerospace flight control problems are not only demanding, but quite diverse. Among the many different phases of a space vehicle mission beginning with lift-off from a launching pad and ending perhaps with a return to a designated landing point on Earth, one can identify three major classes of flight control problems:

Powered flight control - control of the attitude and flight path of the vehicle while thrusting.

Coasting flight control - control of the attitude of the vehicle while coasting in free space.

Atmospheric flight control - control of the attitude and flight path of the vehicle while gliding in an atmosphere.

The nature of the control problems in these different situations is very different. The environment in which the vehicle operates, the requirements on the control systems, the sources of reference information and of control torques are all quite different. But a single space vehicle operates under all of these conditions during the course of a mission and it is to be hoped that the different control problems will not have to be solved one at a time in isolation. The design of an over-all control system which performs well in each phase of the mission, using common equipment wherever possible, and which is in addition integrated efficiently with the guidance and navigation system constitutes a most challenging engineering problem.

In the following chapters, the major characteristics of the control problems in each of these classes is discussed in turn.

CHAPTER VII - 1

POWERED FLIGHT CONTROL

INTRODUCTION

The function of the powered flight control system is to orient the vehicle thrust acceleration vector in response to commands generated by the guidance system. This function is required during each of the powered flight phases of a mission. These phases might include:

Boost from the launching pad into a parking orbit

Acceleration into the destination transfer trajectory

Midcourse velocity corrections

Deceleration into an orbit around the destination planet or moon

Powered descent to the planet or moon

Comparable phases in the return flight

The thrust acceleration vector is, on the average, oriented in the vicinity of the vehicle longitudinal axis. Thus powered flight control is primarily a problem in attitude control. This problem will be discussed first as it relates to two means of generating the required control moments. Then in the following section, some additional considerations involved in acceleration vector control will be discussed.

ATTITUDE CONTROL WITH A GIMBALED ENGINE

With a rocket engine providing a large force which acts on the vehicle, the most evident source of large control moments is the deflection of the direction of this force so it has some moment arm with respect to the vehicle center of mass. This can be conveniently accomplished in the case of a liquid-fueled rocket by mounting the engine on gimbals and rotating it. A solid-fuel rocket requires rotation of the thrust direction with respect to the engine itself. This has been accomplished by different means; among them are jet vanes, jetavators, rotating canted nozzles, and secondary fluid injection. The present discussion supposes a liquid-fueled rocket with a gimbal-mounted engine.

The very first flight control problem encountered in a space mission is probably the most difficult - the control of the unstable, elastic vehicle in its ascent through the gusty atmosphere. In this phase the configuration of the vehicle is largest, the frequencies of body bending and fuel sloshing oscillations are lowest, and the need for control system gain and bandwidth is most critical due to the requirements of stabilizing the aerodynamically unstable vehicle and reducing structural loads due to wind disturbances. In most of the powered flight phases following this, there is no atmosphere to contend with and the vehicle configuration is smaller due to the separation from one or more stages. This would not be true if in-orbit assembly cf separately-boosted payloads were employed.

The most important requirements of the attitude control system during atmospheric exit may be summarized as:

Maintain vehicle stability

Provide adequate performance for execution of commands

Provide adequate alleviation of gust loads

Make reasonably efficient use of control action

Maintain simplicity and reliability

The first requirement dominates the design of this system. For early analysis one might very well model the vehicle and its control system as quasi-stationary and linear. If so, one has only the classical problem of the stability of a linear, invariant feedback system - and the problem would be simple were it not for the very complicated nature of the vehicle being controlled. The basic rigid vehicle is unstable in the atmosphere because of the required distribution of area and mass. It could be stabilized aerodynamically by the addition of fins at the rear of the vehicle, but the additional weight and drag would incur too dear a performance penalty. So the alternative of active stabilization through control system feedback is almost universally preferred. This places a lower bound on control system gain for static stability: the restoring control moment due to a rotation of the vehicle must be greater than the diverging aerodynamic moment due to that rotation.

But this requirement of high gain for static stability is in conflict with the requirement of low gain for dynamic stability. The dynamic stability problem is complicated by the fact that the large vehicle is by no means rigid. The launch configuration consists of a number of vehicle stages coupled with light-weight interstage structure. Significant bending occurs at these coupling points.

Further, each stage is designed as light-weight as possible with the major requirement being to carry axial compressive loads. The resulting structure has appreciable bending elasticity. Still further non-rigid-body behavior is due to fuel and oxydizer sloshing in their tanks and localized bending of the structure between the points of engine gimbal mounting and gimbal actuator attachment.

The primary body bending deflections are decomposed for analytic purposes into a series of normal modes of oscillation. Each mode has a characteristic frequency and mode shape. These resonant modes are very lightly damped. An idealized picture of a fundamental or first order bending mode is shown in Fig. VII-1. The body centerline, deflected according to the first order mode shape, is shown with an undeflected body reference line. The two points which do not translate in this mode of oscillation are called nodes; the point of greatest translation, which is also the point of zero rotation, is called the antinode. The engine gimbal deflection, δ, is measured and controlled with respect to the deflected body. Also, the body attitude is indicated by an instrument located at a particular body station such as that shown on the figure. It will indicate the local body angle with respect to the inertial reference line; in Fig. VII-1, this angle is $\theta + \phi$. Additional higher ordered bending modes may be excited simultaneously. The body translational or rotational deflection at any station is the sum of the deflections in the different modes.

It is clear that if the control system bandwidth extends to the frequency of a particular bending mode, the effect of the control feedback can either tend to stabilize or destabilize the mode. Consider as an example just the effect of attitude rate feedback on the mode shown in Fig. VII-1. We see that for gross vehicle stability the sense of the rate feedback to gimbal angle must be positive. That is, a positive $\dot{\theta}$ should cause a positive engine deflection, δ, which will result in a control moment tending to reduce the angular velocity. But now a positive bending rate $\dot{\phi}$ which is also sensed by the indicator also causes a positive increment in engine deflection, δ, and this produces a normal force component on the body which tends to further bend the body in the same sense. For the configuration shown then, with the engine behind the rear node and the rate indicator forward of the antinode, in-phase rate feedback at the first bending frequency tends to destabilize the mode. If either the engine were forward of the rear node, or the rate indicator were behind the antinode, but not both, such feedback would tend to stabilize the mode. Alternatively, for the configuration shown, rate feedback would stabilize the bending mode if a filter between the rate indicator and engine deflection provided 180 degrees of phase shift at the bending frequency so the engine deflection would be opposite in sign to the indicated bending

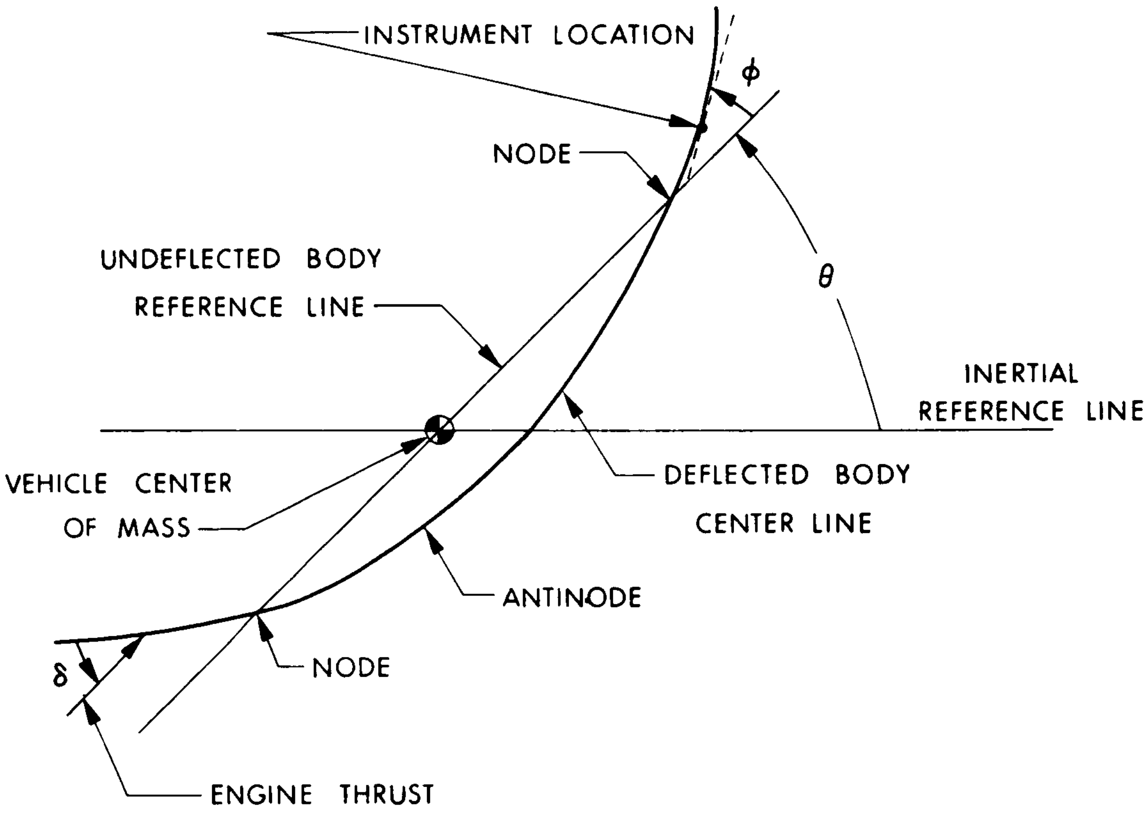


Fig. VII-1 Idealized Body Bending Mode Shape rate. Notice that for the present purpose, it is immaterial whether the 180 degrees of phase shift at the bending frequency is phase lead or lag.

This qualitative view of the effect of the control system on body bending modes may be given quantitative significance using any of the standard methods of studying linear system stability. Consider the system configuration of Fig. VII-2 in which the vehicle is characterized by its rigid body dynamics and first bending mode. The indicated angle is compared with the commanded angle, and the compensated error signal commands the gimbaled engine servo. The lead compensation which is necessary to stabilize even the rigid body dynamics may be thought of as realized either by rate gyro feedback or by an idealized lead network. For a bending mode shape such as that pictured in Fig. VII-1, a simplified root locus plot for this system has the general appearance of that shown in Fig. VII-3. A number of high frequency effects have been omitted from this figure to emphasize the basic problem of stabilizing the aerodynamically unstable vehicle which also suffers elastic deformations. The unstable character of the vehicle is indicated by the rigid body pole in the right half plane. A single pole corresponding to a first order lag approximation to the servo dynamics is shown, as is the zero due to the lead compensation. Were it not for the bending mode, this lead would clearly be adequate to stabilize the rigid body.

The lightly damped bending oscillation is indicated by the pole near the imaginary axis at the bending frequency. The parallel transfer of the rigid body angle plus the bending deformation angle to the attitude indicator gives rise to a zero near this pole. The indicated relation of the zero to the pole corresponds to the situation shown in Fig. VII-1. If it were possible to move the instrument location back along the vehicle, the zero would move toward the pole - lying on top of the pole and cancelling the effect of the mode in the control system when the attitude indicator is located at the antinode. In that location, the attitude indicator does not sense any angle due to the bending oscillation. If the attitude indicator were moved behind the antinode, the zero would appear below the pole in the upper half of the root locus plot and the small locus of closed loop root locations would lie to the left of the pole and zero. The first bending mode would be a stable oscillation for all control system loop gains in that case. However, higher frequency effects not shown in Fig. VII-3 would still be troublesome, so careful placement of the sensors is not generally adequate to solve the problem.

For the configuration shown in Fig. VII-3, which is commonly the case, the design problem results from the fact that for a loop gain high enough to stabilize the rigid body dynamics, which is indicated by the unstable closed loop pole crossing into the left half plane at A, the bending mode pole has already

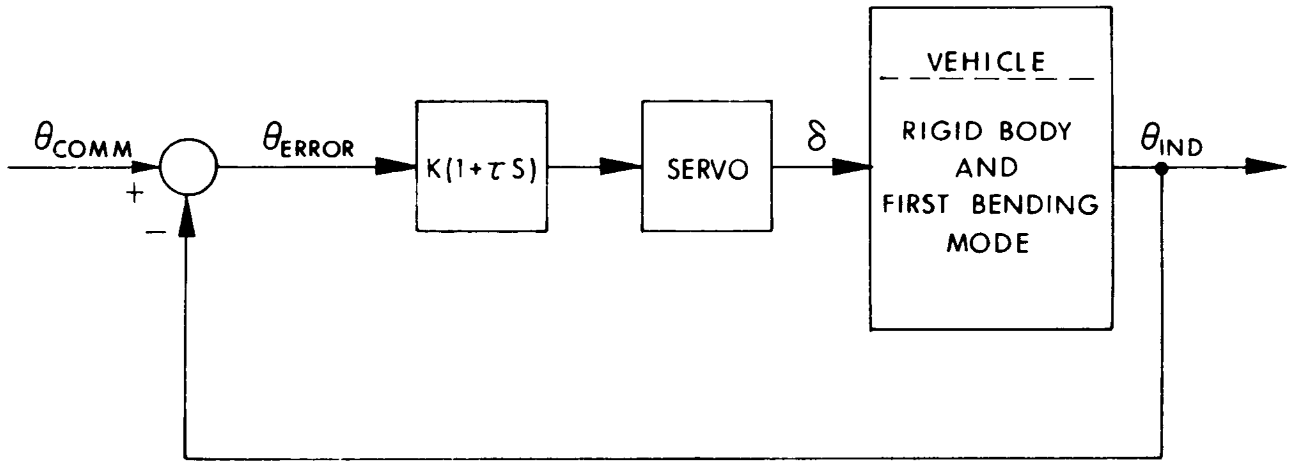


Fig. VII-2 Simplified Attitude Control System Configuration

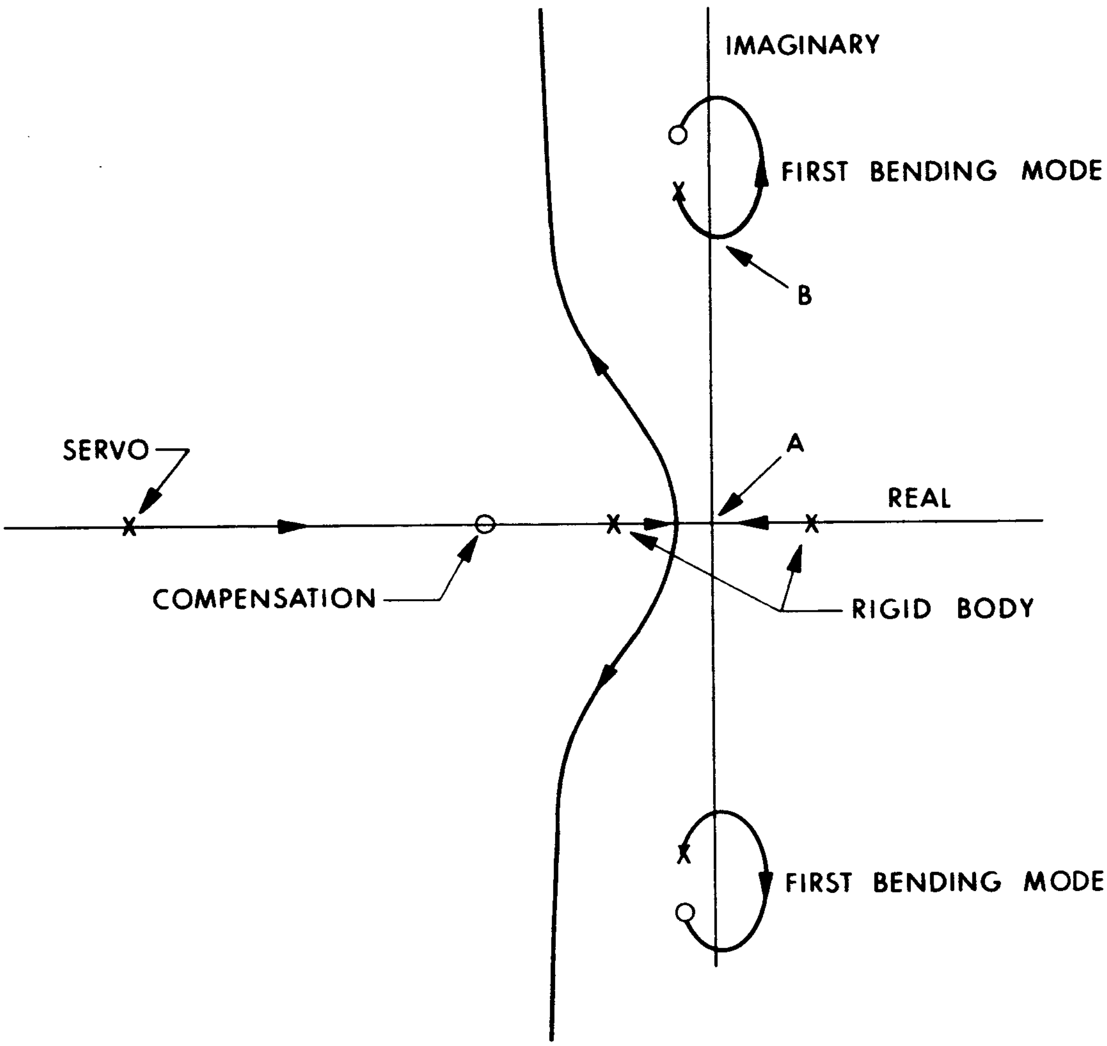


Fig. VII-3 Simplified Root Locus Plot for System with Rigid Body Compensation Only crossed into the right half plane at B. Static stability demands a certain minimum loop gain as noted before, so it is essential to stabilize the bending mode at the required higher gain. This can be thought of as a requirement to turn the bending mode locus from the right to the left of the pole and zero. This can be accomplished by rotating the breakaway angle for the locus from the pole either clockwise or counterclockwise, which corresponds to either lag or lead compensation.

With lag compensation for the bending mode, the root locus has the appearance of that shown in Fig. VII-4. The additional lag has turned the bending mode locus to the left as desired, but it has adversely affected the dominant rigid body mode. There is, however, a range of loop gains for which the system is stable. It has a bandwidth well below the frequency of the bending mode. The neglect of any higher frequency effects may well be justified in this case.

With lead compensation for the bending mode, the root locus has the appearance of that shown in Fig. VII-5. The strong lead compensation required to turn the bending mode locus far enough has significantly influenced the locus of the rigid body mode. Considering only the dynamic effects shown in the figure, wide bandwidth operation would be possible. However, the additional lead maintains the open loop gain at higher frequencies and makes more high frequency effects important in the system. Higher ordered bending modes must then be included in the analysis. The lead at higher frequencies will not be as great, and there may well be instability indicated in a higher frequency mode.

In addition to all this, there is another source of dynamic modes to complicate the picture still further. For the liquid-fueled rockets under consideration here, the propellants are free to slosh back and forth in their tanks. This oscillatory change of momentum of the fluid particles reacts through the tank walls to affect the dynamic behavior of the vehicle - giving rise to additional lightly-damped modes. The analysis of these effects is facilitated by reference to a mechanical analogy to the sloshing fluid. The analogy can be taken in the form of a spring and mass with a degree of freedom normal to the body longitudinal axis, or a pendulum as shown in Fig. VII-6. The parameters of the mechanical analogies have been derived by a number of authors;among them is Lorell[(1)]. One such pendulum is required to simulate each mode of sloshing to be considered in each tank. For a multi-stage vehicle having two tanks in each stage, the complexity compounds rapidly. Fortunately it is often true that only the first sloshing modes in the largest tanks have frequencies in the pass band of the control system.

