1962 IRE International Convention Record

PART 5

Sessions Sponsored by

IRE Professional Groups on

Aerospace and Navigational Electronics

Military Electronics

Radio Frequency Interference

Space Electronics and Telemetry

at

the IRE International Convention, New York, N.Y.

March 26-29, 1962

The Institute of Radio Engineers

1962 IRE INTERNATIONAL CONVENTION RECORD

An annual publication devoted to papers presented at the IRE International Convention held in March of each year in New York City. Formerly published under the titles CONVENTION RECORD OF THE I.R.E. (1953 & 1954), IRE CONVENTION RECORD (1955 & 1956), and IRE NATIONAL CONVENTION RECORD (1957, 1958, & 1959).

Additional copies of the 1962 IRE INTERNATIONAL CONVENTION RECORD may be purchased from the Institute of Radio Engineers, 1 East 79 Street, New York 21, N.Y., at the prices listed below.

Part	Sessions	Subject and Sponsoring IRE Professional Group	Prices for Members of Sponsoring Professional Group (PG), IRE Members (M), Libraries and Sub. Agencies (L), and Nonmembers (NM)			
			PG	M	L	NM
1	8, 16, 23	Antennas & Propagation	$ .70	$ 1.05	$ 2.80	$ 3.50
2	10, 18, 26, 41, 48	Automatic Control Circuit Theory	1.00	1.50	4.00	5.00
3	1, 9, 17, 25, 28, 33	Electron Devices Microwave Theory & Techniques	1.00	1.50	4.00	5.00
4	4, 12, 20, 34, 49	Electronic Computers Information Theory	1.00	1.50	4.00	5.00
5	5, 13, 15, 22, 29, 47, 54	Aerospace & Navigational Electronics Military Electronics Radio Frequency Interference Space Electronics & Telemetry	1.20	1.80	4.80	6.00
6	3, 11, 31, 35, 42, 45, 50, 52	Component Parts Industrial Electronics Product Engineering & Production Reliability & Quality Control Ultrasonics Engineering	1.40	2.10	5.60	7.00
7	30, 37, 43, 51	Audio Broadcasting Broadcast & Television Receivers	.80	1.20	3.20	4.00
8	7, 24, 38, 46, 53	Communications Systems Vehicular Communications	1.00	1.50	4.00	5.00
9	2, 19, 27, 32, 39, 40, 44	Bio-Medical Electronics Human Factors in Electronics Instrumentation Nuclear Science	1.20	1.80	4.80	6.00
10	6, 14, 21, 36	Education Engineering Management Engineering Writing & Speech	.80	1.20	3.20	4.00
		Complete Set (10 Parts)	$10.10	$15.15	$40.40	$50.50

Responsibility for the contents of papers published in the IRE INTERNATIONAL CONVENTION RECORD rests solely upon the authors and not upon the IRE or its members.

1962 IRE INTERNATIONAL CONVENTION RECORD

PART 5 - AEROSPACE AND NAVIGATIONAL ELECTRONICS; MILITARY ELECTRONICS; RADIO FREQUENCY INTERFERENCE; SPACE ELECTRONICS AND TELEMETRY

TABLE OF CONTENTS

Page

Aerospace Radar
(Session 5: sponsored by PGANE)

A New Approach to Radar Cross-Section Measurements . . . . . J. R. Huynen 3
Audio-Radar Monitor Recorder . . . . . A. A. Goldberg 12
A Diplex, Doppler Phase Comparison Radar . . . . . W. D. Boyer 25
Rapid Acquisition of Doppler Frequencies . . . . . W. Ewanus and F. J. Burkitt 32
Optimal Filtering for Radar Doppler Navigators . . . . . J. A. Mullen 40

Military Electronics
(Session 13: sponsored by PGMIL)

The Free World's Largest Radar (Abstract) . . . . . W. L. Kimbell 49
Phased Array Radars for Satellite Tracking . . . . . C. S. Lerch, Jr. 50
A VHF Solar Radar System . . . . . W. G. Abel, J. H. Chisholm, P. L. Fleck, and J. C. James 58
A Motion-Enhancement Display by Time-Compression . . . . . M. E. Crost 67
Frequency Modulation Techniques as Applied to Pulse Doppler Radar . . . . . G. R. Hetrich 76
Application of Space Frequency Equivalence to Radar . . . . . M. I. Skolnik 87

Space Guidance
(Session 15: sponsored by PGANE)

Radar System for Unmanned Cooperative Rendezvous in Space . . . . . H. A. Reuter 95
A Sun-Vector Referenced Earth-Satellite Guidance Scheme . . . . . H. Gabloffsky 105
The Astroguide - A Space-Vehicle Navigation System . . . . . M. M. Birnbaum and P. M. Salomon 137
Emergency and Routine Space Vehicle Recovery . . . . . J. B. Meyer and B. R. Mayo 153

Air and Marine Navigation
(Session 22: sponsored by PGANE)

Ginballess Inertial Reference System . . . . . B. Byrne, W. Murphy, and R. W. Lanzkron 163

Page

Transport Aircraft Navigation with
Airborne Doppler Radar . . . . . R. N. White 175
Propagation-of-Error Equations for Airborne
Doppler Navigation . . . . . N. Braverman and N. Marchand 189
Radar/TV Position Locator Aid to Navigation . . . . . A. Roberts 197
Directional UHF Glide Slope System . . . . . E. R. Hollm and J. L. O'Connor 206
Design Goals, Techniques, and Tests of a
New Instrument Landing Aid . . . . . F. H. Battle, Jr. 210

Panel: Electro-Magnetic Compatibility - Its Significance to Our Survival
(Session 29: sponsored by PGMIL and PGRFI)

Basic Issues of Spectrum Conservation . . . . . D. G. Fink 221
Spectrum Polution . . . . . R. P. Gifford 223
The Cost of Radio Frequency Interference . . . . . F. L. Ankenbrandt 226
System Electromagnetic Compatibility -
DOD's Approach to the Problem . . . . . J. A. McDavid 228

Space Electronics
(Session 47: sponsored by PGSET)

Systems Specifications of a Recovery Control Center
for Hypersonic Vehicles . . . . . J. J. Fishman 233
Manned-Satellite Re-Entry and Landing Guidance
System . . . . . D. K. Breseks, W. G. Green, R. J. Reis,
B. K. Robrock, and J. R. Steinhilber 247
Trends in Space Navigation . . . . . K. N. Satyendra 257
SPACESCAN - A Radio Broadcast System of
Space Navigation . . . . . A. Tatz 272
The Design and Support of a Topside
Sounder (Abstract) . . . . . W. B. Offutt and S. Russell, Jr. 281
Dynamic Analysis of OAO Spacecraft Motion
by Analog-Digital Simulation . . . . . G. Zetkov and R. Fleisig 282

Space Telemetry
(Session 54: sponsored by PGSET)

New Problems in Space Telemetry (Abstract) . . . . . C. Hoeppner 297
A Laser Design for Space Communications . . . . . L. Goldmuntz 298
A Digital Data Telemetry System for an Atomic-Powered
Automatic Weather Station . . . . . A. M. Sullivan, Jr. 306
A Manufacturing Approach to Long Life Requirements
for Unmanned Spacecraft . . . . . R. Waite and C. Balzano 312
CW Redemption - A New Technique for High Velocity
Missile Scoring . . . . . F. C. Lanza and G. M. Hahn 320
A. C. Flashover at Intermediate Altitudes . . . . . D. H. Otto 330

A NEW APPROACH TO RADAR CROSS-SECTION MEASUREMENTS

J. Richard Huynen
Lockheed Missiles & Space Company
Electromagnetics
Sunnyvale, California

Summary

A new method of obtaining radar cross-section measurements is proposed: Complete radar target information is supplied in such a way that the backscattered return from the target can be determined for any given state of transmitted energy. The new technique consists of supplying for a symmetrical target a plot of so-called null-polarizations on a polarization chart and the pattern of so-called maximum polarization amplitude levels. The chart and pattern determine the target at all aspect angles at a fixed frequency and for all possible polarizations. If the target has no rotational symmetry, a second polarization chart is used to locate the position of maximum polarizations. A detailed description of the new technique is presented, and the required instrumentation is discussed. The technique is illustrated with the measurement of the complete characteristic chart for a convex-shaped target.

Introduction

Radar backscatter measurements are usually performed with two linear polarizations, such as horizontal and vertical polarization. The reason for choosing the two linear polarizations is that most radar systems are equipped to produce and receive this type of electromagnetic energy. The question arises whether this technique is adequate to define the target radar-wise in a complete form. It will turn out that this is not the case. A simple example illustrates this fact. For instance, one may ask which pattern is obtained for circular polarization, and this cannot be answered if one has only the patterns for horizontal and vertical polarization. In order to produce the pattern for circular polarization, the target would have to be measured again, using a circularly polarized radar. This fact illustrates an inadequacy in the standard technique of performing radar cross-section measurements, since for every new state of transmitted energy a new pattern would have to be obtained.

A new method is proposed which supplies radar target information in such a manner that for any given state of transmitted energy the return from the target is determined. Thus, complete radar target information becomes available for design or systems consideration.

One advantage of the new method is that performance of a large number of measurements is not necessary because the supplied information is the minimum amount which defines the target radar-wise. Another advantage is that the information becomes independent of the state of the radar at fixed frequency. Further, the information can be used to produce patterns which would be very hard to obtain by other means. For instance, one may produce the pattern of the average of all linear polarization patterns. This has significant application in various systems considerations.

Description of the System

For a target having axial symmetry, the new technique consists of supplying: (1) plot of "maximum null-polarizations" on a polarization chart, and (2) the pattern of so-called maximum polarization amplitude levels. The chart and pattern determine the target at all aspect angles at a fixed frequency and for all possible polarizations.

For a target having no rotational symmetry, a second polarization chart which locates the position of maximum polarizations is necessary. The two charts and the pattern of maxima determine the radar target for the general case without symmetry. Most targets, however, are symmetrical, or almost symmetrical. Consequently, early Lockheed research[16] placed greater emphasis on the property of symmetry.

Polarization Chart

The polarization technique used to present the target data is a slight modification of earlier methods. Although partly published in 1961 IRE papers [1-6], this new technique is not widely known and hence is briefly described below.

Consider at first the description of an elliptically polarized wave as shown in Figure 1. The ellipse described by the electric field is defined by the orientation of the major axis with respect to the X-axis, which is the angle ϕ . The other variable which defines the eccentricity of the ellipse is the angle τ .

A useful technique of representing polarization has been suggested by Poincare'. The method consists of displaying the two angles ϕ and τ as polar angles which determine a point on a unit sphere. The actual polar angles are given by 2ϕ and 2τ . Points on one hemisphere correspond to right-sensed polarizations, while points which belong to the other hemisphere indicate the left-sensed polarizations. In this manner, a unique representation of all possible polarizations is given by points on the unit sphere.

Instead of the spherical representation, a cylindrical projection on a plane will be used in this work. The use of this plane is not to be confused with the stereographic projection of the polarization sphere, which was initiated by V. H. Rumsey in his early 1951 work[3]. The advantage of the cylindrical projection is that a circle on the sphere parallel to the plane of projection will project as a circle, while circles in a plane normal to the projection plane are projected as straight lines.

Figure 2 shows the complete polarization chart. The outside periphery of the chart supplies all linear polarizations, with horizontal polarization on the right

and vertical polarization on the left of the horizontal axis. The position angle of a point on the chart is twice the major axis orientation of the ellipse. The radius vector is a function of ellipticity only. Two such charts – one belonging to the upper, and the other to the lower hemisphere – completely determine all transmitter polarizations uniquely. The center of the chart corresponds to the two circular polarizations. In many applications, it will be more convenient to use one polarization chart instead of two to indicate all points on the polarization sphere where points belonging to differently sensed polarizations are distinguished by some labeling procedure.

The Radar Reception Problem

A radar is defined in this work as a sensor of electromagnetic energy, which uses the same antenna for transmitting and receiving. The mechanism of radar reception may be described as follows: In Figure 3 is shown a state of transmitted elliptically polarized signal which is scattered back by the target such that the returned signal has, in general, a different polarization. The returned signal is decomposed into two orthogonal components: one is the received signal which has a polarization "parallel" to the transmitted signal; the other is the rejected signal which is "orthogonal" to the received polarization. As the names indicate, the received signal contributes to a voltage in the radar receiver.

In mathematical terms, the process is described as follows: Consider the normalized transmitted polarized signal represented by a two-component complex vector h; then the return signal is given by the target scattering operator T applied to the transmitted signal h. The received voltage will then be found as a scalar product of the return signal Th with the transmitted signal h as shown by the equation $V = Th \cdot h$. The received voltage squared: $P = |Th \cdot h|^2$. The scattering operator T which describes the radar backscattering properties of the target at fixed aspect angles is a symmetric operator because of reciprocity. If h is determined by two independent components of electric field in a given frame of reference, the scattering operator T is represented by a two-by-two symmetric matrix.

The quadratic form which determines P is known by the two null-polarizations which make P equal to zero, together with a normalizing factor which can be identified with the maximum received signal for varying polarization at fixed target aspect. Each null-polarization being represented as a point on the polarization sphere is given by two real parameters. Consequently, the general target matrix T has five independent parameters by which it is defined (except for a common phase factor which plays no role with radar amplitude measurements). If the position of the two null-polarizations are known as two points on the polarization sphere, then, for an arbitrary transmitted polarization, one can compute the signal level of return from the target. It is clear from this discussion that nulls and maxima are the parameters which determine the scattering operator in an invariant form, i.e., without dependence on a particular reference frame.

In contrast to the described technique, it has been customary in the past to determine the scattering matrix by measuring its coefficients directly. It is clear that, although this method also would describe the scattering operator (since the reference frame in which the operator is represented is known), the matrix coefficients contain a mixture of information about the target and reference frame, and thus are not pure target parameters. To try to correlate information obtained from two different targets by a comparison of scattering matrix coefficients will be a frustrating task. Large efforts have been expended along this direction in the past, without much apparent success.

Flat Plate Null Characteristic

The theory discussed previously will be illustrated by two simple examples. The first is the radar return at normal incidence from a large flat plate. Consider a circularly polarized transmitted wave as shown in Figure 4. At the plate, the two linearly polarized components in X and Y directions are reversed, and the direction of propagation is reversed so that the returned wave is circularly polarized, but in a sense opposite to that of the transmitted wave. Thus, the transmitted and returned wave are orthogonal to each other. As a consequence, the two polarizations which give a zero voltage in the radar receiver are the two circularly polarized transmitter polarizations. In Figure 4, this case is illustrated by the position of the characteristic nulls at the center of the polarization chart.

Dihedral Polarization Null Characteristic

The second special case to consider consists of the return from a set of dihedral plates with vertical axis. The plates are positioned at 90°, and the aspect angle is symmetrically oriented with respect to the two plates. Figure 5 shows the characteristic null-polarizations for this condition. The figure is self-explanatory and shows that a linearly polarized transmitted wave at 45° inclination with respect to the vertical axis of the plates is returned as a linearly polarized wave at 135° inclination. The transmitted and return polarizations are orthogonal and, consequently, the radar receiver does not register the returned signal. The two characteristic null-polarizations are shown at the periphery of the chart at the 45° and 135° linear polarization positions.

Instrumentation

A special antenna was designed to produce variable polarizations. Figure 6 shows a schematic view of the antenna construction. The reflector consists of an 18-in diameter parabola with an 8-in focal length. The feed consists of a tapered circular waveguide with a specially designed teflon back feed.

A quarter wave vane is positioned within the circular waveguide feed providing a 90° phase lag between the two orthogonal linear polarizations. The orientation of the vane with respect to the linear input polarization determines the ellipticity of the transmitted polarization. If the vane assembly and the linear polarization are simultaneously rotated, a change of major

axis orientation of the transmitted ellipse occurs. The two pointers which determine the settings on the dial providing arbitrary elliptical polarization are shown in Figure 7. Figure 8 shows a block diagram of the CW, X-band system which was employed.

For monostatic operation, the same antenna was used for transmitting and receiving as was mentioned before. This requires that the received signal shall be separated from the transmitted signal. With the CW system, this requirement can be met by utilizing a magic T as a balanced bridge. Isolation between the transmitted signal and the received signal dictates, to a great extent, the accuracy of the minimum detectable return and the operating range of the system. To accurately determine the radar return from small targets, isolation between transmitted and received signals of the order of 100 db must be achieved. This degree of isolation puts exceedingly critical requirements on the system.

Of major importance is the stability of the transmitting source, which is influenced by reflections and mechanical and thermal vibrations in and outside the system. Measurements on targets having radar cross sections exceeding one square foot were accurate to within .25 db.

Symmetrical Target

The general theory which was outlined on the previous pages will now be applied to symmetrical targets. In Figure 9 a target is shown which has a horizontal plane of symmetry through the line of sight. Targets which have an axis of rotational symmetry display this kind of characteristic. For these targets in particular, the description of its target operator at all aspect angles can be obtained through one aspect parameter. The aspect parameter is the rotation angle around a vertical Y axis.

Next we compare the return for two transmitted elliptically polarized waves which are symmetrically oriented with respect to a horizontal plane of symmetry. It is clear that because of the target symmetry, the received signal for these two transmitted polarizations will be the same. From this observation it follows that if one transmitted polarization produces a null, its mirror-image polarization will be the other characteristic null polarization. Consequently, one null-polarization completely determines the second.

This fact has many important consequences. First, we identify the position of the two null-polarizations on the polarization chart. Figure 10 shows a typical locus of null-polarizations for targets with a horizontal plane of symmetry as a function of aspect angle. Next, at fixed aspect angle, one null-polarization is assumed to be located in the upper half of the circular chart. Because of the symmetry, the corresponding second null-polarization will occur as the mirror image about the H-V axis. Thus, as a function of aspect angle, the first null-polarization locus has the corresponding second null-polarization locus as its mirror image about the H-V axis. These two null-loci are shown in Figure 10. Observe that the two corresponding null-polarizations have opposite senses.

If for a symmetrical target one locus as a function of the aspect angle is given, the locus of the corresponding second null-polarization is completely defined. As a consequence, only one locus suffices to define the scattering matrix as a function of orientation angle, if the level of return for the maximum polarization is also known. At any fixed aspect angle we find the polarization for maximum return from the location of the two nulls by a simple construction on the polarization sphere. The two nulls and the maximum polarization are geometrically presented by a "fork" construction. The end points of the two prongs of the fork are the two null-polarizations, while the end point of the "handle" is the maximum polarization. The center of the fork is at the center of the polarization sphere. The fork concept is illustrated in Figure 11. N_1 and N_2 are the two null-polarizations; M is the maximum polarization. By projection of the polarization sphere, the null-locus on the polarization chart establishes the target operator as a function of time. If the polarization nulls for the symmetrical target are located on the right-hand side of the polarization chart, the maximum return will be for vertical polarization. On the other hand, if the nulls on the locus are located on the left-hand side of the chart, the maximum polarization will be horizontal.

Thus, we establish a rather remarkable theorem, which is not obvious at first sight, that for targets with a plane of symmetry along the line of sight, the maximum polarization will always be a linear polarization. This fact has been substantiated by many measurements, one of which is shown in detail in the example to follow.

Measurement Technique for Symmetrical Targets

A target with horizontal symmetry was placed upon a foam tower and rotated along the vertical axis. Four patterns at different polarizations were taken and recorded on transparent paper. The four polarizations used were horizontal, vertical, linear at 45° and right circular. The patterns as a function of rotation angle were superimposed upon each other such that db differences from three different patterns relative to the local maximum were obtained. The polarization null-locator chart (Figure 12) which establishes the position of the nulls was then used to identify the null on the chart. Figure 13 shows the four patterns thus obtained for a dihedral corner reflector.

The colors on the null-locator chart in Figure 12 (these colors are indicated but are not visible due to reproduction limitations) are used as follows:

> Red indicates circular polarization with db differences labeled relative to the local maximum; green refers to horizontal polarization relative to the vertical maximum; purple refers to vertical polarization db differences relative to the horizontal maximum; the blue lines refer to 45° linear polarization db differences relative to the local maximum.

Observe that the null-locator chart is symmetric with respect to the horizontal axis. This makes an ambiguity check necessary. The null positions which have

a cross mark (see Figure 15) indicate those nulls which have been checked individually with the variable polarization antenna.

Experimental Result

Figures 14 and 15 show the experimental results which were obtained on a convex-shaped body and which supply complete radar cross-section information about this target.

Figure 14 gives the four superimposed patterns at different polarizations. Notice that on these patterns the local maximum is indeed attained either at horizontal polarization (green) or at vertical polarization (purple). At no aspect angle does either the red (circular polarization) curve or the blue (45° linear polarization) curve exceed the maximum green or black level. The convex-shaped object can be expected to exhibit a null-locus characteristic (Figure 15) which is distributed around the center of the chart at circular polarization because of specular reflection. It is interesting to note, however, that deviations from the center position occur even for relatively strong signals. This indicates that for convex-shaped objects of a few wavelengths, the specular reflection considered as a local flat plate is no longer a valid assumption at all aspects. Indeed, the null-locus behavior of convex-shaped objects could be used to determine the extent to which stationary phase methods are applicable for predicting diffraction patterns.

This particular target was analyzed for aspect angles from 0° to 180° . The characteristic loop structure for convex shapes is well illustrated in this case.

Measurement of a General Target

The technique thus far presented is suitable for radar cross-section measurements of targets with rotational symmetry. Many targets are of this type, or are approximately symmetric. These targets are particularly suitable to measure, since scattering information obtained for one cut through the rotation axis determines this information for all aspect angles. For a non-symmetrical target, several cuts have to be obtained to cover all aspect angles. Basically, there are two approaches to the measurement technique of non-symmetrical targets. The first consists of a "symmetrization" procedure and involves a rotatable antenna. The second method computes the data from inputs to a stationary antenna. The first method consists of a "maximum following" principle in which the maximum polarization is found either by scanning with linear polarization, or by some other technique. Most targets exhibit the property of having no "helicity" i.e., there is no internal preference for right or left circular polarization. For those targets, the maximum polarization is a linear polarization with orientation angle ψ . By "following" this angle ψ by rotation of the polarization antenna, either manually in the laboratory, or by using a servo device for automatic operation, the angle ψ can be eliminated as a variable parameter. Thus a "symmetrized" target is obtained of which patterns are measured, and from which null-polarizations can be deduced in exactly the same manner as was shown for symmetrical targets.

If the target has appreciable "helicity," this procedure cannot be followed and only a computing scheme will suffice to analyze the data. We plan to discuss elsewhere the theory which is required for the computing scheme. It suffices here to establish the feasibility of complete radar target-scattering operator determination in an invariant form by means of the locus of one "symmetrized" null locus on the polarization chart and the locus and level of maximum polarization response on another chart.

Acknowledgment

The author gratefully acknowledges the use of some unclassified material and photographs of experiment apparatus which were first published in reports submitted to the Electronics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Commander, under Contract AF 19 (604)-5550.[16] Mr. W. F. Fortner of Lockheed Electormagnetics is largely responsible for the successful completion of instrumentation and measurements. To Mr. A. F. Riedel, the author is indebted for many valuable discussions.

References

1. G. Sinclair, "The Transmission and Reception of Elliptically Polarized Waves," Proc. IRE, pp. 148–151, February, 1950.

2. H. G. Booker, "Introduction to Techniques for Handling Elliptically Polarized Waves With Special Reference to Antennas," Proc. IRE, pp. 533–534, May, 1951.

3. V.H. Rumsey, "Transmission Between Elliptically Polarized Antennas," Proc. IRE, pp. 535–540, May, 1951.

4. G. A. Deschamps, "Geometrical Representation of the Polarization of a Plane Electromagnetic Wave," Proc. IRE, pp. 540–544, May, 1951.

5. M. L. Kales, "Elliptically Polarized Waves and Antennas," Proc. IRE, pp. 544–549, May, 1951.

6. J. F. Bohnert, "Measurements on Elliptically Polarized Antennas," Proc. IRE, pp. 549–552, May, 1951.

7. E. M. Kennaugh, Antenna Laboratory, The Ohio State University Research Foundation, "Effects of Type of Polarization on Echo Characteristics," Contract No. AF 28(099)-90 Progress Reports 389-4, 9, 12, 14, 19, June, 1950 to July, 1953.

8. E. M. Kennaugh, Antenna Laboratory, The Ohio State University Research Foundation, "S-band Radar Reflectivity of the Quail," Project Report 601-24 (AD-146145) 24 October, 1957, WADC Contract No. AF 33(616)-2546.

9. Dalmo Victor Company, "Interim Engineering Report on Effect of Polarization on the Radar Return From Ground Targets and Rain," CONFIDENTIAL, Contract No. AF 33(038)-20926, Report Nos. R-135-598, R-135-620, R-135-639, R-135-669, R-135-683, and R-135-697, June, 1951 to May, 1952. Contract No. AF 33(600)-22711 Class 16A, Report Nos. R-135-848, R-135-867, R-135-889, R-135-916, R-135-943, R-135-968, R-135-1008, and R-135-1015, January, 1953 to May, 1954. (Recently declassified.)

10. J. R. Huynen, "Representation of Elliptically Polarized Waves With Exponential Matrices," Dalmo Victor Report, December, 1952.

11. J. R. Huynen, "Effect of Polarization on the Radar Reception From Linear Targets," Dalmo Victor Report, May, 1952.

12. J. R. Huynen, "New Methods in Representation of Polarization," Dalmo Victor Report, September, 1953.

13. H. Gent, "Elliptically Polarized Waves and Their Reflection From Radar Targets: A Theoretical Analysis," IRE Memorandum No. 584, March, 1954.

14. C. D. Graves, "Radar Polarization Power Scattering Matrix," Proc. IRE, pp. 248–252, February, 1956.

15. J. R. Copeland, "Radar Target Classification by Polarization Properties," Proc. IRE, pp. 1290–1296, July, 1960.

16. J. R. Huynen, Lockheed Missiles & Space Company, "Study on Ballistic-Missile Sorting Based on Radar Cross-Section Data," Report No. 4, "Radar Target Sorting Based Upon Polarization Signature Analysis," LMSC-288216, (S), May, 1960.

17. J. R. Huynen, Lockheed Missiles & Space Company, "Radar Measurements on Scattering Operators." Presentation delivered to URSI-IRE May, 1961 Spring meeting, Washington, D.C. Also, presented at 1961 Aerospace Symposium on Ballistic Missile and Space Technology, proceedings published by Academic Press, (vol. II), January, 1962.

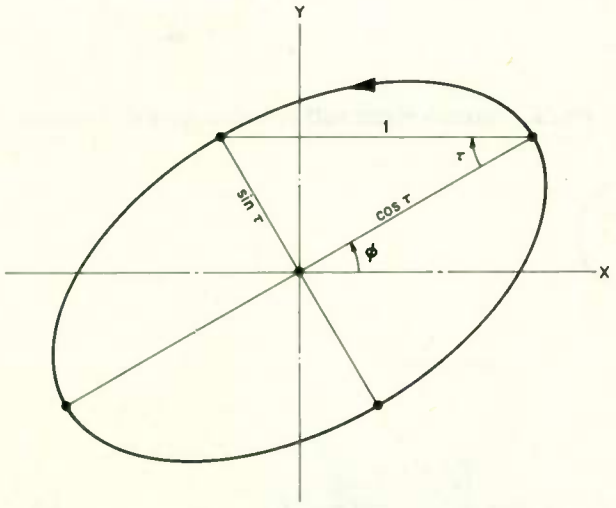


Fig. 1. Normalized polarization ellipse in fixed plane.

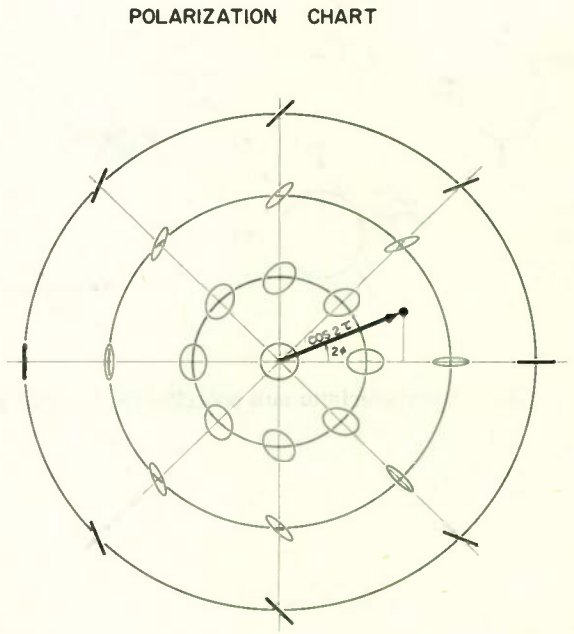


Fig. 2.

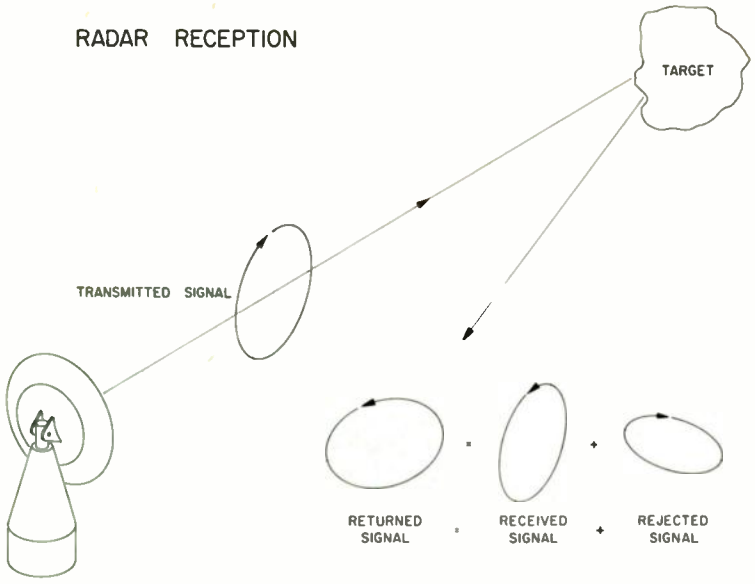


Fig. 3. Principle of radar reception.

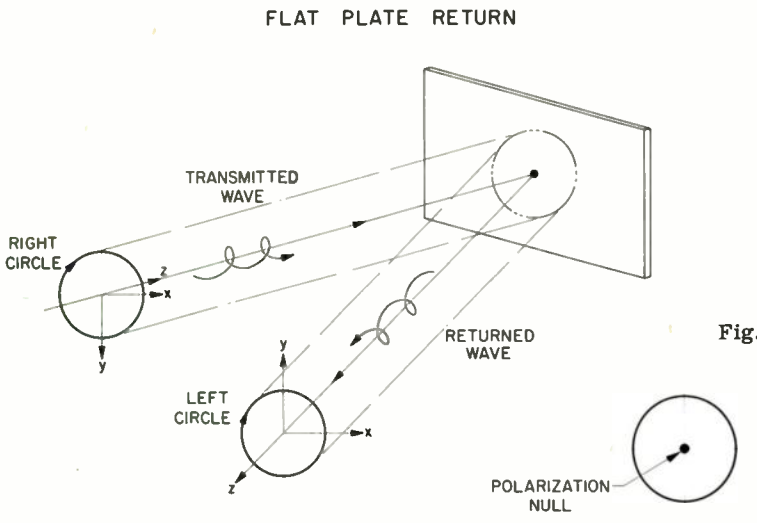


Fig. 4. Characteristic null polarization for flat plate.

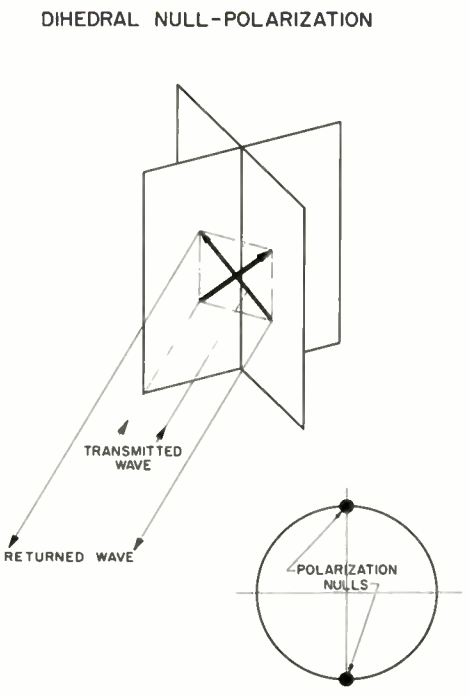


Fig. 5. Characteristic null polarization for dihedral.

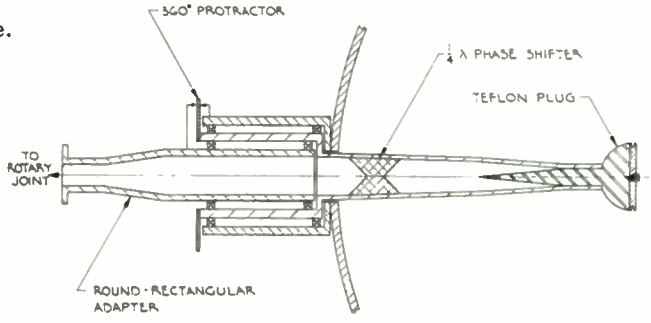


Fig. 6. Schematic diagram of varable polarization antenna.