The coupling of the pendulum oscillation into indicated vehicle attitude is dynamic in this case as opposed to geometric in the case of body bending, so the effect is more difficult to visualize. However, the coupled equations of

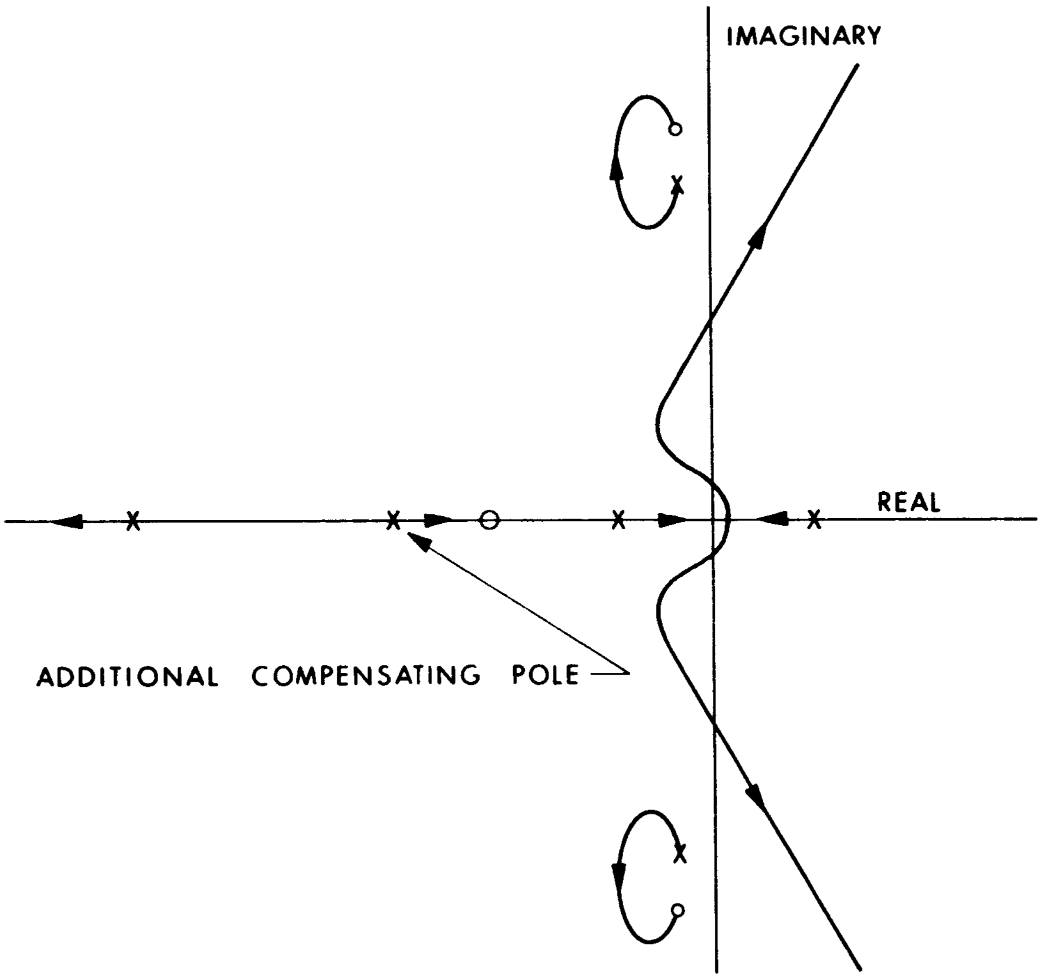


Fig. VII-4 Simplified Root Locus Plot for System with Lag Compensation for Bending Mode

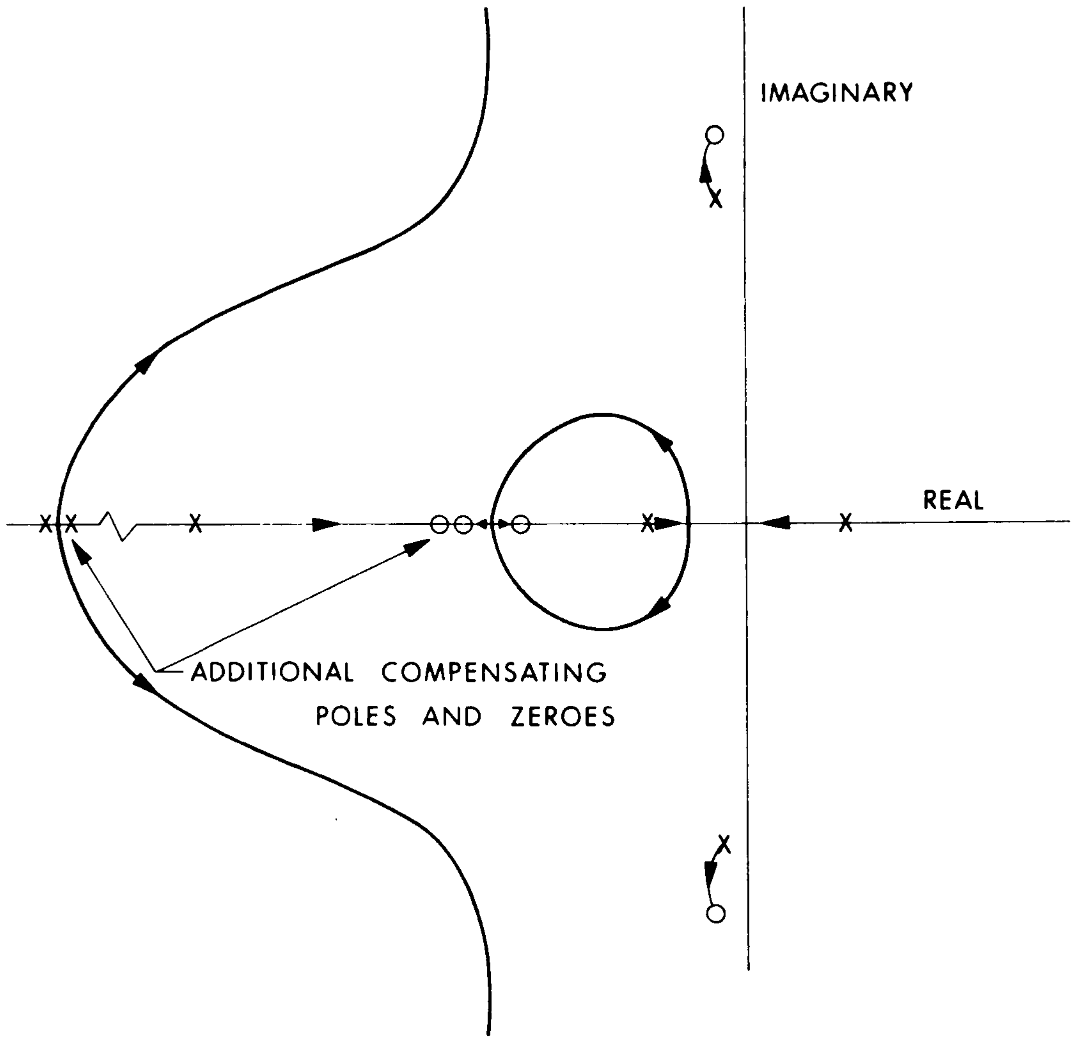


Fig. VII-5 Simplified Root Locus Plot for System with Lead Compensation for Bending Mode

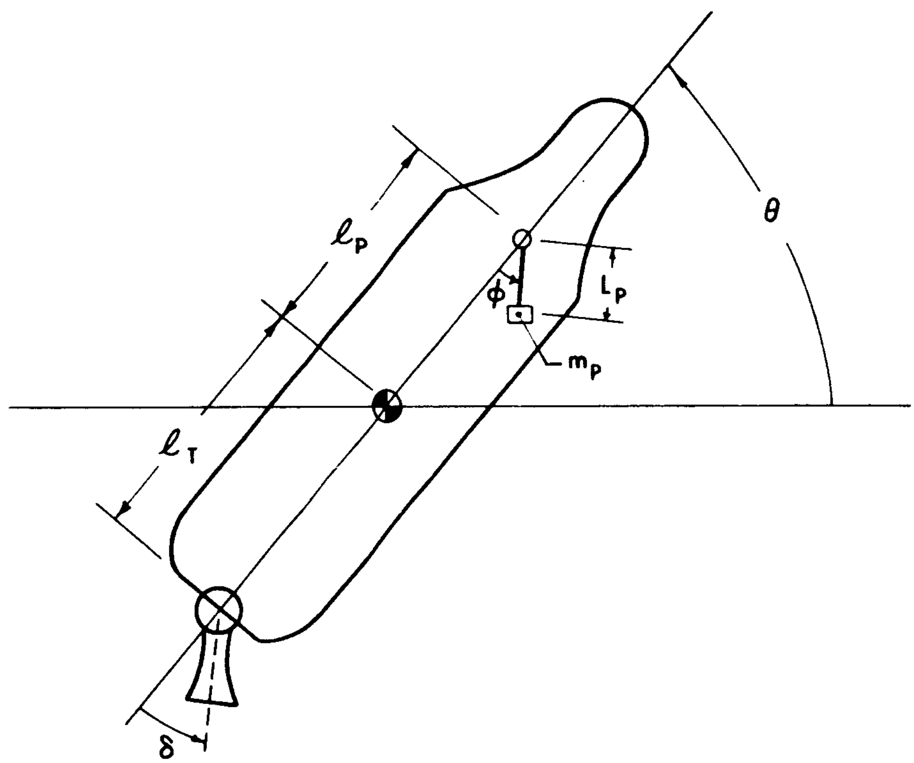


Fig. VII-6 Pendulum Analogy to Sloshing Propellant motion for the vehicle and pendulum produce lightly-damped pole-zero pairs in the system open-loop transfer function just like those due to bending. A single tank located well forward of the vehicle center of mass or to the rear of the center of mass has its upper halfplane zero located below the pole - the configuration which tends to give stable closed loop roots. There is an intermediate range of tank locations which puts the upper half plane zero above the pole and which would tend to give unstable closed loop roots unless compensated. These locations are characterized by the hinge point of the analogous pendulum lying in front of the rigid vehicle center of mass and the pendulum mass lying to the rear of the instantaneous center of vehicle rotation in response to control moments. This rotation center is forward of the rigid vehicle center of mass by the amount k^2/ℓ_T, where k is the rigid vehicle radius of gyration and ℓ_T is shown in Fig. VII-6. There are two cases in which the zero lies on the pole and the slosh mode has no participation in vehicle dynamics: if the pendulum hinge is at the vehicle center of mass, in which case the pendulum reaction causes no moment about the vehicle center of mass, or if the pendulum mass is at the vehicle rotation center, in which case control moments do not excite pendulum motion. Having coupling between slosh modes and vehicle dynamics, the problem of compensation design is the same as for bending modes.

The controlled member is thus characterized by the unstable rigid body dynamics and an assortment of lightly damped oscillatory modes due to body bending and propellant sloshing. The design of compensation which at any one flight condition will either stabilize these modes or leave them essentially uncoupled is somewhat like juggling a large number of balls. Allowing any one mode to be badly behaved is like dropping just one ball - the act is a flop. And to make matters worse, the characteristics of these modes change continuously as the dynamic pressure varies and propellants are consumed. Compensation which works well at one flight conditon may result in unstable operation at the flight condition which exists a half minute later. The least sensitive design is usually one which employs lag compensation and simply cuts off the control system bandwidth below the frequencies of the oscillatory modes. This leaves the bending and slosh modes essentially uncoupled from the control system. They are free to display their characteristic oscillations in response to whatever excitation they receive. For the lowest frequency bending and slosh modes particularly this may allow the existence of continuing oscillations which cause substantial structural loads. If so, the total system design is improved by using the control system to actively damp the lowest frequency modes. This requires a wider bandwidth control system. In fact, from just the point of view of structural loads there may well exist for any vehicle an optimum control system bandwidth: for lower bandwidths the loads due to poorly damped oscillations tend to dominate and for higher bandwidths the loads due to control action tend to dominate. In the case of the propellant sloshing oscillation, a source of damping other than the active effect of the attitude control system is available and often used. At the cost of some weight, structural baffles may be built into the propellant tanks to break up the large-scale oscillations of the fluid near the free surface.

Fortunately, the requirements on the attitude control system, other than stability, are not very demanding in most instances. Very little bandwidth is required to follow the command inputs since the required attitude for efficient powered flights is usually very slowly changing. The requirement for speed of response to alleviate gust loads may be more demanding. In the selection of a system bandwidth to minimize structural loads, as discussed above, the effects of gust and wind shear inputs should be included. The requirement for efficient use of control action is not usually of crucial importance though there is some weight penalty associated with excessive demand for control moments. System simplicity and reliability is always an object of first importance. In the present context this requires the design of simple compensation using parameter values which can readily be realized, and finding simple ways to vary this compensation as needed during the flight.

Another basic choice which bears on the question of reliability is the choice of analog vs digital data processing to close the control loop. Wide band systems are usually more efficiently implemented with analog equipment, but the range of achievable bandwidths in this problem is not very large. On the other hand there is the need for variable compensation and loop gains and the desirability of complex zeros and poles in compensation functions, each of which can more readily be implemented by digital computation. Moreover, there is the fact, usually obscure at the outset, that after a system has been designed to serve its primary function it tends to grow to accommodate the additional requirements of the operational situation. One major cause of this growth is the need for a variety of modes of operation: check-out mode, primary mode, back-up mode, pilot control mode, etc. Mode switching requires changing connections in an analog system, and this increases the component count and decreases the reliability estimate. In a digital system, mode switching is accomplished by branching to another program stored in memory. Since fixed memory is highly reliable, this entails little penalty in the reliability estimate.

The design of a digital attitude control system is quite like the design of a continuous system with the additional requirement to choose the sampling period and quantization of computer input and output variables. The sampling period is chosen primarily on the basis of the desired system bandwidth. If, for example, it is decided that the first bending mode is to be stabilized and higher frequency modes isolated as much as possible, the sampling frequency would be chosen perhaps five times greater than the first bending frequency. If the form of the attitude indicator and A/D convertor permits it, insertion of a low-pass filter before the sampler would be wise in most cases to attenuate the higher frequency noise which might be modulated down into the control system pass band by the sampling process. The signal quantization levels are chosen to cause no more than some tolerable degree of graininess in the operation of the system.

An example of a digital attitude control system for a large vehicle may be cited from the Apollo program. The configuration of the vehicle as it decelerates from the translunar trajectory into the lunar circular orbit is shown in Fig. VII-7. The Lunar Excursion Module is shown attached to the Command Module in the position that allows the astronauts to move from one vehicle to the other. This configuration is rather flexible, most of the bending occurring in the coupling structure between the vehicles. The frequency of the first bending mode is about 2 cps. Fuel sloshing has little effect on the dynamics of this vehicle, the mass of fuel which participates in the sloshing oscillation being only a very small fraction of the mass of the vehicle. The analytic manifestation of this small mass ratio is the very close proximity of the zero to the pole in the sloshing pole-zero pair. Thus the zero nearly cancels the pole and both can be ignored in the analysis of this system. The attitude control system uses a sampling frequency of 25 cps and employs angle information only; no rate gyros are required. The attitude information which is the input to the digital Apollo Guidance Computer (AGC) is quantized at 40 arc sec. The output of the AGC is the command to the engine gimbal servo which is held between the sampling points; this command is quantized at 160 arc sec.

A Z plane locus of roots for the sampled system without compensation is shown in Fig. VII-8. The rigid body dynamics in the absence of atmosphere are just due to the vehicle inertia; this gives rise to the two poles at +1. Pole-zero pairs due to the first two bending modes are included. Without compensation, the first bending mode goes unstable immediately for very low system gain. The locus with lag compensation is shown in Fig. VII-9. In this case a single compensating zero has been used which corresponds to the rigid body compensation mentioned earlier, plus four poles to attenuate higher frequencies sharply. The closed-loop root locus originating at the first bending pole is seen to be pulled well into the stable region. The corresponding locus with lead compensation is

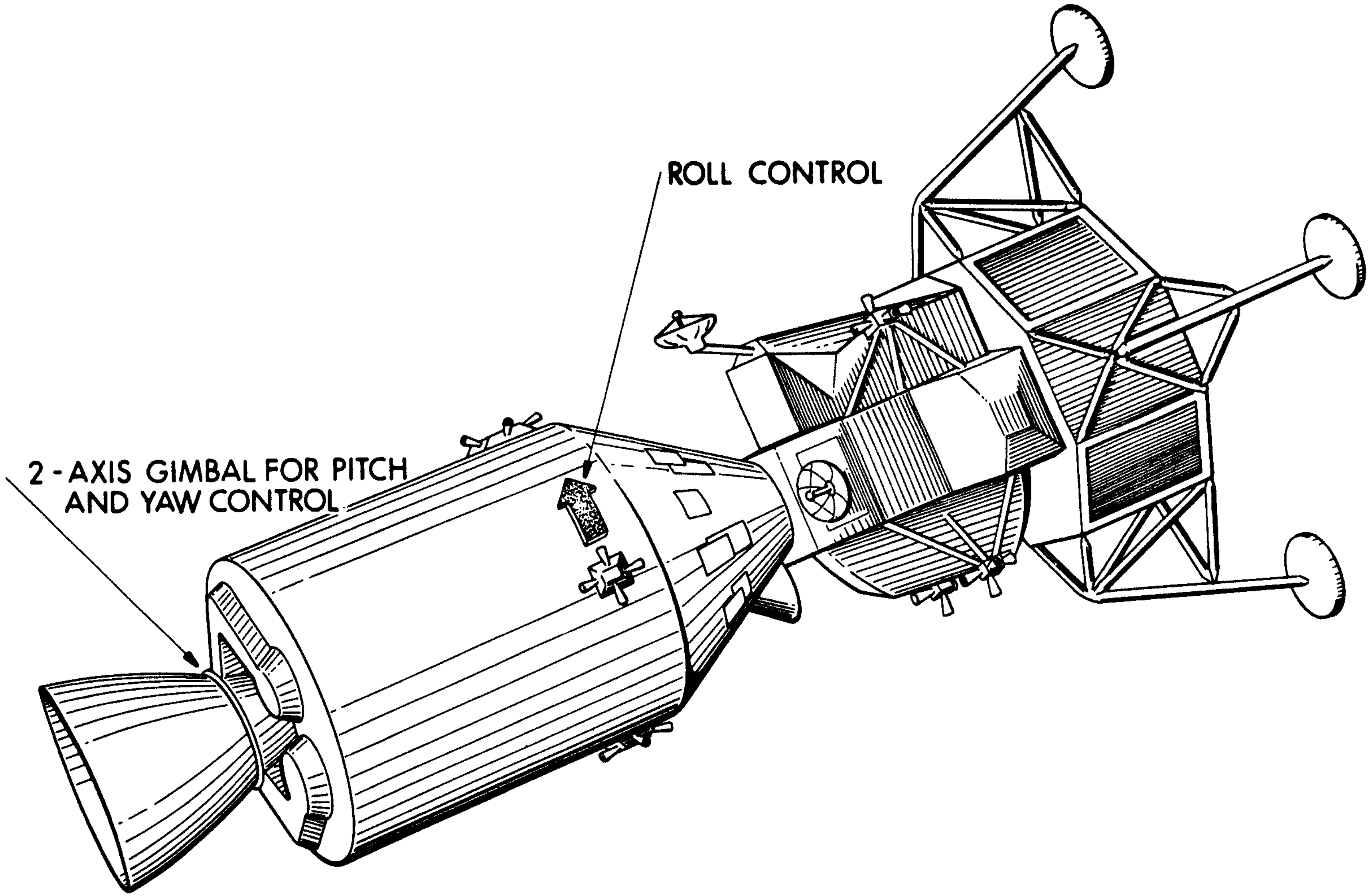


Fig. VII-7 Apollo Lunar Approach Configuration

215

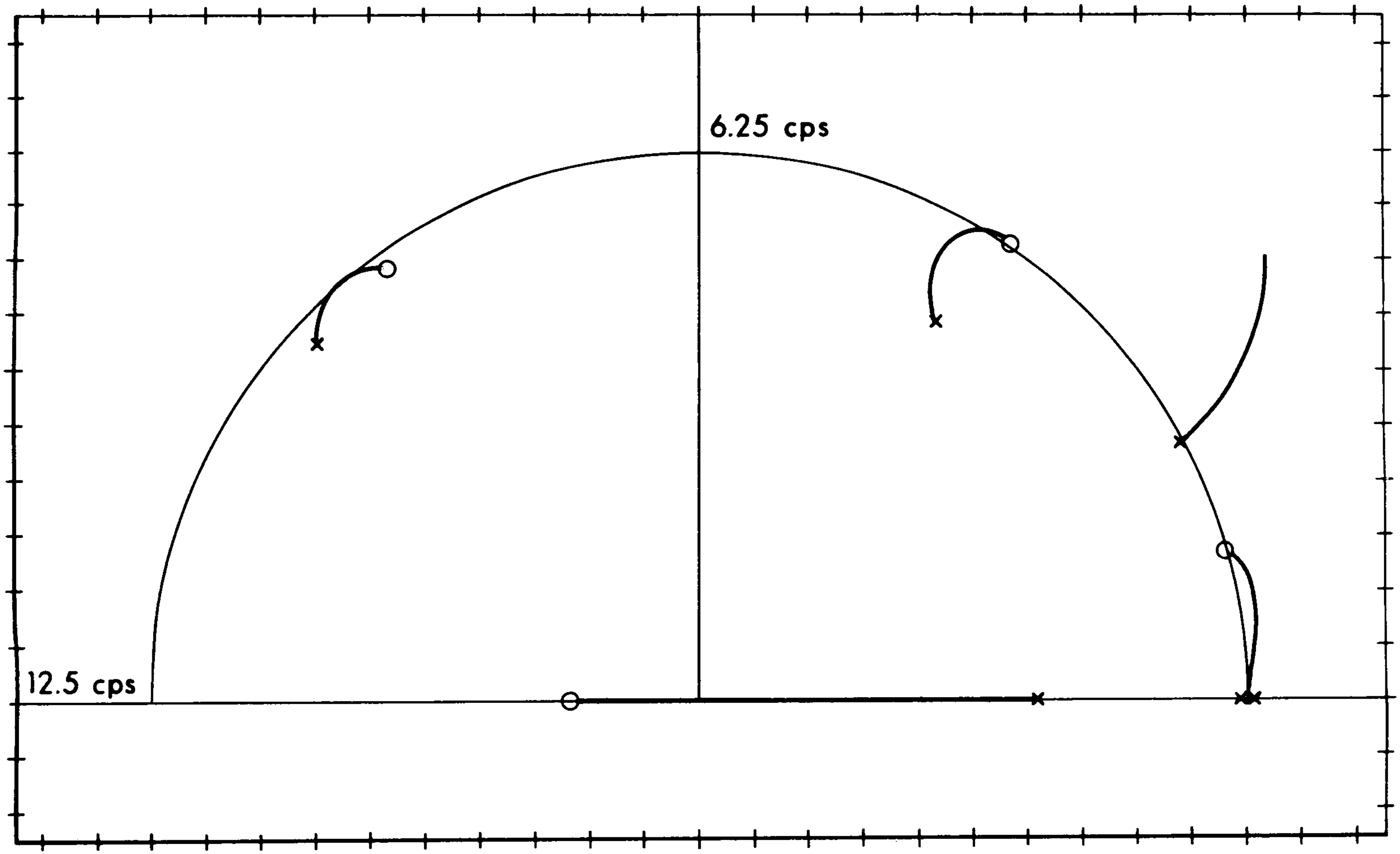


Fig. VII-8 Z Plane Locus of Roots: without Compensation

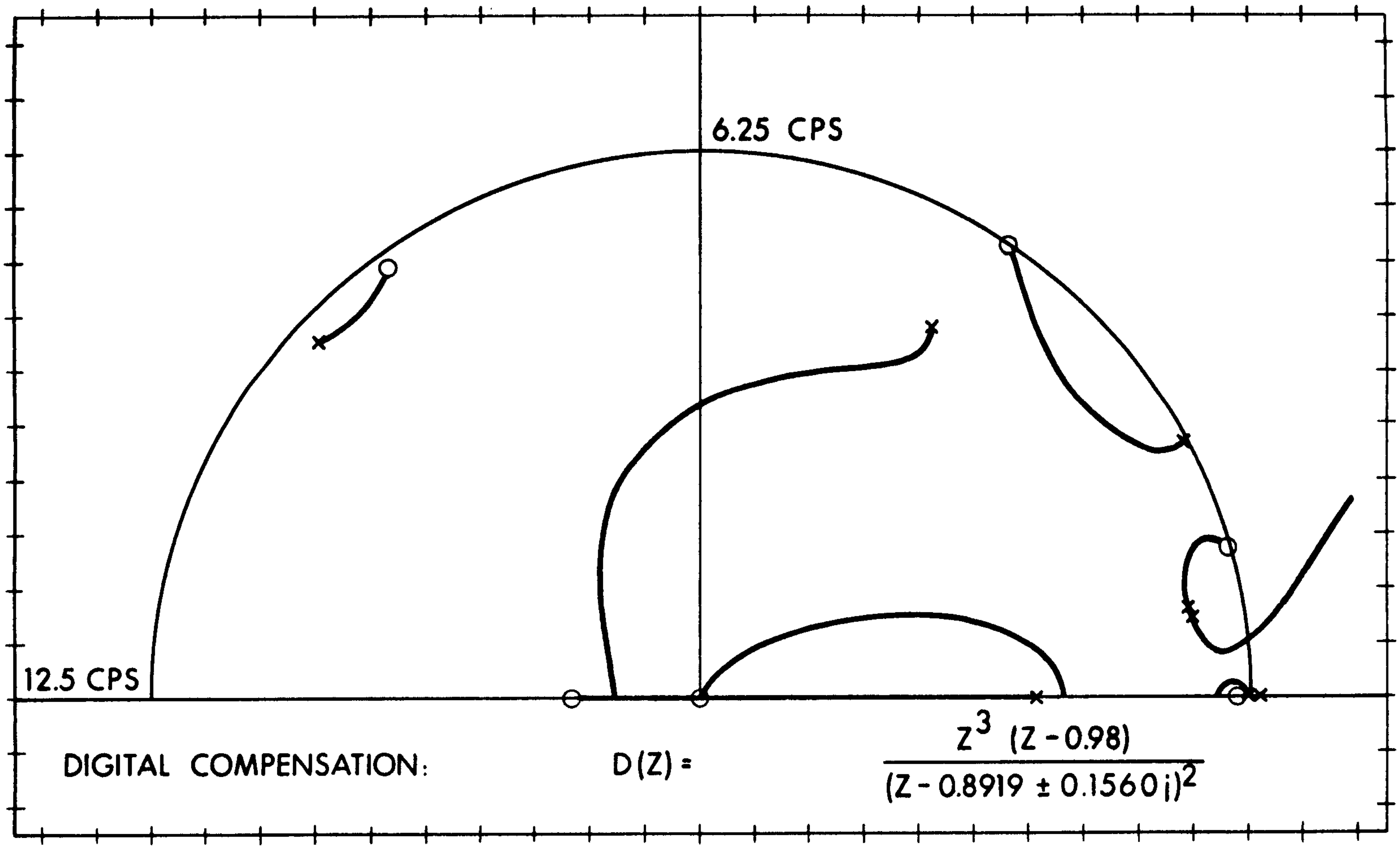


Fig. VII-9 Z Plane Locus of Roots: with Lag Compensation

217

shown in Fig. VII-10. In order to provide active damping of the first bending mode, considerable lead was required. As can be seen in the figure, compensation involving five zeros and five poles is employed for this purpose. The result shows the first bending mode to be not only stabilized but rather well damped. With each of these forms of compensation, the dominant system closed-loop poles have a natural frequency of about 1 rad/sec, roughly one-tenth the fundamental bending frequency. With lead compensation, however, there are closed-loop zeros in the vicinity of these poles resulting in a significantly shorter response time to transient inputs. The bandwidth of each system as measured by the frequency at which the magnitude of the open-loop transfer function is unity is nearly the same - about 1 rad/sec - but the lag compensated system has more phase lag at this frequency and cuts off much more sharply for higher frequencies.

The step response of this system with lead compensation is shown in Fig. VII-11. This response was recorded by a simulator which used an analog computer to simulate the vehicle dynamics but used an actual AGC as the controller. The basic system response is seen to follow quite well the transient response characteristics corresponding to the dominant poles in the Z plane. Moreover, the bending oscillation is seen to be reasonably well damped. An interesting problem is pointed up by the occasional bursts of activity after the major response transient is over. These are occasioned by the attitude error drifting far enough to cause an indicated 1 quantum of error. This is a small error and the actual error rate is very small, so very little control is required to return the error to the zero quantum zone. However, the computer has seen an indicated zero error for many sampling periods, so when one quantum of error is indicated the lead compensator interprets this change as a significant rate of change of error, with a similar interpretation of higher derivatives. This touches off an unnecessarily long period of control activity during which the bending modes are again excited.

So the problem of the attitude control of large, flexible space vehicles is indeed challenging - and it will become even more so in the future. As the vehicles get larger the bending and slosh frequencies become lower, but the performance required of the system does not necessarily decrease. Thus the requirement will be to push the system performance in terms of bandwidth closer to and beyond the bending and slosh frequencies. This can be done, but it makes the stability of the system critically dependent upon the detailed characteristics of these modes - and these characteristics are not altogether predictable in advance of the flight. Thus the interest in self-adaptive control systems for this application. It may also be that modern estimation theory will find some application to this problem. In the meantime, classical linear system analysis techniques remain very useful tools for the system designer, but the complicated character of the controlled member makes hand application of these techniques impractical. A powerful attack on the problem can be launched by preparing a series of digital computer programs to do the tasks that one would otherwise do by hand: tasks such as plotting the frequency response function for an open or closed loop system given the cascade of elements which comprise the open loop system, calculation of the sampled transfer function given the continuous transfer function, plotting the locus of closed loop roots for a sampled or continuous system given the open loop transfer function, etc. Programs* such as these have been used to carry out the design of the Apollo systems at the MIT Instrumentation Laboratory. It is difficult to see how the job could have been done on a reasonable schedule without automation of these analytic operations.

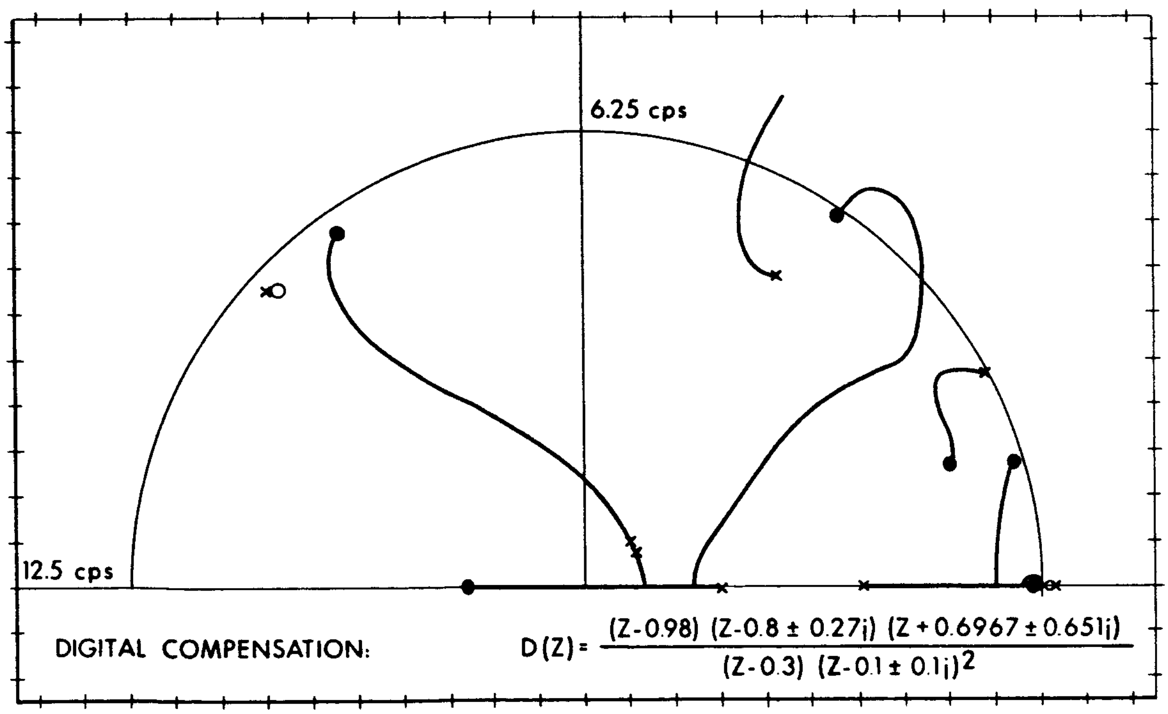


Fig. VII-10 Z Plane Locus of Roots: With Lead Compensation

* Prepared primarily by Donald C. Fraser.

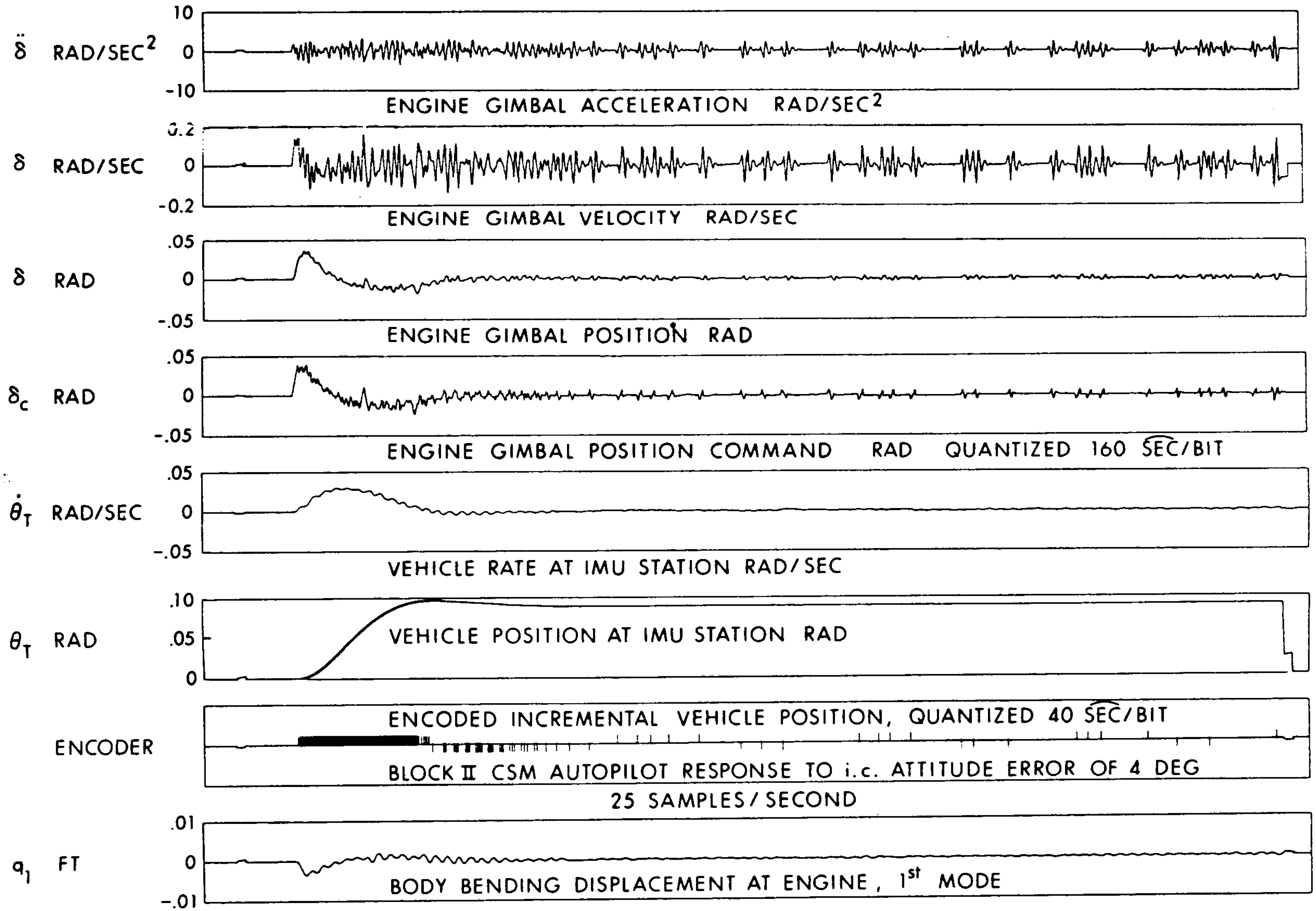


Fig. VII-11 Step Response of Attitude Control System

ATTITUDE CONTROL WITH A FIXED ENGINE

The possible elimination of the massive engine gimbal structure and associated high-power actuators and servo systems is a very appealing step in the direction of system simplicity. The alternative is a control system based on on-off switching of attitude control reaction jets. The control thrusters must in this case be capable of delivering substantial control torque to balance the possible range of thrust axis offsets for the main engine. This offset is due to a number of causes. There is some tolerance in the mounting and aligning of the engine. The thrust axis does not lie along the engine centerline due to unsymmetrical flow and unsymmetrical nozzle ablation. But the major cause is the uncertain location of the vehicle center of mass. During any one mission phase the center of mass may move appreciably as propellants are consumed from off-axis tanks. An interesting compromise design is one which has the main engine gimbaled but provided only with a low-power actuator and very simple servo, such as a constant rate drive, which is switched in sign so the thrust axis tracks the vehicle center of mass. This will permit the use of smaller thrusters for attitude control.

For the present, consider the design of an attitude control system for a vehicle with a fixed engine. This configuration would not be attractive for the control of a very large, high-thrust booster in atmosphere. For a smaller vehicle out of atmosphere the vehicle dynamics are essentially just the inertia effects, body bending and propellant sloshing being very high frequency modes. The parameters which dictate the design in the absence of thrust axis offset are the vehicle acceleration due to control moment, the minimum on time for the control thrusters, and the maximum allowable attitude error. In this situation, with the control moment designed to accommodate the maximum possible thrust axis offset, the acceleration due to control moment is likely to be very large. Also, for any thrusting system, there is a minimum on time which can reliably be commanded. The product of this control acceleration and minimum on time then gives the minimum change of vehicle angular velocity which can be commanded; call it $\Delta\dot{\theta}_{min}$. The most favorable symmetric limit cycle which can be achieved in the absence of thrust offset consists of a coast at the rate of $1/2\ \Delta\dot{\theta}_{min}$ to the specified error boundary, θ_{max}, a minimum control impulse at that point which changes the vehicle rate to $-1/2\ \Delta\dot{\theta}_{min}$, and a coast to the negative θ_{max} boundary. This limit cycle is shown in the phase plane in Fig. VII-12. The angle at which the control moment is switched on and off is

$$\text{Switching angle} = \pm\left(\theta_{max} - 1/8\ \alpha_c t_{min}^2\right) \qquad \text{(VII-1)}$$

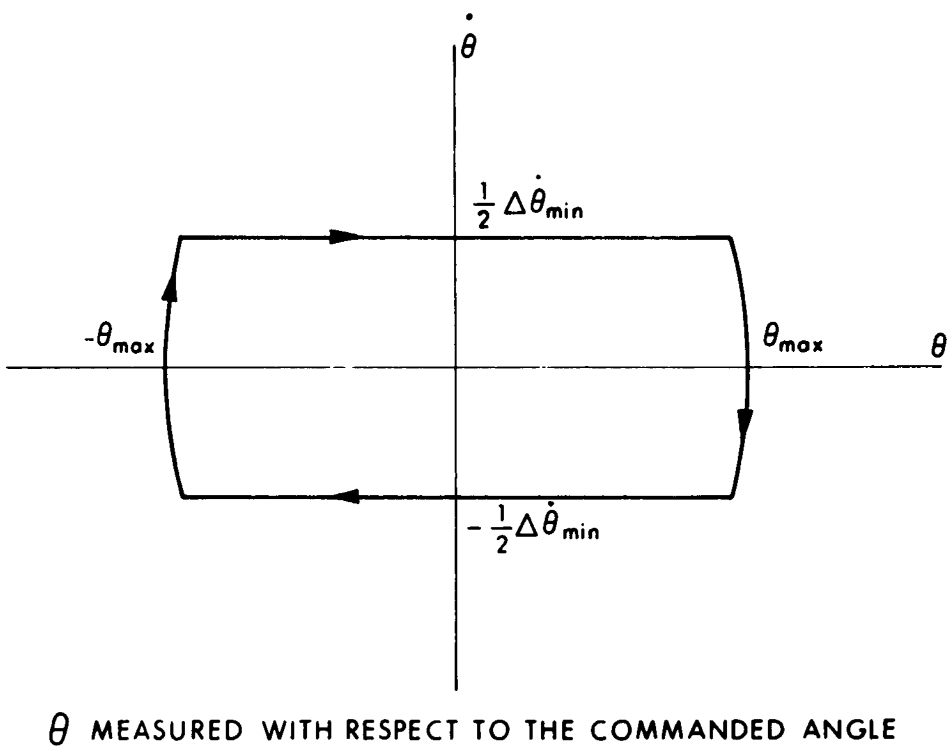


Fig. VII-12 Minimum Symmetric Limit Cycle with no Thrust Offset where α_c is the angular acceleration due to the control moment. The thruster duty cycle is

$$\text{Duty cycle} = \frac{2}{\dfrac{8\,\theta_{max}}{\alpha_c\, t_{min}^2} - 1} \qquad \text{(VII-2)}$$

If, for example, the maximum allowable error is 1 deg, the angular acceleration due to control is 10 deg/sec^2, and the minimum control on time is 10 milliseconds, the switching angle is 0.9999 deg and the duty cycle is 2.5 10^{-4}. With this large a control acceleration, the limit cycle looks rectangular on the phase plane. The thruster duty cycle is quite favorable -- the thrusters being on only 0.025 percent of the time. The vehicle coasts for 40 sec, thrusts for 10 millisec, and coasts for 40 sec again. This is of course an idealization which assumes constant input, perfect information, and no disturbances. The same performance would result with a constant rate input.