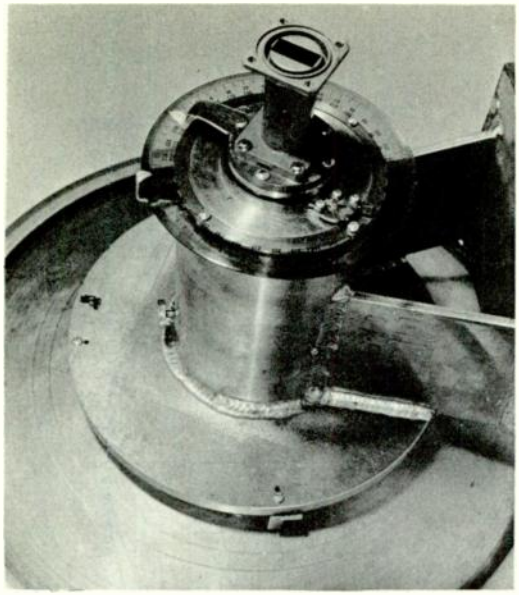


Fig. 7. Pictorial view of polarization antenna.

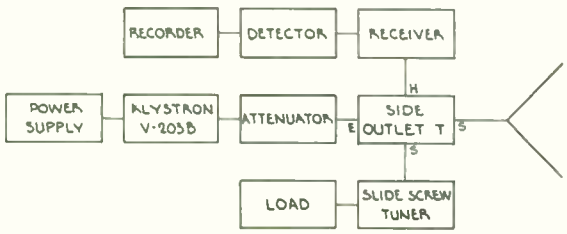


Fig. 8. Schematic diagram of CW X-band system.

TARGET WITH HORIZONTAL SYMMETRY

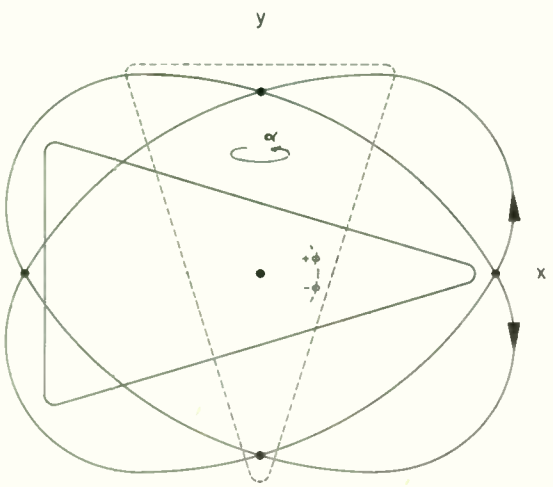


Fig. 9.

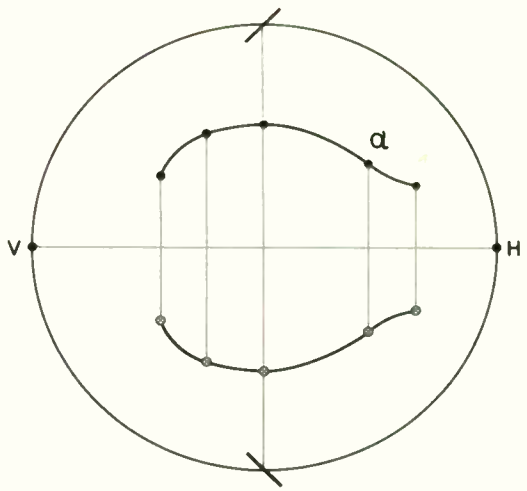


Fig. 10. Null-locus for a symmetrical target.

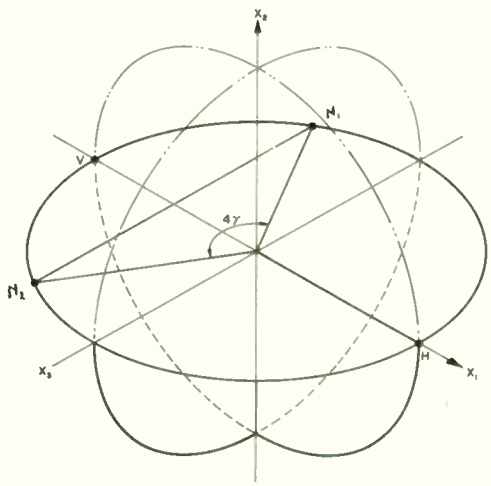


Fig. 11. Polarization fork for gamma target.

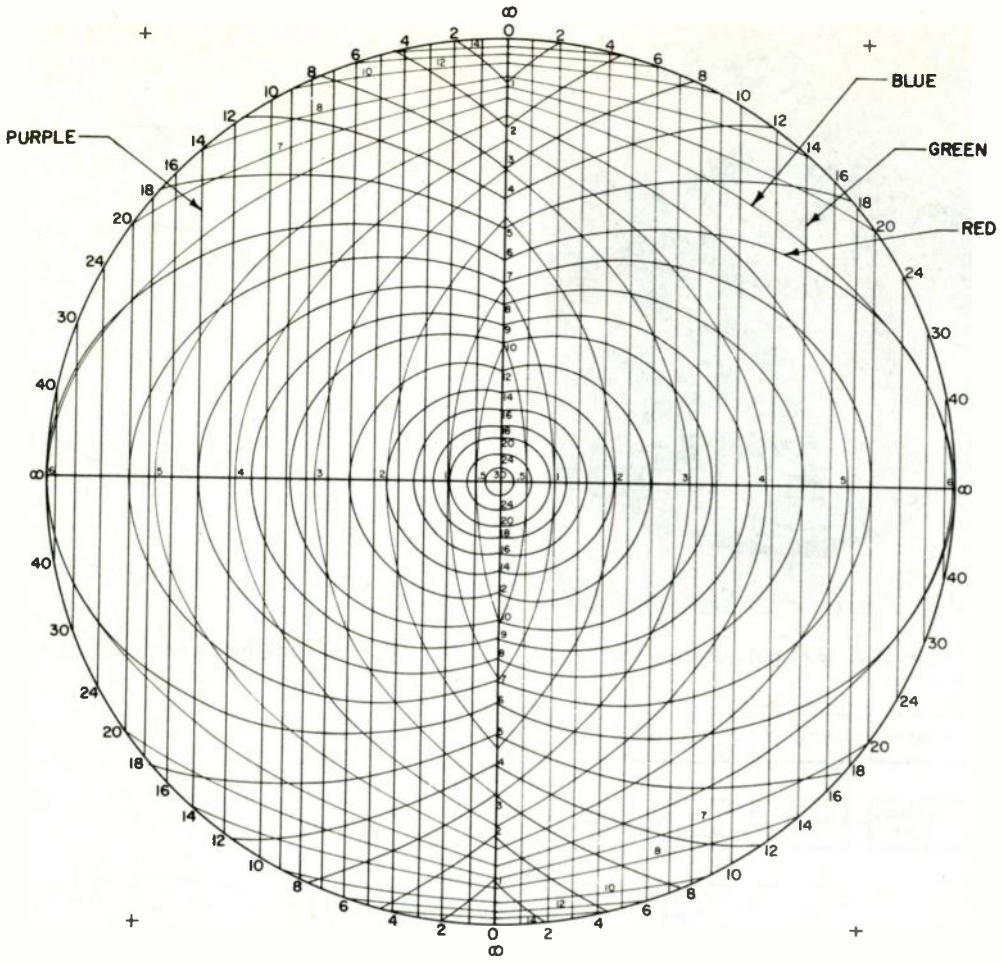


Fig. 12. Polarization null locator chart.

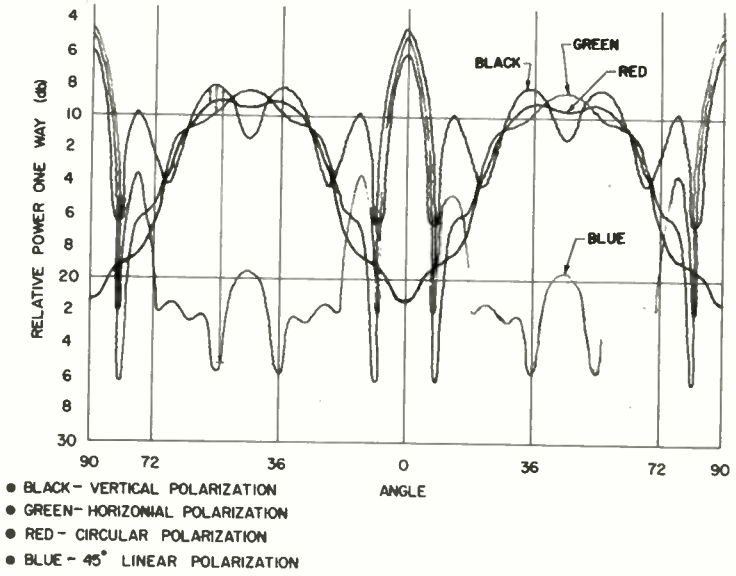


Fig. 13. Four patterns for dihedral plate.

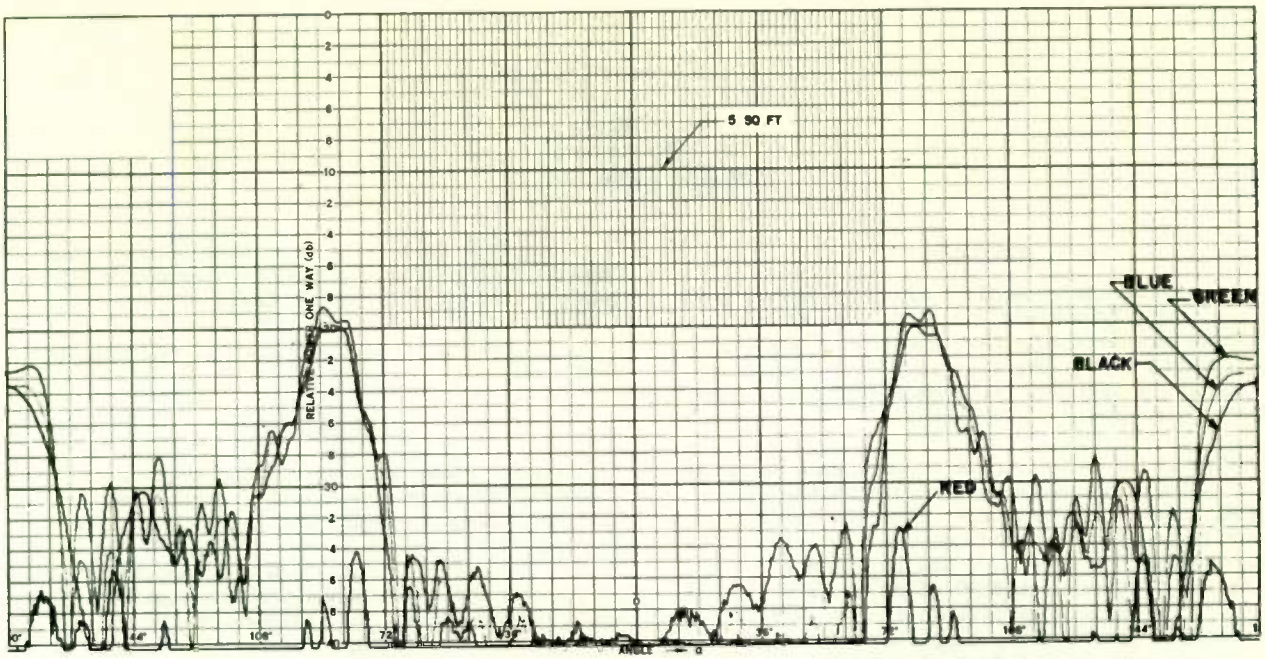


Fig. 14. Four patterns for a convex surface.

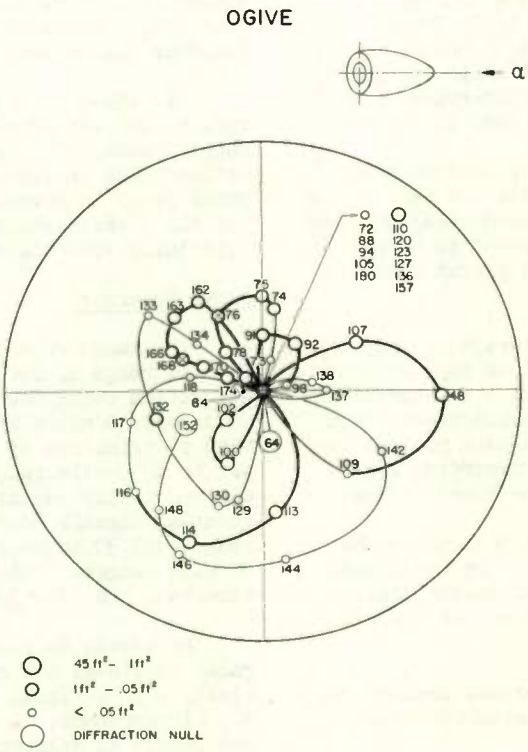


Fig. 15. Null polarization diagram for a convex surface.

AUDIO-RADAR MONITOR RECORDER

A. A. Goldberg
CBS Laboratories
Stamford, Connecticut

Summary

The Audio-Radar Monitor Recorder is designed to continuously monitor and record 16 audio channels and one radar display on a single reel of magnetic tape. The tape is 2" wide and moves at the speed of 15/16 ips. Each reel has a time capacity in excess of 24 hours.

The radar recording is accomplished by means of a quantizing system that separates the video signal into 14 low frequency channels. The ppi radar display is recorded out to a 30 mile range. Range definition between 0 and 20 miles is 0.2 miles and between 20 and 30 miles it is 0.33 miles. The azimuth resolution is 1°. During playback, a ppi display, complete with range marks, is shown on the face of a scope.

Stationary magnetic head and slow-speed tape assure trouble-free operation and long life.

Introduction

Radio and radar are the voice and eyes of the approach air traffic controller. These sounds and blips represent precious cargo; no errors can be afforded here. It is in the interest of all concerned that the audio-radar situation be continuously recorded - just in case!

Major airports like New York International presently record 12 audio channels but have no facility for recording the approach radar picture. The lack of a radar recording proved to be a handicap at the inquiry into the recent crash of two planes over New York City.

Video recorders are characterized by high speed moving parts and a prodigious appetite for magnetic tape. Twenty-four hours a day operation would be a difficult and costly enterprise. The proposed CBS Laboratories audio-radar monitor recorder avoids these problems by recording a quantized picture rather than the direct video.

Aircraft beacon transponder returns cannot be accommodated on the recorder to be described. This, however, is not an inherent system limitation. Beacons could be recorded but at the expense of fewer audio channels.

The CBS Laboratories audio-radar monitor recorder incorporates the following outstanding features:

a. The monitor can record 16 channels of audio and one radar picture on a single tape so that they can be time correlated when played back.

b. Tape reels with a greater than 24 hour capacity can be used. Thus the recorder will require reloading only once a day.

c. System reliability is high and maintenance low.

Description

Direct radar video recording requires a bandwidth of approximately five megacycles. Existing video recorders use 2-inch wide tape running at 15 ips, or a tape area of 30 square inches per second. This high consumption of tape, as well as the maintenance problem imposed by high speed moving parts, has discouraged the use of radar monitor recording.

Examination of the display on the radar scope discloses that the targets themselves require narrow bandwidths if the range definition is limited to a reasonable value. Quantizing target information makes it possible to record the radar on a tape area of approximately 1 square inch per second. It is this small tape area consumption that makes the CBS Laboratories Audio-Radar Monitor Recorder System practical.

The ASR-3 Surveillance Radar rotates at 13 rpm, or 84° per second. The antenna lobewidth is approximately 3°. Thus, each target occupies 36 milliseconds in time (see Figure 1). The 20-mile radar range is normally used for approach control but the 30-mile range is occasionally employed, especially with the newer and faster aircraft.

Radar Recorder

The radar recorder operates as follows: The 20-mile range is recorded in 1% increments, range definition being 0.2 miles. The range between 20 miles and 30 miles is recorded in 3% increments with a definition of 0.33 miles. With the radar set to a 30-mile range, the radar video signal is electronically commutated into 100 separate, sequential signals during the first 20 miles of range, and 30 sequential signals during the 20-to-30-mile ranges. This is done during the 323 us time-base that the 30-mile radar requires.

Referring to Figure 2, the main band of the radar initiates the electronic sweep which sequentially keys a signal into 130 points within the 323 microseconds. A target return causes one of the points to deliver a 1200 pps signal which is the radar pulse rate. These pulses are detected and the output of the detector is a broad pulse of 36 milliseconds in duration which is the time that the echo is written on the ppi scope. The

detector output pulse causes the balanced modulator to pass a sinusoidal burst from oscillator f which is recorded on one track of the magnetic tape. In effect, the radar echo has been quantized into a low frequency audio signal that can be recorded on slow-moving magnetic tape.

By a system of frequency multiplex the 130 range increments are recorded on 13 magnetic tape channels, shown in Figure 3. Ten range increments are recorded on each magnetic track. Since it is intended that the tape be used to record audio up to 4000 cps, the oscillator frequencies in Figure 3 are restricted to below 4000 cps. The lowest frequency is that which can include at least 10 cycles within 36 milliseconds or approximately 280 cps. Frequencies f1 to f10 should not be closely harmonically related and are: 300, 500, 700, 1100, 1300, 1600, 1800, 2200, 2600 and 2900 cps. Outputs of all 10 balanced modulators are linearly added and recorded on one magnetic track. In order to prevent intermodulation, an ac biased magnetic recording system is used and each frequency burst is 20 db below maximum record level. The signal-to-noise ratio is adequate due to narrow-band audio filters which discriminate against noise during playback.

All 13 record channels are similar. Oscillators f1 through f10 are stable tuning fork devices and are used in common for the other 12 record channels. In the event of a magnetic tape dropout in one channel, every 13th range point will momentarily disappear; this is less objectionable than experiencing the disappearance of a large contiguous block of range points.

Reproducer

Figure 4 represents one of 13 playback channels. The playback head signal is amplified and separated into frequencies f1 through f10. Each tone is detected and the resultant broad pulses are fed to one point of an electronic switch. The 130-point electronic switch sequentially passes each detector signal to a common video amplifier and then to the intensity grid of a long persistence ppi scope. The deflection yoke of the scope is energized by a 1200 cps sawtooth current which is synchronized with the playback electronic switch. Simultaneously, the playback scope deflection yoke rotates mechanically at the same speed and phase as the original radar antenna. The resulting display closely approximates the original radar.

A 14th recording channel is used to synchronize the rotation of the playback scope deflection yoke with the radar antenna. Minor tape speed variations will not effect system azimuth accuracy because the target and synchronization tracks are on one tape. The synchronizing channel also contains a brightening pulse at zero degrees as an aid to checking the azimuth accuracy. The playback electronic switch need not be synchronized to the original radar either in phase or repetition rate.

Original radar range circles can be reproduced but may obscure targets superimposed upon them. To prevent this from happening, the radar range circles are recorded and reproduced only in alternate quadrants of alternate azimuth sweeps. Thus, a target that may normally be obscured by a range marker can be clearly seen on alternate azimuth sweeps. This is shown in Figure 5.

The electronic radar map is not recorded because of its greater bandwidth requirements. It can, however, be reinserted during playback or it can take the form of a plastic overlay on the face of the playback scope.

Audio Recording

CBS Laboratories has demonstrated that magnetic recording densities of 8000 cps/inch/second are practical. The audio-radar monitor records 3000 cps at a tape speed of 15/16 ips. This is less than 3000 cps/inch/second and excellent reliability can be predicted.

Sixteen separate audio channels, each capable of 200 to 3000 cps response are available in the audio-radar monitor recorder. The radar recording requires 14 parallel tracks and these, added to the 16 audio tracks, make a total of 30 magnetic tracks on the 2-inch wide tape.

The audio reproducer contains 4 audio playback channels that can be connected via pushbutton to any of the 16 audio playback magnetic heads. Means are provided to mix these 4 signals into a single channel if desired.

Magnetic Tape

The magnetic tape is 2 inches wide, 7000 feet long, and records for 25 hours. Total tape thickness is 0.9 mil; it includes a 0.6 mil tensilized polyester backing and a 0.3 mil thick coating. Unlike video tape, the magnetic domains are longitudinally oriented. Each 12 1/2 inch diameter reel of tape can be used thousands of times without deterioration in quality. The specification of the magnetic tape tracks is shown in Figure 6.

Magnetic Heads

Figure 7 shows the geometry of the record and playback magnetic heads. Two staggered stacks of 15 heads each are used. All 30 heads are prealigned and securely mounted to a single base; a multiconnector provides convenient connection to the electronic circuitry.

The record head gaps are approximately 5 microns wide and the playback head gaps are approximately 1.5 microns wide. The absence of any moving heads or pole pieces in this recording system results in excellent reliability and low maintenance.

Apparatus

The Audio-Radar Monitor Recorder System consists of two main components; one recorder and one reproducer.

Recorder Apparatus

Each recorder occupies a standard 19 inch wide rack which is approximately 6-1/2 feet tall. As shown in Figure 8, the tape deck occupies three feet of vertical space; the balance being used for electronic circuitry.

The recorder tape deck, as seen in Figure 8 consists of a supply reel, supply idler, record heads, capstan, take-up idler and take-up reel. Separate torque motors and mechanical brakes are used for both the supply and take-up reels. A hysteresis motor insures constant capstan speed. Although the tape deck has provisions for rapid rewind, it is intended that the recorded tape be stored on the take-up reel and rewound on the reproducer only in the event of playback. An automatic device that does not depend on any signal from the magnetic tape will warn when the tape is within one hour of being exhausted. Additional alarms will be sounded if for any reason the tape stops moving or breaks.

Figure 9 is a symbolic diagram of the 30-channel recorder circuit. All channels are identical. Inputs are bridged across 600 ohm lines, 0 db ($\pm$ 10 db). AGC amplifiers effectively maintain constant record level with input variations of 10 db, control action being accomplished without audible distortion.

One 50 kc bias osciallator is used in common for all channels. Each channel contains a bias amplifier to prevent interaction between record channels.

A decibel meter may be switched into one of three modes: input level, record level and bias level. A 30-position switch connects the meter to any one of the record channels.

The first 16 Channels are allocated to audio recording and the incoming audio lines are connected to inputs 1 to 16.

Channels 17 through 30 are allocated to radar recording and these inputs receive signals from the Radar Quantizer which is shown in Figure 10.

Radar Quantizer

The radar quantizer converts the radar information to a form which can be recorded in 14 channels of the recorder. Referring to Figure 10, the following signals are required from the radar:

a. Video
b. Range Markers
c. Time Base
d. Azimuth Servo
e. 0° Azimuth Pulse

The radar video is amplified and sent to an adder where it is summed with the processed range markers and the 0° azimuth pulse. The summed signal feeds the pole of the 130-position electronic switch. The electronic switch is actuated from the radar time base signal via a switch scan generator. (Only one of the 13 quantizing channels is shown.) Every 13th position of the electronic switch feeds a separate detector that converts the 1200 pps signal to 36 millisecond pulses. These pulses are fed to a balance modulator and gate the audio tone "on." The audio tones are derived from oscillators f1 through f10 and amplified in separate audio amplifiers. All ten balance modulator outputs are combined in an adder and the resultant audio signal is amplified and fed to a 600 ohm, 0 db line.

The other positions of the electronic switch excite 12 other identical quantizing channels. The single set of oscillators (f1 - f10) is common to all thirteen channels.

The azimuth servo signal controls a quadrant switch that gates the range markers "on" and "off" as previously described. The azimuth servo signal also feeds an azimuth signal converter which changes the azimuth signal into audio tones for recording on the 30th channel of the recorder. Frequencies F6, F8 and F10 are used for this conversion.

The outputs of the Radar Quantizer are 14 audio lines at 600 ohms impedance and 0 db level.

Reproducer

The reproducer consists of two main elements, the audio section and the radar section. Both are visualized in simplified form in Figure 11.

Audio Section

Tape tracks 1 through 16 are played by their respective magnetic heads and amplified through separate preamplifiers. The 16 preamplifier outputs are connected to a push-button matrix that allows any combination of audio channels to be switched into any combination of transcribe lines, A, B, C, and D. The transcribe line signals are amplified and appear at the four 600 ohm, 0 db lines, across which is bridged a mixing switch that combines the A, B, C, D lines for earphone listening, loudspeaker listening, and for transcribing.

Radar Section

Tape tracks 17 through 30 are played by their respective magnetic heads and amplified in separate preamplifiers. With the exception of channel 30, each preamplifier feeds an audio filter detector network consisting of 10 resonant pass filters F1 through F10 and ten detectors. The combined audio tones on each playback channel are separated by the audio filter, and the detector generates a 36 millisecond pulse for each target. The output of each of the ten detectors is

connected to one associated position of the 130-position electronic switch. As this switch is scanned at a 1200 cps rate, the output of the switch resembles the original radar video signal. The video signal is amplified and applied to the intensity grid of a 10" ppi scope.

A 1200 pps free-running pulse generator excites the switch scan generator as well as the deflection scan generator for the ppi scope.

The preamplifier output of channel 30 is separated through filters f6, f8, and f10 for reconstituting the azimuth signal of the original radar. This azimuth signal controls the motor that rotates the deflection yoke of the ppi scope.

The ppi scope high voltage is generated by the customary high voltage supply.

Mechanical Design

The reproducer is self-contained and is shown in Figure 12. The reels of tape are threaded on the vertical tape deck and the operator sits in front of the console and looks down at the ppi scope as though he is operating a radar. All controls for operating the tape deck functions as well as electrical circuit controls are handy to the operator and are designed in accordance with relevent human engineering considerations. A loudspeaker is located above the tape deck. Large caster wheels and a guide handle are furnished for convenience in moving the reproducer from one area to another.

Conclusion

The audio-radar monitor recorder provides a convenient, reliable and inexpensive means for making a permanent record of the air traffic control situation. Although the system has not been built, experience and study has proved its feasibility. New requirements for recording the transponder returns can be accommodated in future designs.

B. B. Bauer, Vice President of CBS Laboratories contributed to the study. The cooperation of J. E. Grambart, ARDS, Federal Aviation Agency is gratefully acknowledged.