This limit cycle can be achieved in the absence of thrust offset using standard switching logic. Figure VII-13 shows a common system configuration in which a linear combination of error and error rate controls the thrusters through a trigger circuit or switching logic which includes a dead band and hysteresis. The indicated derivative can be thought of either as an idealization of a lead network or the effect of rate gyro feedback. Given the magnitude of the control torque, this system is designed by choosing values for the free parameters k, δ_1, and δ_2. The switching logic is assumed symmetrical. Realization of the limit cycle shown in Fig. VII-12 places two constraints on these free parameters; the two switching lines on each side of the origin must pass through the corners of the desired phase trajectory. There remains a degree of freedom yet to be specified. This freedom can be used to achieve fast recovery from step changes in command of probable magnitude. If k is very small, and δ_1 and δ_2 are chosen to yield the desired steady-state limit cycle, the response to a change in command is poorly damped. Such a phase trajectory is shown in Fig. VII-14. If k is very large, and again δ_1 and δ_2 are chosen to yield the desired steady state limit cycle, the response to a change in command is slow and sluggish as indicated in Fig. VII-15. For any given initial error there is a choice of k (and δ_1, δ_2) which will carry the response into the limit cycle immediately as indicated in Fig. VII-16. The magnitudes of command changes are not predictable, of course, but some are more likely than others -- and k (and δ_1, δ_2) can be chosen to yield good response in the most probable cases.

Having such a system designed to yield good step response for inputs of likely magnitude, and an efficient limit cycle, it remains to be seen how the performance changes with thrust offset. The performance can be quite adversely

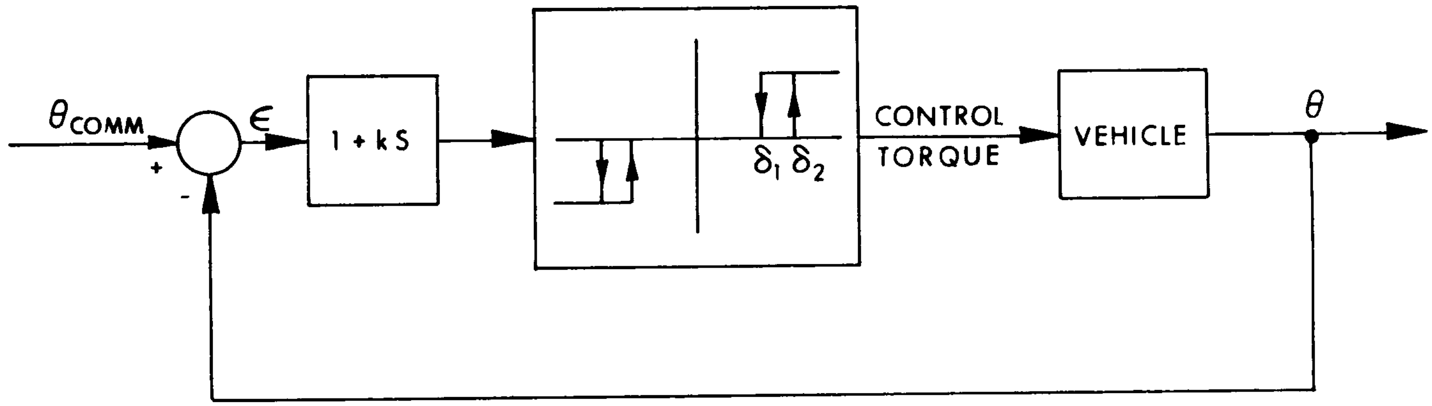


Fig. VII-13 Standard On-Off Control System

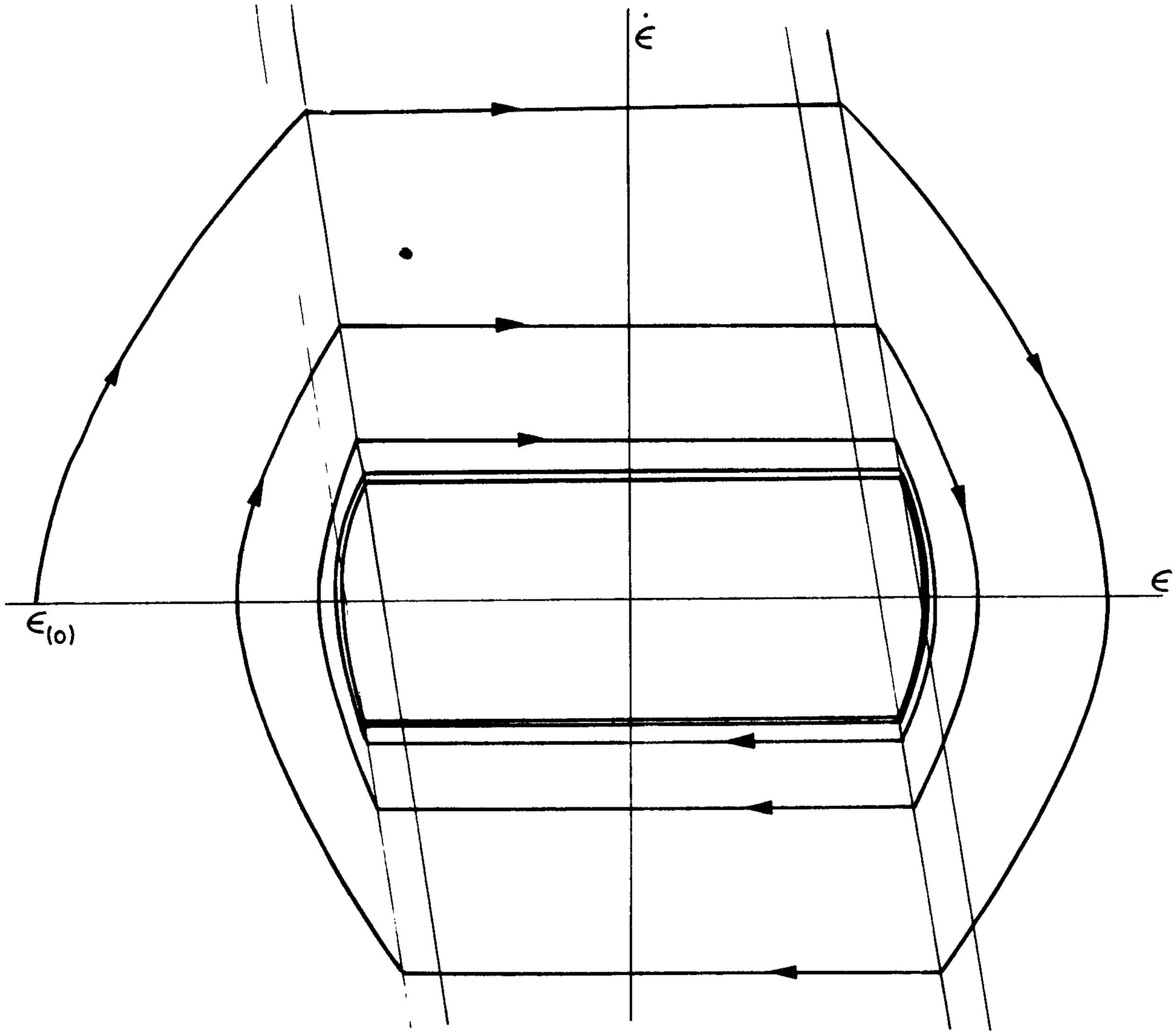


Fig. VII-14 Step Response with Small Rate Gain

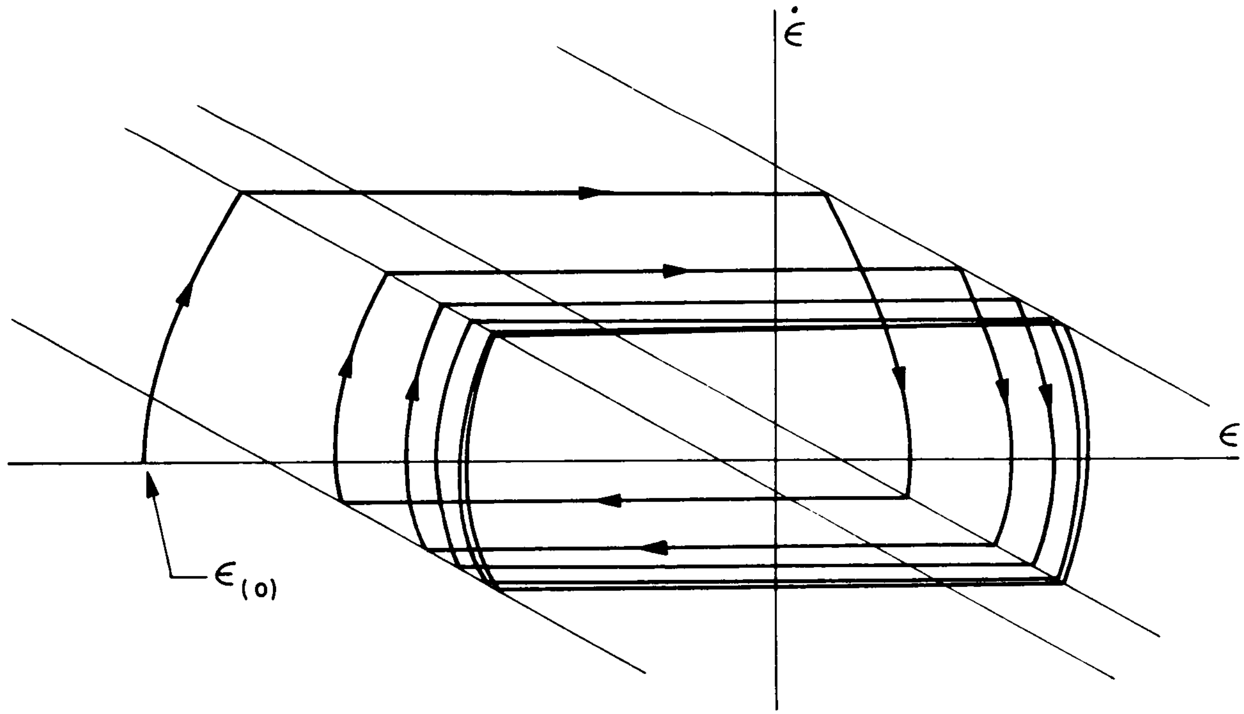


Fig. VII-15 Step Response with Large Rate Gain

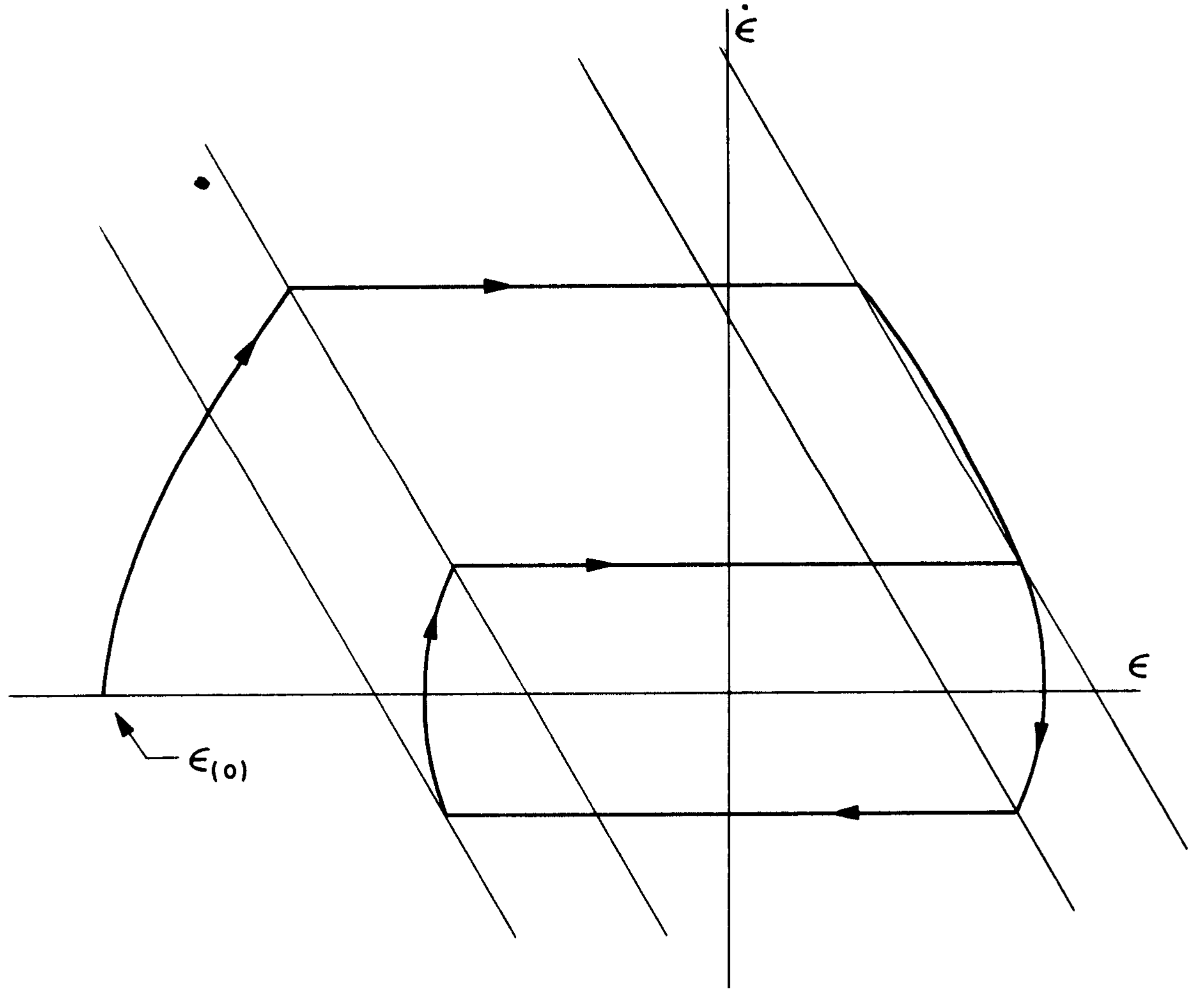


Fig. VII-16 Step Response with Proper Rate Gain affected as suggested by Fig. VII-17. This figure shows one form of limit cycle which may result from a positive offset moment -- which tends to drive the error rate negative. The trajectory is shown starting from the switching line in the first quadrant. A minimum positive control impulse is commanded at that point, which acting with the offset torque drives the vehicle to a large negative rate. The negative acceleration then continues due to the offset torque alone until the trajectory hits the lower switching line. From that point the phase trajectory "chatters" between the switching lines moving to the left and approaching a limiting closed cycle. This steady-state operation is undesirable both because of the negative error bias and because of the rapid on-off cycling of the thrusters which is wasteful of fuel.

If the amount of the thrust offset were known, it would be easy to command a much more desirable limit cycle. The thrusters would never be used to add to the offset torque unless the offset were below some threshold. Rather, one portion of the cycle as seen in the phase plane would be the parabola due to the offset moment alone which just stays within the specified error bound. The other portion would be the parabola due to the control moment acting against the offset moment and which also stays just within the specified error bound. Such a trajectory is shown in Fig. VII-18. With the vehicle dynamics modeled as just an inertia, the required conditions for switching the control moment on and off are easily and simply expressed. The remaining requirement is an indication of the offset moment in addition to the error and error rate. It is clear that information about the offset moment is contained in the time history of attitude which results from the known history of control moment. If, for example, one took an estimate of the offset moment and applied it together with the known control moment to a model of the vehicle dynamics, and later found that the attitude indicated by the model tended to grow more positive than the attitude of the actual vehicle, this would be evidence that the estimated offset moment was in error in the positive sense. This information could then be fed back to adjust the estimated offset moment.

This is the physical concept which underlies optimal linear estimation theory as formalized primarily by Kalman[2]. The form of the Kalman estimating filter is shown in Fig. VII-19 where only one quantity is shown as being measured and several parameters may be estimated. It is also possible to process more than one measurement at a time. A model of the system is used to generate a prediction of the quantity being measured based on the current estimates of all parameters being estimated. The difference between the predicted measurement and the actual measurement feeds back through gains to alter the parameter estimates. A gain computer varies the estimate adjustment gains in an optimal manner depending on the nature of the measurement and the statistics of the measurement errors and of the uncertainties in the prior knowledge of the estimated quantities. In the present application, the

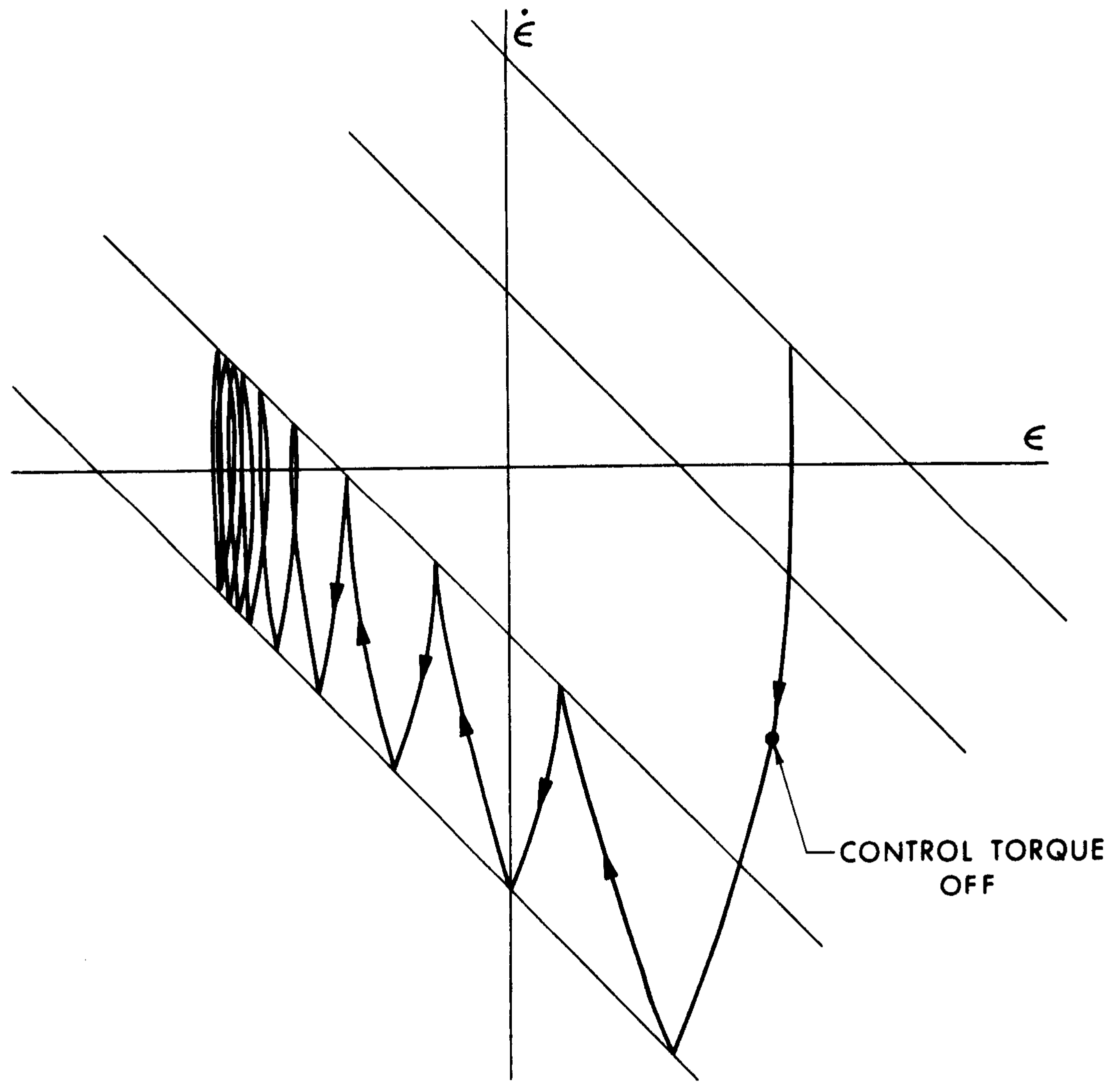


Fig. VII-17 Limit Cycle with Thrust Offset

229

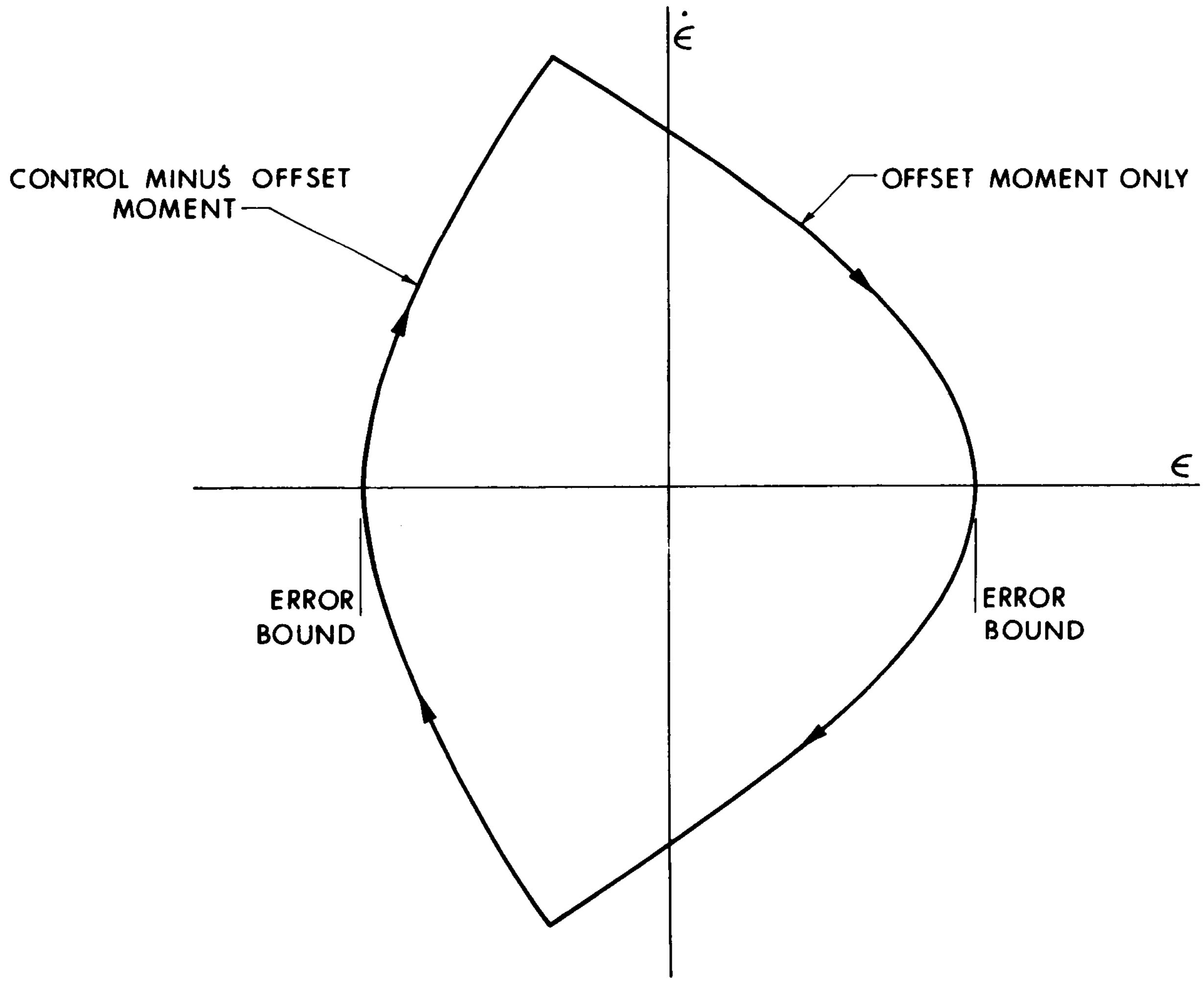


Fig. VII-18 Desirable Limit Cycle with Thrust Offset

220

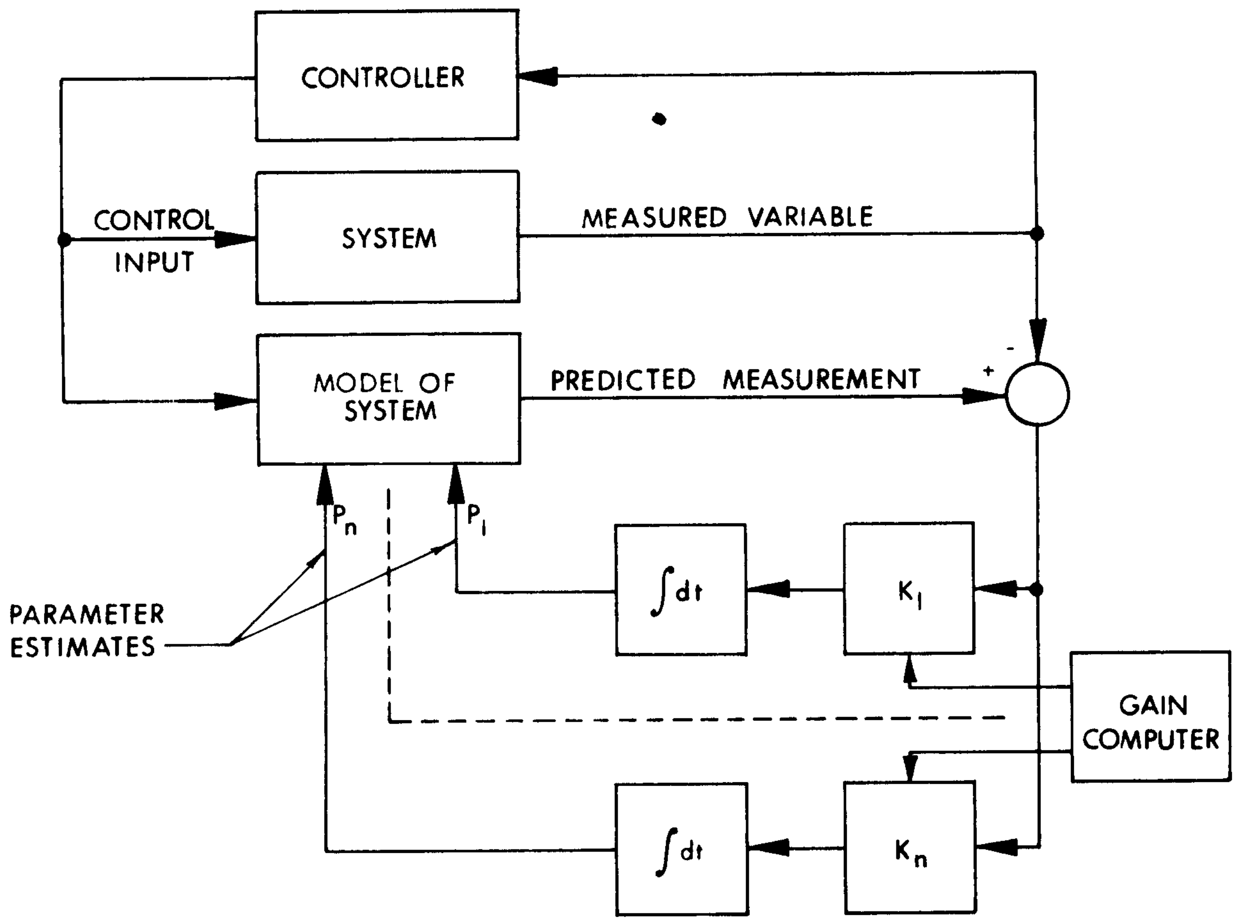


Fig. VII-19 Form of the Kalman Estimator measured quantity is vehicle attitude and the estimated parameters would include vehicle attitude, attitude rate, offset moment, and possibly vehicle inertia. Vehicle attitude is included as an estimated parameter even though it is the measured variable since the measurement is subject to noise or uncertainties. With an appropriate model of the system being of such simple form in this case, the major computational load is the gain computation. But the optimally varying gains converge quickly to steady-state values, and in many mission situations, little would be lost by using constant gains. If that is the case, then estimates of all the quantities needed to permit efficient control with arbitrary thrust offsets can be generated with very little required computation.

The resulting system can of course be instrumented with either analog or digital computation. Because of the rapid response of the vehicle to control and offset moments, rather high sampling frequencies would be required if a digital computer were used in the standard way in which control action can be taken only when the computer samples the attitude and processes the control equations. For example, in a limit cycle of the form of that shown in Fig. VII-18 corresponding to a control acceleration of 50 deg/sec^2, an offset acceleration of 10 deg/sec^2, and an error tolerance of 0.5 deg, the control torque is on for a period of 0.2 sec each cycle. If this time is to be resolved with no more than 10% error by the basic computer samples, a sampling frequency of 50 samples/sec would be required. This could impose an appreciable load on the computer. However, if the computer is organized (as is the Apollo Guidance Computer) to count input pulses while processing other calculations, and also can be interrupted briefly when a certain count has been reached to issue a discrete output, then control action can be taken at times other than the basic computer sampling times. With this capability, a slower sampling frequency can be used. Each time the attitude is sampled and the estimates of parameters updated, the computation can predict whether control action (either on or off) should be taken before the next sampling point. If so, the time till that event or the change in attitude till that event can be read into a counter and counted down. Each time a thruster is turned on, the time it should be on is calculated and read into a counter so it can be turned off again between sampling times.

A fixed-position rocket engine is used in the ascent stage of the LEM. For the purpose of standardization, the same hypergolic thrusting system used on the Command and Service Module is also used on the Lunar Excursion Module. Toward the end of the ascent boost the response of the LEM to this control moment is quite lively -- about 50 deg/sec^2. The moment due to thrust axis offset from the vehicle center of mass may be as much as half the control moment. Vehicle attitude is derived from the IMU gimbal angles and is indicated by the Coupling and Display Unit with a quantization of 40 arc sec. No rate gyros are required by the primary attitude control system. Simulation of this system using digital control of standard switching logic has indicated that a sampling frequency of about 40 samples/sec would be required. Even at that sampling rate the effects of timing errors are clearly evident. A system using a Kalman filter to estimate attitude, attitude rate, and thrust axis offset has also been simulated.* The model of vehicle dynamics used in the filter is just an inertia for each axis. This is an incomplete model of three-axis vehicle dynamics, but it is quite adequate. Different modes of control are programmed depending on the magnitude of the estimated error and the estimated offset. For steady-state operation with the offset above a threshold value, limit cycles of the desired form shown in Fig. VII-18 are commanded. In each case, when a control moment is commanded, the time interval during which it should be on is calculated and read into a counter to be counted down by clock pulses. With this form of control, a sampling rate of 10 per sec yields excellent performance.

ACCELERATION VECTOR CONTROL

The purpose of the control system during powered flight is not just to control the vehicle attitude but rather to control the thrust acceleration vector, $\underline{a_T}$, in response to commands from the guidance system. In fact, it is possible to effect the $\underline{a_T}$ control directly without even feeding back attitude information, but a system which uses velocity information only requires complicated compensation which must be varied rather carefully with the changing flight condition and vehicle characteristics. It seems clearly preferable in most cases to use attitude feedback to stabilize the vehicle. Having a well-behaved attitude control system to work through, the design of the $\underline{a_T}$ controller is not difficult.

One approach is open-loop in nature. Having the desired direction of a_T as an input from the guidance system, one can interpret this simply as an attitude command and orient the longitudinal axis of the vehicle in the desired direction for $\underline{a_T}$. The discrepancy in so doing is that the thrust acceleration vector is not necessarily aligned with the longtudinal vehicle axis. In addition, if the vehicle center of mass moves off the longitudinal axis, the attitude control system develops a forced error large enough to point the average $\underline{a_T}$ direction through the offset center of mass location. For a vehicle with a gimbaled engine, the resulting angular error about one axis is

$$\begin{array}{l}\text{Angle between actual } \underline{a_T} \text{ direction} \\ \text{and desired } \underline{a_T} \text{ direction}\end{array} = (1 + \frac{1}{K})\,\delta_{ss} \qquad \text{(VII-3)}$$

where K is the attitude control system forward gain - the gimbal angle per unit attitude error - and δ_{ss} is the steady-state engine deflection required to point $\underline{a_T}$

* This work done by George W. Cherry.

through the vehicle center of mass. The use of integral control in the attitude control system makes K in effect infinite, but the angular deviation between the direction of $\underline{a_T}$ and the vehicle longitudinal axis cannot be accounted for directly by an $\underline{a_T}$ control system which commands attitude in an open-loop manner. The resulting offset $\underline{a_T}$ direction feeds back through the guidance system to influence subsequent $\underline{a_T}$ commands, but there does result from this a forced guidance error at cutoff proportional to the $\underline{a_T}$ directional error indicated above. In some mission situations this forced error may be tolerable. If the powered flight phase under consideration is followed by another guided phase, such as a translunar or transplanetary midcourse phase, the cost of the forced error will very likely be measured in terms of the fuel required to accommodate the error in the subsequent phase. In other situations the inaccuracy resulting from the forced guidance error is of more direct consequence.

Such a forced error can be eliminated, if desired, by commanding the attitude control system through a closed-loop scheme in which the actual $\underline{a_T}$ direction as indicated by the IMU accelerometers is compared with the desired $\underline{a_T}$ direction and the vehicle commanded to rotate at a rate proportional to the angular deviation. This can be instrumented conveniently by noting that the cross product of the indicated unit $(\underline{a_T})$ with the desired unit $(\underline{a_T})$ is a vector which gives in magnitude and direction the rotation required to carry the indicated $\underline{a_T}$ into the direction of the desired $\underline{a_T}$. If the command is taken to be an angular rate proportional to this angular error, the resulting control law is

$$\underline{W_c} = S(\underline{1a}_{ind} \times \underline{1a}_{com}) \tag{VII-4}$$

where $\underline{W_c}$ is the commanded angular rate, S a sensitivity or gain to be designed, $\underline{1a}_{ind}$ a unit vector in the direction of the indicated $\underline{a_T}$, and $\underline{1a}_{com}$ a unit vector in the commanded or desired direction for $\underline{a_T}$. If a rate-responding autopilot is used, this rate command can be transformed into body coordinates and applied as the command input. Only pitch and yaw components of the commanded rate in body axes would be computed and used. If an attitude autopilot or attitude control system is used, this commanded angular rate is transformed into the corresponding rate of change of direction for the vehicle longitudinal axis using

$$\underline{\dot{1}_\ell} = \underline{W_c} \times \underline{1_\ell} \tag{VII-5}$$

which is then integrated to the desired direction for the longitudinal axis. $\underline{1_\ell}$ is a unit vector along the longitudinal axis of the vehicle. It is compared with actual vehicle attitude to provide pitch and yaw attitude errors for the attitude control systems to null. In many mission situations the computing axis system can be chosen favorably with respect to the general direction of desired $\underline{a_T}$ so some of the indicated computation can be abbreviated. Toward the end of the powered flight phase when the cut-off condition is approached, the computed direction for the desired $\underline{a_T}$ tends to change rapidly. However, $\underline{W_c}$ can be limited in magnitude or even clamped to zero for a brief period of time prior to cut-off with little loss in guidance accuracy.

This closed-loop $\underline{a_T}$ control renders the system insensitive to any static thrust axis offset or offset center of mass location. The vehicle is simply commanded to rotate until the desired condition is achieved, regardless of what vehicle attitude is required to achieve it. The forced guidance error is inthis case proportional to the rate of change of the offset angle rather than the angle itself. It is then primarily the rate of change of vehicle center of mass which designs the system sensitivity, S. This sensitivity can often be quite low; for example, in the Apollo lunar approach configuration, a gain of 0.06 rad/sec/rad is adequate.

CHAPTER VII-2

COASTING FLIGHT CONTROL

During periods of coasting flight when there are no significant forces acting on the spacecraft, control over the vehicle implies attitude control only. In mid-course flight through free space the most common requirements for attitude control result from the requirements for:

Sun pointing - Very often a space vehicle is designed to operate most of the time with one face pointed to the sun. This allows efficient operation of solar cell panels and heat radiators and provides a predictable heat load on the vehicle which eases the thermal control problem. In addition, the sun line constitutes the most easily acquired and identified reference direction in solar space.

Antenna pointing - A communication or telemetry antenna oriented toward Earth is an almost universal requirement. This requirement can be accommodated by vehicle roll control around the sun line plus one degree of freedom of the antenna with respect to the vehicle.

Orienting for navigation sights - Depending upon the design of the on-board optical equipment, vehicle reorientation may be required in taking navigation sightings. Measurement of the angle between a star line and the line of sight to a landmark on a near body, for example, requires orienting the sextant precision drive axis normal to the plane of the measurement. This is a two-degree-of-freedom specification. If for reasons of mechanical simplicity and instrument accuracy only one degree of freedom of the precision drive axis with respect to the vehicle is provided, the additional degree of freedom must be provided by vehicle reorientation about one axis.

Orienting for thrusts - Prior to starting the spacecraft propulsive engine for a thrusting period, the vehicle is oriented with its longitudinal axis in the direction along which velocity is to be gained. This requires a two-axis vehicle reorientation.

The attitude references usually used during midcourse coasting flight are Sun and star lines during quiescent periods and an inertial reference during re-orientations. A two-axis Sun sensor serves as the reference for control of one axis of the vehicle along the Sun line. The degree of freedom in roll about the

Sun line is controlled against a reference provided by a star tracker. The star Canopus is a popular choice for the reference star because of its location well out of the ecliptic plane. The Canopus line thus makes a large angle with respect to the Sun line as seen from a vehicle travelling near the ecliptic, and it is not confused by neighboring stars of comparable magnitude. Orienting for a navigation sight or a thrust requires giving up one or both of these optical references. An inertial reference must then be provided. The accuracy required of such a re-orientation is not so demanding as to require the use of the navigation system IMU. A set of body-mounted single-degree-of-freedom integrating gyros is a convenient and simple alternative. If these gyros are used as pulse-rebalanced instruments, communication with the central digital computer is particularly easy. A two-axis reorientation is accomplished by rotating from the reference orientation through a prescribed angle about one axis followed by a prescribed rotation about a second axis. The vehicle control system holds the third gyro at null during this process. The reference orientation can then be recovered by executing negative rotations in the reverse order.

The sources of control moments during midcourse space flight are reaction jets or momentum exchange systems. Reaction jets can employ either cold or hot gas - hypergolic thrusters being preferable for the larger vehicles. Cold gas systems, most commonly employing pressurized nitrogen, have been used in many missions, but they have a restricted specific impulse cability - about 60 to 80 sec - and require even more weight in tankage than the weight of control gas stored. On-off control systems for vehicles using reaction jets are of the same type as those discussed in the preceding chapter under "Attitude Control with a Fixed Engine". As described there, the thruster duty cycle in steady-state limit cycling with no external moments acting on the vehicle depends on the minimum impulse which can reliably be commanded and the maximum attitude error which can be tolerated. The presence of an external moment, such as a moment due to unbalanced solar pressure forces, can actually be used to advantage to further reduce the rate of control fuel consumption. But this saving can only be achieved if the control system is capable of recognizing the presence of the moment and only thrusting against it - waiting for the external moment to return the attitutde error to the limit.

Momentum exchange systems derive moments either by accelerating wheels about their spin axes or by precessing wheels which are maintained continuously spinning as gyros. In either case it is possible to design a linear control system around such a torque generator. Steady-state operation of such a system does not necessarily include a limit cycle, and very accurate attitude control - essentially

37

as accurate as the basic attitude reference - can be maintained if needed. On the other hand, there is an upper limit to the amount of angular momentum such a system can exchange. The saturated condition is represented by an inertia wheel spinning at its maximum speed in the case of the accelerating wheel system, and by the gyros having turned through 90 degrees in the case of the precessing gyro system. Control saturation must be prevented either by adjusting the attitude or configuration of the vehicle to prevent an external moment from acting in the same sense over a long period of time, or by applying a moment to the vehicle with a reaction jet system in such a sense as to move the momentum exchange controllers away from the staturated condition.

A number of forms of control logic have been devised for use with reaction jet controllers in an effort to achieve some approximation to linear, or proportional, control. The simplest of these are pulse rate modulation and pulse width modulation. In the former case, control torque pulses of constant width are commanded to recur at a rate proportional to a control signal which may be attitude error or vehicle-rate-damped attitude error. The major disadvantage of this logic is the high pulse repetition rate at large errors when it would be less wearing on valves and more efficient in fuel use to simply leave the control jets on. In the latter case, control torque pulses are commanded to recur at a fixed rate - the width of these pulses being proportional to the control signal. The disadvantage of this scheme is the repeated operation of the thrusters even at small errors when it would be preferable to leave them off.