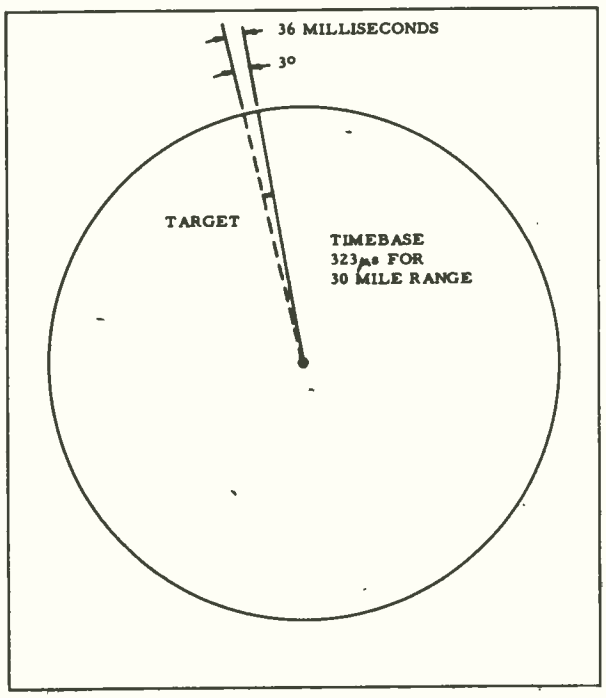


Fig. 1 ASR-3 P.P.I. Scope

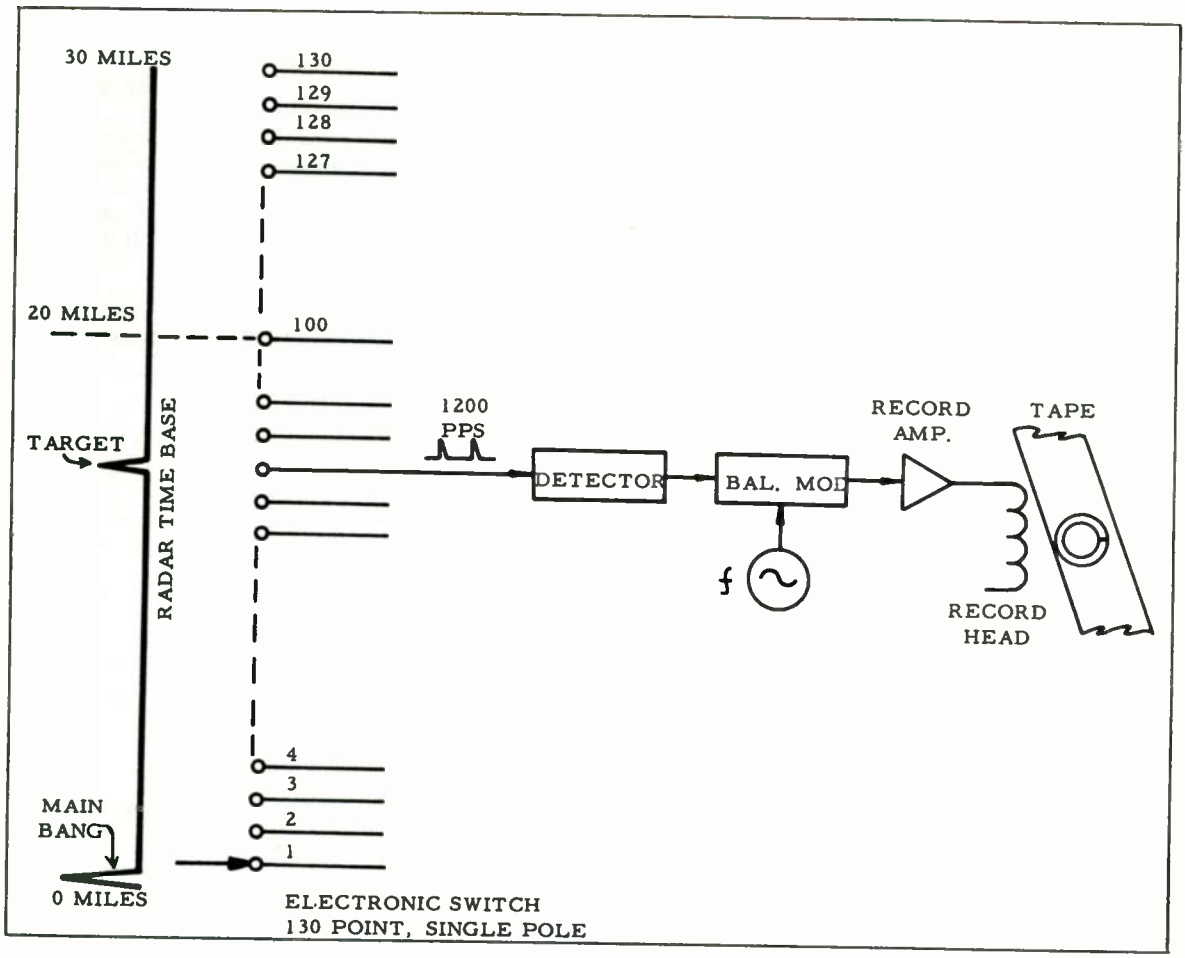


Fig. 2 Basic Radar Recording System

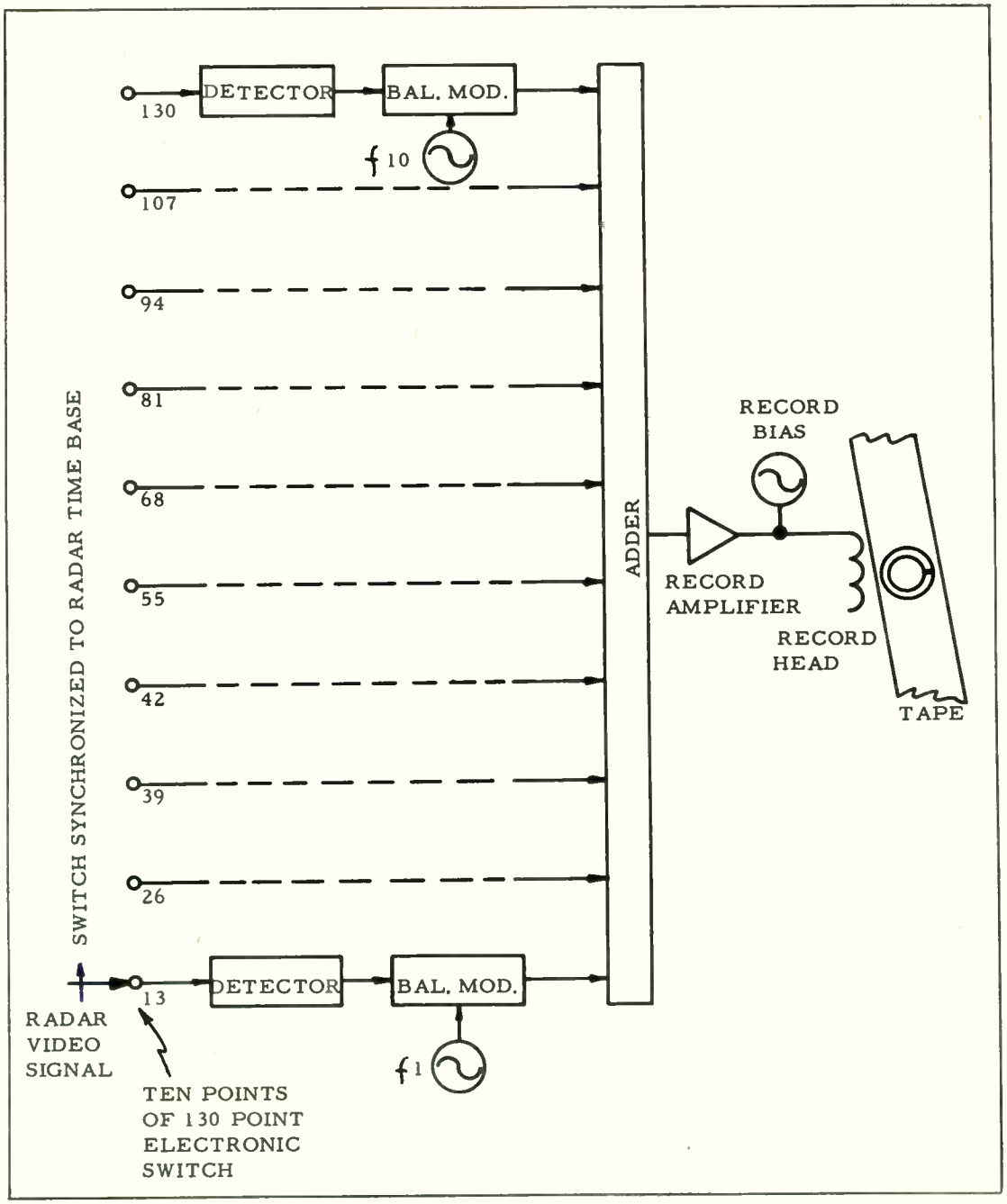


Fig. 3 One of 13 Radar Recording Channels

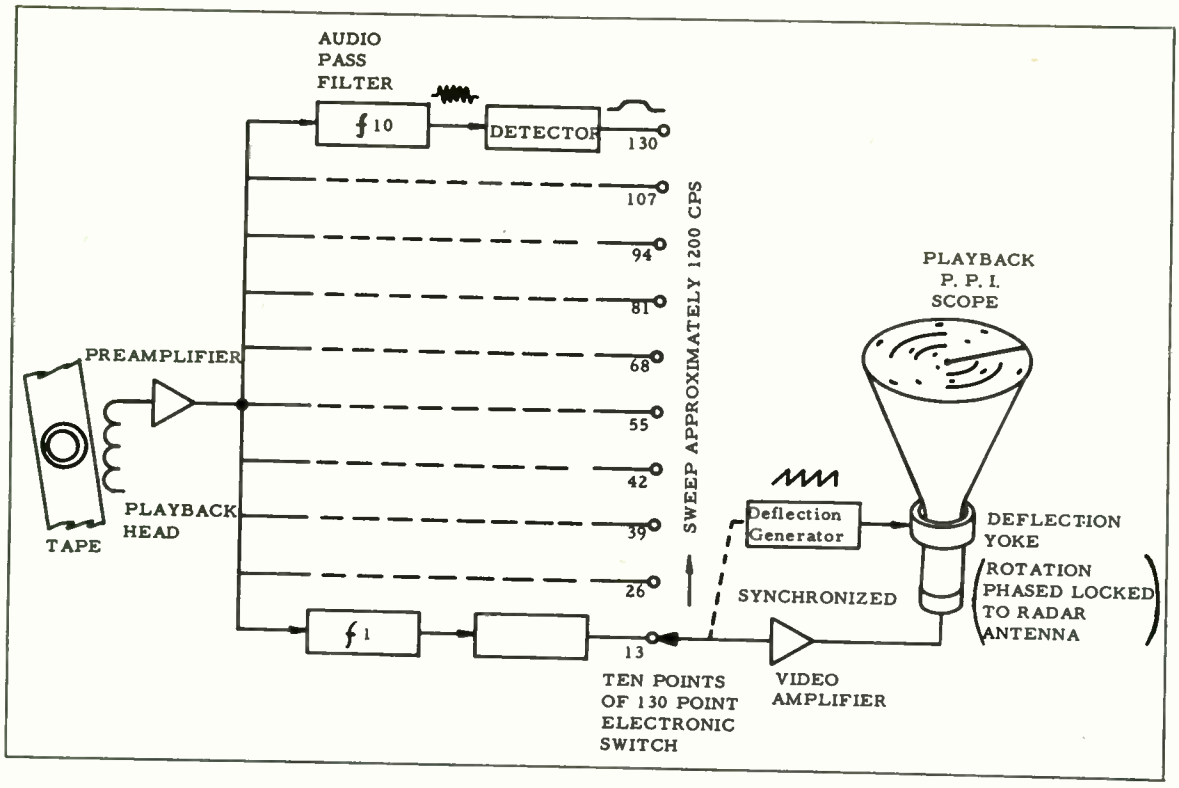


Fig. 4 Playback Channels

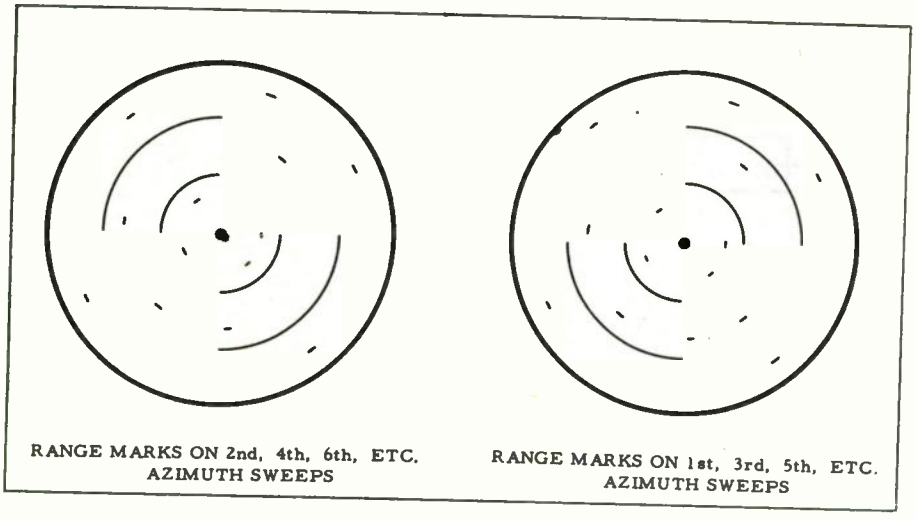


Fig. 5 Range Markers on Playback Scope

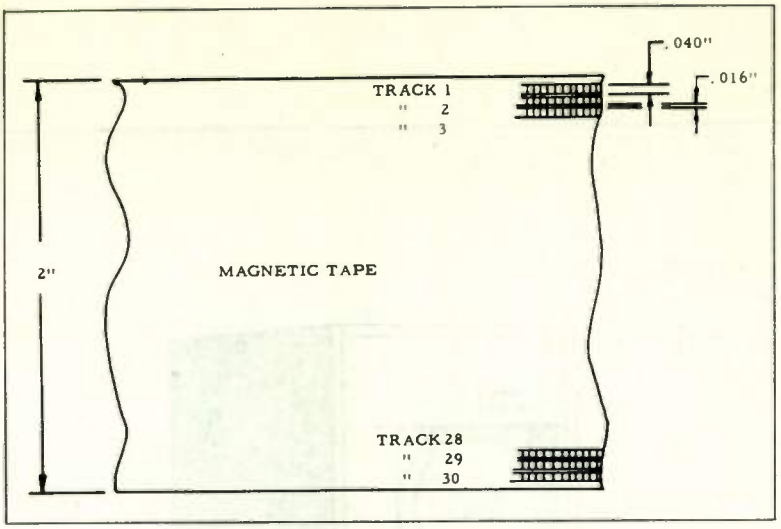


Fig. 6 Magnetic Tape Track Specification

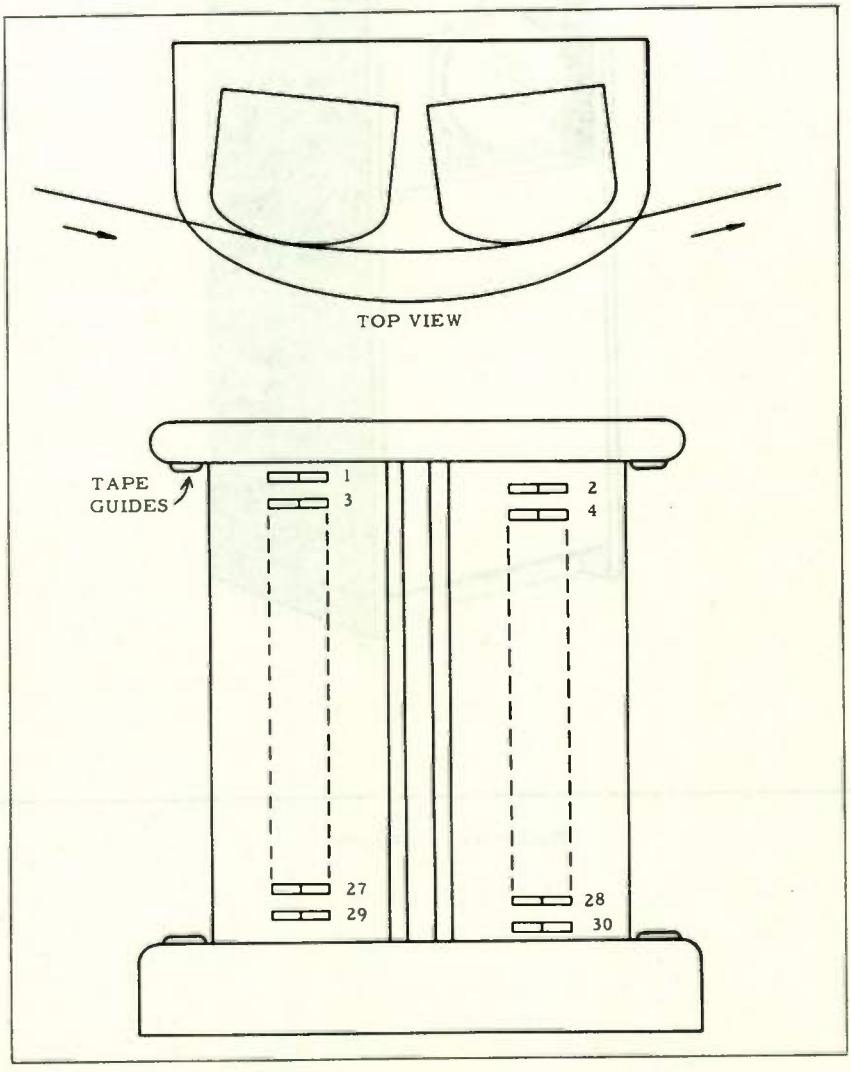


Figure 7. Magnetic Head

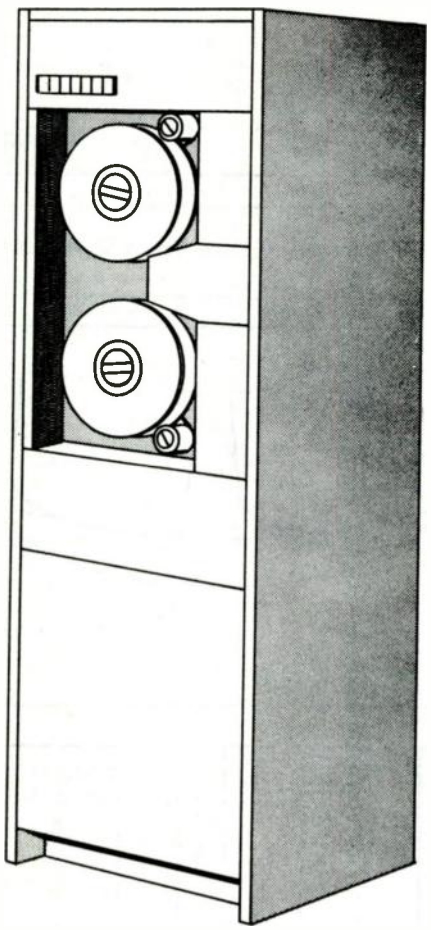


Figure 8. Recorder

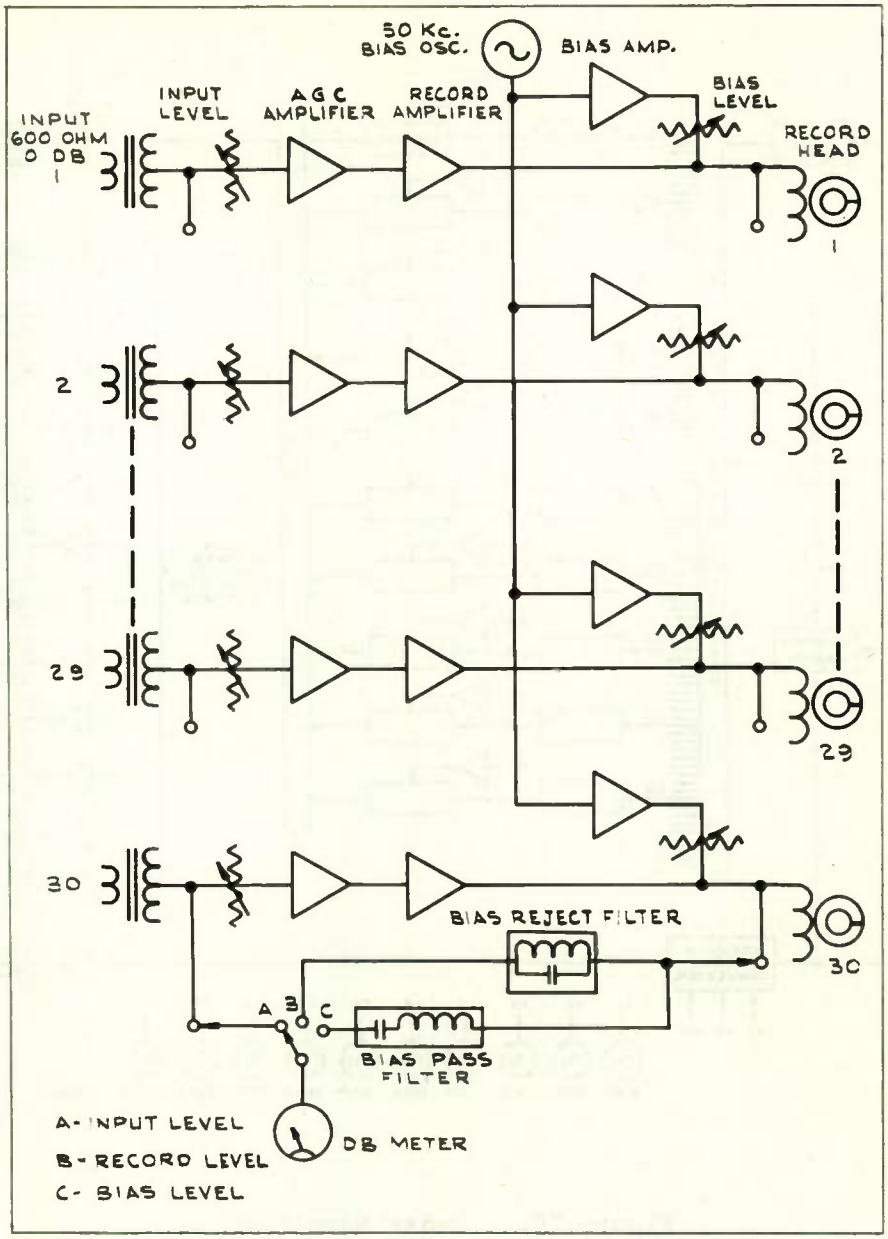


Figure 9. 30-Channel Recorder

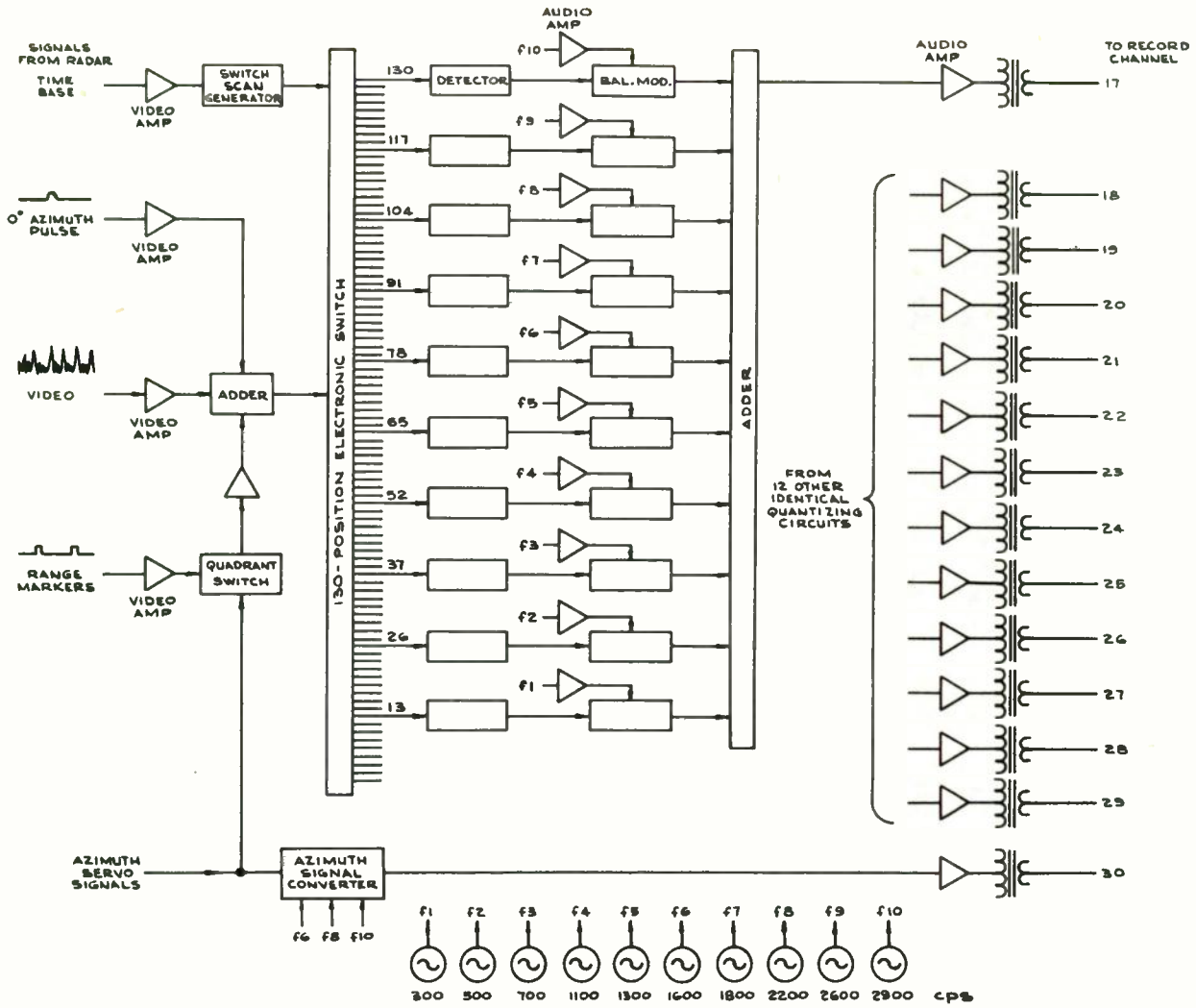


Figure 10. Radar Quantizer

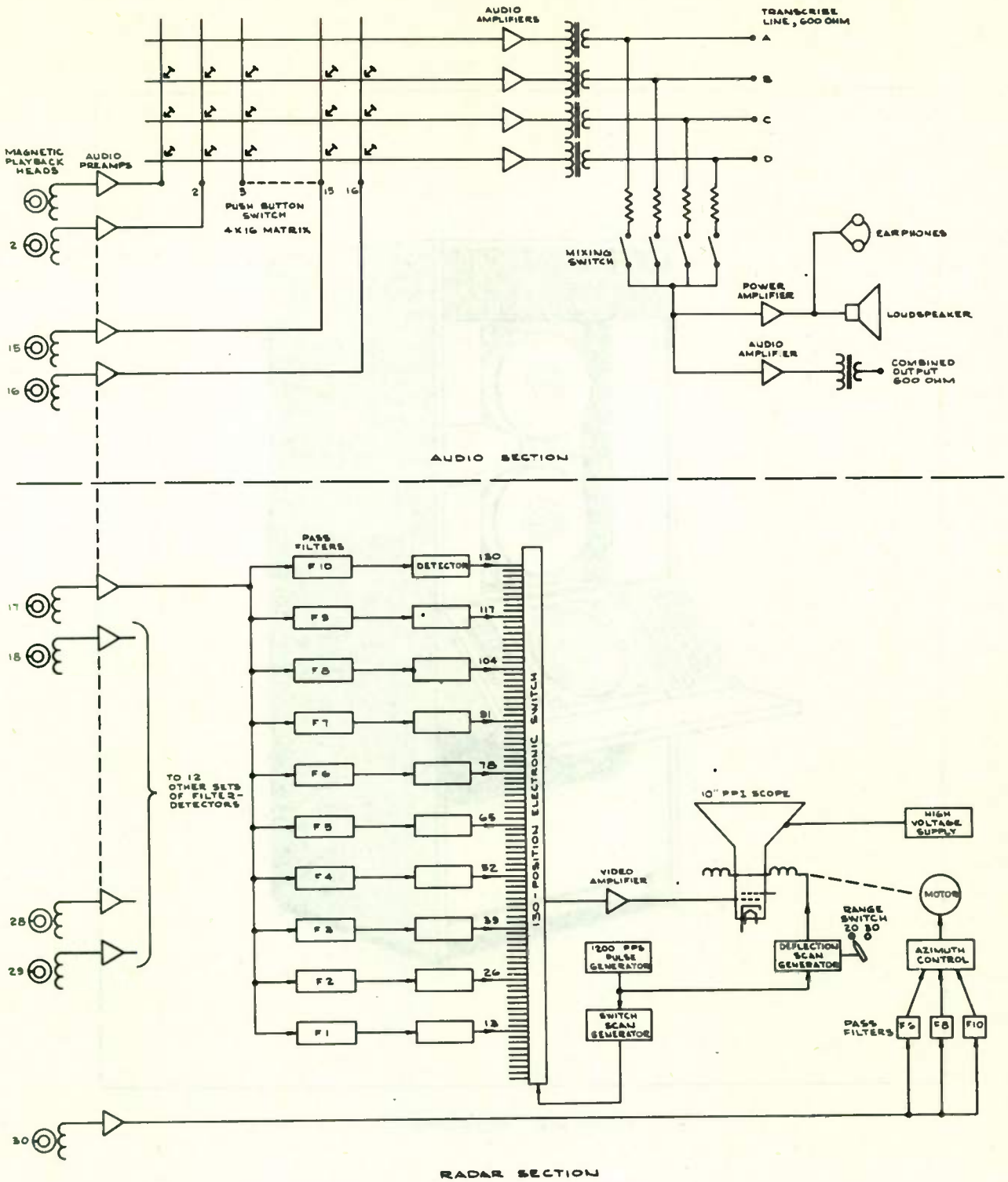


Figure 11. Reproducer Block Diagram

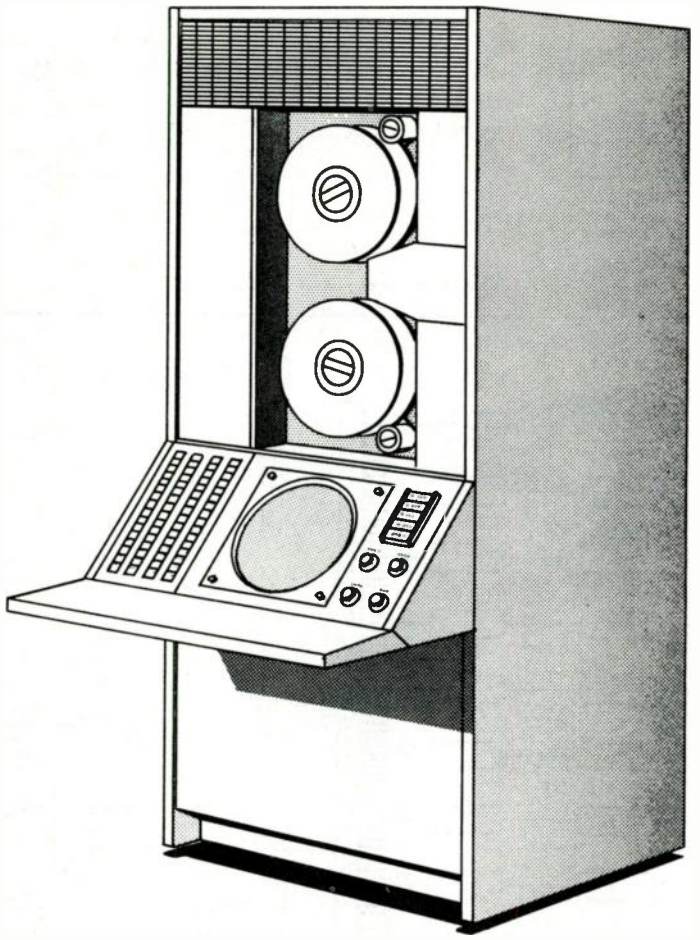


Figure 12. Reproducer

A DIPLEX, DOPPLER PHASE COMPARISON RADAR

Wesley D. Boyer
Scientific Laboratory
Ford Motor Company
Dearborn, Michigan

Summary

A simple method of obtaining range and velocity information has long been sought for numerous applications[1]. One such application concerns itself with detecting the presence of an obstacle in the path of a moving vehicle, the distance to it, and the rate of closure. Since the ranging capabilities of such a system must include an approach to zero distance and extend to several hundred feet, most conventional radar systems can be dismissed as impractically complex and/or requiring large radio frequency bandwidths.

A new and simple radar is presented for ranging to a moving target at close distances. Intended for a range interval of from zero to five hundred feet, the practice of operation may be extended to several times that distance; or on the other hand, very close ranges (on the order of inches) may be measured. An important feature is this system's ability to distinguish a single moving target among a multitude of stationary ones and to correctly indicate the direction of motion--approaching or receding. The target's speed may be determined directly from its Doppler frequency.

Extremely simple RF "front end" and detector circuits, and unusually narrow bandwidths utilized by the solid state receiver, are salient features of a portable system which has been built to operate from an automobile battery.

Introduction

This paper reviews the characteristics of a radar based upon the radiation and reception of two closely spaced microwave frequencies each of which, when reflected from a moving target, gives rise to a corresponding Doppler signal. The Doppler signals are nearly identical within a given range interval except for a phase shift of one with respect to the other. For a given frequency difference of the transmitters the phase difference is linearly dependent on the distance to the target.

The new method of ranging based upon the phase comparison of two Doppler signals utilizes a single transmitter which is diplexed (or switched) with a receiver to simulate the results of independent transmitters and receivers.

Principle of Ranging by Doppler Phase Comparison

The principle of ranging by phase comparison of Doppler signals is not new[2,3], but a review of the fundamentals of a two-frequency radar is desirable.

A single unmodulated continuous wave, radio frequency signal may be represented as a phasor

$$S_{T,1} = A_1 \epsilon^{j\omega_o t} \qquad (1)$$

the real part of the complex expression being understood. If this signal is transmitted in a continuous medium, reflected from a target with a possible phase change θ, and received back at the transmitter position it will have been attenuated (by the factor α) and delayed in time by distance (r) and target (s) dependent functions.

$$S_{R,1} = \alpha_{r,s} A_1 \epsilon^{j(\omega_o t - \theta - \omega_o \tau)} . \qquad (2)$$

The total delay time τ is a function of the variable position only:

$$\tau = \frac{2r}{c} ; \qquad (3)$$

r being the one way range; and c, the speed of wave propagation.

A signal is obtained by mixing the received signal $S_{R,1}$ with a sufficiently large portion of the transmitted signal $S_{T,1}$. The resulting signal $S_{D,1}$ may be looked upon as having an amplitude the same as $S_{R,1}$ and as being periodic in r for every half wave-length of the transmitted signal: thus

$$S_{D,1} = \alpha_{(r,s)} A_1 \epsilon^{-j(\theta+\omega_o\tau)} \qquad (4,a)$$

$$= \alpha_{(r,s)} A_1 \epsilon^{-j(\theta+2\frac{r\omega_o}{c})} . \qquad (4,b)$$

$S_{D,1}$ is periodic in time and truly the Doppler signal only is r is a continuous function of time during a particular time interval of interest. For a target moving with a constant velocity v_r relative to the transceiver the derivative with respect to time of the phase angle of $S_{D,1}$ in Equation (4,b) yields the conventional Doppler angular frequency.

$$\omega_{D,1} = \frac{d}{dt}(\theta+2\frac{r\omega_o}{c}) \qquad (5,a)$$

$$= 2\frac{\omega_o}{c}\frac{dr}{dt} \qquad (5,b)$$

$$= 2\frac{v_r}{c}\omega_o . \qquad (5,c)$$

A second unmodulated CW, RF signal whose

frequency differs from ω_o by a relatively small ω_1 gives rise to the following relations corresponding respectively to Equations (1), (2), (4,b), and (5,c).

$$S_{T,2} = A_2 \epsilon^{j(\omega_o+\omega_1)t} \quad (6)$$

$$S_{R,2} = \alpha_{(r,s)} A_2 \epsilon^{j[(\omega_o+\omega_1)t-\theta-(\omega_o+\omega_1)\tau]} \quad (7)$$

$$S_{D,2} = \alpha_{(r,s)} A_2 \epsilon^{-j[\theta+2\frac{r}{c}(\omega_o+\omega_1)]} \quad (8)$$

$$\omega_{D,2} = 2\frac{v_r}{c}(\omega_o+\omega_1) \quad (9)$$

Figure 1 shows how the two Doppler signals $S_{D,1}$ and $S_{D,2}$ differ as a function of range. The absolute phase angle of each is given by the exponent of the phasor such that

$$\varphi_{D,1} = -\theta-2\frac{r}{c}\omega_o \quad (10)$$

and

$$\varphi_{D,2} = -\theta-2\frac{r}{c}(\omega_o+\omega_1) \ . \quad (11)$$

Although θ was included in this development to account for an unknown phase shift of the reflected signal from the target it is easily eliminated because both RF signals are affected similarly. Taking the phase difference or measureing the relative phase shift of one Doppler wave with respect to the other gives the desired range equation:

$$\varphi_{D,1}-\varphi_{D,2} = \frac{2\omega_1}{c} r \quad (12)$$

or

$$r = \frac{c}{2\omega_1}(\varphi_{D,1}-\varphi_{D,2}) \ . \quad (13)$$

Distinguishing Approaching from Receding Targets. If the Doppler components $S_{D,1}$ and $S_{D,2}$ are drawn as in Figure 2 for a distant target as a function of position, a very interesting feature of this radar scheme becomes evident. The individual signals vary sinusoidally in amplitude as a function of range, but which one leads or lags the other in time-phase depends upon the direction of target movement. The portion of the Doppler signals to the right in (a) beyond an arbitrary range R_a is shown in (b) as the target recedes tracing out the waves. Notice that the signal resulting from ω_o lags the one from $\omega_o+\omega_1$. In (c) the target approaches from R_a tracing out the left hand portion of (a) as time runs on. In this case the wave from ω_o leads the one from $\omega_o+\omega_1$. This information along with a measure of the Doppler frequency (Equations 5 and 9) preserves the sense of direction of the relative motion and completes the radial velocity vector component. Figure 3 shows the phase relationships as a function of range for a practical phasemeter capable of reading both plus and minus 180°. The results are linear and unique within a range interval of $\pi c/2\omega_1$. Between $\pi c/2\omega_1$ and $\pi c/\omega_1$ a receding target appears to give the same phase reading as one approaching in the first two quadrants. This region of non-unique indication is a result of the phase meter's inability to distinguish larger angles than ±180°. The first actual ambiguity, in which an approaching target at some distant point appears to be closer, occurs at a range equal to half the wave length of ω_1; i.e. $\pi c/\omega_1$.

Principle of Diplexing

The principle of diplexing is introduced here to obtain results similar to those obtained by more complex phase comparison systems, but with much greater simplicity. The novelty of the proposed solution lies with the use of a single tube transmitter on a time shared basis between two frequencies. This approach permits pseudo steady state analysis as presented earlier if we place a restriction on the shortness of the period with which the transmitter is switched. Specifically we will limit the switching frequency f_s to a value such that $1/f_s$ is much greater than the time delay of the signal returning from the maximum range of interest. Such a restriction is not severe, for the period of a 100 kc wave is ten times the time delay of an echo from a target at 500 feet. To accurately range to a larger distance a lower switching frequency or other techniques must be used.

Mechanical Frequency Switching Analogy. To illustrate the principle of diplexing consider a transmitter whose frequency is switched from ω_o to $\omega_o+\omega_1$ and back again at a rate of f_s; Figure 4. According to our previous restriction on maximum range the echo signal is received while the transmitter is still emitting a pseudo CW signal. The output of the Doppler receiver is simply the Doppler signal which may be delivered synchronously to separate channels corresponding to the transmitter frequency states. If the Doppler frequency is much lower than the switching frequency the signal in each channel will appear accurately formed but "chopped" at f_s. Simple filtering removes f_s and higher frequencies from the outputs and leaves only the accurately reconstituted Doppler frequency sine waves. A phase meter capable of accepting variable amplitude signals with the spread of anticipated Doppler frequencies and comparing their phases can be calibrated in feet or meters to read out the range directly.

Diplexing by Electronic Modulation. Switching by mechanical methods at rates as high as a hundred kilocycles would be rather difficult and was shown only as an analogy to electronic diplexing. Reflex klystron oscillators possess several very desirable characteristics that lend themselves nicely to this new radar system as a whole. Being microwave devices they are capable of producing the high frequencies necessary for obtaining usable Doppler frequencies from targets moving at moderate and slow speeds--hundreds of miles per hour down to a fraction of a foot per second. Having rather low (but perfectly adequate) power output their power supply requirements are easily met. And having nearly linear

control of the oscillator frequency by changing the repeller electrode potential they are easily frequency modulated. By applying a square wave of voltage to the repeller the frequency is switched an amount proportional to the peak to peak voltage--about one megacycle per volt for a typical low power reflex klystron at X-band.