More sophisticated schemes attempt to relieve these disadvantages through additional data processing. One such scheme due to R. A. Schaefer[3] is known as the Schaefer modulator or pulse ratio modulator. The control logic is pictured in Fig. VII-20. The variable x(t) shown in that figure is a function of the rate-damped attitude error signal; it must range between 0 and + 1, and is often taken in the form shown in Fig. VII-21. In operation, the control logic moves repeatedly around the flow graph of Fig. VII-20, turning the appropriate jets on long enough for the integral of 1 - x(t) to accumulate to T_{min}, then turning the jets off long enough for the integral of x(t) to accumulate to T_{min}. Clearly, if the control signal remains within the dead zone of Fig. VII-21, the jets would remain continuously off, and if the control signal remains in the saturated regions of that figure, the jets would remain continuously on. For intermediate values of the control signal, the control jets are pulsed on and off, the pulse width, pulse rate, and pulse duty ratio all varying with the control signal. If x(t) is treated as quasi-stationary, the operating characteristics are:

$$\text{Pulse width} = t_{on}$$

$$= \frac{T_{min}}{1 - x} \qquad \text{(VII-6)}$$

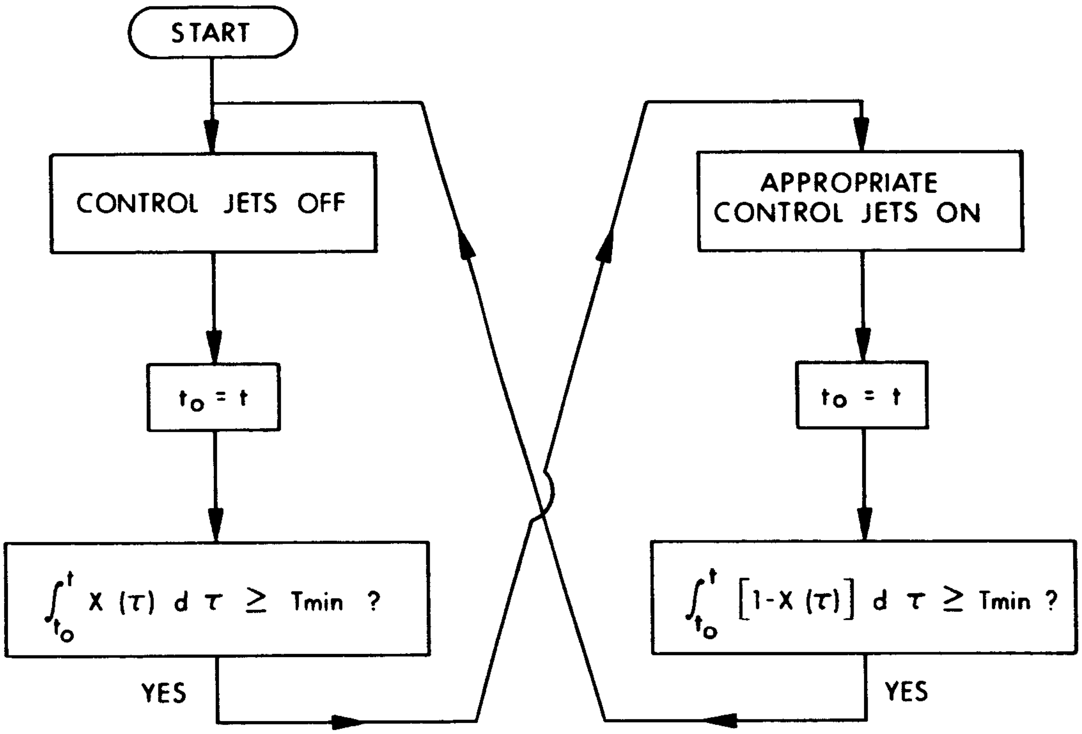
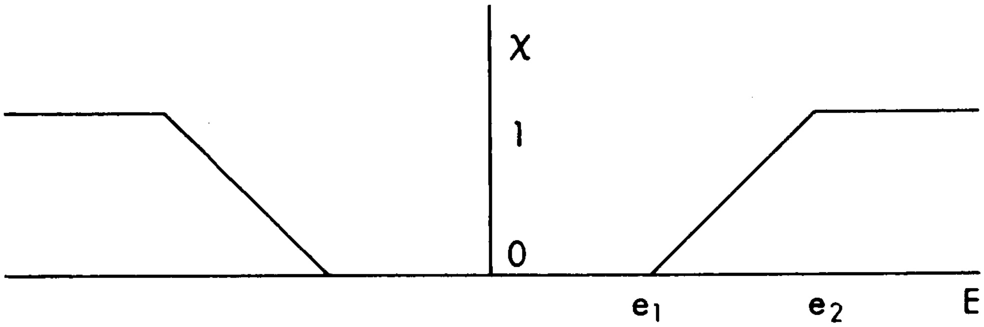


Fig. VII-20 Control Logic for Schaefer or Pulse Ratio Modulator



$$E(t) = \theta_{comm} - \theta(t) - k\,\dot{\theta}(t)$$

Fig. VII-21 Typical Definition of x(t)

$$\text{Pulse rate} = \frac{1}{t_{on} + t_{off}}$$

$$= \frac{1}{T_{min}} \; x(1 - x) \qquad \text{(VII-7)}$$

$$\text{Duty ratio} = \frac{t_{on}}{t_{on} + t_{off}}$$

$$= x \qquad \text{(VII-8)}$$

The variable x thus measures directly the duty ratio of the control history. The pulse width and pulse rate are plotted in Fig. VII-22 for T_{min} = 10 millisec. It may be seen from these curves that for small values of x (control signals near the dead zone) the pulse ratio modulator acts essentially as a pulse rate modulator using minimum width pulses. For larger errors, the pulse width increases as desired to prevent very large pulse rates. The maximum pulse rate for quasi-static x is $\frac{1}{4\,T_{min}}$.

Even the performance of such a control logic leaves something to be desired. A typical response to a command input using the pulse ratio modulator is shown in Fig. VII-23. The control jets are turned on and off unnecessarily often before reaching the final limit cycle. If the performance of the system (its acceleration level and time lags) is known well enough, only two control pulses are needed to enter the limit cycle. A set of switching curves which achieve this is shown in Fig. VII-24. The slope of the straight-line torque-off switching lines may be varied as desired to compromise between fuel used and response time. If a digital controller is employed, switching logic such as this is easy to implement.

Choice of a momentum exchange or reaction jet system, or both, should be based on an analysis of system weight, performance, and reliability. If high accuracy attitude control is required, such as in the case of a vehicle carrying a large telescope fixed to the body, a momentum exchange system would seem to be a clear choice. Most space missions do not require exceptional attitude accuracy, however, and a reaction jet system may be selected on the grounds of reliability. The reliability advantage would be especially significant if one interpreted successful operation of a momentum exchange system to require the operation of a reaction jet system used for desaturation as well. If it is understood at the outset of a system design that accurate attitude control is costly, the vehicle and its subsystems can often be designed with a view to avoiding the need for accurate control. For example, it is easy to imagine designs for an optical measurement unit which would require precise vehicle control to track some reference line. But it may also be possible to design that unit, as in the Apollo system, to tolerate a slow drift in vehicle attitude. Rather than controlling the vehicle precisely to a reference condition, the vehicle is allowed to drift through the reference condtion - and the event of passage through the reference is noted.

An additional factor bearing on the choice of a momentum exchange or reaction jet system, especially for space missions of extended lifetime, is difficult to assess quantitatively. This is the possibility of consuming all available fuel for a reaction jet system. A momentum exchange system has the advantage of consuming a quantity which can be replenished in space - electrical energy. There is no obvious threat of running out. But a reaction jet system consumes mass, which cannot yet be replenished in space. So although the analysis may show a very low probability of exhausting all control fuel, the threat of this possibility hangs over the mission during its entire lifetime.

During its midcourse flight, the Apollo spacecraft[4] effects attitude control with a system of hypergolic rocket engines which are also used for vernier translational control when needed. They employ hydrazine and nitrogen tetroxide, or variations of them, as fuel and oxidizer. They are capable of reliable operation in pulses as short as 10 milliseconds. Sixteen of these engines are mounted on the sides of the service module in quadruple sets at 4 locations. They are normally fired in pairs to produce control couples. A variety of operational modes can be selected by the crew. These are suggested by the simplified block diagram shown in Fig. VII-25. In the primary mode, the Apollo Guidance Computer operates the jet solenoid valve drivers directly based on attitude reference information only. The analog Stabilization and Control System is used as a back-up mode. It employs both attitude reference and rate gyro information. Crew-operated modes include attitude-hold and rate-command modes in addition to direct actuation of the reaction jets by the pilot through a three-axis hand controller.

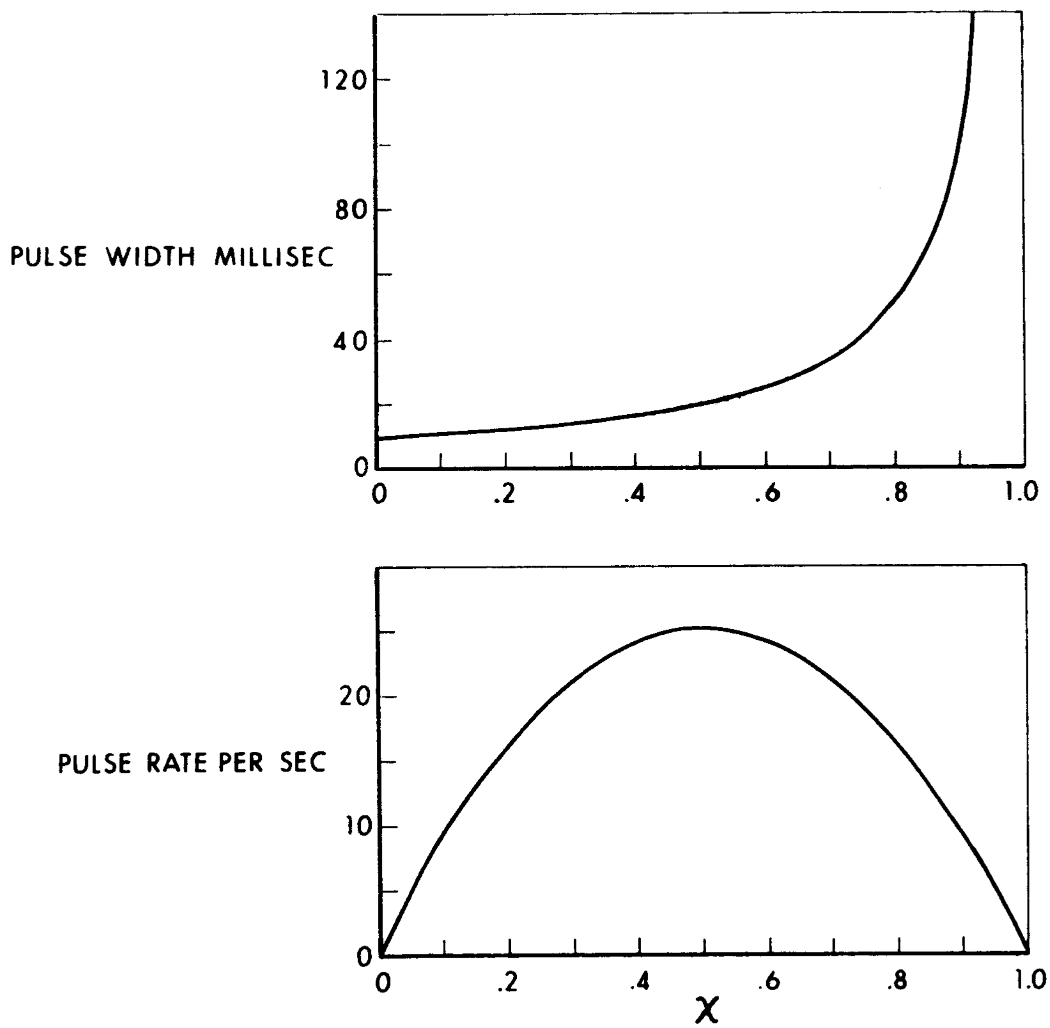


Fig. VII-22 Performance of the Pulse Ratio Modulator

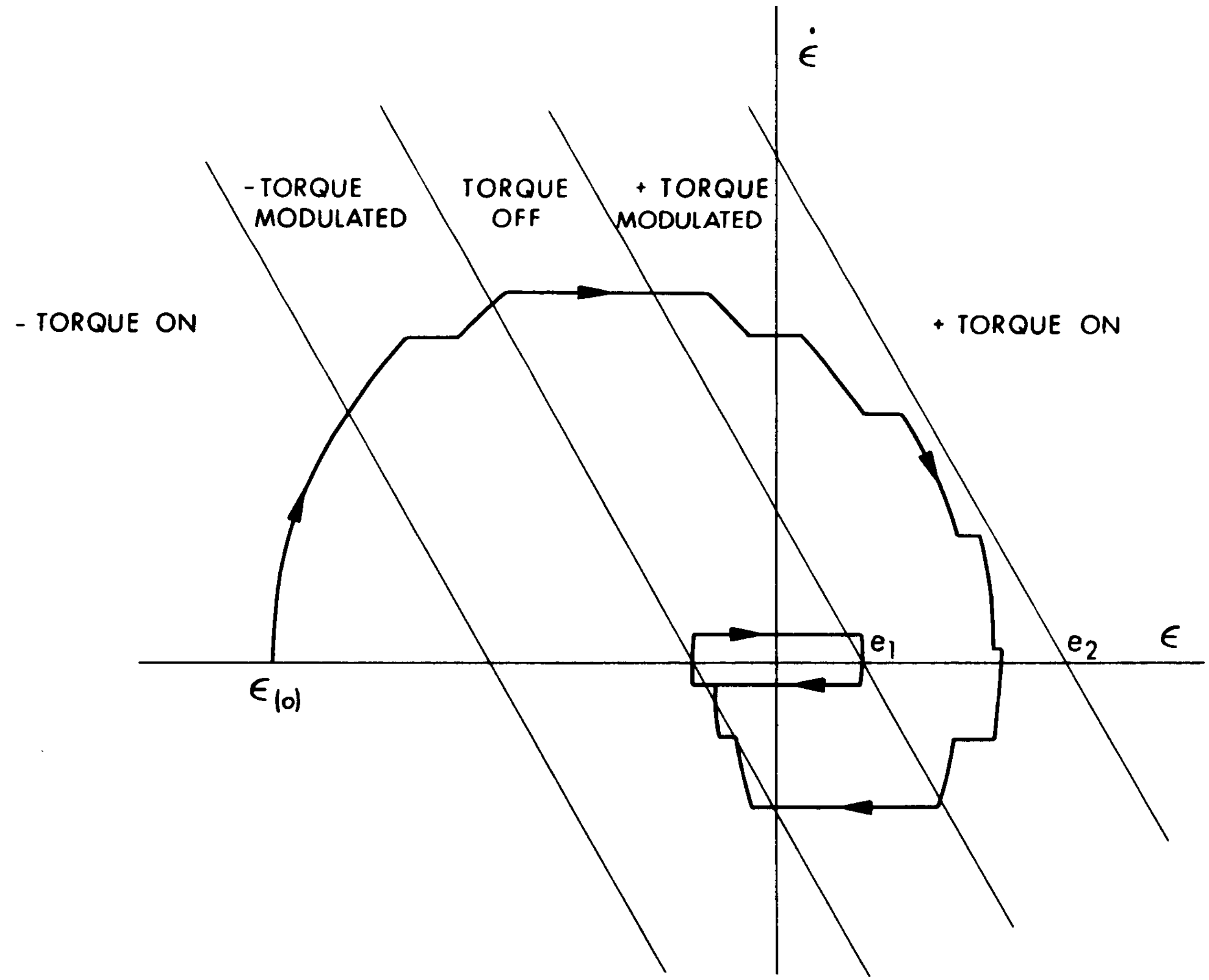


Fig. VII-23 Typical Step Response of Pulse Ratio Modulated System

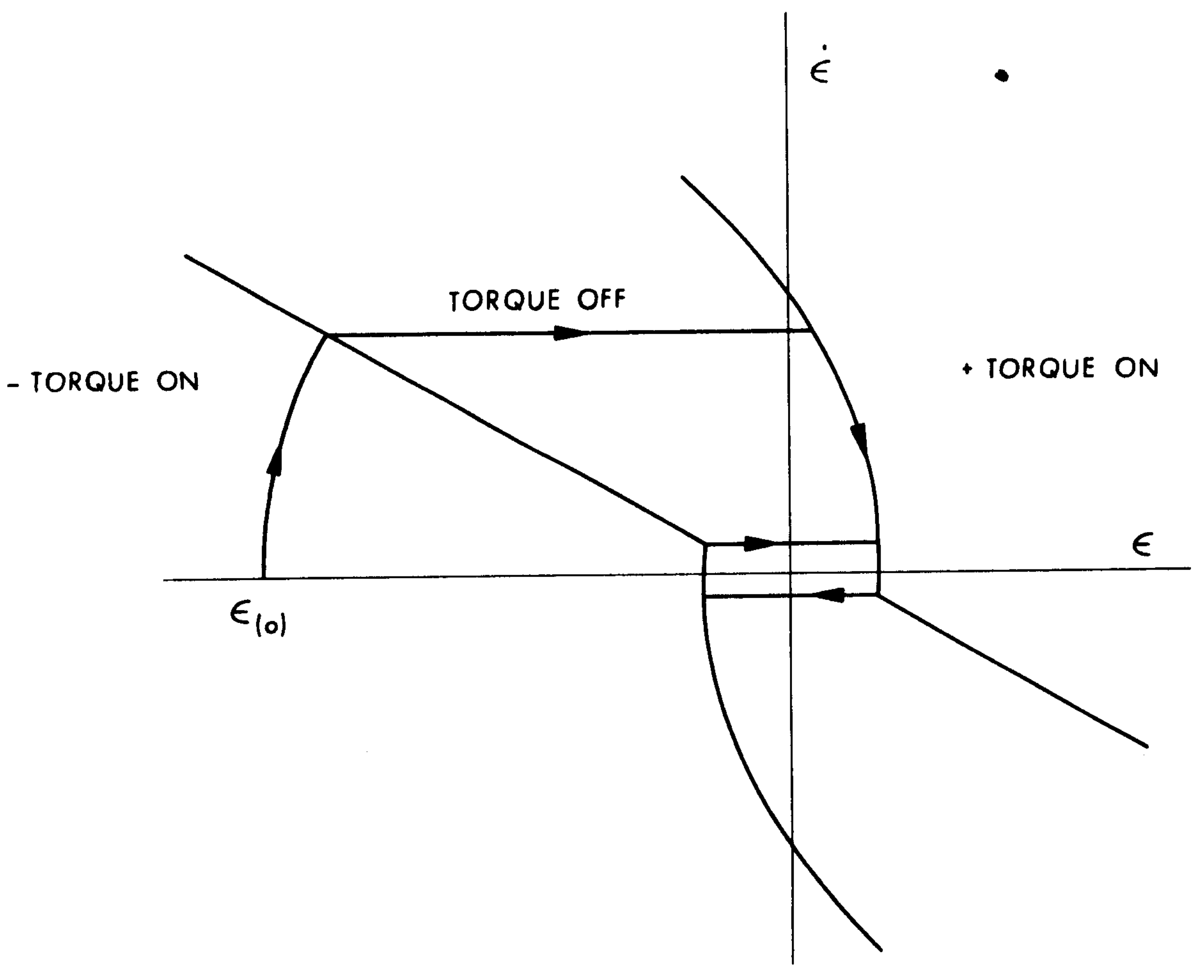


Fig. VII-24 Step Response with Straight-Line and Parabolic Switching Curves

245

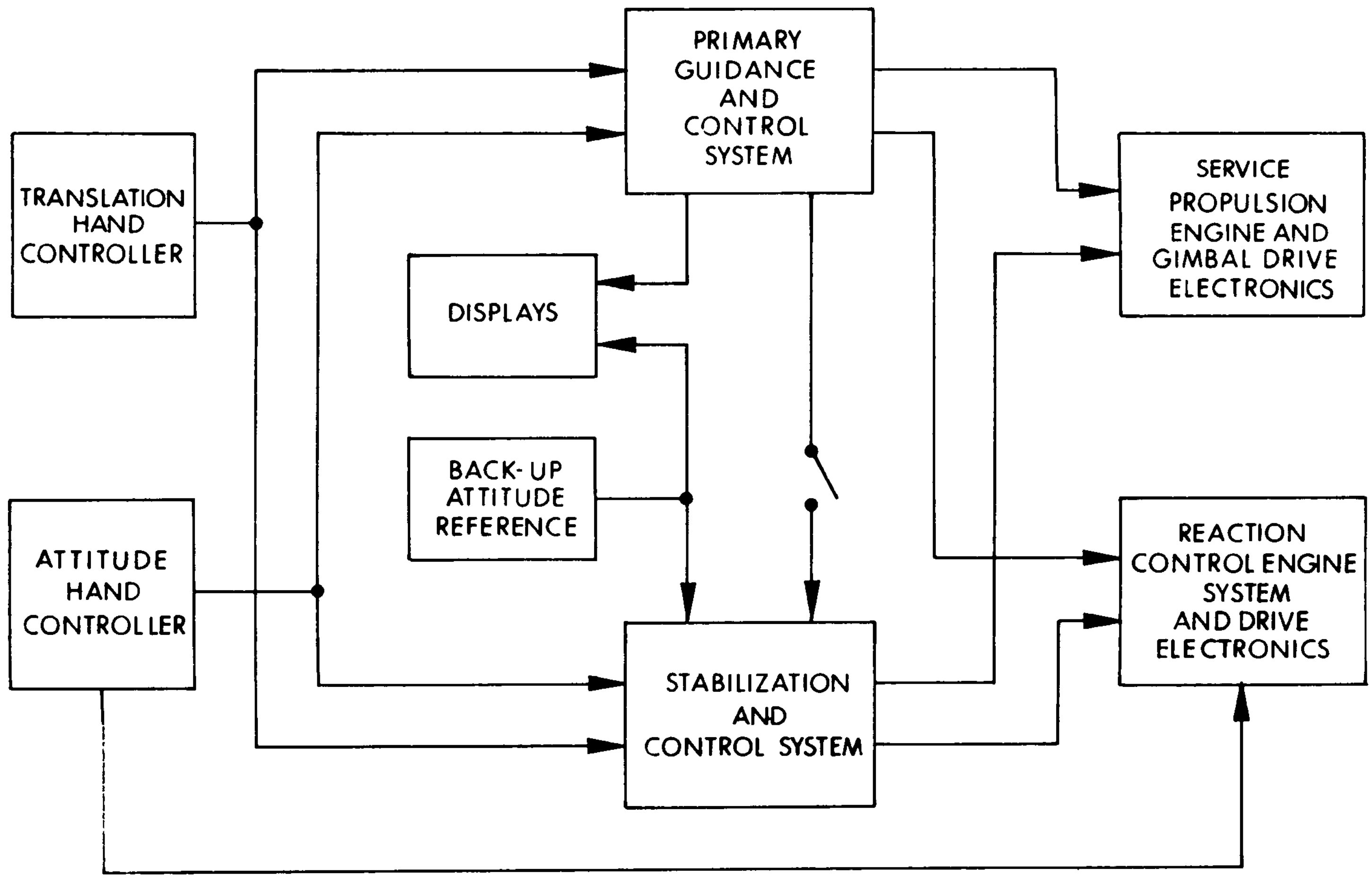


Fig. VII-25 CSM Guidance and Control System Block Diagram, Block II Configuration

CHAPTER VII-3

ATMOSPHERIC FLIGHT CONTROL

INTRODUCTION

A significantly different control problem is encountered in the high-speed atmospheric flight phase or phases of a space mission. Attention is here centered on flight following entry into a planetary atmosphere in which the path of the vehicle is controlled entirely or primarily by control of the aerodynamic force acting on the vehicle. This would include the final phase of an Earth-return mission - re-entry into Earth's atmosphere and control of the vehicle to a desired landing point. It would include a pass through the atmosphere of a planet for the purpose of reducing the energy of the vehicle's orbit relative to that planet prior to a propulsive transfer into a planetary orbit. We do not consider in this chapter the brief period of atmospheric flight which occurs at the beginning of a mission - the boost from the pad out of Earth's atmosphere. The control problem during boost is dominated primarily by the propulsive force rather than the atmospheric force and was considered earlier in Chapter VII-1.

Control over the flight path of the vehicle as well as attitude control of the vehicle itself is discussed here. Flight path control is also called guidance, and as such could well have been treated in Part III. However, the distinction between guidance and control is arbitrary, and it was thought better to consolidate in one section the discussion of the unique problems of guidance and control in the high-speed atmospheric flight situation.

FLIGHT PATH CONTROL

The discussion of this section will center on the re-entry and landing control problem. An atmospheric braking pass which involves entry into and subsequent exit from a planetary atmosphere is in fact a truncated version of a landing trajectory, which often requires a controlled skip out of the atmosphere to achieve the required range. The geometry of entry trajectories leading to a landing point in indicated in Fig. VII-26. The direction of approach relative to Earth is constrained by the nature and timing of the mission. However, when the spacecraft is a substantial distance from Earth its velocity vector is oriented nearly along the direction to Earth's center, so it is possible with little expenditure of fuel to rotate the plane of the approach trajectory arbitrarily about the local vertical line. Thus it is possible

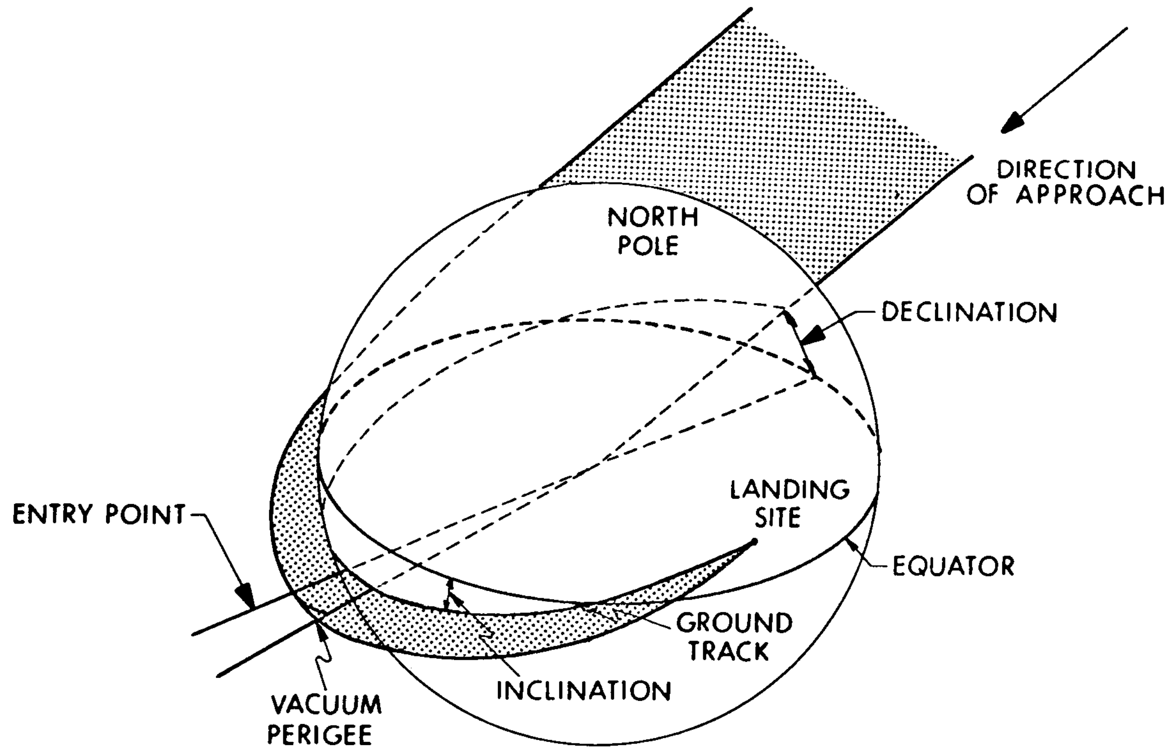


Fig. VII-26 Geometry of Entry Trajectories to orient the trajectory plane so the vehicle will fly to any selected landing point with no lateral control under nominal conditions. The trajectory plane required to achieve this is not immediately obvious because the landing point is moving due to the rotation of the Earth and the vehicle without lateral control does not fly in an inertial plane. Prior to entering significant atmosphere, the vehicle's trajectory does lie essentially in a single inertial plane. But except for winds, the atmosphere rotates with the Earth and the vehicle is gradually captured by the rotating air mass, and eventually flies essentially in a plane which rotates with the Earth. None-the-less, it is possible to determine by iteration a plane for the approach trajectory such that no lateral control would nominally be required to fly to the landing point. The required range after entry is then given by the distance from the entry point to the landing site. The entry point is defined as that point at which the approach trajectory passes through an arbitrary altitude - often taken to be 400,000 ft. The entry point thus occurs slightly before the perigee point for the approach trajectory computed as if there were no atmosphere.

This problem is further constrained by the desire to hold the azimuth of the approach to the landing site within certain bounds. This requirement is due to the desire to fly over well-instrumented areas and the desire to avoid hostile areas. This hostility may be either natural or political. Especially in a manned mission one would want to avoid over-flying certain countries and would want to avoid the colder regions of the Earth. If it were possible to select and reach a landing site nearly along the line of the approach direction, then by rotation of the approach trajectory plane about the local vertical line while the vehicle is some distance from Earth it would be possible to approach that landing point with any desired azimuth. But in many cases this would represent a severe re-entry range requirement. The other extreme is a landing point near the plane normal to the approach direction. In that case the plane of the approach trajectory must be taken to give a reasonable lateral range requirement after entry, and very little freedom remains to adjust the approach azimuth. These requirements may very well conflict to such an extent that it is impossible to select one or even several landing sites which can be reached by the vehicle with the desired azimuth or orbital inclination under all conditions of timing of the mission. In that case an alternative is to use a water landing and move the landing point continuously with the changing mission situation. In the Apollo mission, for example, the direction of return to Earth depends on the declination of the moon which changes from day to day. A band of possible landing sites is chosen at low latitudes in the Pacific Ocean so the actual landing point toward which the re-entry system guides is moved from day to day if a take-off is delayed.

Another concept associated with atmospheric entry trajectories is that of the entry corridor. This is simply recognition of the fact that a given vehicle cannot fly an acceptable atmospheric flight for arbitrary initial conditions at the entry point. If the flight path angle at that point is too steep, the vehicle will later suffer excessive aerodynamic loading even if its maximum lift is directed upward. Or if the flight path angle at entry is too shallow, the vehicle will exit the atmosphere again with a supercircular velocity even if its maximum lift is directed downward. These boundaries of entry flight path angle are often taken to define the corridor of acceptable entry conditions. The corridor can also be specified in terms of the range of acceptable virtual perigees - the perigee altitude computed as if there were no atmosphere. The corridor depends on the specific definitions of its boundaries, such as 10 g's for the undershoot boundary, and on the entry velocity and vehicle L/D capability. For a lunar return with an entry velocity of about 36,000 ft/sec and $L/D = 1/2$, the 10 g entry corridor is about 2.5 degrees wide in terms of entry flight path angle. For a planetary return with an entry velocity of perhaps 50,000 ft/sec the same corridor shrinks to about 0.7 degree.

For lunar return conditions, the range control which is available to vehicles of different hypersonic L/D capability is shown in Fig. VII-27. (Data from Ref. 5) For each L/D, the short range limit is determined by the 10 g constraint. The maximum range increases essentially without limit with decreasing entry flight path angle. This is indicative of trajectories which turn upward after entry and exit the atmosphere for a long ballistic skip. Such long range performance is not necessarily usable in practice due to the extreme sensitivity of range to errors in exit conditions. The data of Fig. VII-27 is for constant L/D flight. The minimum range for a given initial flight path angle and L/D capability can be improved somewhat by controlling the lift during the flight. The gain is realized by reducing the lift as the vehicle approaches a 10 g acceleration so as to reduce the total aerodynamic load and prevent a violation of the 10 g limit. The lateral control capability of vehicles of various hypersonic L/D as a function of the downrange distance travelled is shown in Fig. VII-28 (Ref. 5)

Some of the basic requirements which must be placed on an atmospheric flight control system are the following:

Avoid excessive aerodynamic loads.

Avoid uncontrolled skip-out in the presence of navigation, vehicle, and atmospheric uncertainties.

Achieve the necessary range and crossrange.

Maintain a suitably low heating load, perhaps heating rate.

Achieve adequate accuracy at landing point.

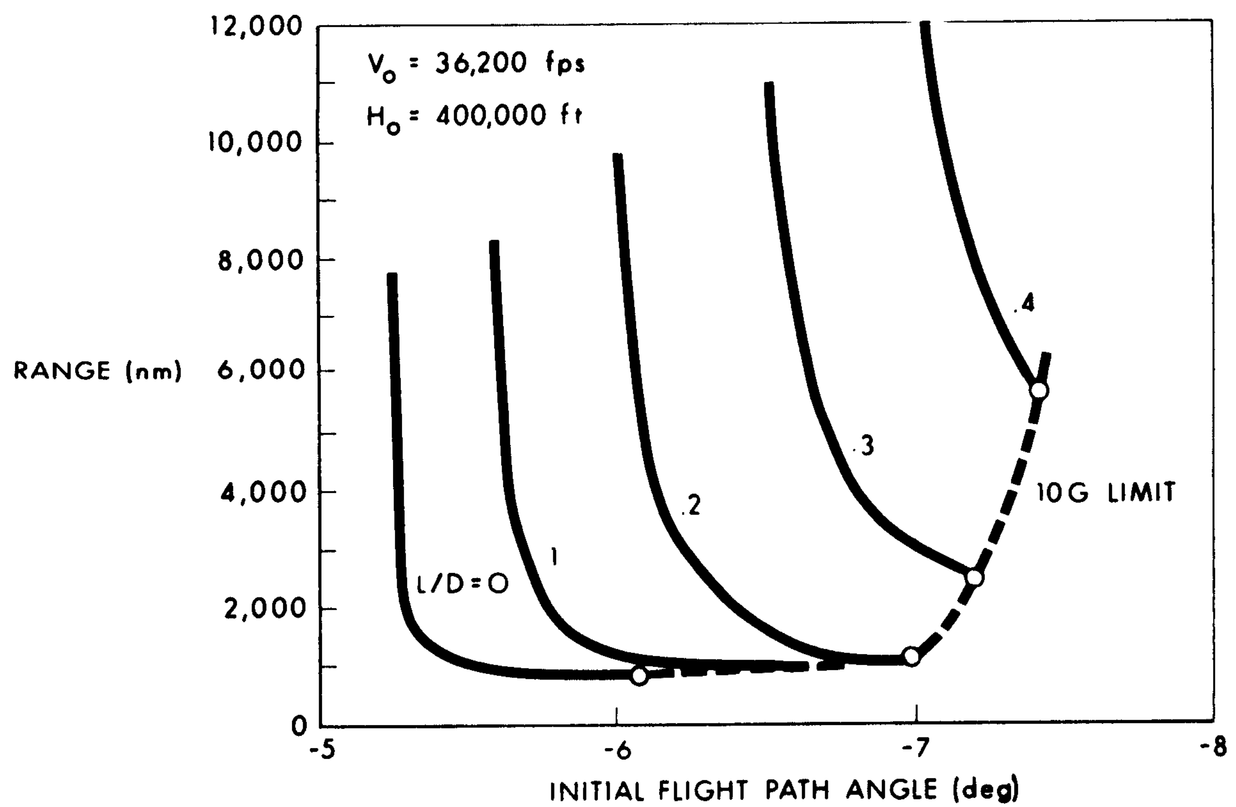


Fig. VII-27 Re-entry Range Capability for Lunar Return Conditions, Constant L/D

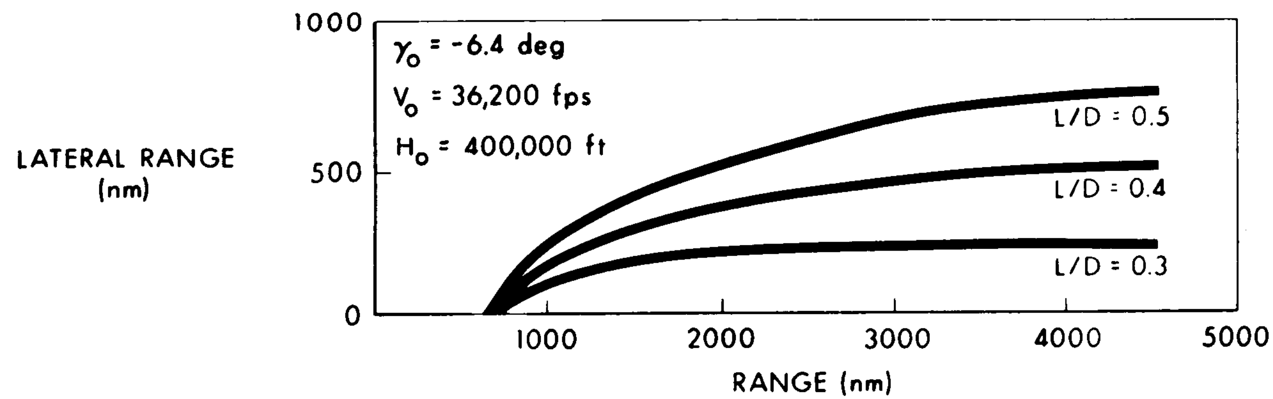


Fig. VII-28 Lateral Range Capability for Lunar Return Conditions, Constant L/D

252

These requirements will be considered within the context of a standard flight plan as indicated in Fig. VII-29.

Pre-Entry - A large part of the responsibility for meeting the first two requirements above rests with the midcourse guidance system. Throughout the mission each guided phase has been followed by another, so the cost of guidance inaccuracies is measured primarily in terms of the fuel required to accommodate the errors in the next mission phase. A more serious cost, however, is associated with the last midcourse correction prior to entry into an atmosphere. That correction must yield entry conditions within the acceptable corridor or it will be beyond the capability of the atmospheric flight control system to fly a proper trajectory. For reasons of simplicity and reliability, the midcourse corrections prior to the last one may be made under the control of an abbreviated inertial system - such as one based on three body-mounted pulse-rebalanced gyros for attitude control and a single body-mounted longitudinal accelerometer for engine cut-off. But due to the premium on accuracy for the last correction, it will be delayed as long as possible so as to benefit from the best possible navigation information, and be executed under the control of the primary instruments of the guidance and navigation system. The IMU will have been aligned for this purpose, and prior to atmospheric entry the inertial navigator is provided with initial conditions.

Initial Pull-Up - During the initial portion of the entry trajectory, attention may well be centered not on reaching the landing point but on avoiding excessive g's or uncontrolled skip-out. Control during this phase may be considered logical in nature - the decision to direct lift up or down depending on the indicated entry conditions and continuing trajectory. If not all the lift of which the vehicle is capable is required to avoid excessive g's or skip-out, and if substantial lateral range is required to reach the landing point, at least some component of lift may be used in the lateral direction. To achieve near-maximum lateral range, it is essential to begin turning the trajectory as early as possible. The end objective of this phase might be taken to be a horizontal flight path angle at a suitably low g level and a low enough velocity so the vehicle can maintain capture in the atmosphere. This means the velocity and altitude at the end of pull-up must be such that the maximum lift of which the vehicle is capable is nearly enough to balance the excess of centripetal over gravitational acceleration.