A particular repeller voltage V_R may be chosen to give the same output power (essentially the maximum) at each frequency as shown in Figure 5. The range equation (13) developed earlier indicates that only the frequency deviation ω_1 need be held constant for a given range resolution. If the dc repeller voltage were to drift to V_R' the output would change and appear at different carrier frequencies ω_o' and $\omega_o'+\omega_1$ perhaps many megacycles from the original ω_o. The linear frequency modulation characteristic, however, maintains the same peak deviation (ω_1) so that the ranging accuracy is not impaired, even though some amplitude modulation has been introduced.

To continue with the illustration of a system designed for a 500' maximum range, the necessary frequency deviation of ±123 kc ($\omega_1 = 1.55 \times 10^6$ sec^{-1}) would be obtained by modulating the klystron repeller voltage with a 246 millivolt peak to peak square wave[4]. A 100 kc repetition rate allows the audio frequency Doppler signals to be easily recovered from the diplexing and provides the necessarily long switching period.

In the practical case degradation of the square wave modulating voltage to a trapezoidal wave is not seriously detrimental to the range measuring capabilities of this system. In fact limiting the frequency response of the modulation circuit to the third or fourth significant component (fifth or seventh harmonic) of the "square" wave frequency may require less transmitted bandwidth. Certainly bandwidths of half to one-and-a-half megacycles at 10,000 to 30,000 megacycles are not extravagant! The receiver needs only to pass the Doppler signals and the significant diplexing (or chopping) frequency components (500 to 700 kc) until the audio frequency Doppler waves are extracted and presented for phase comparison. The final bandwidth is a compromise between the lowest anticipated Doppler frequency and the highest dynamic range rate of the fastest moving target it is desired to follow.

Application of Principles

An experimental radar system employing the foregoing principles has been built in the form of Figure 6. The major component blocks are the microwave circuit, power supply/diplexer, Doppler separator, phase meter, and optionally an AGC driver. The microwave circuit is a simple low power X-band transmitter using the same antenna as the Doppler mixer receiver. The power supply/diplexer is a two transistor converter operating at 100 kc from an automobile battery to supply the high voltages to the klystron, as well as providing the diplexing signal to modulate the transmitter and gate the Doppler separator. The Doppler separator amplifies the composite received signal, separates the two Doppler channels in synchronism with the transmitter modulation, and filters the audio output signals. The phase meter accepts the Doppler signals, amplifies and limits them, and performs the phase comparison which is displayed on a meter or recording device as range.

Obtaining the Doppler Signals

Microwave Circuit. The microwave circuit consists of klystron oscillator, directional coupler, circulator, antenna, tuner, crystal detector, and optionally, a variable attenuator connected as shown in Figure 7. This X-band circuit permits operation at low power with minimum losses. The klystron oscillator produces the diplexed CW signal at a level of about 100 milliwatts which is transmitted through the antenna with typically less than half a decibel loss. A known sample of the generated signal (down 20-30 db) is directed into the crystal mixer for use as the local oscillator signal in the homodyne process. The ferrite circulator allows nearly all of the generated microwave power to be delivered to the antenna thereby isolating the receiver from the transmitted signal. Reflection from an antenna mismatch is usually greater than the inherent leakage through the arms of the circulator but simple tuning with a slide screw tuner will keep the total leakage contribution to the LO signal to an acceptably small level. It can be seen here that reflections from stationary objects illuminated by the microwave act only to add vectorially to the LO injection and can have no effect on the phase difference of the two similarly derived Doppler signals.

Provision is made by means of an electrically variable attenuator to dynamically adjust the level of the received signal as a form of automatic gain control. The ferrite attenuator has negligible effect on the received signal until current is forced through its control coil in response to Doppler signals with amplitudes near the upper limit of the system's dynamic range; i.e. the received microwave signal is delivered to the detector crystal diode at a level much lower than the local oscillator signal in order to assure proper mixing.

Power Supply/Diplexer. High voltages for the operation of the microwave transmitter tube and ac gate signals for the receiver are produced by the power supply/diplexer. This multifunction unit is basically a dc to dc converter with the unusually high conversion frequency of 100 kc being used to obtain the necessary voltages without hum or ripple that would interfere with the small desired modulating voltage applied to the klystron repeller. The intermediate ac signal is used both as the transmitter modulation and as the receiver gate. Being far removed from the Doppler frequencies of interest and coherent with the modulation, it

does not contribute to the overall noise in the information output.

Two high-frequency power transistors operate as a push-pull oscillator delivering power at the 100 kc rate from a 12 volt storage battery to a toroidal transformer. The wave shape at the transistor collectors is roughly square (more trapezoidal due to finite rise, fall and storage times which are an appreciable fraction of the five microsecond semi-period of switching). A pair of Zener diodes is used to limit a portion of the ac signal for use as the internal, regulated modulation voltage determining the frequency deviation of the transmitter. No other regulation was incorporated in the experimental power supply. "Square" waves of voltage, one in phase with the repeller modulating voltage and another of opposite polarity, are available from the Zener diode wave shaping network to be applied in diplexing or gating the Doppler separator. A diagram of the functional parts of the power supply/diplexer and the Doppler separator is shown in Figure 8.

Separating the Doppler Signals

The signal from the microwave mixer diode was shown in Figure 7 to be a composite of both Doppler signals chopped at the 100 kc diplexing rate. Since this signal is very weak it is necessary to amplify it and separate the two signals before a phase comparison may be made. A common transistor preamplifier is used for the composite signal because it is desirable to treat the Doppler signals the same lest one become phase shifted from the other in the processing channels. (After separation the individual Doppler signals are treated alike in two identical channels minimizing the chance for differential phase distortion.) The bandwidth of the transistor preamplifier is about 500 kc since significant frequency components as high as that may be present in the composite signal.

Doppler Separator. The function of the Doppler separator is to extract the two individual Doppler signals from the amplified composite signal preserving their phase relationship to each other. This is accomplished by feeding the same signal to two identical gated amplifiers which are alternately turned on and cut off in synchronism with the transmitter diplexing square wave. The gating is done by driving the emitter of a normally Class A biased, common emitter transistor amplifier to cutoff.[5] Cutoff pulses alternating between the two channels separate the composite signal into the desired two channels, each of which contains only the Doppler wave corresponding to one of the transmitter frequency states.

Clean Doppler signals are recovered after separation by filtering away the unnecessary high frequency components. Simply passing the outputs through low band pass filters and audio amplifiers reduces the signals to audio sine waves whose phases are to be compared.

Measuring the Phase Difference

Perhaps the most important part of the practical range measuring system is the phase meter. It has been shown that the range information is obtained in terms of the phase difference of two audio signals. The method by which the measurement is made determines the accuracy of the result. An experimental circuit was devised to accommodate a wide dynamic range of amplitudes with moderate accuracy for system evaluation. See Figure 9.

Amplification of the very weak signals which result from targets at the maximum range is required before phase comparison can practically be made. Linear amplification is not necessary inasmuch as only the timing or phasing needs to be preserved. Identical diode limiters and amplifiers are cascaded in a series to permit overloading of following stages and to trigger squaring amplifiers (modified Schmidt triggers) at corresponding phases of each input wave. The squaring amplifiers serve a dual purpose: They serve as amplitude gates allowing only signals above their triggering amplitude to pass, and they give a rectangular output wave of uniform amplitude and rise time regardless of the amplitude and frequency of the input sine wave.

To obtain the display of range, characteristic to this system as illustrated in Figure 3, the outputs of the squaring amplifiers are used to set and reset a flip flop whose output wave will vary in symmetry according to the timing of the set and reset pulses. For example, assume that the inputs to the squaring amplifiers are perfectly in phase, but that one of them feeds an inverted wave to reset the flip flop. The outputs of the flip flop will be set and reset symmetrically by the in phase and inverted pulses so that there is zero average voltage between the flip flop outputs. If, however, the phases are different the output will be unbalanced and there will be a net voltage between the outputs. The magnitude of this unbalance will depend on the distance to the target, and its polarity will be determined by the direction of motion, Figure 9, inset. The full scale sensitivity may be adjusted for any practical resolution of range between zero and the maximum unambiguous distance determined by $\pi c/\omega_1$. If the full scale reading of the meter is set to indicate $\pi c/4\omega_1$, i.e. a $\pm$ 90° phase shift, a long buffer zone is established in which no reading is obtained. The amplitude sensitivity of the squaring amplifiers may be adjusted for an average target to the point where it is very unlikely that any target beyond $3\pi c/4\omega_1$ would produce a return strong enough to be construed as a closer target. The problem of ambiguity is thereby reduced significantly, but at the expense of range resolution.

Conclusion

Experimental Test Results

The overall system was built using the equipment as described. All microwave components were commercially available and posed no special problems. The experimental equipment was held on a portable bench with an end-fed, parabolic reflector antenna directed along a straight stretch of roadway toward an automobile used for a test target. The dynamic range-measuring capability of the system was tested by placing hydraulic hoses at one hundred foot intervals in the test range and recording the range/phase against these markers. The milliammeter readout indication could be calibrated linearly in terms of range and worked well out to 500 feet. In a limited number of trial runs the distance to the car at the markers could be established well within 10% from 100 feet out to 500 feet. Accuracy closer than 100 feet was not measured but phase readings were obtained from as close as a few feet. It is at these very close distances where it becomes necessary to incorporate the electrically variable attenuator described as optionally useful. The return echo must be maintained at an amplitude level much weaker than the local oscillator injection for proper Doppler mixing and phase comparison.

Other experimental tests applying the system to extremely close ranging were conducted on a 12' indoor range. The moving target in this case was a toy train on a model railroad track which approached to within a few inches of the transmitting antenna. Using a frequency deviation of about $\pm$ 5 megacycles ($\omega_1 = 2\pi 10^6 \text{sec}^{-1}$) permitted a range resolution of about four centimeters per degree.

Extension of Results

It was not intended to establish quantitative accuracy measurements but rather to apply the principles involved and prove that such a system would work. There is room for much sophistication if it is desired to make accurate near as well as far measurements. It appears possible to combine the results of the two experimental tests described by a simple extension of the principle of diplexing to the use of more than two frequencies or multiplexing. Thus a system with both coarse and fine resolution of range is possible.

It is conceivable that the microwave circuitry could be simplified considerably by designing components of multiple function, for example combining circulator, directional coupler and variable attenuator into a single device. Although tests were made at X-band around 10 Gc there is no reason why higher "ultra-microwave" or even optical frequencies may not be used just as successfully or to greater advantage because of higher Doppler frequencies for very slowly moving targets. Furthermore, smaller more directional antennas would replace large parabolic reflectors and give greater gain, too.

Places where a simple system such as this may be used to advantage are in aircraft landing altimetry, range gating, proximity fuzing, or for a variety of moving target locators where there may be many stationary reflectors within the range of interest. Land or sea vehicle obstacle detection systems may be considered with due respect for the fact that all "stationary" objects become moving targets when the system is in motion, but the relative velocity and distance between similarly moving objects may be determined by frequency discernment. A plurality of moving targets adds greatly to the complexity of this (or any other) Doppler range measuring system and hence lies beyond the present scope of this study.

References

1. An extensive bibliography of recent radar developments is given by W. K. Saunders, "Post War Developments in Continuous-Wave and Frequency-Modulated Radar," IRE Transactions on Aerospace and Navigational Electronics, Volume ANE-8, Number 1, March 1961.

2. W. W. Hansen, "Radar System Engineering," L. N. Ridenour Ed., McGraw Hill Book Co., Inc., New York, New York, ch. 5, 1947.

3. R. H. Varian, W. W. Hansen and J. R. Woodyard, U. S. Patent 2,435,615, applied for September 30, 1941; granted on February 10, 1948.

4. A brief analysis of this square-wave frequency modulation technique, but in terms of trangular-wave phase modulation, is given by O. K. Nilssen, "New Methods of Range Measuring Doppler Radar," Ford Motor Co., Scientific Laboratory, to be published.

5. An alternate approach is to add a large portion of the modulating square wave to the composite signal and perform both positive and negative peak detection of the resulting envelope thereby separating the two Doppler signals.

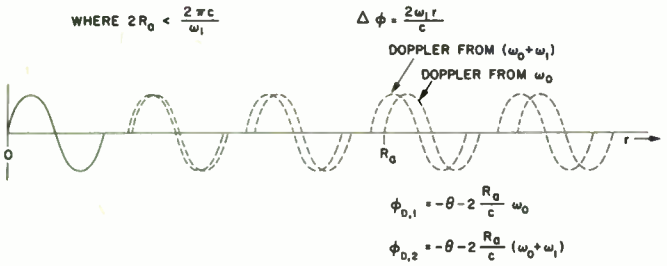


Fig.1 Ranging by phase comparison of doppler signals

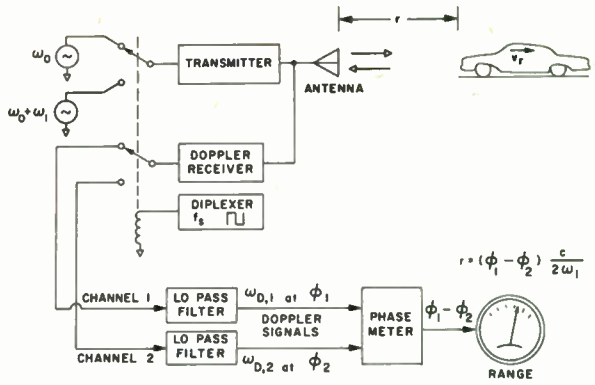


Fig. 4 The principle of diplex, phase comparison ranging

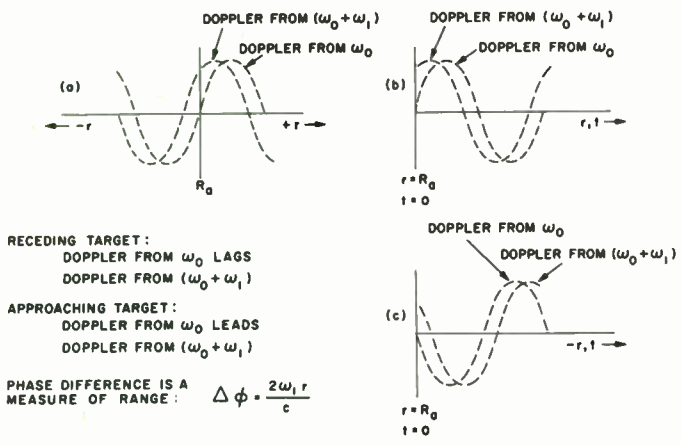


Fig. 2 Distinguishing approaching from receding targets

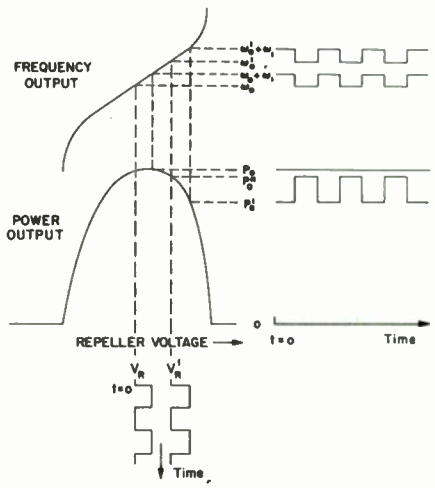


Fig. 5 Diplexing by frequency modulation of a reflex klystron oscillator

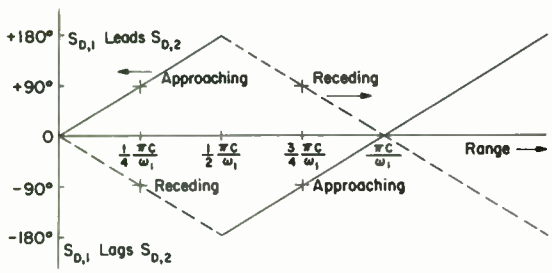


Fig. 3 Phase relationships of doppler signals by phasemeter capable of reading ± 180°. $S_{D,1}$ is doppler signal resulting from ω_0 ; $S_{D,2}$ results from $\omega_0+\omega_1$

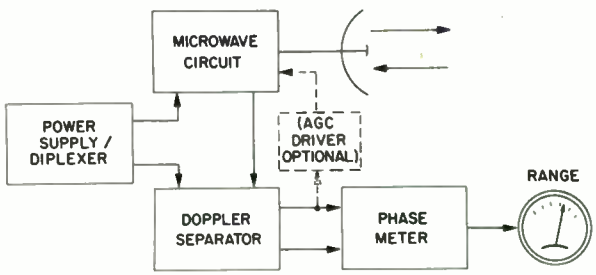


Fig. 6 Experimental radar system block diagram

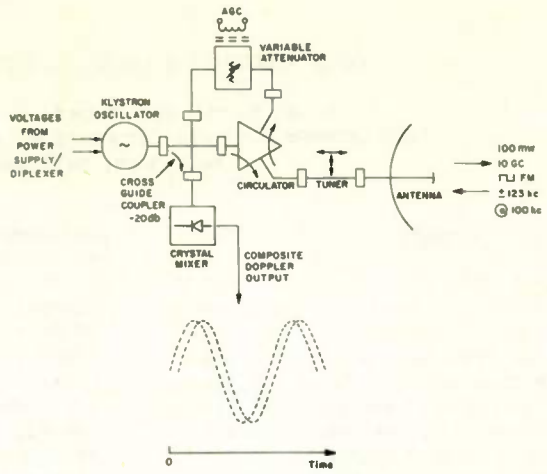


Fig. 7 Microwave circuit for diplex, doppler radar

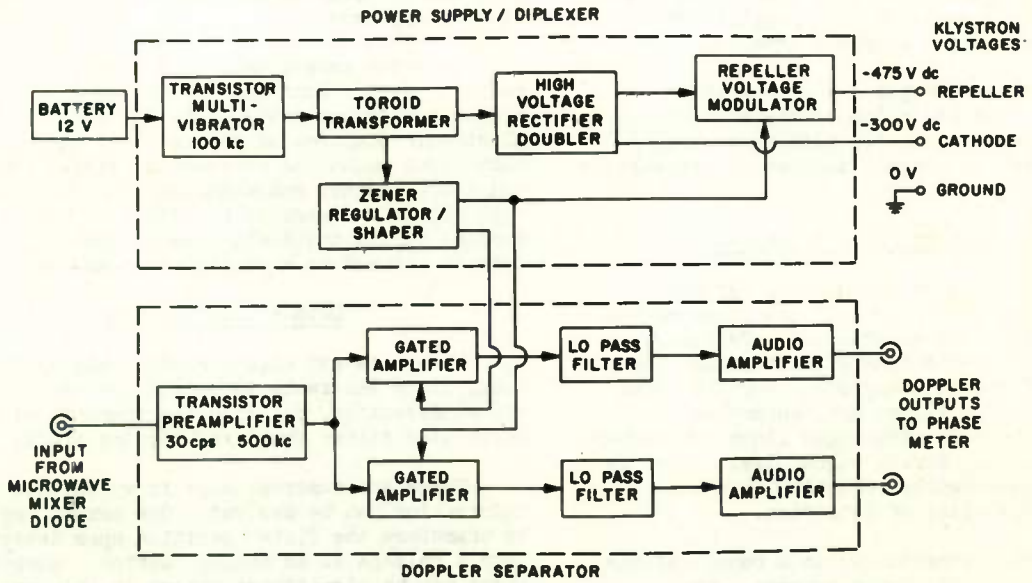


Fig. 8 Functional diagram of system components

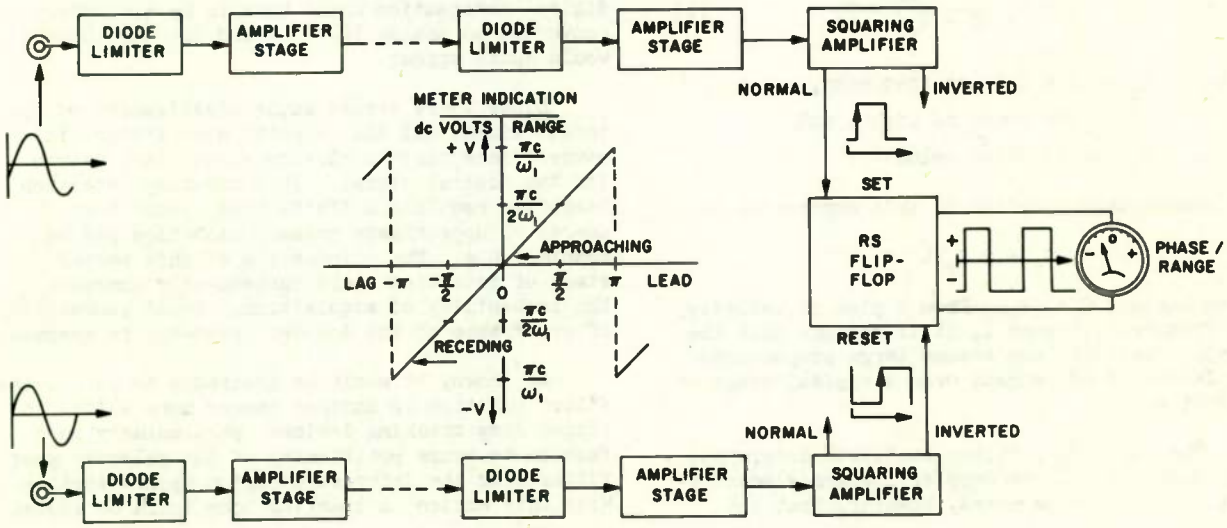


Fig. 9 Functional diagram of phase meter

RAPID ACQUISITION OF DOPPLER FREQUENCIES

Walter Ewanus and Francis J. Burkitt
Westinghouse Electric Corporation, Air Arm Division
Baltimore, Maryland

Summary

This paper describes a system which performs rapid acquisition of doppler signals over wide bands of frequency uncertainty. This uncertainty, resulting from the nature of transmissions eminating from moving sources, dictates incompatibility with narrow band tracking devices. The acquisition process utilizes high speed, electronic scanning of a contiguous bank of filters for primary detection; and it processes the result in a manner which is applicable to narrow band frequency or phase tracking.

The technique proposes a method which provides significant reduction of the acquisition time of unknown frequencies with an attendant improvement in the overall cumulative probability of acquisition.

Doppler Radar Techniques

Doppler detection techniques, whether pulsed or CW, utilize spectrum analyzing methods to measure frequency uncertainty. Sweeping filters or contiguous bank of filters are the most common forms of analyzers. For the radar detection of signals, the contiguous bank of filters provide some advantages since the number of signal samples, during signal dwell time, is increased significantly resulting in a more favorable probability of detection.

The doppler uncertainty is a result of the relative velocity of moving sources. The frequency shift is uniquely described by the equation

$$f_d = f_c \left(\frac{C + V_r}{C - V_r}\right) - f_c \qquad (1)$$

where f_c is the carrier frequency,

C is the speed of light, and

V_r the relative velocity.

The common approximation to this expression is

$$f_d = 2 V_r/\lambda$$

which assumes $C \gg V_r$. From a plot of velocity vs. frequency, Figure 1, it is obvious that the doppler frequency can assume large proportions for detection of targets over a typical range of velocities.

The individual filter bandwidth determines the resolution of the doppler frequency measurement. It should be noted, however, that the designer must also consider other factors, such as dwell time and signal-to-noise ratio in order to obtain the highest probability of target detection. To permit a wider latitude for basis of filter selection, post detection integration (PDI) is commonly incorporated to reduce the overall effective noise bandwidth. PDI is an effective means for removing some of the severe restrictions on this filter. The optimization of filter and PDI parameters is relatively complex and can only be determined by an integrated systems approach.

In order to evaluate target presence, a decision device must analyze the detected outputs of the contiguous filter bank (or comb filter). Electronic scanning of the detected outputs is a convenient means for determining signal presence and its frequency characteristics. This scanning rate should be large to provide a number of samples during any dwell time so that scanning loss is reduced to a negligible quantity.

Doppler Acquisition

When a target signal excites the filter bank, it is desirable to obtain not only the signal detection, but also the identity of the particular filter that exhibits the signal.

There are numerous ways in which this information can be derived. One common way is to transform the filter position upon detection into a voltage in an analog fashion. However, there can be significant errors in this analog transform and subsequent storage. It is also possible to code this filter position information digitally, but this is not desirable since the digital information would have to be ultimately converted to analog (D/A) -- and the inaccuracies would again appear.

Since these errors cause misalignment of the target signal and the velocity gate filter, it becomes necessary to perform a secondary search for the desired signal. This two-step detection technique requires a finite time, which for coarse or approximate primary detection can be considerable. The elimination of this second stage of detection would subsequently increase the probability of acquisition. Equal probability of occurrence of the doppler frequency is assumed.

As shown, it would be desirable to derive the filter position in another manner more suited to closed loop tracking devices, particularly in a fashion to cause positioning of the velocity gate filter over the interested region upon detection. With this action, a tracking loop could be closed

on the detected signal immediately without recourse to a secondary search phase. With an electronic scanning switch, or interrogator, it is feasible to cause the interrogator to acquire the excited filter and cease further scanning (primary detection). By acquiring the filter of interest, we now have provided the unique means for identification of the proper filter. This scan detection describes the doppler signal within an accuracy of $\pm \Delta f/2$, where Δf is the filter bandwidth. The requirements necessary to close the tracking loop upon detection are twofold:

1. The velocity gate bandwidth must be equal to the resolution of the filter bank analyzer, and
2. The detected signal frequency (F_d) requires unique translation of its spectrum into the center frequency of the velocity gate (F_h).

The necessary requirement of condition two at this point is the generation of a group of frequencies, each frequency of which is associated in some prescribed manner with a filter in the contiguous filter bank, illustrated in Figure 2A. These frequencies could be represented by a group of discrete oscillators which are synchronously sampled by the scanning switch. Upon detection the switch selects one of these oscillators (oscillating at F_{sg}) so that this frequency could be uniquely associated with the filter which had exhibited the signal. It now becomes possible to hetrodyne the input signal F_d and F_{sg} to a new frequency $F_{sg} - F_d$. The mixing process generates $F_{sg} \pm F_d$ for which a low pass filter selects only $F_{sg} - F_d$. If another limitation is placed on F_{sg}, namely that F_{sg} is equal to some frequency $F_h + F_d$, it is obvious that the hetrodyned signal $F_{sg} - F_d$ is equal to $F_d + F_h - F_d$ or just F_h. The center frequency of the velocity gate was chosen to be F_h so that we have achieved the initial step in the acquisition of the input signal; namely, placing the signal in a narrow band filter.

It is evident, however, that a large number of filters in a bank would require an equal number of oscillators. It would be advantageous to derive a channelized, discrete spectrum by other means to alleviate this burden. For example, suppose the oscillators were replaced by an auxiliary filter bank identical to the signal channel, except the entire bank is offset by an amount equal to the center frequency of the velocity gate. If this auxiliary bank of filters was excited by a frequency synthesizer (hereafter called a spectrum generator) so that one frequency exists for each filter, the individual frequencies could be channelized and selected in synchronism with the scan of the signal bank, as illustrated in Figure 2B. Selection of a frequency, F_{sg}, in this manner would uniquely determine a frequency offset from F_d by $F_h \pm e$ where $\pm e \equiv \pm \Delta f/2$, or the velocity gate bandwidth. The velocity gate filter frequency is derived by simple mixing of the doppler velocity gate filter frequency F_d and the selected frequency F_{sg}. Acquisition is then accomplished by simple switching upon detection.

It now becomes necessary to describe a means for generating a spectrum with the above requirements. If a frequency source were modulated by a signal equal to the filter spacing, we could reduce the required number of oscillators by a factor determined by the number of sidebands which were generated. It is obviously advantageous to utilize one oscillator and generate sufficient sidebands to cover the entire bandwidth of interest. There are several ways which such a spectrum could be generated. For example, pulse modulation of a sine wave produces a $\frac{\sin X}{X}$ amplitude function with discrete frequencies appearing at the modulation frequency centered around the carrier. If the modulation frequency is selected to be the filter spacing, this requirement could be obtained. To keep the amplitude from changing radically from filter to filter, it would be judicious to choose a pulse width (T) so that $\frac{1}{T} > \Delta F$ where ΔF is the entire filter bank bandwidth.

Another modulation scheme which would meet this requirement is linear FM of a sinusoid. Although generation of the spectrum by either technique appears to be straight forward, the designer must consider amplitude and phase variations over the filter pass band for reasons stated in subsequent paragraphs. For the moment let it be assumed that ideal filters are considered.

An additional refinement can now be considered which eliminates the auxiliary contiguous bank. After the initial detection is made and the interrogator acquires the excited filter, the main bank is no longer required to perform any other function. Therefore, if at the time of detection, we switch the spectrum gnerator output into the main filter bank, we will obtain a discrete signal which is within $\pm \frac{\Delta f}{2}$ of the original signal, refer to Block Diagram, Figure 3. Now all that is required is to hetrodyne this discrete signal with a translating oscillator whose frequency (F_h) is centered at the velocity gate filter. The mixing process generates $F_{sg} \pm F_h$ which is filtered to obtain the single sideband spectrum $F_{sg} + F_h$. The mixing of this frequency with the original target signal (F_d) results in a frequency $F_h \pm \frac{\Delta f}{2}$ which is the expected result. Since:

$$F_{sg} = F_d \pm \frac{\Delta f}{2} \qquad (1)$$

and

$$F_{sg} + F_h - F_d = F_h \pm e \qquad (2)$$

where e is the error term. Substituting for

F_{sg} from (1),

$$(F_d \pm \frac{\Delta f}{2}) + F_h - F_d = F_h \pm e \qquad (3)$$

$$F_h \pm \frac{\Delta f}{2} = F_h \pm e \qquad (4)$$

then

$$e = \frac{\Delta f}{2} \qquad (5)$$

The acquisition error as shown is the half bandwidth of the velocity gate filter which is a heuristic proof of signal acquisition in a narrow band filter.

Since ideal filters were discussed in the filter bank, the degradation from ideal must be considered. Due to the finite skirt selectivity of individual filters, the adjacent channels will cause apparent sidebands to appear on the frequency F_{sg}. These spurious signals appear as amplitude modulation of F_{sg} so that the output frequency could be described as a carrier with sidebands, $F_{sg} + (F_{sg} \pm \Delta f)$. Hard limiting of the output is utilized to suppress the spurious sidebands to a non-interfering level. The second order sidebands ($F_{sg} \pm 2\ \Delta f$) are neglected.

To show the effects of hard limiting on the spurious signals for amplitude and phase distortions, a curve (Figure 4 - Sideband Power vs. Frequency Displacement 'P_{SB} vs. X') is derived in Appendix A. Several cases of spectral spacing are shown where the factor V is the normalized frequency spacing in terms of half bandwidth. For illustration, taking an error of 10% in filter displacement and V = 2, which is the realistic case of adjacent filters, P_{SB} is found to be 32 db down from the desired signal. Experimentally sideband rejection with hard limiting has been measured with filters not expressly designed for this purpose, i.e., poor skirt selectivity. The resulting suppression with one stage of limiting, was measured to be 23 db. Therefore, if care is exercised in specifying the filter and spectrum characteristics, it is evident that a practical system could be achieved with little difficulty.

Conclusions

In determining the improvements derived from the use of this fast acquisition technique, the cumulative probability of acquisition was calculated for three cases. These three cases are illustrated in Figure 5.

The first, Curve A, represents the case where the velocity gate performs a narrow search in order to detect the signal. Using a false alarm time of 30 minutes, a 500 millisecond sampling time, a 400 cps velocity gate bandwidth, and 12.5 samples, the method of Marcum[1] was applied to obtain the detection curve A, Figure 5.

The second case, Curve B, was calculated for velocity gate positioning by the rapid acquisition techniques. All parameters were kept identical with the first case for comparison. A 2.7 db improvement in the signal to noise ratio (S/N) was noted for a 50% cumulative probability of acquisition.

The third case, Curve C, Figure 5, was calculated for a higher false alarm rate since a velocity gate search is not required and the total number of signal samples could be reduced to the same number as in Case A. In effect then, the sampling time was reduced from 500 milliseconds to 25 milliseconds resulting in 12.5 samples and the false alarm rate was increased by a factor of 100. The net result was a relatively minor decrease in the probability of acquisition from Case B.

Therefore, to summarize, the use of rapid acquisition techniques can improve the overall probability of acquisition of a doppler system and give a substantial reduction in the time of acquisition (milliseconds or less).

APPENDIX A

The problem is to determine the spectrum of an infinitely limited output of a filter excited by a comb spectrum:

Input Signal:

$$s(t) = \sum_n \cos\ (\omega_o + n\ \omega_n)\ t \qquad (A1)$$

The filter bandwidth is such that only one spectral line and two weaker adjacent lines are predominant at the filter output. The significant output of the filter prior to limiting can be expressed as:

$$\begin{aligned} r(t) = a_o \cos \omega_o t + \cos \omega_o t \big[& a_1 \cos(\omega_n t + \phi_1) + \\ & a_2 \cos\ (\omega_n t - \phi_2)\big] + \\ & \sin\ (\omega_o t) \big[a_2 \sin\ (\omega_n t - \phi_2) - \\ & a_1 \sin\ (\omega_n t + \phi_1)\big] \end{aligned} \qquad (A2)$$

where

$$\phi_1 = \phi_{(\omega_o + \omega_n)} - (1 + \frac{\omega_n}{\omega_o})\ \phi_{\omega o}$$

$$\phi_2 = \phi_{(\omega_o - \omega_n)} - (1 - \frac{\omega_n}{\omega_o})\ \phi_{\omega o}$$

It is desirable to represent r(t) as the sum of a carrier, a symmetric sideband pair, and an anti-symmetric sideband pair

$$\begin{aligned} r(t) = & +2\ b_s \cos \omega_o t \cos\ (\omega_n t + \phi_2) - \\ & - 2\ b_a \sin \omega_o t \sin\ (\omega_n t + \phi_a) \end{aligned} \qquad (A3)$$

If $a_1^2, a_2^2 \ll \frac{a_o^2}{4}$, then r(t) can be considered as a composite of a narrow band phase modulated signal with weak amplitude modulation, the symmetric sideband pair being AM. Infinite limiting will eliminate the symmetric sidebands. The resulting signal after infinite limiting will be approximately

$$p(t) = a_o \cos \omega_o t + \sin \omega_o t \left[a_2 \sin (\omega_n t - \phi_2) - a_1 \sin (\omega_n t + \phi_1)\right] \quad (A4)$$

equating coefficients and solving for b_a, we get

$$b_a = \frac{1}{2}\sqrt{a_1^2 + a_2^2 - 2 a_1 a_2 \cos (\phi_1 + \phi_2)} \quad (A5)$$

Setting

$$a_1 = a_2 (1 + \Delta a)$$

and

$$(\phi_1 + \phi_2) = \phi_{(\omega_o + \omega_n)} - 2\phi_{\omega_o} + \phi_{(\omega_o - \omega_n)} = \Delta\phi$$

we get

$$b_a \approx \frac{1}{2} a_2 \sqrt{\Delta\phi^2 + \Delta a^2} \quad (A6)$$

A narrow band phase modulated signal can be approximated by

$$J_o(u) \cos \omega_o t + J_1(u) \cos\left[(\omega_o + \omega_n) t + \gamma\right] - J_1(u) \cos\left[(\omega_o - \omega_n) t - \gamma\right]$$

The error in this approximation is less than $\frac{u^2}{8}$ and for a small index of PM, $J_o(u) \approx 1$ and $J_1(u) \approx \frac{u}{2}$. Therefore

$$u \approx 2\frac{b_a}{a_o} = \frac{a_2}{a_o}\sqrt{\Delta\phi^2 + \Delta a^2} \quad (A7)$$

for

$$\Delta\phi^2, \Delta a^2 \ll a_2^2$$

The spectral power will be approximately

$$P_{sb} = \left(\frac{a_2}{a_o}\right)^2 \frac{\Delta\phi^2 + \Delta a^2}{4} \quad (A8)$$

in each sideband.

Example:

Assume n cascaded single tuned stages. For mean resonance, we have

$$A = \left[\frac{1}{1 + \left(\frac{\omega_r - \omega}{\Delta\omega}\right)^2}\right]^{\frac{n}{2}} \quad (A9)$$

$$\phi_\omega = n \tan^{-1}\left(\frac{\omega_r - \omega}{\Delta\omega}\right) \quad (A10)$$

where $\Delta\omega = \frac{1}{2}$ bandwidth of one stage

$\omega_r = \omega_s + \omega_o$

ω_s = the error from ω_o

Therefore, for n = 2

$$\Delta a = \frac{a_1}{a_2} - 1 \quad (A11)$$

or

$$\Delta a \approx \frac{-2\frac{\omega_n^2 \omega_s^2}{\Delta\omega^4}}{1 + \left(\frac{\omega_n}{\Delta\omega}\right)^2} \text{ for } \omega_s^2 \lll \omega_n^2 \quad (A12)$$

and

$$\Delta\phi = 2\left[\tan^{-1}\frac{\omega_s + \omega_n}{\Delta\omega} + \tan^{-1}\frac{\omega_s - \omega_n}{\Delta\omega} - 2\tan^{-1}\frac{\omega_s}{\Delta\omega}\right] \approx \frac{4\omega_s}{\Delta\omega}\left(\frac{-1}{1 + \frac{\Delta\omega^2}{\omega_n^2}}\right) \quad (A13)$$

Let

$$v = \frac{\omega_n}{\Delta\omega} \text{ and } x = \frac{\omega_s}{\Delta\omega}.$$

Then

$$P_{sb} = \frac{v^4 (4x^2 + x^4)}{(1 + v^2)^4} \quad (A14)$$

for $v = 2$ or $2 = \frac{\omega_n}{\Delta\omega}$, $2\Delta\omega = \omega_n$, or when the adjacent sideband is removed by twice the $\frac{1}{2}$ bandwidth of the filter.

$$P_{sb} = \frac{16}{625}(4x^2 + x^4) \quad (A15)$$

where the x^2 term is due to $\Delta\phi$ and

x^4 term is due to Δa.

In this case, the contribution due to Δa can be neglected. However, any amplitude distortion of the input comb spectrum must be interpreted in the Δa term of Equation (A8). If it is deterministic, it should be added as a scalar quantity

$$P_{sb} = \left(\frac{a_2}{a_o}\right)^2 \cdot \frac{\Delta\phi^2 + (\Delta a_F + \Delta a_c)^2}{4} \quad (A16)$$

where Δa_c is due to difference in amplitude of input comb spectrum and Δa_F is due to the filter, if it is taken to be a random quantity of some mean square average, the sideband power becomes:

$$P_{sb} = \left(\frac{a_2}{a_o}\right)^2 \cdot \frac{\Delta\phi^2 + \Delta a_F^2 + \Delta a_c^2}{4} \qquad (A17)$$

Note: The expressions used for the tuned stage, Equations 9 and 10, are only approximate and x and v are in units of individual stage $\frac{1}{2}$ bandwidth, $\Delta\omega$. The results are shown in Figure 4. The parameters x and v have been normalized to be in units of the overall bandwidth of two cascaded stages and are represented by the symbols X and V respectively.

$$x = 1.3\ X \qquad v = 1.3\ V$$

It will be noticed that for V = 2, the sideband power, P_{sb}, is 32 db removed from the carrier power for a 10% error in alingment of the spectral line from the center of the filter.

REFERENCES

1. Marcum, J.I., "A Statistical Theory of Target Detection by Pulsed Radar," Rand Corporation Report No. RM-754.

ACKNOWLEDGMENT

The authors are grateful for the analytical assistance and suggestions provided by J. H. Mims and D. H. Mooney of the Westinghouse Electric Corporation where this experimental work was performed.

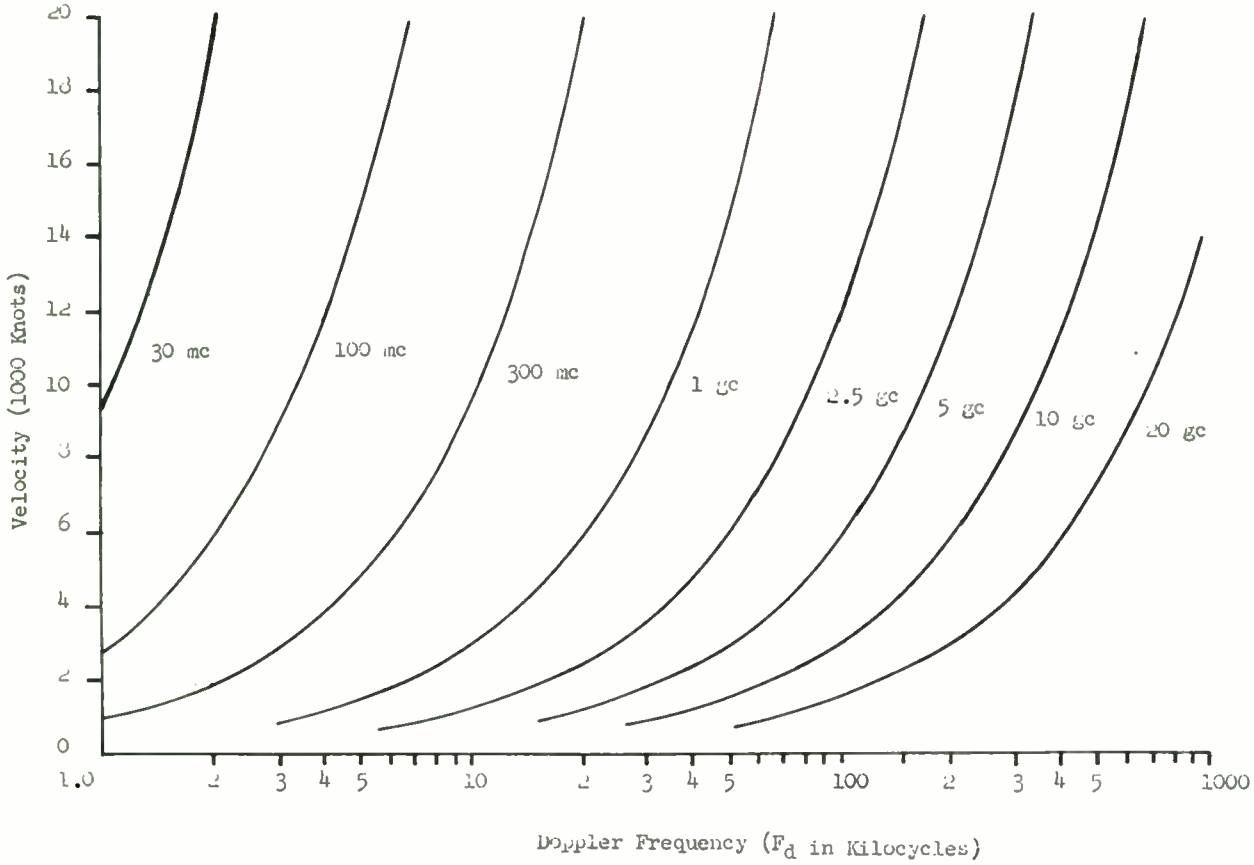
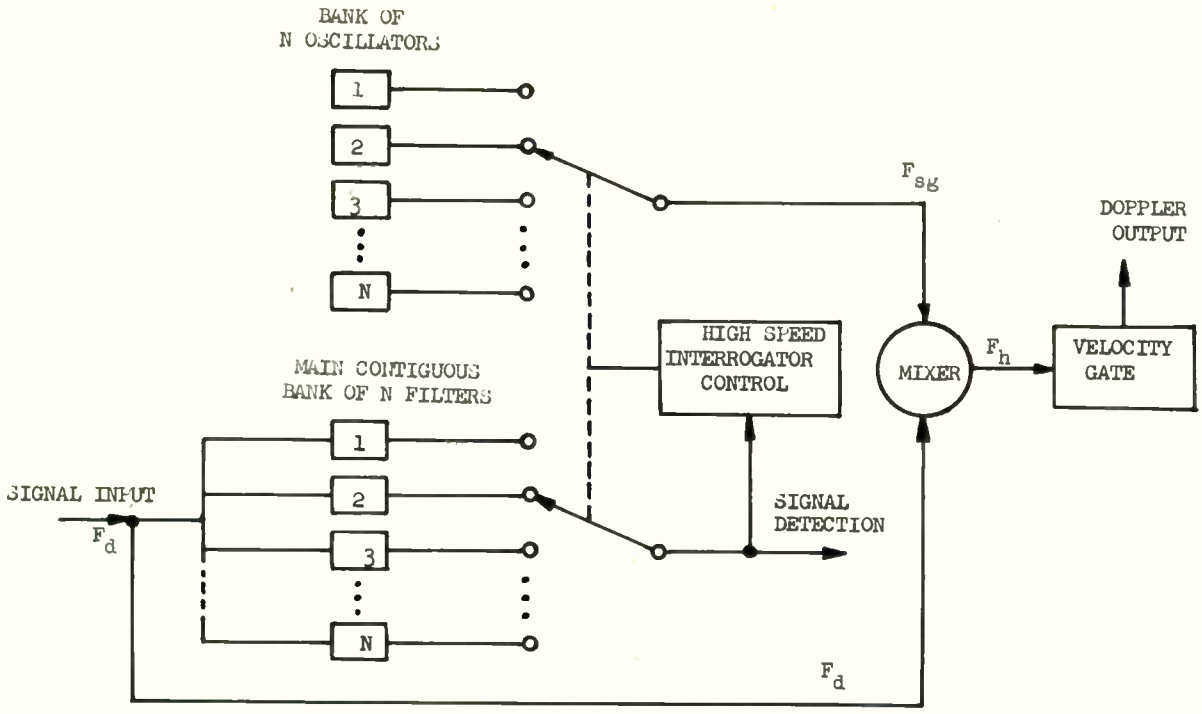
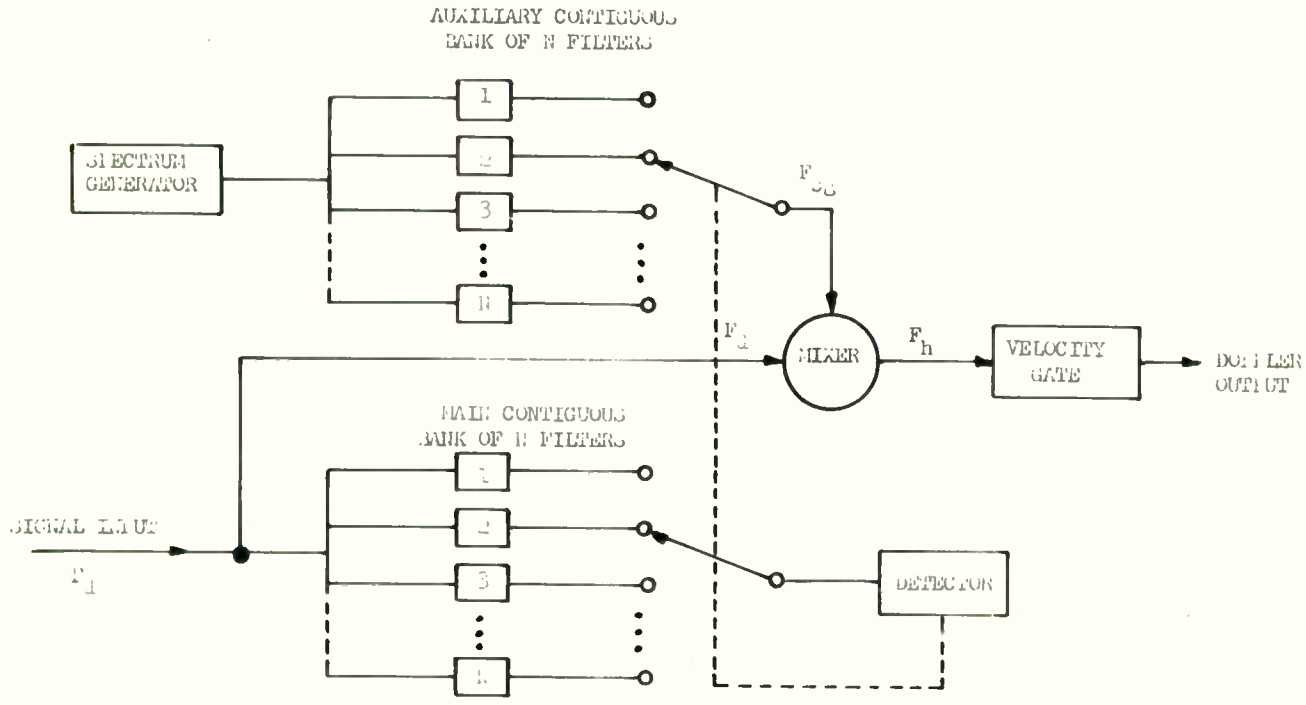


Fig. 1. Velocity vs. Doppler frequency.



(a)

MODIFIED DOPPLER ACQUISITION



(b)

Fig. 2. Doppler acquisition and modified Doppler acquisition.

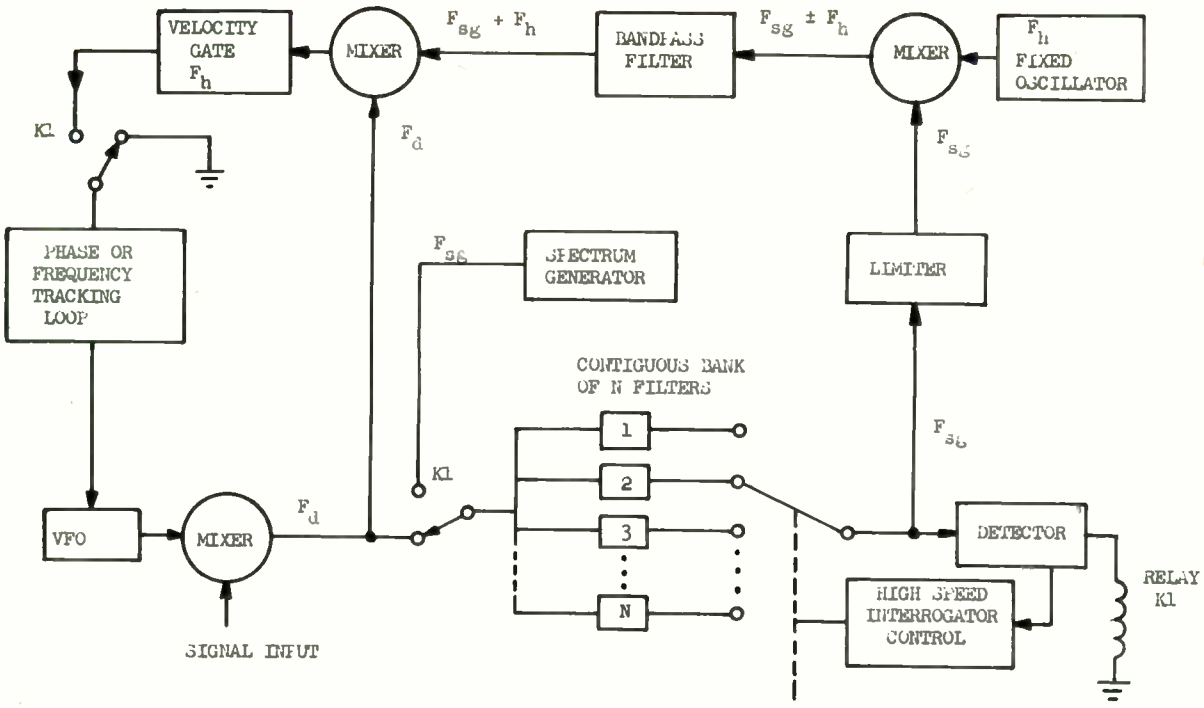
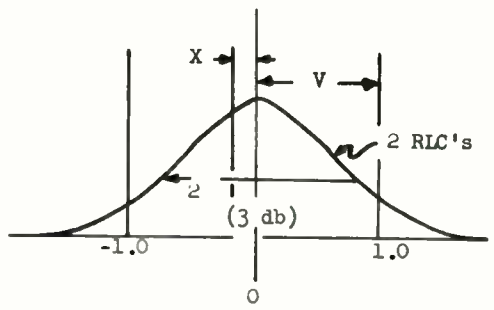


Fig. 3. Block diagram of rapid Doppler frequency acquisition system.

$$P_{sb} \quad \frac{v^4 (4x^2 + x^4)}{(1 + v^2)^3}$$

$v = 1.3\ V$

$x = 1.3\ X$



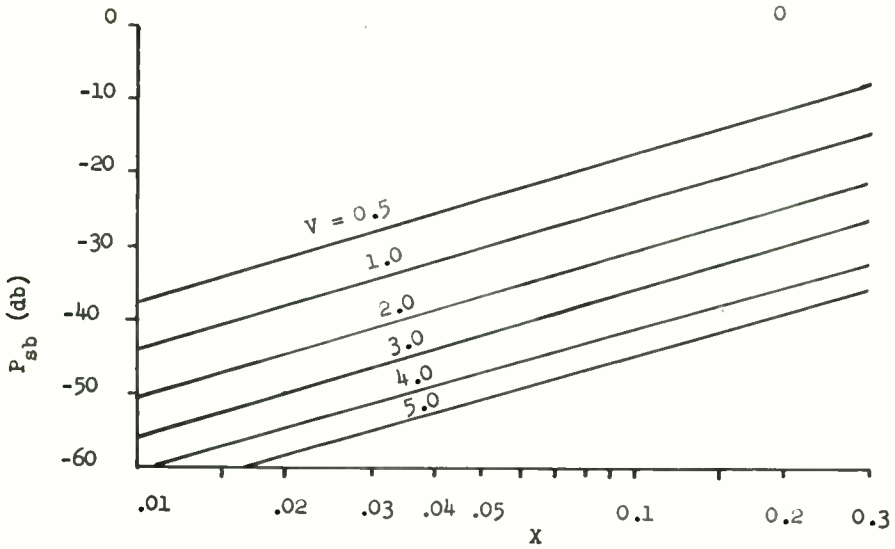


Fig. 4. Sideband power vs. spectral error, P_{sb} vs. X.

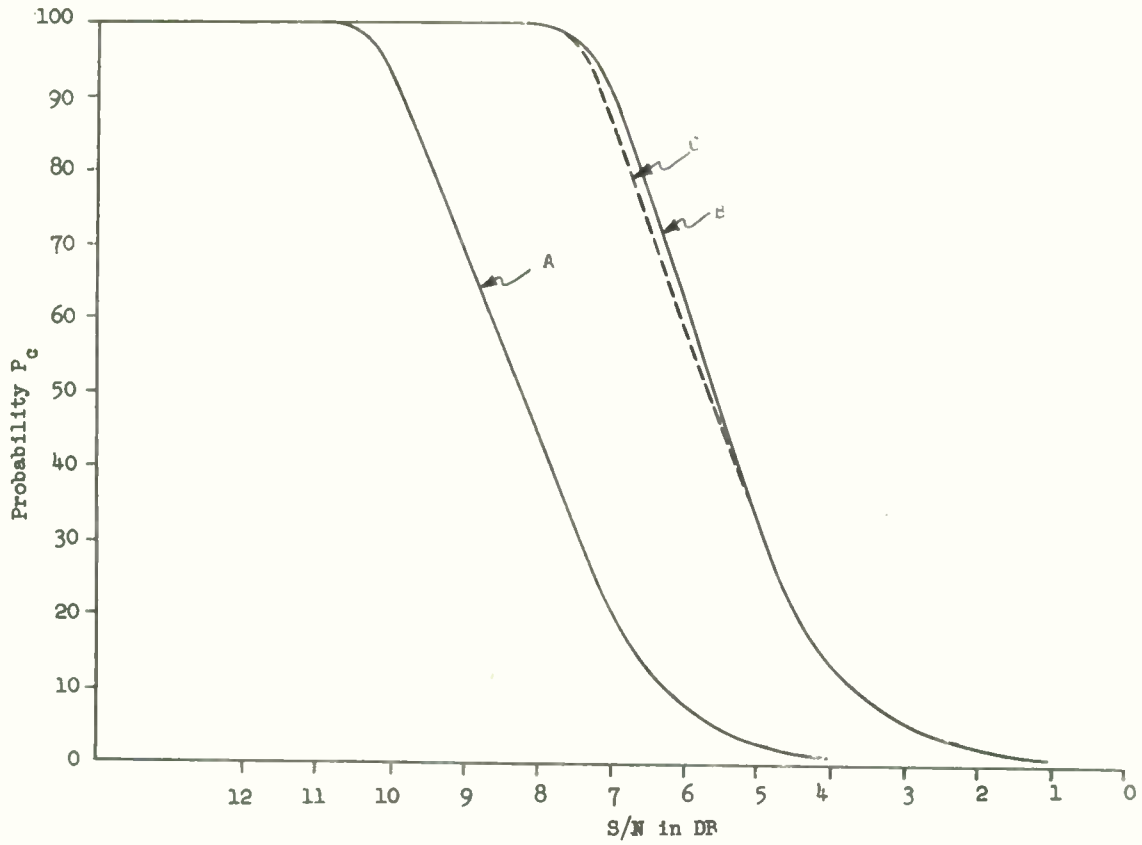


Fig. 5. Cumulative probability vs. signal-to-noise ratio, $(P_c$ vs. S/N$)$.

OPTIMAL FILTERING FOR RADAR DOPPLER NAVIGATORS*

James A. Mullen
Research Division
Raytheon Company
Waltham 54, Mass.

Abstract

The optimal filter to estimate the center frequency of a gaussian signal in a gaussian noise background is found under conditions appropriate for Doppler navigator application. The optimum estimation error is given for a class of signal spectra of differing rates of fall-off on the skirts. The optimum is about 3 times better (on a fluctuation error comparison) than currently available discriminators.

Introduction

The stringency required of Doppler navigation system parameters for advanced aerospace missions dictates that the ultimate limitations on performance need be determined. While advancing the state-of-the-art has been satisfactory in the past, the time has come to consider the maximum extent to which improvement in performance may be expected from improvements in filtering and tracking equipment. It is realized that there will always be ineradicable errors in the system output due to the noise-like character of the input signal and the constraints on averaging time.

This analysis considers the optimum filtering to be performed on the Doppler return spectrum as it enters the receiver in order to minimize the mean square error of estimating the Doppler shift.

The problem of filtering a random signal to obtain an estimate of the center frequency of the input has been previously considered by P. Swerling.[1] Since he has provided the mathematical solution, we here emphasize the physical meaning of the minimization analysis, and discuss and illustrate the performance of the prescribed discriminator.

*This work was supported in part by the Navigation and Guidance Laboratory, ASD-ARDC, Wright-Patterson AFB, Ohio, under contract AF33(616)7388.

Principles Involved in the Optimization

In actual practice, the Doppler shift is an order of magnitude larger than the spectrum bandwidth, but the "frequency tracker" (which performs the estimation operation) is designed as a closed-loop servo. The output, when the loop is closed, is very nearly at the proper frequency, so that the error signal (which is to be found by the estimating filter) is very small. The optimum filter for the tracking problem can then safely assume the frequency shift to be small compared to the width of the spectrum.

Since it is desired to find the optimum mean square filter for the spectrum, the well-established theory of optimum filtering could be applied were it not for the fact that the spectrum is not available. The spectrum is an average property of the input, and its value from the random waveform which is available must be estimated.

The estimation of the spectrum is obtained by the methods of Blackman and Tukey,[2] which consists of finding the correlation function of the input wave, and then taking the locally averaged Fourier transform. The advantage of the method is that one can fairly readily evaluate the estimate of error in the spectrum (on which the error of the estimate of center frequency depends) in terms of various relevant characteristic times. In the case of a moderately long averaging time (which is the practical one for navigator use) an equivalent implementation could be made by filtering the upper and lower frequency bands of the input waveform, then squaring them separately, and subtracting them. This corresponds more closely to an ordinary discriminator or, particularly, to the "single line discriminator",[3, 4] both of which have been previously used in Doppler navigation radars.

Discussion of the Optimum

The analysis on optimum filtering shows that the proper discriminator power characteristic is $D(f) = \text{const.}\ P'(f)/P^2(f)$ where $P(f)$ is the power spectrum of the signal plus the background noise.

Several points about this deserve mention. In the first place, the phase of the discriminator characteristic is unspecified by the optimization process. This presumably reflects lack of concern for the waveform and concentration on centering the power spectrum.

The optimum discriminator will have a power characteristic linear with frequency (which is different from discriminators usual in other applications whose voltage characteristics are linear with frequency). The fall-off of the characteristic should follow the derivative of the signal spectrum on the reasonable assumption that the background noise is white.

As the noise background becomes weaker, the weighting which the discriminator gives to the tails of the spectrum increases. It can often occur in mathematical models that in the absence of background noise perfect determination of the center frequency is possible. In particular, for gaussian or $(\sin x/x)^2$ spectra, which are common models, this singular case occurs. The existence of singular cases of precisely this kind is a familiar feature of the general theory of estimation and detection of random signals.[5,6,7] A convenient family of curves, which always avoids the singular case and yet allows differing rates of cutoff, is the Butterworth shapes, $P_{sig}(f) = B^{2p}/(f^{2p} + B^{2p})$.

The shape of the signal spectrum is due almost entirely to the transmitting and receiving antenna patterns, and therefore it is very natural to use gaussian, six x/x shapes, etc. These fail to satisfy the nonsingularity requirement, which can usually be recast into a Paley-Wiener[8] requirement. In this respect it is plain that the gaussian shape is unsatisfactory because the tails of the spectrum decrease too rapidly to be physical, and the $\sin^2 x/x^2$ shape is unsatisfactory if we take the frequency over the range $(-\infty$ to $\infty)$, although over a finite range (which is more natural if the angle of arrival were taken as the independent variable) no singular case occurs. It is not proposed that actual antennas have Butterworth patterns, but that the Butterworth family has the proper general features of the real patterns, namely a bandwidth and a tolerable, i.e. physical, rate of fall-off.

Results for Butterworth Signal Spectra

The results of optimal filtering of Butterworth signal spectra in white background noise is shown graphically. It is apparent that the error diminishes monotonically toward a constant value as the signal-to-noise ratio increases, and that higher fall-offs permit more accurate estimates. For large signal-to-noise ratios, the residual error is associated with the signal spectrum alone, and only a relatively small difference occurs between different drop-offs. For small signal-to-noise ratios, on the other hand, a more rapid drop-off enables a stronger weighting against the noise background, and a big improvement occurs as the roll-off increases.

The normalized error remains finite as signal strength increases, tending to an asymptotic value [given by Eq. (16) of the Appendix] which is less as the drop-off increases. The performance of the optimum discriminator appears to be nearly 10 db better than those available in practice, and about 10 db better than the performance of nonoptimal discriminators previously analyzed,[9*] and matches the intuitive analysis of Berger.[11]

The Physical Realizability Condition

In the derivation of the optimal discriminator, we have not imposed the requirement of physical realizability, nor have we provided error bounds for

*The figure of merit of Schultheiss et al.[9] for several nonoptimal discriminators is $.008 - .009\ f_o^2/BT$, where f_o is the Doppler frequency, and B and T are the same as in this paper. If the Doppler spectrum is 10 percent wide which is the usual aircraft case, the optimum is:

$$(10B)^2/200\pi BT \sim .0015(10B)^2/BT.$$

the degree to which a slightly asymmetric spectrum or a window function of non-negligible width will degrade the optimum. The physical realizability condition and the window function condition we regard as effectively satisfied through the incorporation of the frequency tracker unit in a navigation system, so that very long averaging times are available and a substantial lag is permissible. As is well known,[12] the physical realizability condition is only restrictive when a small lag is required. The difficulties of providing an up-to-date velocity estimate are more likely to lie in acceleration errors than in the realizability requirement. The changes caused by an asymmetric spectrum deserve consideration at some future time.

Example of a Fixed Design

For the present, we note that the optimal discriminator involves continual adjustment of the characteristic to suit the signal bandwidth and the signal-to-noise ratio. These are awkward to build into a system, and therefore it is worthwhile to see whether the performance of the system is substantially degraded by choice of a fixed design. Here we treat only one example of a signal-to-noise ratio adjustment and suppose the discriminator always matched properly for the signal bandwidth.

The performance of three fixed designs optimum for a particular value of signal-to-noise ratio with a signal spectrum of single-tuned resonant circuit shape is presented in Figure 2.* One can see that there is very little degradation except at such small signal-to-noise ratios that the performance of even the optimum is intolerably poor. It should perhaps be mentioned that the suboptima are actually worse than the optima for large S/N, and that the curve for 0 db is above that for 5 db which is above that for 10 db which is above the optimum. All four curves lie so close together that the figure cannot show this.

*The general insensitivity to signal-to-noise ratio confirms the analysis of nonoptimal systems (Figure 11 of Reference 9).

Concluding Remarks

During the course of this investigation, the optimum filtering analysis to find the center frequency of a signal noise band in a white noise background has been applied to the frequency tracker portion of a Doppler navigator system. It is found that:

a. The currently used discriminators do the proper functional processing to be optimum.

b. The frequency characteristic of the discriminator must be adjusted to match to the signal spectrum and noise level for the optimum according to the prescription of Eq. (7) of the Appendix.

c. The optimum discriminator has an rms fluctuation error about $1/\sqrt{10}$ times that of currently available discriminators.

d. Although the discriminator characteristic should in principle be continuously varied to match the signal-to-noise ratio, in actuality the degradation in performance from the use of a fixed design appears to be so slight that a fixed design for the weakest signal intended to be used is preferable.

Acknowledgment

It is a pleasure to acknowledge the valuable discussions with J. C. Rand, G. Marcus, and J. C. Hanson of the Missile and Space Division, Raytheon Company, which took place during the course of this investigation.

Appendix

Detailed Optimal Filtering Analysis

Details of Optimization

The mathematical formulation of the optimum filtering problem follows. Given an input waveform belonging to a random process whose power spectrum is $P(f - f_0)$; i. e. centered at the unknown error signal of the Doppler frequency, one can filter the spectrum by a discriminator with frequency weighting $D(f)$, and take the value of the output as a measure of the shift. For convenience, the value for zero Doppler shift is subtracted,

and divided by the incremental discriminator gain at zero Doppler shift, resulting in

$$\hat{f}_e = -\frac{\int D(f)\,\hat{P}(f - f_e)\,df - \int D(f)\,P(f)\,df}{\int D(f)\,P'(f)\,df} \qquad (1)$$

where the integrations extend over all frequencies, and the primes indicate derivatives.

The caret indicates an estimate. The left-hand side is an estimate of the error signal and depends on the estimate of the shifted power spectrum.

The mean of the error signal estimate is

$$\langle \hat{f}_e \rangle = -\frac{\int D(f)\,P(f - f_e)\,df - \int D(f)\,P(f)\,df}{\int D(f)\,P'(f)\,df} \qquad (2)$$

where the pointed brackets stand for averaging, and the average of the estimated spectrum is the true spectrum.

Since the error signal shift is small, the power spectrum can be expanded in a Taylor series, giving

$$\langle \hat{f}_e \rangle = f_e - \left(\tfrac{1}{2} f_e^2\right)\left(\frac{\int D(f)\,P''(f)\,df}{\int D(f)\,P'(f)\,df}\right) + \left(\tfrac{1}{6} f_e^3\right)\left(\frac{\int D(f)\,P'''(f)\,df}{\int D(f)\,P'(f)\,df}\right) + \cdots \qquad (3)$$

In order to make the estimate unbiased in the sense of statistics (i.e., the average value of the estimate is the true value), it is plain that one should require, as additional constraints on the problem, that several of the higher order terms of the Taylor series should vanish, until the remainder term is so small that it may be neglected.