Controlled Climb to Atmospheric Exit - In most instances, if the vehicle is to make the required range it must climb and do an out-of-atmosphere skip to a new entry point close enough to the landing point so the remaining range can be covered in a steady glide. This is the most sensitive maneuver of the re-entry flight. The range achieved in the ballistic portion of the flight is shown in Fig. VII-30

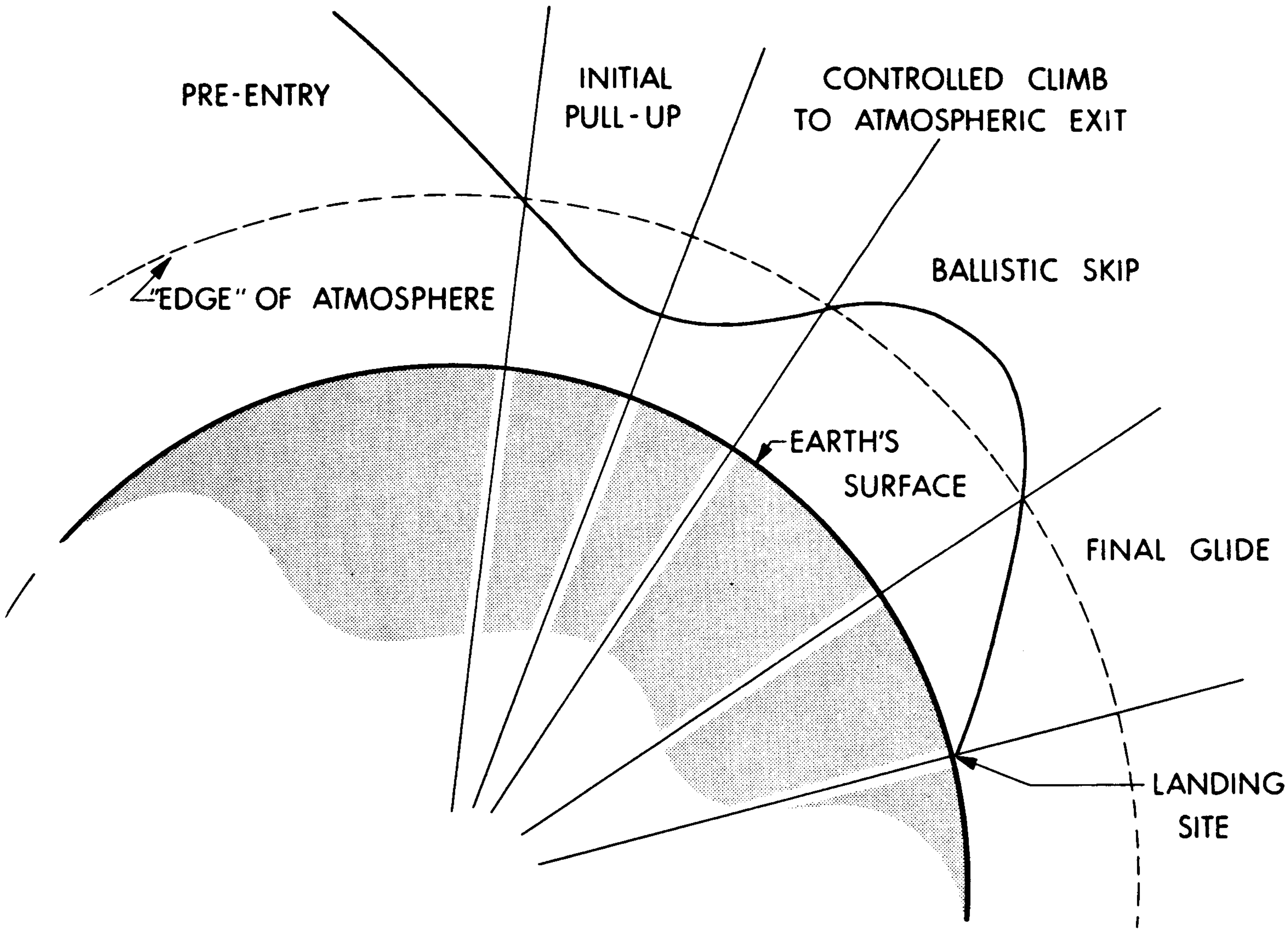


Fig. VII-29 Typical Re-entry Trajectory

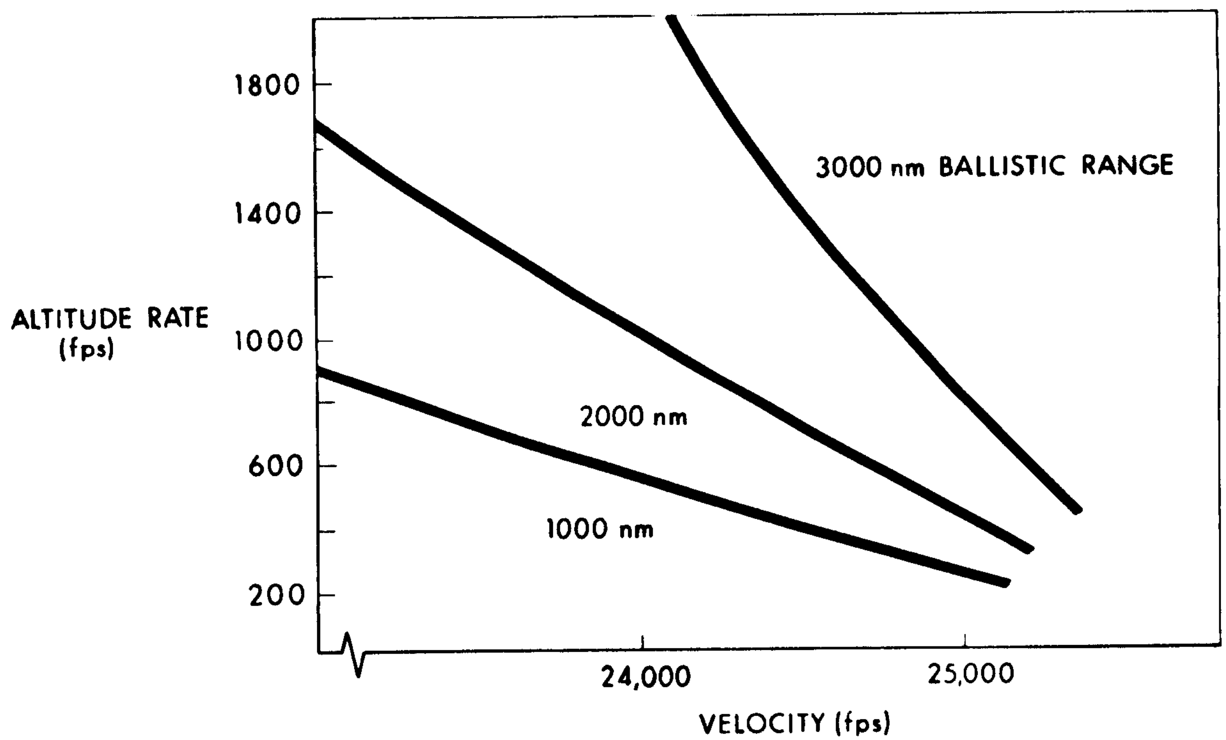


Fig. VII-30 Ballistic Range for Different Exit Conditions as a function of velocity and altitude rate at exit which is taken as 400,000 ft altitude. It is clear that there is a one parameter infinity of exit conditions which will realize the desired ballistic range. For a given required range, a lower exit velocity can be compensated for by a steeper flight path angle resulting in a larger altitude rate. These different trajectories resulting in the same ballistic range have very different sensitivities to errors in knowledge of or control over exit conditions. This is indicated in Fig. VII-31 which shows the sensitivity of ballistic range to exit velocity and in Fig. VII-32 which shows the sensitivity of ballistic range to exit altitude rate. In each case the sensitivity decreases sharply with velocity, so a slow, steep exit is preferable to a fast, shallow exit from the point of view of error sensitivity. For example, if 2000 nm range is required in the ballistic portion of the flight, it can be achieved with an exit velocity of 24,600 fps and an altitude rate of about 680 fps or an exit velocity of 25,400 fps and an altitude rate of about 230 fps - among other combinations. The range derivatives with respect to both velocity and altitude rate are roughly 3 times larger in the second case than in the first. Since the navigation information used by the guidance system may have appreciable errors, due chiefly to the imperfect initial conditions supplied to the navigator at the end of midcourse flight, control over sensitivity to errors must play a dominant role in the design of the flight control system for this phase of the flight.

Two fundamentally different approaches to this flight control problem are often considered: predicted final value control and nominal-following control. The final value control scheme involves prediction of the effect of an assumed control history on the conditions at the end of flight. This trajectory prediction can be done analytically if possible, or otherwise by high speed computer runs starting from the indicated present state. If the terminal conditions are not as desired, in this case if the terminal range is not the desired range, the assumed form of the control is altered and a new predicted trajectory is computed. This iterative process converges on a suitable control history. This form of control can accommodate large off-nominal perturbations in initial conditions, it does not attempt to fly back to a pre-selected nominal trajectory which may be a costly maneuver, and with proper choice for the form of control history it can direct the vehicle along trajectories which are desirable for other reasons than terminal accuracy - such as low heat load. A nominal-following control scheme would be undesirable if it referred to a nominal trajectory selected prior to the start of the mission. This would be a severe constraint for a flight control system which must be able to bring the vehicle home under the wide variety of circumstances resulting, for example, from aborts at all possible times during the mission. However, for this climb-to-exit phase of re-entry flight it has been demonstrated [5] that some simplifying approximations to trajectory equations permit on-board calculation of a reference trajectory which is based on the existing

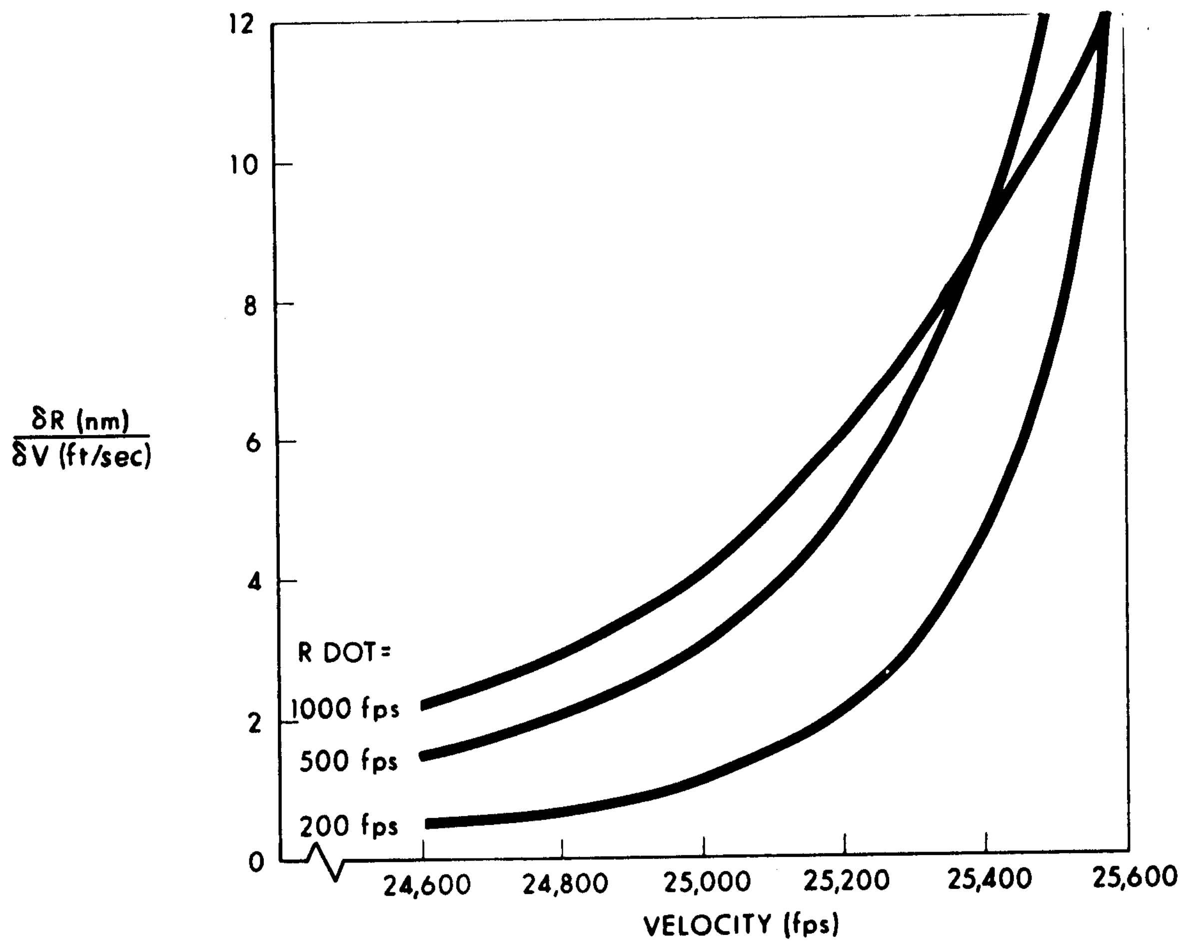


Fig. VII-31 Sensitivity of Ballistic Range to Exit Velocity

257

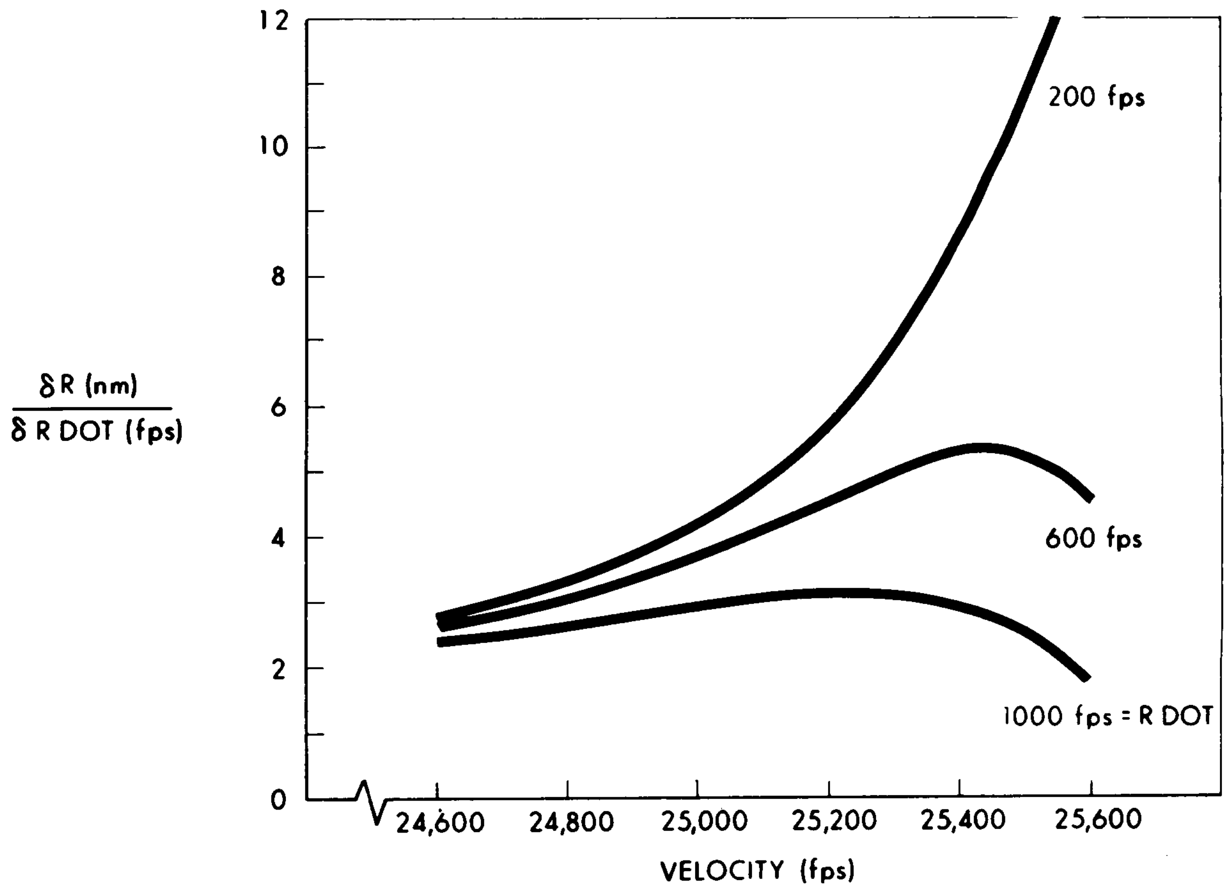


Fig. VII-32 Sensitivity of Ballistic Range to Exit Altitude Rate conditions at the end of the initial pull-up and which reflects the conditions desired at exit so as to achieve the required ballistic range. With such an analytical reference trajectory available, the nominal-following scheme has complete flexibility. Moreover, with the feedback gains in the nominal-following system time programmed to minimize the mean squared range perturbation at the end of ballistic flight due to the ensemble of expected errors in navigation information, a surprising degree of insensitivity to these errors is achieved.[6] In spite of a range sensitivity to altitude rate which is typically about 3 nm/fps, this system yields trivial control errors with errors in knowledge of altitude rate as large as 200 fps. This insensitivity to error is probably the most important criterion by which to judge the merit of a flight control system for this critical phase of the flight.

Ballistic Skip - No effective aerodynamic control can be exercised during this phase of the flight. For a vehicle with no propulsive capability remaining, this skip is then uncontrolled. This just serves to emphasize the importance of achieving accurate exit conditions - or what is closer to the case, achieving a combination of errors at exit such that a desired ballistic trajectory results.

Final Glide - Again during the final glide to the landing point either predicted final value control or nominal-following control can be employed. If the vehicle has a reasonable lifting capability, say L/D of 0.5 or greater, a particularly simple range predictor is available for use in a final value control scheme. If the control history is taken to be just a constant L/D, the range to which the vehicle will glide is given to good accuracy by the simple equilibrium glide relation

$$\text{Range angle} = \frac{1}{2}\,\frac{L}{D}\,\ln\left(\frac{1-\overline{V}_f^2}{1-\overline{V}_i^2}\right) \qquad \text{(VII-9)}$$

where $\overline{V}_i$ and $\overline{V}_f$ are the initial and final velocities divided by the circular satellite velocity at some mean altitude. In fact, the form of this range expression is so simple as to allow direct calculation of the required L/D given the range to go. This expression for L/D, modified by correction terms to account for the fact that the initial altitude and flight path angle may not be consistent with equilibrium glide at the required L/D[7], results in a direct closed-loop flight control system for the final glide.

The Apollo vehicle has an L/D of only 0.3. With such little lift, the assumptions underlying the equilibrium glide relations are not well satisfied, and the nominal-following philosophy seems preferable. This scheme is especially amenable to this last phase of the flight since the ballistic skip was intended to bring the vehicle into the final glide at a nominal range from the landing point. Thus trajectory data calculated before the start of the mission and stored in the on-board computer is perfectly usuable in this phase. The nominal distance flown by the

Apollo vehicle in this final glide phase is about 640 nm and the control capability for accommodating range errors at the end of the ballistic skip is about $\pm$ 200 nm. It is expected that the errors will be no more than about 20 nm.

VEHICLE CONTROL

The result of the flight control schemes discussed above is a desired or commanded L/D. It still remains for a vehicle control system to execute these commands. For many reasons it would be desirable to have full aerodynamic control available; that is, to be able to roll to any desired bank angle and to trim the vehicle to any desired lift or L/D. But aerodynamic trimming poses a most difficult problem. This requires something like control surfaces or trim tabs which will not burn up when deflected into the hot gas flow, will not jam if heat-protecting material from the vehicle forebody flows back and condenses, which are reasonably small and lightweight, have modest power requirements for actuation, and so on. It seems fair to say that this problem has no satisfactory solution at present.

A perfectly reasonable alternative is the use of a fixed vehicle trim at some L/D and the use of roll only to control the flight. In this case the L/D referred to throughout the discussion of in-plane flight is interpreted as the vertical component of L/D. The vehicle is then rolled to whatever bank angle is required to achieve the commanded vertical component of L/D. The resulting lateral component of lift can be directed either to the right or left. An evident logic for lateral control is to reverse the direction of roll when the predicted lateral error exceeds some limit. This limit can be set large initially and decreased during flight - perhaps in proportion to the vehicle's lateral control capability. Roll control in the Apollo mission is exercised by on-off operation of hypergolic thrusters using attitude information derived from the IMU gimbal angles. The vehicle is aerodynamically stable in pitch and yaw; additional control engines are used for rate damping about these axes.

The re-entry flight control problem for lunar return missions seems well in hand. The problem becomes much more challenging as one looks ahead to planetary return missions. It may be that new techniques will be required to satisfactorily solve that problem. The entry corridor will be considerably narrowed; perhaps large area controllable drag devices may be useful to broaden the corridor or propulsion may be useful to rotate the flight path somewhat at a strategic point during entry. The vehicle energy which must be dissipated in the atmosphere will be much greater, thus complicating considerably the problem of heat protection. Perhaps different flight plans can be used to assist in this problem - very long flights around the Earth at high altitude to dissipate the energy under radiative equilibrium may be useful. Trajectory sensitivity to error will be greater, control responsiveness will have to be faster and more accurate. But one thing is certain: there will always be space missions ending with atmospheric flight to a landing point. So workable solutions to the re-entry flight control problem will have to be found for all such missions.

BIBLIOGRAPHY

1. Lorell, J.: Forces Produced by Fuel Oscillations, Report 20-149, Jet Propulsion Laboratory, Pasadena, Calif.; Oct. 1951.

2. Kalman, R.E.: "A New Approach to Linear Filtering and Prediction Problems," Transactions of the ASME, Journal of Basic Engineering; March 1960.

3. Schaefer, R.A.: "A New Pulse Modulator for Accurate D.C. Amplification with Linear or Nonlinear Devices," IRE Transactions on Instrumentation, Vol. I-11, No. 2; Sept. 1962.

4. Chilton, R.G.: Apollo Spacecraft Control Systems, Presented at IFAC Symposium on Automatic Control in the Peaceful Uses of Space, Stavanger, Norway; June 1965.

5. Lickly, D.J., Morth, H.R., and Crawford, B.S.: Apollo Reentry Guidance, Report R-415, MIT Instrumentation Laboratory, Cambridge, Mass.; July 1963.

6. Crawford, B.S.: Optimization of Reentry "Up-Phase" Guidance, Space Guidance Analysis Memo #35-64, MIT Instrumentation Laboratory, Cambridge, Mass.; Sept. 1964.

7. Rosenbaum, R.: Longitudinal Range Control for a Lifting Vehicle Entering a Planetary Atmosphere, ARS Paper 1911-61; August 1961.