The mean square error (E^2) is then

$$E^2 = \left\langle (\hat{f}_e - \langle \hat{f}_e \rangle)^2 \right\rangle$$

$$E^2 = \frac{\iint D(f)\,D(g)\left[\langle \hat{P}(f - f_e)\,\hat{P}(g - f_e)\rangle - \langle \hat{P}(f - f_e)\rangle\,\langle \hat{P}(g - f_e)\rangle\right] df\,dg}{\left(\int D(f)\,P'(f)\,df\right)^2} \qquad (4)$$

The average under the double integral represents the cross variance of the spectral estimate.

In general, estimating the spectrum requires the use of extra smoothing filters, called window functions, as discussed fully in Blackman and Tukey. The necessary frequency separation for two estimates to be independent is the bandwidth of the window function. This bandwidth is about an order of magnitude larger than the reciprocal of the averaging time. If the averaging time is long enough, e.g., two or three orders of magnitude larger than the reciprocal of the spectrum bandwidth, the finite resolution which the window function imposes can be ignored, and the limiting case assumed. In the frequency tracker case, the loop bandwidth corresponds to the reciprocal of the averaging time, and is sufficiently large relative to the spectral bandwidth that the limiting case applies.

With finite resolution, the elementary bands are B_T cycles wide, corresponding to the window function. The estimate has a chi-square distribution so that its variance is proportional to the square of the estimate divided by the number of degrees of freedom, $B_T T$ (T is averaging time). In the limiting case, the interval estimate becomes a point estimate, and a δ function, whose normalization is found by requiring that the integration of the point estimates over an elementary band should equal the interval estimate, appears. The presence of B_T is cancelled out, and the cross variance of spectral estimates becomes

$$\langle \hat{P}(f)\hat{P}(g)\rangle - \langle \hat{P}(f)\rangle\,\langle \hat{P}(g)\rangle = \frac{1}{T} P^2(f)\,\delta(f - g) \qquad (5)$$

The δ function enables one integration to be performed immediately. In addition, remembering the constraints on the Taylor series, the quantity to be minimized is found to be

$$E^2 = \frac{\int D^2(f)\,P^2(f)\,df}{T\left(\int D(f)\,P'(f)\,df\right)^2} \qquad (6)$$

under the constraints

$$\int D(f)\,P''(f)\,df = 0, \quad \int D(f)\,P'''(f)\,df = 0, \text{ etc.}$$

For symmetric spectra, which (apart from second order effects) is the usual case, the first constraint is always satisfied. Since the shift is small, a good approximation is forthcoming without the constraints.

Then the Euler-Lagrange equation gives immediately

$$D(f) = -\frac{\lambda}{2}\frac{P'(f)}{P^2(f)} \tag{7}$$

and the error in the optimum case is

$$E^2_{opt} = \frac{1}{T\int \frac{P'(f)^2}{P(f)^2}\,df} \quad . \tag{8}$$

Explicit Results for Butterworth Spectra

For a Butterworth signal shape and a white noise background,

$$P(f) = S\frac{B^{2p}}{f^{2p}+B^{2p}} + N = N\frac{f^{2p}+C^{2p}}{f^{2p}+B^{2p}}, \tag{9}$$

where $C^{2p} = (1+\frac{S}{N})B^{2p}$.

The optimum discriminator characteristic is

$$D(f) = \frac{\lambda}{2}\frac{S}{N^2}\frac{2pB^{2p}f^{2p-1}}{(f^{2p}+C^{2p})^2} \quad . \tag{10}$$

The estimation error becomes, after normalizing the variable of integration,

$$E^2_{opt} = \frac{B}{4p^2\frac{S^2}{N^2}T}\left[\int \frac{x^{4p-2}\,dx}{(x^{2p}+1+\frac{S}{N})^2\,(x^{2p}+1)^2}\right]^{-1} \tag{11}$$

The integral is related to

$$I(\alpha,\beta) = \int \frac{x^{2p}}{(x^{2p}+\alpha^{2p})(x^{2p}+\beta^{2p})}\,x^{2p-2}\,dx, \tag{12}$$

which can be reduced to a tabulated integral[10] (or evaluated by residues) by the use of partial fractions, giving

$$I(\alpha,\beta) = \frac{\pi}{p\sin\frac{\pi}{2p}}\frac{\alpha^{2p-1}-\beta^{2p-1}}{\alpha^{2p}-\beta^{2p}} \quad . \tag{13}$$

The mixed second order partial derivative of I is proportional to the integral required. Substituting

$$\alpha = 1 + \frac{S}{N}, \quad \beta = 1,$$

we get

$$E^2_{opt} = \frac{B\sin\frac{\pi}{2p}}{4\pi pT}\left[\left(1-\frac{1}{2p}\right)\left(1+\left[1+\frac{S}{N}\right]^{-\frac{1}{2p}}\right) + 2\frac{N}{S}\left(1-\left[1+\frac{S}{N}\right]^{1-\frac{1}{2p}}\right)\right]^{-1}, \tag{14}$$

which is the result plotted in the figure. For large signal-to-noise ratios:

$$E^2_{opt} = \left(\frac{B}{2\pi T}\right)\frac{\sin\frac{\pi}{2p}}{2p-1+O\left[(S/N)^{-\frac{1}{2p}}\right]} \tag{15}$$

and for small signal-to-noise ratios:

$$E^2_{opt} = \left(\frac{B}{2\pi T}\right)\frac{6\sin\frac{\pi}{2p}}{\left(1-\frac{1}{4p^2}\right)\left(\frac{S}{N}\right)^2 + O\left[\left(\frac{S}{N}\right)^3\right]} \quad . \tag{16}$$

Formulation of the Sub-Optimum Problem

If the discriminator is not optimum, then the estimation error is given by:

$$E^2 = \frac{\int D^2(f)\, P^2(f)\, df}{T\left[\int D(f)\, P'(f)\, df\right]^2} \quad . \tag{17}$$

Here one proposes to evaluate the performance of a discriminator which is optimum for a particular value of bandwidth and signal-to-noise ratio, when the actual spectrum is different, but still has the proper degree of roll-off. The discriminator characteristic is given by the same formula, Eq. (7), except that the spectrum which appears is the design spectrum, $P_d(f)$, which is given by

$$P_d(f) = N \frac{f^{2p} + C_d{}^{2p}}{f^{2p} + B_d{}^{2p}} \quad . \tag{18}$$

Introducing normalized variables leads us to:

$$\frac{J_{22}(\alpha, \beta_d; \beta)}{I_{22}^2(\alpha, \beta_d)} \tag{19}$$

where I_{mn} is like $I(\alpha, \beta)$ of Eq. (12) except for having the denominator factors raised to the m^{th} and n^{th} powers respectively. J_{mn} is like I_{mn} except for having an extra factor $(x^{2p} + \beta^{2p})^2 / (x^{2p} + \beta_d{}^{2p})^2$ in the integrand. The numerator can be expressed as the sum of three I_{mn}'s by expressing $[x^{2p} + \beta^{2p}]^2$ as $[x^{2p} + \beta_d{}^{2p} + (\beta^{2p} - \beta_d{}^{2p})]^2$, and expanding by the binomial theorem, giving

$$J_{22}(\alpha, \beta_d; \beta) = I_{22}(\alpha, \beta_d) + 2(\beta^{2p} - \beta_d{}^{2p})\, I_{23}(\alpha, \beta_d) + (\beta^{2p} - \beta_d{}^{2p})\, I_{24}(\alpha, \beta_d) \tag{20}$$

The integrals can again be found although now third and fourth order derivatives are needed.

After much algebraic manipulation, the answer is found as:

$$E^2 = \frac{B}{4\pi p\left(\frac{S^2}{N^2}\right) T} \times \left[\frac{\sin\frac{\pi}{2p}\left(1 - \frac{C_D}{B}\right)^2}{\left(1 - \frac{1}{2p}\right)} \Big/ \left(1 - \frac{1}{2p}\right)\left(1 + \left(\frac{B}{C_D}\right)^{\frac{1}{2p}}\right) - 2\,\frac{1 - \left(\frac{C_D}{B}\right)}{1 - \frac{C_D}{B}} \right]$$

$$\times \left[1 - \left(\frac{C}{C_D} - 1\right)\left\{ \frac{3}{1 - \frac{C_D}{B}} + \left(1 - \frac{1}{2p}\right)\frac{K}{L} \right\} + \left(\frac{C}{C_D} - 1\right)^2 \left\{ \frac{2}{\left(1 - \frac{C_D}{B}\right)^2} + \frac{2p - 1}{2p\left(1 - \frac{C_D}{B}\right)}\,\frac{K}{L} + \frac{2p - 1}{24p^2\left(\frac{C_D}{B}\right)^{\left(1 + \frac{1}{2p}\right)}}\,\frac{2p - 3 + (2p - 1)\frac{B}{C_D}}{L} \right\} \right] \tag{21}$$

where

$$K = 2p - (2p - 3)\left(\frac{B}{C_D}\right)^{\frac{1}{2p}} - \left(\frac{B}{C_D}\right)^{1 + \frac{1}{2p}},$$

$$L = (2p - 1)\left[\left(\frac{B}{C_D}\right)^{\frac{1}{2p}} - 1\right]\left(1 + \frac{C_D}{B}\right) + 2\left[1 - \frac{B}{C_D}^{1 - \frac{1}{2p}}\right]$$

This equation is the one which has been used, for p = 1, to obtain the curves for the suboptimum design fixed with respect to signal-to-noise ratio.

References

1. P. B. Swerling, "The estimation of Doppler shifts in noise spectra," I. R. E. International Convention Record 1960, Part 4.

2. R. B. Blackman and J. W. Tukey, "The Measurement of Power Spectra," Dover Publications, Inc., New York, N. Y., 1958, also in Bell Sys. Tech. J. 37, 1958.

3. T. B. Pickard, "A new method of frequency measurement," Trans. I. R. E. Information Theory 4, pp. 83-88, 1958.

4. J. A. Mullen, "The discriminator rule of the frequency tracker," [which depends on some early (1954) work of J. A. Sheehan, now of M. I. T., Lincoln Laboratory, Lexington, Mass.] Memorandum No. M-288, Government Equipment Division, Raytheon Company, March 9, 1959.

5. D. Slepian, "Some comments on the detection of gaussian signals in noise," Trans. I. R. E. Information Theory 4, pp. 65-68, 1958.

6. D. Middleton, "On singular and non-singular optimum (Bayes) tests for the detection of normal stochastic signals in normal noise," Trans. I. R. E. Information Theory 7, April 1961.

7. I. Good, "Effective sample rates for signal detection or can the gaussian model be salvaged," Information and Control 3, pp. 116-140, 1960.

8. R. E. A. C. Paley and N. Wiener, "Fourier Transforms in the Complex Domain," Am. Math. Soc. Chap. I, Theorem 411, 1934.

9. P. M. Schultheiss, C. A. Wogrin, F. Zweig, "Short-time frequency measurement of narrow-band random signals in the presence of wide-band noise," J. Appl. Phys. 25, pp. 1025-1036, 1954.

10. W. Gröbner and N. Hofreiter, "Integral tafeln, II. bestimmte integrale," Springer-Wien, No. 141. 14, 1950.

11. F. B. Berger, "The nature of Doppler velocity measurement," Trans. I. R. E. Aero Nav. Elec. 4, pp. 103-112, Eq. (30), September 1957.

12. N. Wiener, "Stationary Time Series," John Wiley and Sons, New York, N. Y. sec. 3.4, 1949.

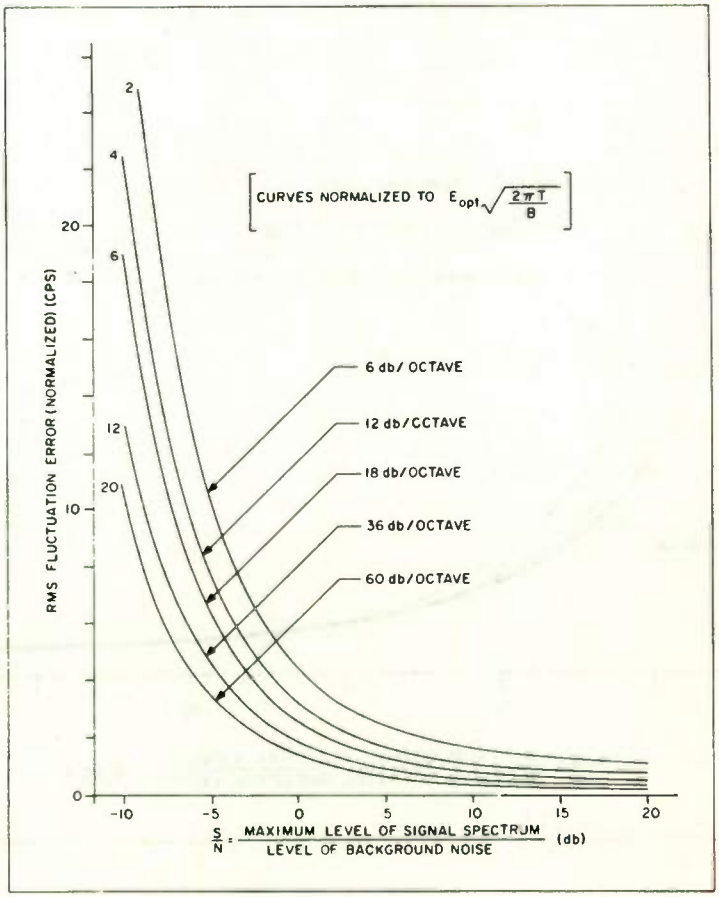


Fig. 1. Minimum estimation error for Butterworth spectal shapes.

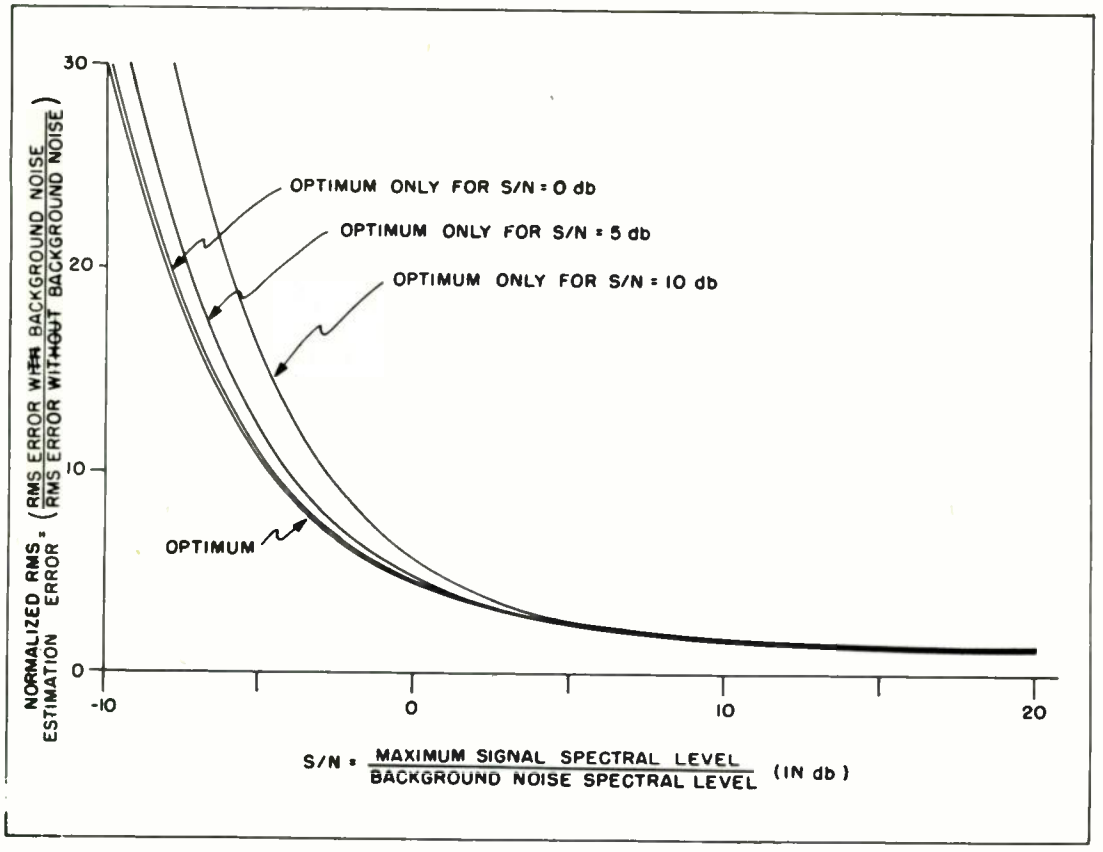


Fig. 2. Fixed S/N designs evaluated with respect to the optimum (6 db) octave Butterworth signal spectrum.

THE FREE WORLD'S LARGEST RADAR

W. L. Kimbell
GE Company
Syracuse, N. Y.

Abstract

This paper describes the detection radar for the Ballistic Missile Early Warning System. This radar uses huge torus reflectors with organ pipe scanners to scan in a series of 40° azimuth sectors with four beams at two elevation angles in each sector. These beams have half power beamwidths of 1° and are separated vertically by an angle of 3.5° and are separated in azimuth by 20°.

Independent receivers are used on each beam with the radar video output feeding both automatic data processing equipment and oscilloscope displays. Waveguide, coax and crossbar switches allow switching in standby units with semiautomatic control to provide rapid switching. The combination of the physical size of the RF components of this low frequency radar and the large number of equipments in the system leads to many unique problems.

PHASED ARRAY RADARS FOR SATELLITE TRACKING

C. S. Lerch, Jr.
Bendix Radio Division
The Bendix Corporation
Baltimore 4, Md.

SUMMARY

This paper describes the features of phased array antennas which make them particularly suitable for long range, high density radar applications such as satellite tracking. A hypothetical system is discussed which can detect and track orbiting space vehicles. An experimental phased array radar (ESAR) is described.

I. INTRODUCTION

The use of a different technique to do a job cannot be justified on mere novelty alone. It must be justified by some other reason, such as its ability to perform better or more economically, or it must accomplish something not possible with the older techniques. Since the application of phased array techniques to radar is relatively new, the first part of this paper presents such a justification by describing the theory of a phased array radar, how it works, and its advantages. This is followed by a description of a hypothetical phased array which could perform the satellite tracking mission. After this, the experimental work which backs up the theory is described, including some details on a large scale phased array model which has been in operation for more than a year.

In a brief paper such as this it is not possible to develop the theory in detail but adequate references are available, some of which are listed in the bibliography.

II. PHASED ARRAY DEFINITION

A phased array is a set of two or more antennas radiating or receiving electromagnetic energy with a known and controllable relationship between the time phase of the energy in each element. It is this coherence which gives the phased array some of its most valuable properties. Figure 1 illustrates the most elementary array, which consists of two radiators fed from a common signal source and a phase or delay control on each element. As can be seen, the signals will add vectorially in space, and the direction of the maximum will be controlled by the time delay or phase of the signals on each element. The shape of the beam is controlled by the number and spacing of the elements and the relative amplitude and phase of the currents.[1,2,3] The strength of the signal can be increased by putting an amplifier ahead of each antenna, as shown by the dotted box in Figure 1.

This paper is limited to a class of radar arrays which have undergone considerable development in recent years and which are characterized by the properties discussed in the following paragraphs.

Multi-element Planar Array

The multi-element planar array consists of many thousands of individual antenna elements located on a flat metallic ground plane with about one-half wavelength spacings between adjacent elements. Although it is also possible to locate the antenna elements of a phased array on other types of surfaces, such as spheres, cones, or cylinders, and while such arrangements are of interest, they will not be considered further in this paper.

Individual Transmitters and Receivers

By individual transmitters and receivers, we mean that a transmitter and a receiver are connected to a fairly large percentage of the elements, if not to all. This use of many small transmitters makes possible radiated powers which are thousands of times greater than those attainable from conventional antenna-transmitter systems.

Electronic Phase Control

By electronic phase control, we mean that the phase of the signal in each channel is controlled electronically. This electronic control of phase permits changing beam position in microseconds as compared to seconds with mechanically scanned antennas.

III. PHASED ARRAY FEATURES

Some of the features of phased arrays which make them particularly suitable for the space detection and tracking problem are obvious from the properties discussed above; however, some other features which may not be immediately apparent will now be described.

Reliability

With many thousands of units operating in parallel, the failure of a few channels will not seriously detract from the radar's performance. In addition, the tolerances needed on the phase and amplitude control of the signals in each channel become looser as the number of active elements is increased. With adequate monitoring circuits, defective parts can be located and replaced without turning the set off. Hence, if the channel failure rate is kept reasonably low, the set can operate continuously as long as prime power is supplied. This assumes that those few units which are common to all channels are multiplexed. An analysis of the effects of out-of-tolerance and failed parts on performance has been published[3,4].

Utilization of Beam Mobility

The beam mobility afforded by the fast electronic scan feature allows the design of a single radar set which can perform a number of functions. With a given set of installed components, the radar set has the ability to transmit a number of pulses per unit of time depending on the peak and average power limitations of the tubes. These pulses can be utilized in a variety of ways as illustrated by Figure 2. They can all be transmitted in a sequence of directions for establishing a search pattern, or they can be transmitted in specific directions for tracking targets previously detected, or they can be used in any combination of search and track desired. The way in which the radar can track a number of targets at once is shown in Figure 3. This is

accomplished by transmitting a group of pulses on each pulse repetition interval which might be spaced 20 or 25 microseconds apart. The group of pulses can be transmitted all in one direction for long-range targets, or in different directions for shorter range targets. They would be transmitted in the direction of targets previously acquired in a search mode.

The receiving operation can be understood by again referring to Figure 3. Immediately after the last pulse is transmitted, the received beam is positioned in the direction of the earliest expected return (target 1). After this return is received, the beam is moved to the position of the next return and remains there until number 2 is received. This action is continued until all target returns are received.

The advantage here is evident; one phased array radar can perform the function of a number of trackers. Further, it may not be necessary to track every target with a data interval equal to the pulse repetition interval, hence, each pulse can be time-shared between 1, 2, 5, 10, or more targets.

In the foregoing example the beam mobility gives a phased array a time efficiency which a mechanical radar cannot have due to its inertia.

The speed and operating complexity of an array are such that human operators or mechanical controls are inefficient and computers would be used for radar control. This gives an extra degree of flexibility since a change of operating modes can be accomplished merely by changing programs, and no mechanical or circuit changes are required. In fact, by providing sophisticated control circuitry, we can have a truly "adaptive" radar whose operating modes change automatically as the tactical situation changes!

Multiple Beam Forming Ability

By a technique known as post amplifier beam forming, any number of simultaneous receiving beams, each using the full aperture and each pointing in a different direction, can be formed. These beams can be steered as a cluster by the phasing control. A simplified explanation of this operation for the case of two antenna elements is shown in Figure 4. Note that each receiver feeds a delay line that is tapped as shown. Signals arriving in beam 2 add up in phase only on the number 2 output line. A signal arriving in beam 1 reaches the top antenna a short time before it reaches the lower one, and this delay is exactly compensated by the delay line for the beam 1 output.

The post amplifier beam-forming technique extends the utility of phase arrays—the most obvious is its use for monopulse tracking. Historically, the first use of this type of operation was in the MUSA System in the early 1930's[5].

One can also have a system in which receiving beams are formed looking in all directions at once and which does not need to be steered. Only those beam outputs would be used which were pointing in directions that power had been transmitted. A linear receiver array using this principle has been built for an air traffic control application and is now being tested[6].

Coverage

While it is generally considered necessary for a radar set to have hemispherical coverage, there are many applications where this may not be a necessity. Planar phased arrays by their nature are generally constrained to cover a solid angle of about 90 by 90 degrees.

The coverage of a phased array is determined by the antenna element spacing and the antenna element pattern. The limitation imposed by the spacing factor is due to the formation of grating lobes. These grating lobes are spurious beams that appear as the array is scanned away from boresight and are due to the periodic nature of the radiated energy. Consider a linear array which is fed with signals whose phase shift between elements is 180 degrees ($\lambda/2$). These signals will add up in two directions in space to form beams pointing in either direction along the axis of the array. References 1, 2, and 3 listed in the bibliography cover this topic in detail.

For the spacings generally employed in planar arrays (about 0.57λ), it is possible to scan about 45 degrees along the principal diagonals and much further along any of the others. However, as the spacing is reduced towards 0.5λ, it becomes possible to scan more than 80 degrees before grating lobes appear along any diagonal when relatively narrow beams are used, although the beam broadens considerably, but predictably, at large scan angles.

The second and more important factor is the coverage which can be obtained from one antenna element when imbedded in an array; since the gain of the array at any pointing angle is the product of the array gain and the effective element gain at the angle of interest. The effective element gain, or coverage, is far different from that measured on an isolated element due to the effects of mutual coupling. One of the more difficult array problems has been the effects of mutual impedance on antenna pattern, but they are now well understood and can be controlled.

This angular coverage can be placed wherever desired by appropriately tilting the plane of the array. To cover a full hemisphere will, in general, require three or more faces. This feature, however, is not the severe limitation that one would expect at first sight, since a search radar uses a fixed amount of average power to cover a given volume of space to detect targets. This power requirement depends primarily on the range and the target size and not on the type of radar used. Hence, three phased arrays would need only one-third the power per array required by another type of radar which covered the whole area by itself. Studies have shown that the controlling factor in the cost (in high power long range radars) is the average power required and that type of radar (mechanical, phased array, etc.) has much less effect. In fact, there are some instances where it is less costly to use a phased array because of power economy. This is particularly true when very high powers are needed because the voltages required for single tube power generation are extremely high and they are difficult to generate and control.

Complexity and Economics

When first considered, the thought of the many thousands of transmitters and receivers which must be built, installed, and checked out is, to say the least, horrifying. Further reflection shows, however, that this is an asset instead of a liability. A phased array is essentially a number of identical and relatively simple radar sets operating in parallel. Consider the fact that each channel of a typical array has a size of a two-way mobile radio and complexity about that of a

television set. Television sets can be produced in quantities of several hundred thousand per year in a relatively small production facility and at a reasonable cost. In addition, a uniformity and reliability is obtained which is difficult to match in small quantity, job-lot operations. This basic simplicity and redundancy is the factor that makes phased array radars practical in many applications.

IV. BEAM STEERING METHODS

A large number of beam steering methods have been proposed for arrays and several have been investigated thoroughly. Consequently, the designer has a choice of methods with which to work. The best method for any given array depends on many factors and a detailed account of steering methods and their relative merits is beyond the scope of this paper. Instead, two typical schemes will be illustrated. More details can be found in the references listed in the bibliography[2,7,8].

Direct Phase Shift

Figure 1 illustrated a simple array which had a phase shifter in each channel. This phase shifter can be any device whose electrical phase shift can be varied by a control signal. Figure 5 is a simplified drawing of a device reported by Lincoln Laboratories which uses semiconductor diodes to control the effective length of a set of cables[7]. When a diode switch is off (open circuit) for a segment, the signal goes through the cable loop and is delayed. When the switch is on (short circuited), the signal bypasses that segment of cable and suffers no delay or phase shift. By providing a number of different length cables, a reasonable degree of phase control fineness can be provided. Analysis indicates that five to seven steps are sufficient for many applications.

Indirect Phase Shift

Because the phase of a signal is preserved in a mixing or heterodyne operation, there are ways of generating phase shifts outside of the signal channel and introducing them through a mixing operation such as is shown in Figure 6, where the phase shift signals are carried by the local oscillator signal of the receiver. A convenient way to generate the phase shift signals by means of a delay line is shown in Figure 7. The taps are equally spaced, and as the control frequency is varied, the phase delay varies at each tap[3]. By means of the second mixing process, the frequency variation is subtracted and the outputs from the delay taps are a set of constant frequency signals carrying the phase information. These can be used for the local oscillator signal shown in Figure 6.

V. SATELLITE TRACKING ARRAY

Having discussed some of the properties of a phased array, it is now possible to discuss a tentative application. The space detection and tracking problem requires many orders of magnitude more of performance than that needed for detecting and tracking aircraft targets. Detecting targets at 4000 miles rather than at 200 requires an increase in performance of 160,000 times if the targets are the same size. This is due to the fact that the detection range of a search radar is proportional to the fourth root of the product of the average power and receiving aperture. To get this increase, we can only enlarge the aperture by a relatively small factor (due to size limitations), and for any further increase in performance we must raise the transmitted power. This is where the phased array comes into its own because it can get this power increase easily. Figure 8 shows the kind of peak power which can be achieved at a frequency of 400 megacycles using tubes available today. One can readily see that the power increases needed are available.

Figure 9 shows how a space track radar could perform the dual functions of detecting and tracking. Using the time-shared principle described earlier, a planar array lying on its back would use a portion of the transmitted pulses to form a detection search fence. As satellites cross the fence and are detected, the other pulses would be used to track them. If they are old satellites, the information would be used to update catalogue files. If they are new, as could rapidly be determined, the tracking process would supply enough data to predict their orbit. Using two radar sets and a short range gap filler with their planes facing up, practically all objects crossing the United States could be detected and tracked; see Figure 10.

One problem that exists in satellite detection is due to the varying altitudes of the targets. The scan period must be short compared to the time it takes a satellite to cross the fence. Table 1 shows the approximate time it takes a satellite to cross a 1.5-degree beam as a function of the altitude of a vehicle crossing at right angles to the scan direction.

TABLE 1. TIME REQUIRED FOR SATELLITE TO CROSS 1.5-DEGREE BEAM FOR SEVERAL ALTITUDES

Altitude (NM)	Time in Fan (SEC)
100	0.65
200	1.3
500	3.7
1000	8.2
2000	17.0

If the period of scan is made small enough to get even one look at the 100-mile satellite as it crosses, the scan rate becomes completely impractical. However, this problem can be eliminated easily by making use of the fact that the return signal from a 200-mile target is 40 DB higher than that from a 2000-mile target. By providing extra receiving beams at right angles to the scan direction or fence, the transmitting beam sidelobes can be used to illuminate the low altitude targets. These are about 30 DB down, and there is more than enough energy for the purpose. Use of this technique allows the scan period to be set for the object at the highest altitude. A scan period or frame time of 5 seconds to cover 90 degrees will give a minimum of three looks per pass on the 2000-mile targets.

The equipment for a radar such as described could be housed in a square structure about 125 feet on a side. With an operating frequency of 400 megacycles, about 3200 transmitters and receivers would be needed. The average power from each transmitter would be only 30 watts. The antenna elements would be located on the roof of the structure in a circular aperture about 115 feet in diameter.

VI. EXPERIMENTAL RESULTS

During the past few years a number of groups, both in the United States and Great Britain, have accelerated the amount of work done in the large scale phased array field. Some of the earliest work in phased arrays was in the sonar field with acoustic arrays. The theory and results from these acoustic arrays carries over directly to the radar field. A complete and fair coverage of the various programs would be impossible, so this paper will describe the work done on the Electronically Steerable Array Radar (ESAR) project with which the author is most familiar.

The first item on the ESAR project was the construction and testing of a 90-element linear array which is shown in Figure 11. The radiating elements are located on the roof and the array forms a fan beam looking up, and it is steerable 45 degrees from vertical in either direction along the main axis of the building. The array is a complete operating radar, including transmitters, receivers, modulators, beam steering circuits, and associated controls. Figure 12 shows a set of patterns taken by flying an airplane along the major axis of the array and transmitting a signal to the array. After the plane flew through the nose of the beam at the far left, it was advanced to the next step and held there until that position was recorded and then moved again. By this means it was possible to record on one pass, a number of beam positions and demonstrate that the beam could be steered without and degradation in beam shape and without serious increase in sidelobe levels. The solid line in the figure represents the constant signal strength line, and it is not flat due to the changing distance from the aircraft to the array. The dotted line is the 3-DB down reference. The beam width changes as the inverse of the cosine of the steering angle as predicted by theory.

The next step in the ESAR program was the construction of a large scale planar array operating at L-band. A photograph of this radar is shown in Figure 13. It is a five-story structure with more than 8000 antenna elements located on the sloping face. The design of this array was started on November 1st, 1959, and it was constructed and placed in operation in just under a year. Its coverage sector is so placed that it can survey air traffic on the New York-Washington Airway and also track the missile firings from the NASA Station at Wallops Island, Virginia.

The face of the building is approximately 50 by 50 feet and the antenna elements are imbedded in a polyurethane foam for protection from the weather. Figure 14 shows one of the antennas which is called a logarithmically periodic antenna. Only 760 of the antenna elements are connected to transmitters and receivers for reasons of economy. The active elements are located in a fashion similar to that shown in Figure 15. This distribution gives the same effective main beam shape and width as would be obtained from a nearly full array at the expense of somewhat higher sidelobes and lower gain. It is, however, eminently suited for a development model since it has all the operating characteristics of a full power array except range.

A photograph of a scale model of the building is shown in Figure 16, which gives an idea of how the equipment is located in the building. The structure on the left houses the air-conditioning equipment which provides the constant temperature environment necessary to maintain phase control. Space is provided for maintenance and operating functions.

The platform on the front of the structure (Figure 13) can be raised to provide access to the face for installation and maintenance of the antenna elements.

Connected to each active antenna is a module, shown in Figure 17, which contains a transmitter, receiver, and duplexer. The steering control signals are provided by a delay line system similar to Figure 7. The outputs of the receivers are summed in a set of networks which also provide the multiple beam outputs. Computing equipment is provided to program the beam positions to make the radar search a volume and also place tracking beams on multiple targets through a time sharing arrangement such as described previously in the paragraph headed Utilization of Beam Mobility.

The phased array radar described above has been in operation for more than a year and a large amount of data has been collected which demonstrates that it is operating according to its design specifications and that phased array radars are practical.

VII. CONCLUSION

Phased array radars have been shown to have features which make them particularly suitable for long range, high traffic density applications such as satellite tracking. A practical design which is based on solid experimental evidence can be synthesized from components available today.

VIII ACKNOWLEDGEMENT

The work on ESAR has been performed for Project Defender, sponsored by the Advanced Research Projects Agency (ARPA), under a contract administered by the Rome Air Development Center (RADC). Important contributions to this effort have been made by all the personnel on the ESAR Project team at Bendix Radio and by the personnel at RADC and ARPA who have guided the work since its inception.

BIBLIOGRAPHY

1. "Microwave Antenna Theory and Design," Silver, S., Massachusetts Institute of Technology, Radio Laboratory Series, McGraw-Hill Book Co., Inc., New York, N.Y., Volume 12, 1949.
2. "Properties of Phased Arrays," von Aulock, W.H., Proceedings of the Institute of Radio Engineers, Volume 48, 1960, pp 1715-1727.
3. "Steerable Array Radars," Ogg, F., Institute of Radio Engineers, Transactions on Military Electronics, Volume MIL-5, 1961, pp 80-93.
4. "The Effect of Aperture Errors on the Antenna Radiation Pattern," Ruze, J., Supplement Nuovo Cimento (Italian), Volume 9, No. 3, 1952, pp 364-382.
5. "A Multiple Unit Steerable Antenna for Short-Wave Reception," Friis, H.T. and Feldman, C.B., Proceedings of the Institute of Radio Engineers, Volume 25, 1937, pp 841-917.
6. "New Techniques in Three-Dimensional Radar," Simpson, M., Institute of Radio Engineers Transactions on Military Electronics, Volume MIL-5, 1961, pp 146-153.
7. "Phased Array Radar Studies, 1 July 1959 to 1 July 1960," Technical Report No. 228, Lincoln Laboratory, Massachusetts Institute of Technology.
8. "A Fast Electronically Scanned Radar Receiving System," Davies, D.E.N., Journal British Institute of Radio Engineers, 1961, pp 305-318.

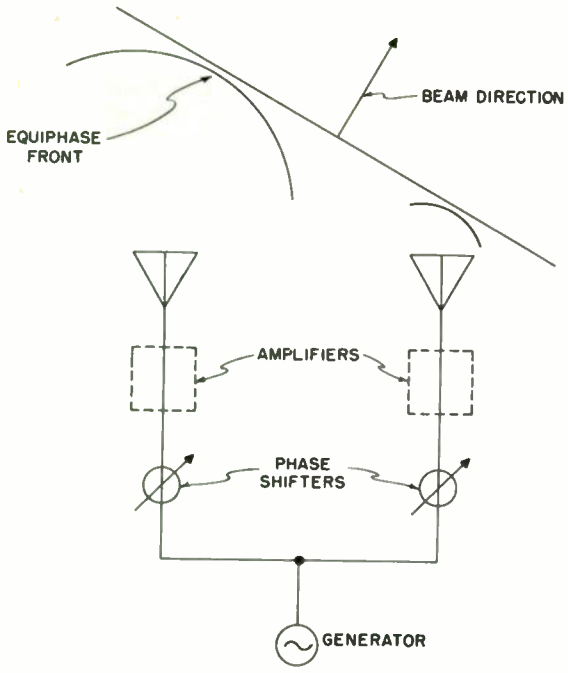


Figure 1. Elementary Phased Array

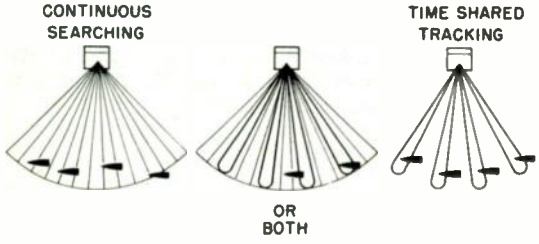


Figure 2. Time Share Feature

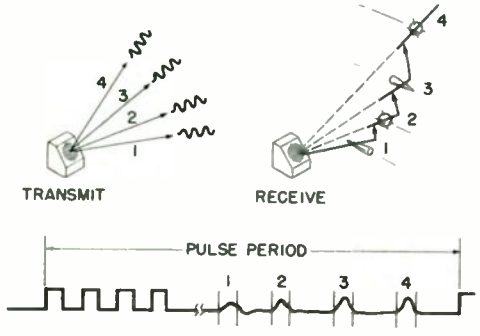


Figure 3. Multiple Target Tracking

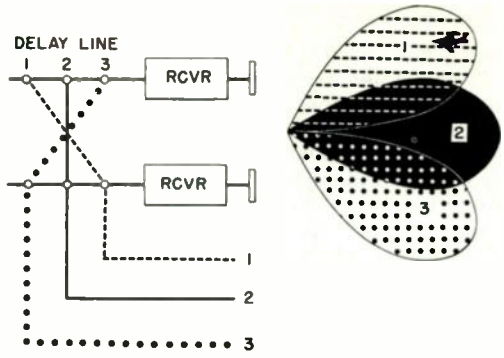


Figure 4. Multiple Beam Formation

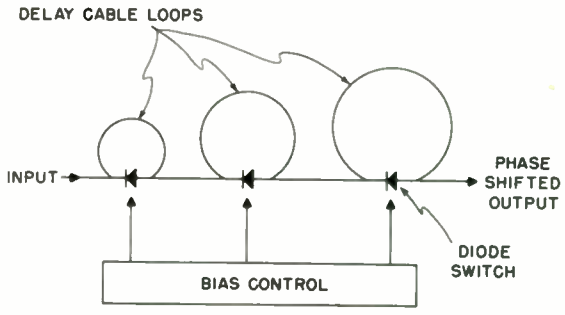


Figure 5. Direct Beam Steering Methods

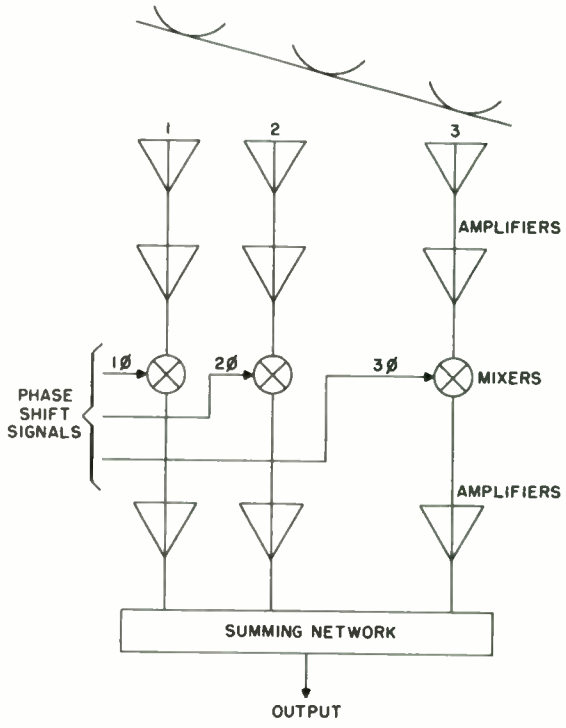


Figure 6. Indirect Beam Steering

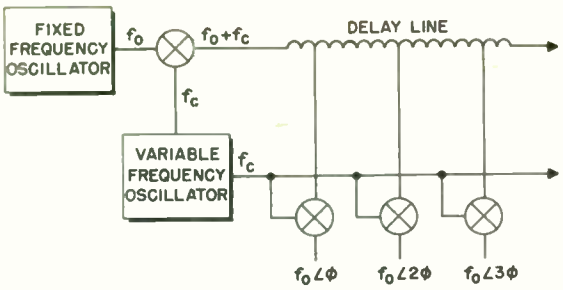


Figure 7. Delay Line Phase Shifter

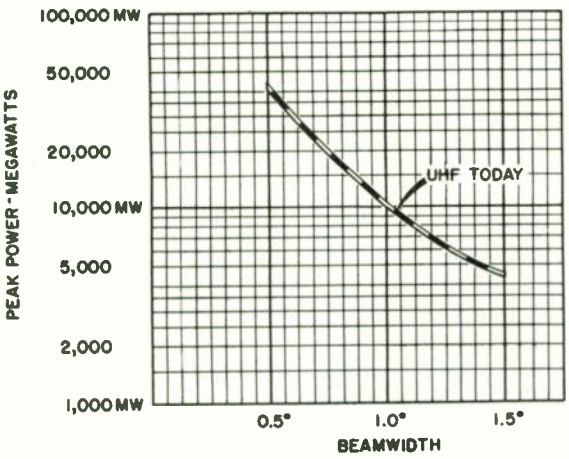


Figure 8. Total Peak Radiated Power

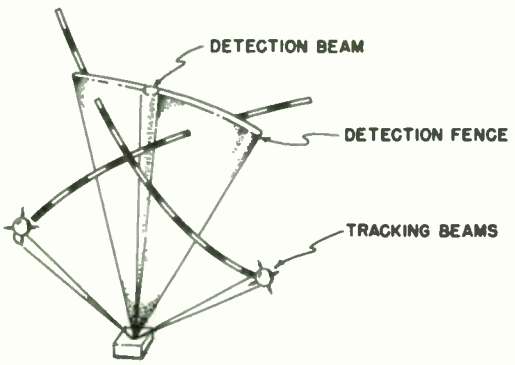


Figure 9. Operating Modes

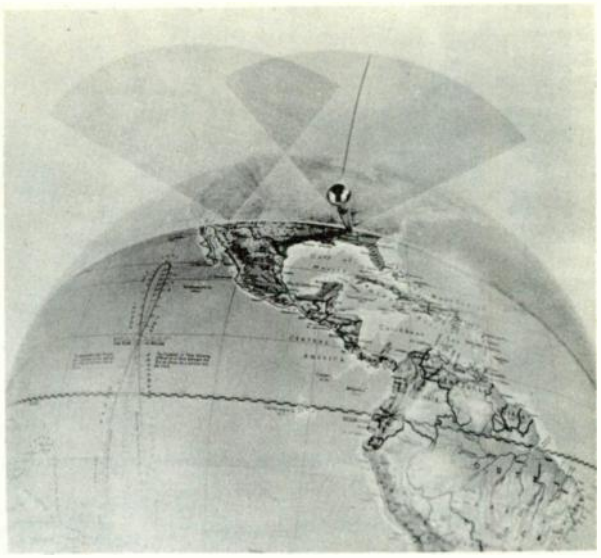


Figure 10. ESAR deployment—with gap filler.

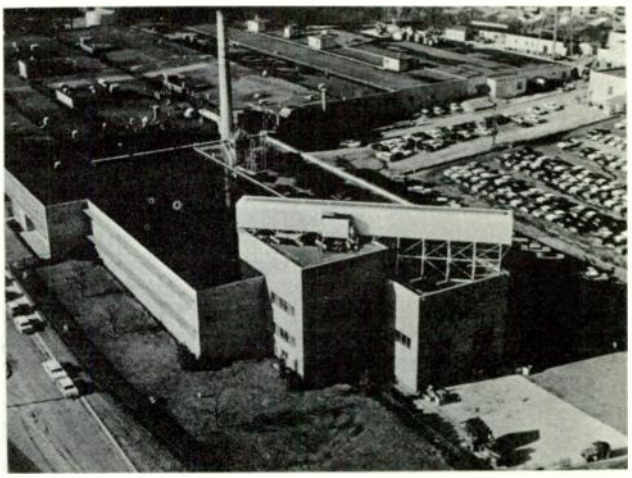


Figure 11. UHF linear array.

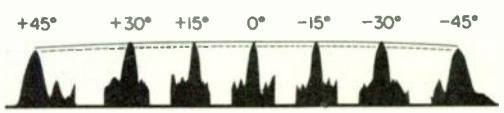


Figure 12. Multiple Beam Recording

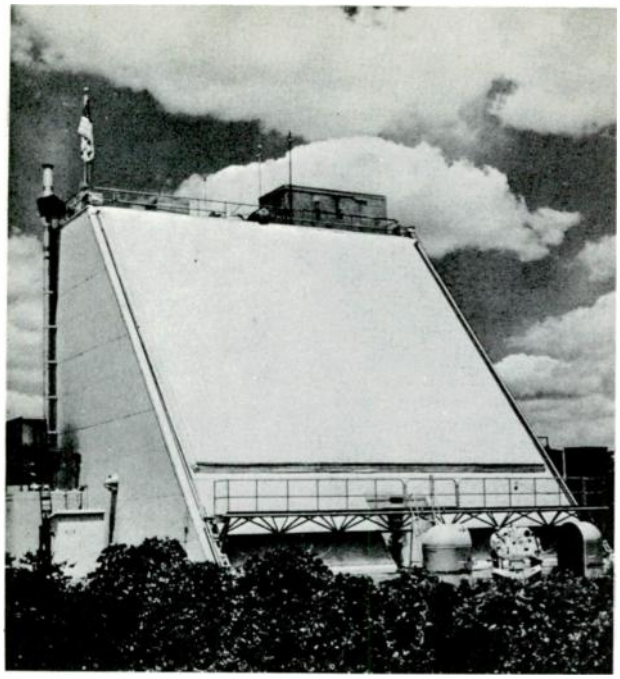


Figure 13. Photograph of L-band model.

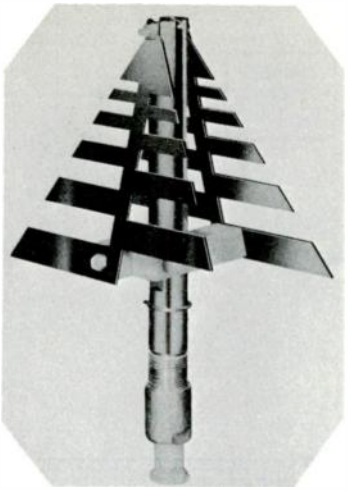


Figure 14. Photograph of log periodic antenna.

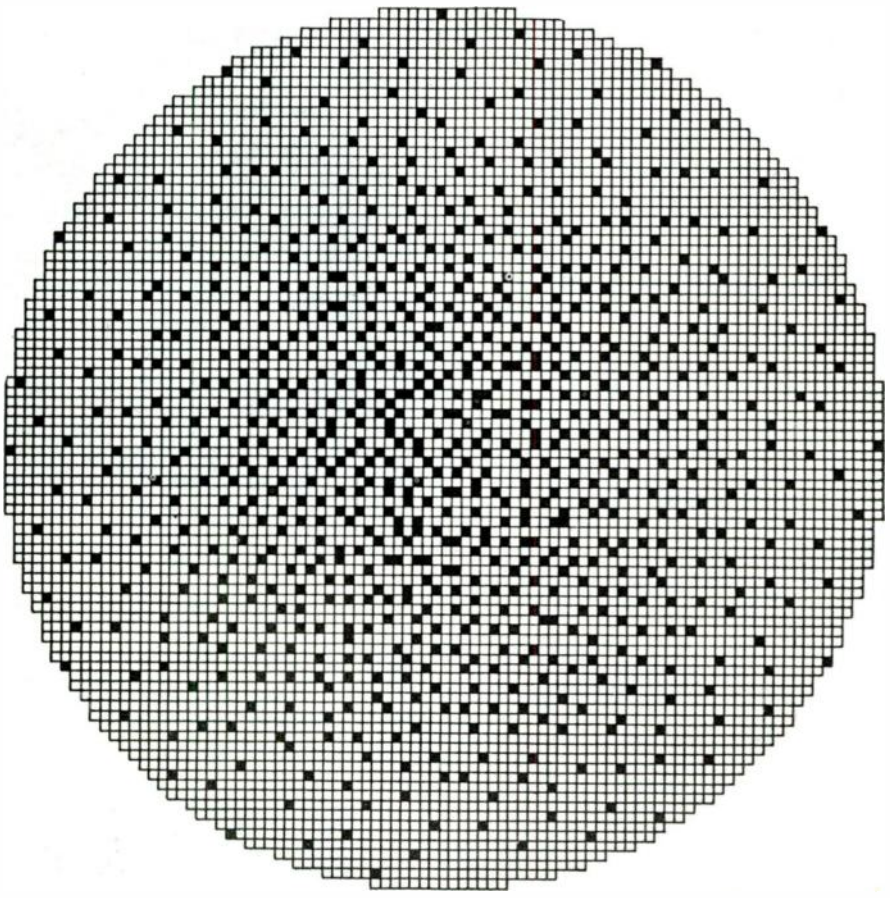


Figure 15. 760 Element Space Taper

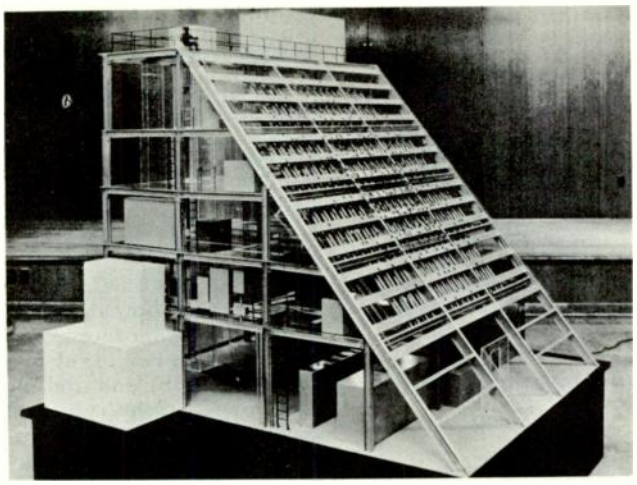


Figure 16. Photograph of L-band scale model.

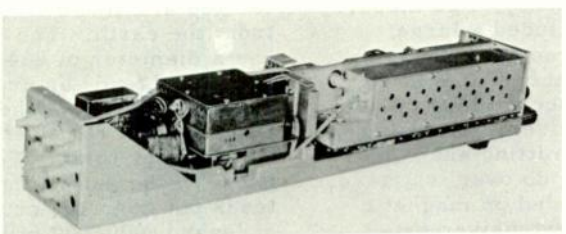


Figure 17. Photograph of L-band module.

A VHF SOLAR RADAR SYSTEM

W. G. Abel, J. H. Chisholm, P. L. Fleck and J. C. James
Lincoln Laboratory, * Massachusetts Institute of Technology
Lexington, Massachusetts

Summary

Radar echoes from the sun at a frequency near 38 mcps have been obtained regularly since April 19, 1961 at the MIT Lincoln Laboratory radar site near El Campo, Texas. These echoes are buried in solar and cosmic noise and can be detected only after cross-correlation with the transmitted signal and integration over at least several minutes. The equipment and techniques employed at the El Campo site are described here together with a brief summary of data analysis to date.

Introduction

The first serious study of the possibility of obtaining radar reflections from the sun was made by Kerr[1] in a paper published in 1952. A more recent study of the subject appeared in the Russian scientific literature in 1957 in a paper by Bass and Braude.[2] Both articles were in agreement that radar reflections could be obtained from the sun and listed the necessary system parameters as follows: transmitter, 250 kilowatts peak power; antenna gain, 34 db over isotropic; pulse length; at least 7 seconds; frequency range, 30-35 mcps. Using the above parameters in the standard radar equation and assuming a radar cross section at 30 mcps slightly larger than the projected disk of the visible sun, Kerr computed an expected echo-to-noise ratio of 3.7 db for quiet sun conditions.

Radar contact with the sun was first established in 1959 when Eshleman[3] and his associates at Stanford University obtained echoes using transmissions at a frequency of 26 mcps. The Stanford radar system included a large antenna made up of 8 rhombic antennas in a broadside array, a transmitter operated at 40 kilowatts peak power with 30 second long on-off pulses and a conventional receiver. The antenna, which was used for both transmitting and receiving, ha a computed gain of 25 db over isotropic. The receiver output was recorded on magnetic tape and analyzed later using extensive post-detection integration with an electronic computer. The Stanford data revealed that radar reflections from the sun had indeed been obtained and that these reflections were from the sun's atmosphere or corona at a distance of 1.5 - 1.7 solar radii from the center of the sun.

A program of radar probing of the sun was begun by the MIT Lincoln Laboratory in April 1961 using transmissions at a frequency near 38 mcps from a site near El Campo, Texas, at 29 degrees north latitude. Radar observations of the sun have been made regularly since then on almost a daily basis.

The El Campo antenna consists of 1024 half-wave dipoles and has a theoretical gain of 36 db above isotropic. The transmitter is operated at a power level of about 500 kilowatts. A cascode front end and converter together with a communications type receiver are used for receiving. The solar echoes obtained using this system are below the level of received noise and, even under the most favorable noise conditions, several minutes of post-detection integration is required before they can be recognized in the noise background.

At this time, only a partial and preliminary analysis of the data obtained from all the solar observations has been made. However, some of the more important results obtained to date may be summarized as follows: 1. for frequencies near 38 mcps, radar echoes are obtained from the sun's corona at a distance of about 1.5 solar radii from the center of the sun, 2. the radar cross section of the sun is not constant but undergoes a day-to-day variation, 3. the spectrum of the solar echoes at this frequency is spread over at least 12 kcps of bandwidth.

The Sun as a Radar Target

The sun is a rotating, hot, gaseous body located at a mean distance of 93,000,000 miles from the earth. The visible sun, the photosphere, has a diameter of 864,000 miles and from the earth, subtends an angle of 1/2 degree. Immediately above the photosphere lies the cromosphere, a relatively homogeneous layer about 9,000 miles thick. The outer layer or atmosphere of the sun is known as the corona. It extends out from the cromosphere for a distance of at least 1,000,000 miles.[4] The corona has a temperature of about 1,000,000 degrees Kelvin and is completely ionized. Electron densities decrease radially from about 10^8 electrons per cm^3 in the lower corona to 10^5 electrons per cm^3 at a distance of two photospheric radii from the center of the sun.[5] Local fluctuations, both temporal and spatial, are believed to be superposed on this spherically symmetric distribution of electron densities.

* Operated with support from the U.S. Army, Navy and Air Force.

Radio waves impinging on the sun are not reflected in the same way that they would be from a hard target, such as one of the planets. For wavelengths greater than about 50 cm they do not penetrate to the visible sun but are bent back within the solar corona. The depth to which the rays penetrate the corona is a function of radio frequency, angle of incidence at the sun and electron density distribution. The ray paths computed for a frequency of 38.25 mcps are shown in Figure 1. At this frequency, the direct or zero angle ray penetrates to a distance of just over 1.4 radii from the center of the sun. Other rays penetrate less deeply and at 0.2 degree the penetration is to about 1.5 solar radii. The surface formed by the family of turning points is approximately paraboloidal in shape, but it is highly irregular or "rough" because of local fluctuations in electron density.

Radio waves propagated in the sun's corona undergo both attenuation and group retardation. Values computed for a frequency of 38.25 mcps show that group retardation varies from 1.6 seconds for the ray directed toward the center of the sun to less than 0.4 second at 0.5 degree from the center; attenuation changes from slightly more than 6 db to about 0.25 db over the same angular range.

Frequencies which may be used for radar probing of the sun fall into a rather narrow band. The upper limitation is imposed by absorption losses in the sun's corona. Absorption increases with frequency and the highest useable frequency with presently available transmitter powers is less than 100 mcps. The lower frequency limitation is controlled by the earth's ionosphere, which must be traversed twice by radar signals beamed toward the sun. The F-layer of the ionosphere has critical frequency values up to 15 mcps at vertical incidence and 30 mcps at 30 degrees elevation angle. The lower frequency limit is, then, above 25 mcps for year around operation from sites in the southern United States. This frequency will change somewhat with sunspot cycle.

Radar signals returned from the sun will be spread over a spectrum of frequencies because of the rotation of both the sun and the earth. They will also be shifted in frequency by the displacement velocity of the sun and earth along the line joining their centers. The expected doppler spread of signals at 38.25 mcps for reflecting points on the sun's equator at a distance of 1.5 radii from the center was computed to be $\pm$715 cps. The spread due to earth rotation was computed for the El Campo latitude to be about 100 cps. The maximum shift due to the rate of displacement along the earth-sun axis is 130 cps.

The noise which limits the detectibility of radar signals from the sun is external to the receiver and is identified as galactic noise and noise from the sun itself. The level of galactic noise is determined by the portion of the celestial sphere which is "seen" by the antenna. For the El Campo location and considering elevation angles between 30 degrees and the zenith the galactic noise level at 38.25 mcps varies from approximately 10 to 20 db above 290° K over a siderial day for 8 kcps bandwidth. Superposed on the galactic noise background is the noise from the sun itself, which has an average level of about 16 db. However, the limiting noise is predominantly from the sun, except during the month of December when the galactic noise peak occurs near the time of solar transit. During disturbed solar conditions the quiet sun noise level may be increased by as much as 10-15 db. This increase in noise level may be of a continuous nature and last up to several days or it may be of an intermittant nature with burst of noise enduring for periods of a second or so to several minutes.

The El Campo Solar Radar System

The El Campo solar radar system comprises a high power transmitter, a large dipole antenna array, receivers, digital correlating equipment, integrators and a magnetic tape recording system. The transmitter is operated at a frequency of 38.25 mcps, although it can be operated over the range of 20-50 mcps. The final amplifier consists of two triodes in resonant cavities operated 180 degrees out of phase and capable of producing over 250 kilowatts average power each. The outputs from the final amplifier are fed separately to the antenna through two runs of 6 1/8 inch coaxial line.

The antenna, which is used for both transmitting and receiving, consists of 1024 horizontally polarized half-wave dipoles located approximately 5 feet above ground. The dipoles are arranged in 8 rows of 128 elements each, with the rows extending along a north-south line. The spacing between dipole centers is slightly more than 0.5λ (13 feet 9 inches) in both the north-south and east-west directions. In overall size, the antenna measures 1750 feet long by 110 feet wide and covers almost 5 acres of land. An aerial photograph of the antenna, looking south toward the transmitter building, is shown in Figure 2.

The eastern and western halves of the antenna are fed separately by the two coaxial lines from the transmitter. Each coaxial line branches into four lines in a "Christmas tree" type power dividing system. Both the input and output impedance of this branching system is maintained at 50 ohms by suitable impedance transformations in the 6 1/8 inch coaxial line. The eight outputs of the two power dividing systems are connected through transitions to 7 inch by 5 inch rectangular trough feedlines, which run the length of each row of the antenna.

The dipole elements of the antenna are connected to the trough line through RG-8 coaxial cables and a capacitive coupling. Approximately equal amounts of power are fed to each dipole by varying the degree of coupling

between the dipoles and the line. Fifteen different degrees of coupling are used over the length of each row, from about -20 db at the feed end to a direct coupling of the last element at the far end.

The beam produced by the antenna is fan-shaped and has a computed half-power width of 12 degrees in azimuth and 3/4 degree in elevation for angles near the zenith. The beam position is changed in elevation by changing the phase delay between dipoles along the length of each row. Phase changes in terms of $1/16\lambda$ are made by using 4 sections of RG-8 coaxial cable of length $1/16\lambda$, $1/8\lambda$, $1/4\lambda$ and $1/2\lambda$. These cables are joined together to form the proper lengths and inserted between the dipole feed line and the capacitive coupling in the trough line. A photograph of a section of the antenna showing the dipoles, trough line, phasing cables and coupler housing is shown in Figure 3. Phasing the antenna is a manual operation and requires about 16 man hours of labor.

The antenna beam can be moved either north or south from the zenith to about 30 degrees elevation. No change in the azimuth position of the beam can be made for transmitting; however, small changes in azimuth are made for receiving by changing the length of one of the two receiving feedlines. In order to maintain the VSWR of the antenna at less than 1.2 for transmitting it is necessary to rematch the dipoles several times as the beam is changed in elevation from the zenith to 30 degrees.

The gain of the antenna and its elevation pattern have been determined by measuring the signals from an airborne beacon-transmitter moving along the north-south axis of the antenna. The half-power beam width was found to be 1.2 degrees and the gain, 30 db over isotropic for angles near the zenith. The gain of the antenna and its pattern over a few degrees in azimuth were obtained from measurements of the flux density of the radio star, Virgo-A. These measurements yielded a half-power beam-width of 11 degrees and a gain of 30 db. From the two series of measurements, aircraft and Virgo-A, it was determined that the beam broadening in elevation was caused by progressively increasing phase delay along the trough lines. After a recomputation of the dipole phasing based on these measurements, more recent measurements show the antenna gain to be 32 db and the half-power beam-width in elevation, 0.88 degree.

A block diagram of the solar radar system is shown in Figure 4. In order to use the full output power of the transmitter, it is pulsed by switching between two frequencies equally offset from the central frequency of 38.25 mcps. The separation between frequencies is adjustable in steps from 250 cps to 16 kcps. Keying between the two frequencies is controlled by pulses from a shift register, which generates either a square wave or a pseudo-random pulse sequence.

The receiver comprises a cascode front end, IF amplifier, two variable bandwidth filter channels and two square law detectors. Each filter is tuned to one of the two transmitted frequencies and is adjustable in bandwidth from 50 cps to 16 kcps. The converter now in use has a bandpass of 200 kcps and the IF amplifier, 35 kcps; however, in the initial equipment, which was used to obtain most of the data analyzed to date, the bandpasses were 85 kcps and 16 kcps, respectively.

The outputs of the two square law detectors, one containing the radar-signal plus noise and the other, only noise, are fed to a subtractor circuit where their voltage difference is taken. This difference voltage is then applied either directly or through an inverter to each of 20 range box integrators. The input to each integrator is switched alternately between the direct and inverted outputs of the subtractor circuit by a square wave or pseudo-random pulse pattern generated during the receiving period. This pattern is identical with the one generated during the transmitting period which switched the transmitter between the two offset frequencies.

Short term variations of received noise are suppressed by causing the receiver to have a voltage gain which is inversely proportional to the received rms voltage (-1 slope). This gain control is accomplished by driving the receiver automatic-gain-control with a several millisecond average of noise received in a separate receiver. Hard limiting is applied to reduce the effect of ignition noise and similar noise spikes.

A simplified block diagram of the timing and control system is shown in Figure 5. The 100 kcps basic frequency is obtained from a frequency standard and fed to a divide-by-100 counter (A) to obtain one millisecond pulses. These pulses provide the basic time unit used in all the timing operations. The transmitted pulse-width is derived in a pulse rate counter (B), which is adjustable in powers of 2 from 2^0 to 2^{15}, i.e., from one millisecond to 32.768 seconds. The pseudo-random pulse pattern used during transmission is obtained from the shift-register (C), which provides pulses varying in duration from 1 to 5 unit pulse lengths. Either the square-wave pulse pattern from (B) or the pseudo-random pattern from (C) may be used to activate the transmitter keying relay.

The cross-correlation time difference, which determines the range interval of each integrator range box, is derived in a separate pulse-rate counter (F). This time is adjustable in steps from 2^0 to 2^{15} milliseconds and provides the basic pulse-width of either the square-wave or pseudo-random pulse pattern used during the receiving cycle. The pseudo-random pulse-pattern is obtained from a shift-register (G) separate from that used during the transmitting period. It is delayed in the 20-step delay shift-register (H) and then applied in turn to each of the 20 integrator switching relays. The delay

between successive outputs of the delay shift-register is equal to the cross-correlator time difference.

The length of the transmitting period is controlled by the roundtrip time counter (D), which is adjustable in intervals of one millisecond. Control pulses from this circuit activate the relays which change the radar system from the transmitting to the receiving mode of operation.

The solar echo is programmed to be in a particular range box integrator by choosing the appropriate setting for the first cross-correlator time delay (E). The range box number is determined from the relation:

$$N = \frac{T-C}{D} + 1$$

where N = range box number
T = length of transmitting period
C = first cross-correlator time delay
D = cross-correlator time difference

The tape recorder system provides a means of storing the radar signals received from the sun so that they can be reprocessed for additional information. The tape recorder has provision for recording 7 channels of information and is run at a tape speed of 30 inches per second. The accessory circuitry used with the recorder consists of a tape indexing control and a frequency converter.

The tape indexing control operates on signals from the radar timing and control circuit. Its function, during the recording cycle, is to provide automatic starting of the tape drive mechanism prior to the beginning of the radar receiving period and to provide timing pulses for recording on one of the tape channels. During the playback cycle it again provides timing pulses for comparison with those previously recorded on the magnetic tape.

The frequency converter heterodynes the 455 kcps IF signal from the receiver down to 35 kcps for recording. It has a bandpass of 20 kcps. During playback, it reconverts the magnetic tape signals back to 455 kcps for injection ahead of the two filter channels in the receiver. Provision is made in the converter for shifting the local oscillator frequency by means of an external signal generator so that the radar signals can be studied for doppler-shifted components.

Operation

The beam of the dipole antenna array is essentially fixed in azimuth and only one observation is scheduled for each day, near the time of solar meridian transit. The antenna beam is positioned in elevation so that on the first day of a given series of observations the beam is slightly offset from the sun's elevation at meridian transit. On successive days, as the elevation of the sun changes, the beam effectively scans the face of the sun in a series of overlapping observations. Ordinarily, scanning of the sun has been limited to $\pm$ 0.5 degree (2 solar radii), although several series of observations were begun with the beam offset by slightly more than one degree.

The transmitting period is scheduled to begin at a time which preceeds the time of solar meridian transit by an interval equal to the computed radar propagation time to the sun. This interval varies from day-to-day but has a median value of about 16 1/2 minutes. During this time the earth rotates through approximately 4 1/8 degrees so that the azimuth center of the antenna beam in in line with the center of the sun at the end of the transmitting period. The receiving period begins at this time and continues for another 16 1/2 minutes or so.

Signals are integrated over the entire length of the receiving period. Provision is made in the equipment for integration of the signal and noise received on one of the two transmitted frequencies, as well as for integration of the difference of the signal and noise received on the two frequencies. Since the signal and noise averaged over the receiving period should be equal on the two frequencies, their integrated difference is an indication of the gain stability in the two receiving channels. The single channel integrator output voltage is used in the data analysis to determine the theoretical standard deviation of the rms noise about the mean.

During the receiving period, recordings are made on a Sanborn chart recorder of the following information: 1. sampling readouts of all the integrators at closely spaced time intervals, 2. the pseudo-random pulse sequence which controls the switching relays of the 20 range box integrators, 3. the noise level at 38.25 mcps. The recorded readouts of the range integrators shows how the integrating technique suppresses the received noise while allowing the signal to build up in one or more of the range boxes. In the case of a doubtful signal, it is possible to judge by the way the voltage builds up in a particular integrator whether the final voltage reading is due to signal or to noise. The noise record is a measure of the noise level during the receiving period. It is used in correlating the solar radar cross section determined from the data analysis with noise conditions on the sun.

The output of the receiver IF amplifier is recorded on magnetic tape for later playback and processing. During playback, the output from the tape recorder replaces the output of the receiver IF amplifier. The center frequency of the tape recorder oscillator is shifted so that a search of the solar echoes can be made for signals which have been doppler shifted by $\pm$2 kcps, $\pm$ 4 kcps and $\pm$ 6 kcps from the transmitted frequencies. Two additional playbacks are also made, one at the filter bandpass used during the original recording and the other at 1 kcps.

The receiver integrator system is calibrated by inserting weak test signals from a signal generator together with received noise (cosmic) or noise from a noise generator. The frequency of the test signal is switched in a pseudo-random sequence with an appropriate time delay to simulate a radar return in a particular range box. For a 1,000 second integration period, the selected range box produces an output of about 4 standard deviations above the mean of the other 19 integrators for a signal -25 db with reference to the noise level in a 2 kcps filter bandpass.

The position of the range box in which the solar echo is programmed to appear is adjusted daily to correspond to the range of a discrete radar reflection from a point in the solar corona 1.5 photospheric radii from the center of the sun toward the earth. A computed delay of 2 seconds is included to allow for the slower group velocity of the signal in the corona.

Since the beginning of the solar observations in April 1961, various pulse widths, filter bandwidths, frequency separations and positions of the expected solar echo range box have been used. Recently, however, parameters were standardized for a 4 second pulse width, ± 8 kcps frequency separation, 8 kcps filter bandwidth and solar radar echo in range box 13.

Data and Analysis

The data obtained during each solar observation consist of the 20 range box integrator voltages and a record of noise conditions during the time when radar signals were impinging on the sun. The rms value of the received noise is determined from the voltage readings of the range boxes which contain only noise. The strength of the solar echo relative to the noise level is then obtained in terms of the number of standard deviations above the rms noise background. This ratio of solar echo to received noise is determined for each solar observation.

Additional integration of the radar signal is achieved by summing the voltage in each range box for a number of days. Before making this summation it is necessary, however, to weight the range box voltage by a factor which eliminates the effect of day-to-day changes in the level of transmitted power and of noise background. The results obtained by adding up 32 radar observations of the sun using 8-second pulses during the period of April 19 to July 7, 1961, are shown in Figure 6. The predominant radar return in this Figure is in range box 16, with approximately 9 standard deviations above the mean of the other 19 range intervals. With the 8-second pulse width used in this series of observations the radar effectively searches for echoes over a range interval of 34 solar radii.

The 8-second pulse widths were used principally during the earlier solar observations while pulse widths of 4, 2 and 1 seconds have been used during some of the more recent ones. The narrower pulse widths should provide better range definition and possibly more information concerning the distribution of reflecting centers in the solar atmosphere. Figures 7 and 8 show the results obtained using 2-second and 1-second pulse widths respectively. Both Figures show that the earliest radar return from the sun is obtained at about 1.5 radii from the center, assuming a 2-second group delay in the corona. They also show that the echoes obtained during these observations are distributed over a range interval of about 4-seconds. This is equivalent to 0.9 solar radii. However, partial analysis of more recent data using 4-second pulses indicates that the radar echo may be distributed over more than 4 seconds.

Analysis of a number of solar observations indicates that the radar cross section of the sun for frequencies near 38 mcps ranges from 0.1 to 0.01 of the projected area of the solar photosphere with an average value of about 0.3. These values are preliminary and will be adjusted later to account for some uncertainty about the gain and beamwidth of the El Campo antenna, the absolute position of the antenna beam and the weighting that should be applied due to the relative position of the antenna beam with respect to the sun. There is also some indication that the sun's radar cross section undergoes day-to-day variations and that this variability may be associated with changes in the solar noise background.

Studies of the doppler spread and doppler shift of the solar echoes have also been made. A comparison of the signal energy received in an 8 kcps filter bandpass with that in a one kcps bandpass is shown in Figure 9. The echo is expected in range box 13. The total signal energy received in the 8 kcps bandpass is more than 5 times greater than that received in the one kcps bandpass.

The results obtained in a doppler search of the solar echo over the frequency interval of ± 6 kcps from the transmitted frequency are shown in Figure 10. The filter bandwidth is one kcps. It is at once apparent that the spectrum of return signal is at least 12 kcps wide. This wide doppler spread of the solar echo is much greater than the one kcps calculated for solar and earth rotation alone and must be explained in terms of motion of the ionized particles which make up the sun's corona.

Present equipment design was based upon theoretical models of the sun's corona from which an expected doppler spread of about one kcps was computed. Doppler search with this equipment is limited to about ± 7 kcps by the maximum spread of 16 kcps between the two transmitted frequencies. Modification and redesign of the El Campo equipment is presently underway to accomodate a doppler spread in the solar echo of about 35 kcps.

Conclusions

It is believed that the radar system and preliminary results described in this paper have demonstrated the feasibility of regular systematic radar studies of the solar corona. In addition, the results indicate the minimum capability required to obtain solar echoes and should serve as a reliable guide for future radar systems of improved sensitivity. In this paper, the emphasis has been placed upon system parameters and system techniques with only a few typical samples from some 200 radar observations which have been obtained in the program to the present date. The variability of the daily solar radar observation and the association of the radar results with other radio and optical indices of solar activity is currently under investigation and the results of these studies will be reported in detail at a later date.

References

1. F. J. Kerr, "On the Possibility of Obtaining Radar Echoes from the Sun and Planets," Proceedings of the Institute of Radio Engineers, Vol. 40, pp. 660-666 (June 1952).

2. F. G. Bass and S. I. A. Braude, "On the Question of Reflecting Radar Signals from the Sun," Ukranian Journal of Physics, Vol. 2, pp. 149-163 (1957).

3. V. R. Eshleman, R. C. Barthle and P. B. Gallagher, "Radar Echoes from the Sun," Science, Vol. 131, pp. 329-332 (February 5, 1960).

4. G. P. Kuiper, Editor, The Sun, Vol. 1 (The Solar System), The University of Chicago Press, Chicago, Illinois, 1953.

5. S. F. Smerd, "Radio-Frequency Radiation from the Quiet Sun," Australian Journal of Scientific Research, Series A, Vol. 2, No. 2 (1957).

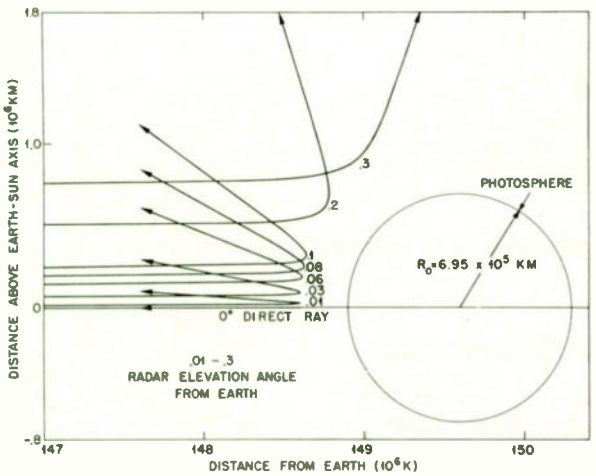


Fig. 1. Computed ray-paths in solar corona at 38.25 mcps.

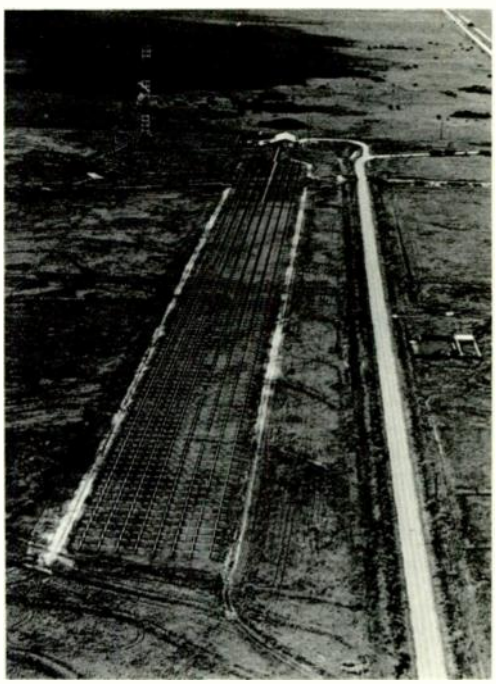


Fig. 2. Aerial view of El Campo antenna looking south.

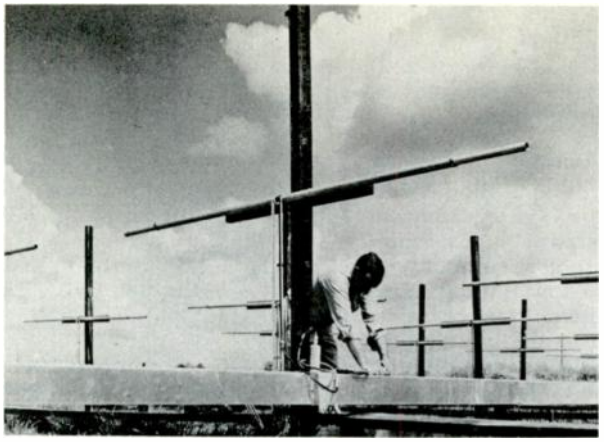


Fig. 3. El Campo antenna, dipole elements and feed line.

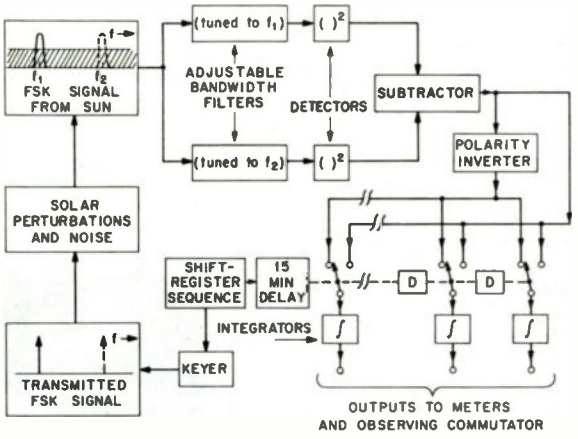


Fig. 4. Block diagram of El Campo solar radar system.

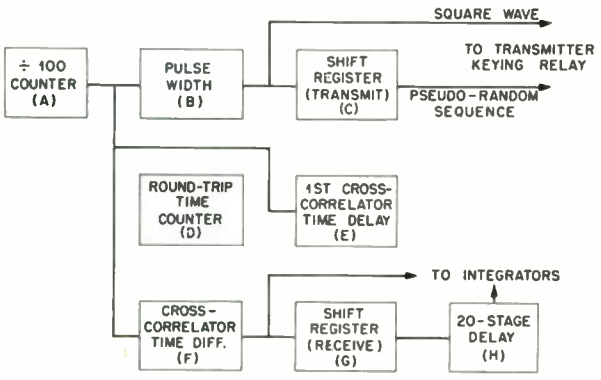


Fig. 5. Block diagram of timing and control circuit, El Campo solar radar system.

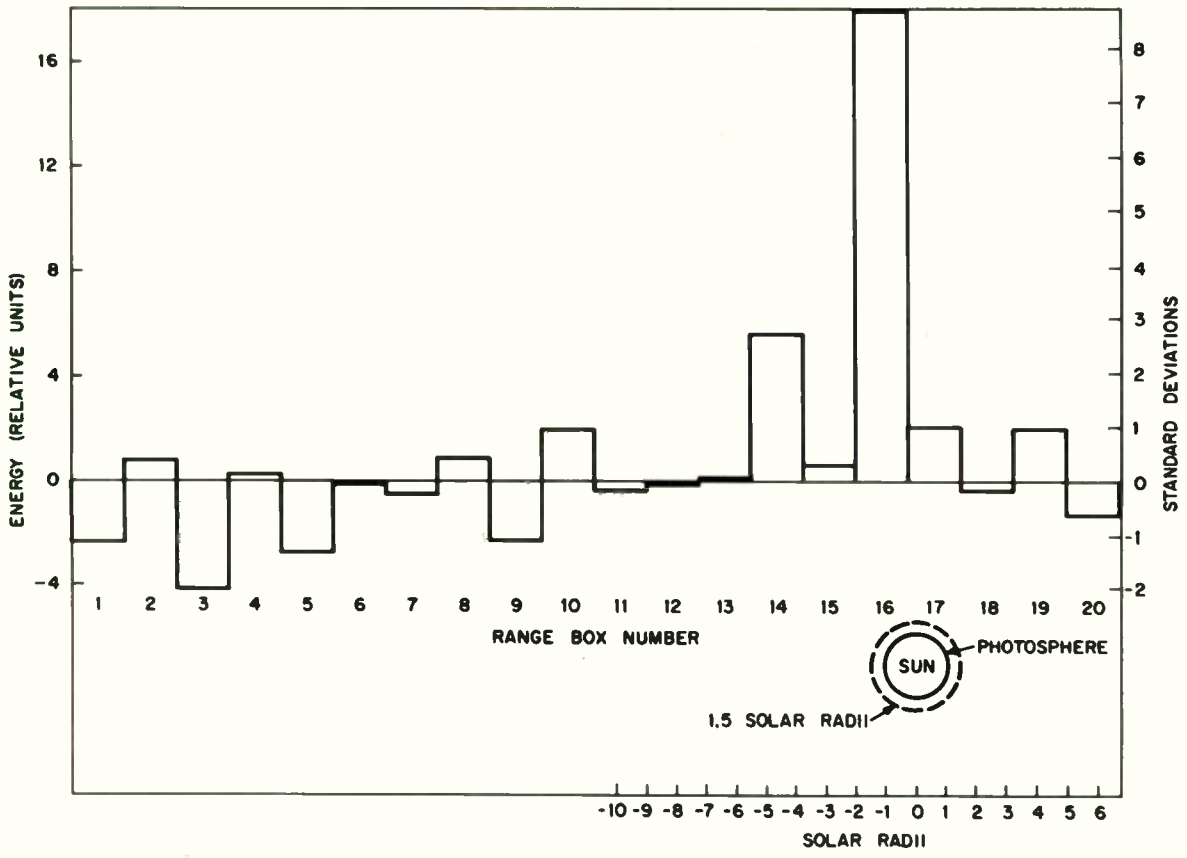


Fig. 6. Solar radar echo and noise received at El Campo, using 8-second pulse width.

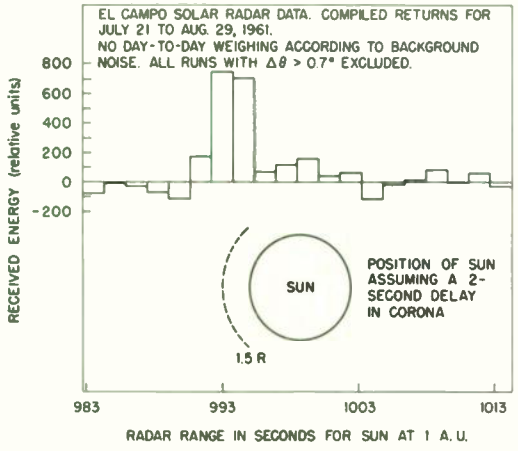
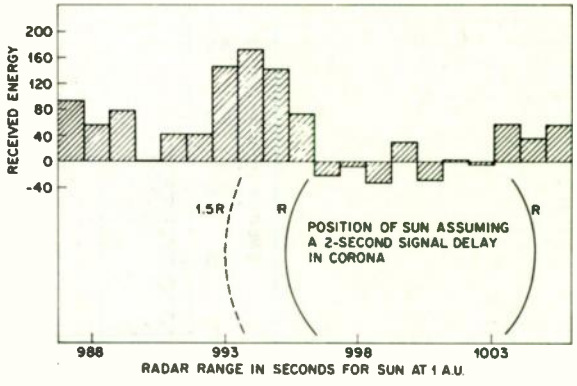


Fig. 7. Solar radar echo and noise received at El Campo, using 2-second pulse width.



COMPILED SOLAR RADAR DATA FROM AUGUST 30 THROUGH SEPTEMBER 9, 1961 AT 38.25 Mc. EL CAMPO, TEXAS

Fig. 8. Solar radar echo and noise received at El Campo, using 1-second pulse width.

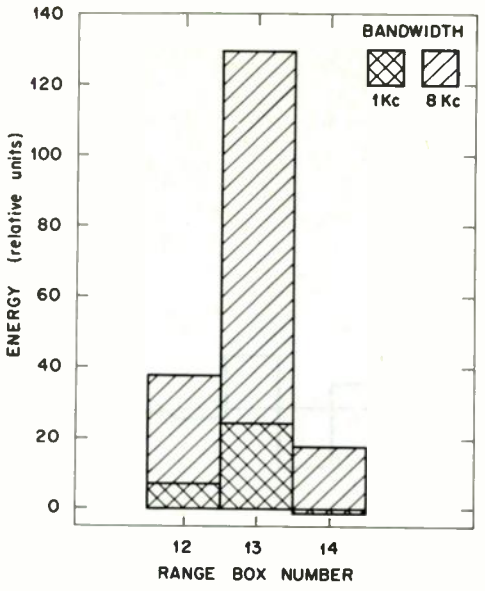


Fig. 9. Comparison of solar radar echo strength in 8 kcps and 1 kcps receiver bandwidths.

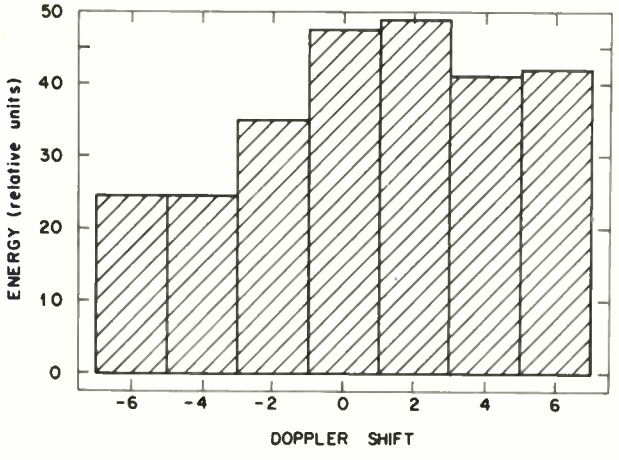


Fig. 10. Doppler-shifted solar radar echo.

A MOTION-ENHANCEMENT DISPLAY BY TIME-COMPRESSION*

Munsey E. Crost
U. S. Army Signal Research and Development Laboratory
Fort Monmouth, New Jersey

Summary

To demonstrate the enhancement of detectability of moving targets in the presence of severe clutter, a display system of electronic time-compression was devised and constructed at USASRDL. The principle of operation is that a number of immediate-past-history scans are stored within the system and then displayed sequentially at a much faster rate than acquired.

The Westinghouse Permachon storage vidicon was chosen as the storage tube type around which the system would be constructed. To provide subjectively-satisfactory continuity of motion, at least six previous-history images should be stored. The system was designed around the minimum number of six images. Each Permachon required essentially its own vidicon camera. The six cameras were arranged in a regularly-spaced hexagon surrounding a central PPI cathode-ray tube, which receives the real-time radar information. Each camera includes its own lens and video preamplifier. All sweep and blanking signals are synchronized at standard television rates.

The cameras are exposed sequentially to complete PPI scans by means of an intermittently-rotating plane mirror and an apertured cone in front of the lenses. Immediately before a new scan is to be recorded in a Permachon, the previously-stored information is erased by means of a light flash, triggered by a switch on the mirror shaft as the shaft begins to rotate. Since the 60° rotational step of the mirror is not instantaneous, the video signal corresponding to this period will be lost. However, after a flash erasure the Permachon also requires a finite time to recover. In order that each Permachon shall receive a full PPI sweep, the intermittent rotation of the mirror is so synchronized with the PPI sweep that 60° of PPI rotation are lost for the erasure and recovery period, and the stepping action occurs at each 420° of PPI rotation.

Simultaneously with the recording and storage processes, the stored information is read out at a much faster rate into a viewing monitor. All the Permachons are scanned simultaneously, and the six preamplifiers feed video signal into six gated amplifiers having a common output circuit. An "on-gate" is applied successively to each gated amplifier in proper sequence, so that the output signal is composed of the successive video outputs of the amplifiers in a repeating sequence. The novel circuitry provides that in dynamic operation only one gate may be "on" at a time. The gates are timed by means of the 60 pps vertical driving pulses from the television synchronizing-signal generator. The frequency is halved to provide frame-rate pulses, and these pulses are counted in a three-stage binary counter so designed that any preselected number of frames from one to eight may be counted. Each output pulse from this counter switches the "on-gate" from one amplifier channel to the next. Therefore, the video signal from each Permachon may be viewed for one to eight frames before the next Permachon's signal is viewed. The output video signal from the gated amplifiers is further amplified before sending the signal to a video monitor.

In the laboratory demonstration system a PPI rotation period of six seconds was chosen for convenience. The stepping period is then seven seconds, or one complete cycle of recording requires 42 seconds. The reading cycle may be selected from 6/30 to 48/30 seconds. The time-compression ratio may thus be between 210 and 26¼.

The moving signals are readily visible as a series of six spots progressing in the direction of target motion to essentially the real-time position of the target and then repeating. Whereas on a real-time monitor the moving target is almost indistinguishable from the clutter signal, the same target on the time-compression monitor is immediately detectable. The selectable gate-widths allow adjustment of the time-compression ratio for best conspicuousness of targets of various speeds.

Several methods have been investigated for preventing the signal from the Permachon being erased from interrupting the desired display. The method currently in use involves generating an additional trigger pulse for gate-advancement whenever there is a coincidence between any reading gate and another gating signal derived from a microswitch on the mirror rotation mechanism. The video signal from the erased camera is then skipped without display until the coincidence period is completed. During this period there will appear only five spots for each moving target instead of the usual six.

An electrical-input and electrical-output storage tube based on the use of a Permachon target, presently being developed, will permit conversion of the system from an optical-mechanical-electronic system to an all-electronic system.

Introduction

In electronic displays showing targets moving through areas of clutter or fixed targets, the moving targets may easily be lost to view if the display is presented in real time. If moving target indicator filters, such as delay lines or storage tubes, are used, the display will show the presence of the moving target, but with very little

*The views of the author do not purport to reflect the position of the Department of the Army or the Signal Corps.

indication of its direction or speed of movement. In addition, the display loses its positional reference signals.

It is a frequent practice to photograph electronic displays for later study and analysis. When a time-sequential series of such photographs is projected at a rate considerably faster than the rate at which the actual displays appear, it has been noted that the increased motion of the targets through the clutter enhances their detectability to a marked degree. Paths of targets may be followed through areas of dense clutter where they might otherwise be completely lost in real-time displays, even with long-persistence phosphors.

While this system is effective, it is unsuitable for displays in which the present positions and directions of motion of moving targets are required to be noticed immediately. There will always be a delay for processing the photographic film after the display is recorded, and, generally, additional time will be lost in accumulating a sufficient number of frames to make the motion-enhancement sequence possible.

Several workers in the display field have independently proposed methods whereby such photographs or recordings could be used to produce a final display including motion-enhancement, with a minimum of lost time between the real-time display and the compressed-time display. Basically, these schemes consist of shortening the series of recordings to a number of the most recent frames and displaying these frames repetitively at a time-sequential rate faster than the recording rate. The most recent real-time recording is added to the end of the series, and simultaneously the oldest frame is removed. This results in the display of a series of dots for any given target, progressing in the direction of the target and ending with the latest position of the target. The series is presented at a rate sufficiently slow to permit the observer to recognize motion.

Engineering Considerations

In the Pickup, Display, and Storage Devices Section of Special Tubes Branch, at the United States Army Signal Research and Development Laboratory, two methods were devised to provide such a display by electronic means. The esthetically more desirable method involved the use of scan-converter, electrical-output storage tubes. A series of such tubes would successively record sequential PPI frames from a radar at the real-time rate. The stored information would then be read from these tubes by a TV scan, in the same sequence but at a considerably faster rate, to provide on a TV monitor the final display previously described. Immediately before a new PPI frame is to be recorded in a storage tube in which old information is recorded, the recording in this tube would be erased electronically. The second method involved the use of the then newly-disclosed Permachon developed by Westinghouse Electric Corporation. The Permachon is a camera tube basically of the vidicon type, but which has a sensitive target that can record and retain images for many minutes while being read by an electron beam. This system requires an optical link between the PPI display tube and the individual Permachon, with sequential exposure of the Permachons to the PPI image. The stored image in the Permachon may be erased quickly by exposure to an intense light flash. Otherwise, the system is similar to the fully-electronic system.

In order to demonstrate the advantages of the time-compression motion-enhancement display, it appeared desirable to construct a model of the display system in our Laboratory. Consideration of the final display indicated that at least six PPI frames should be stored in order to get a useful sequence of dots. The commercially available scan-converter storage tubes are quite large in size, especially with the required external components, and they are individually very expensive. Therefore, this method was considered infeasible at the time we started this experiment for a comparatively inexpensive display. On the other hand, when the Westinghouse engineers and management were informed of the plan to make a model of this display system with Permachons, they offered to make available at low cost as many experimental Permachons as needed for the display. It was therefore decided to engineer the display model around the Permachon tubes.

System Details

The system as presently engineered consists principally of the opto-mechanical camera assembly, the electronic timing unit, the monitor, and the radar simulator assembly, as shown in the overall system block diagram (Fig. 1).

The opto-mechanical camera assembly consists of the six Permachon camera units clustered symmetrically about a central PPI display tube (Fig. 2). Each Permachon has its individual deflection and alignment yokes and focusing coil, with external electrical adjustments. Each camera unit has its own lens and erasing lamps within the lens housing (Fig. 3), and individual preamplifiers are also provided (Fig. 4). The PPI tube in the center provides the real-time display for the Permachons to view, using a short-persistence phosphor to prevent unequal exposure of sequential areas of the Permachon target.

Facing the PPI tube is a 15° inclined mirror that rotates about the extended axis of the PPI tube. This mirror is adjusted to reflect the image of the PPI display into the individual Permachon cameras, which are mounted at 30° angles to the axis. The mirror assembly is synchronized to advance one position with the completion of each PPI scan. At present the mirror shaft is indexed through each 60° rotational step by means of an interrupted-gear and pawl mechanism, during 60° of rotation of a continuously-rotating driving shaft. Since this arrangement imposes high accelerations on the mirror assembly, the interrupted-gear mechanism is being replaced by a modified six-

position Geneva gear, also indexed during 60° of rotation of the driving shaft. Rotating with the mirror there is a shield cone having a single aperture, which obscures all the camera lenses except the one receiving the image at any given time.

A cam on the mirror shaft actuates one of a ring of six microswitches just as the indexing rotation begins. This switch discharges a condenser through the erasing lamps in the lens housing toward which the mirror is moving, erasing the signal stored in the associated Permachon. Since a finite time is required for the Permachon to recover its sensitivity after the flash erasure, no useful signal is available from the Permachon during that period, and no recording can take place.

Since the advancement of the mirror requires 60° of rotation of the driving shaft, this part of each PPI frame would be lost to the Permachon if the driving shaft and the PPI scan rotated at the same speed. To permit recording of complete PPI frames, the driving shaft is synchronized to rotate 360° during 420° of rotation of the PPI scan. Thus, each Permachon sees a complete PPI frame, but 60° of PPI rotation elapses between the end of one Permachon's view and the beginning of the next Permachon's view.

In the laboratory demonstration model a typical PPI rotation period of six seconds was chosen for convenience. The stepping period is then seven seconds, or one complete cycle of recording requires 42 seconds.

The one second per step lost for mirror motion is sufficient for erasure and complete recovery of the Permachon sensitivity.

Several other functions occur during the stepping period. Since the lenses of the cameras are always open, the PPI tube is electronically blanked off during the stepping period to prevent smearing of the departing and arriving images. This function is accomplished by means of a microswitch and cam on the driving shaft, to permit bracketing the entire stepping period.

Since the spurious video signal read out from the Permachon being erased and recovering would otherwise appear on the final monitor, provision is made to interrupt the video signal from that Permachon by means of another microswitch and cam on the driving shaft and a ring of six selecting microswitches driven by a cam rotating at one-sixth the speed of the shaft. This function will be further explained later.

Simultaneously with the recording and storage processes, the stored information is read out at a much faster rate into a viewing monitor. All the Permachons are scanned continuously, and the six preamplifiers feed video signals continuously into six cyclically-gated video amplifiers having a common output circuit. The electronic timing circuitry controls the sequence of video signals read out of the gated amplifiers. Fig. 5 is a photograph of the camera control circuitry on the table-top and the cyclically-gated video amplifier at the lower right.

In normal operation only one amplifier at a time may transmit information to the output circuits. The amplifiers are gated "on" in a continuous cyclic sequence controlled by the timing circuitry. Fig. 6 is a block diagram showing the signal flow. Fig. 7 shows the schematic details of the circuit. All timing is controlled by means of the standard 60 pps vertical driving pulses from the television synchronizing-signal generator. The frequency is halved to provide frame-rate pulses, and these pulses are counted in a three-stage binary counter so designed that any preselected number of frames from one to eight may be counted. This preselection is based upon a principle our group developed about ten years ago and have used in many similar applications. The output pulse at the end of the counter is used to gate "on" momentarily a set of resetting tubes directly connected to the counter stages to preset these stages to the complement of the desired counting number for the "zero-th" count. This system permits counting by arbitrarily chosen numbers. In the present system the counter determines the number of frames each of the gated video amplifiers will transmit before it is switched "off" and the next amplifier is switched "on". The pulse from the counter is connected to six coincidence amplifiers in parallel. The other event of the coincidence is the "on-gate" of one of the six bistable gates connected to the video amplifier. The output of the coincidence tube switches its own gate "off", which, in turn, switches the succeeding gate "on" and prepares the associated coincidence tube for the next switching pulse.

When the counter is first energized, it is possible for several gates to start in the "on" condition. Therefore a "lockout" pulser has been provided to switch all other gates "off" whenever gate "A" is switched "on". This does not affect the gates in normal operation and ensures single-gate video signals. In the very unlikely event that all gates should start in the "off" condition, a momentary push-button switch will turn gate "A" "on".

By this counting and gating system the reading cycle may be selected from 6/30 to 48/30 seconds. The time-compression ratio may thus be between 210 and 26¼. The best ratio for greatest conspicuousness of the moving signals may be selected at will by the operator at any time during operation.

The sequentially-gated video signals are then further amplified, blanking and synchronizing signals are added, and the composite signal is displayed on a TV monitor.

Several methods have been investigated for preventing the signal from the Permachon being erased from interrupting the desired display. Short-circuiting the output from the selected preamplifier produced a bright field that diluted the

contrast of the total display. Another method, the circuitry for which is still included in the system and on the diagram, provided a synchronized reestablishment of the black level in the composite video. In contrast to the previous system, this method produced a blackout of the video signal from the selected channel, which was visually unpleasant. The circuitry is retained, however, because it permits selecting combinations of particular channels to be displayed for sweep, position, and signal line-up purposes.

The method currently in use involves generating an additional trigger pulse to advance the gate through one channel whenever there is a coincidence between any reading gate and another gating signal derived from one of the six cam-operated microswitches associated with the driving shaft, as mentioned earlier. The video signal from the erased camera is then skipped without display until the coincidence period is completed. During this period there will appear only five spots for each moving target, instead of the usual six.

In order to be able to test and demonstrate this display system without reliance on an actual operating radar, a simulator has been assembled. A set of simulated targets is generated by small fluorescent-painted tabs moving on a string around a pulley system in a black box. Background signals are simulated by other, stationary fluorescent tabs in random groups on the background in the box. The tabs are illuminated by means of ultraviolet lamps. This scene is viewed by a vidicon television camera and is transmitted in a closed TV system to the PPI tube in the opto-mechanical camera assembly, as well as to a comparison TV monitor with a long-persistence P-7 screen. The PPI scan is simulated by rotating a circumferentially-driven black disc with a narrow wedge-shaped aperture at the front of the black box.

While the detection of target motion on the real-time monitor was very difficult, whether with PPI or TV-type scan, the motion of the moving targets was very conspicuous on the compressed-time monitor with either type of display.

Because of the attractiveness of the Permachon principle for storage tube application, the U. S. Army Signal Research and Development Laboratory has contracted with Westinghouse to develop an electrical-input, electrical-output, scan-converter storage tube based upon a Permachon-type storage target. When tubes of this type become available, the time-compression display system could be converted to an all-electronic system of much greater convenience and compactness.

Potential Applications

The motion-enhancement display is expected to be particularly advantageous in the detection of aircraft flying at a low altitude in order to be lost in ground clutter. A further military application would be to increase the detectability of aircraft by long-range radar, where the real-time motion of the return signal on the display might be poorly visible. In addition to the predominantly military applications of this type of display, an important application is expected in air-traffic control, where both direction of motion and speed would be instantly apparent. With the addition of altitude discrimination, as might be obtained with a color presentation, a rather complete picture of air traffic in a certain block of air space could be obtained.

Acknowledgements

The author wishes to acknowledge the invaluable efforts of his co-workers on this experiment: Dr. Friedrich Dinhobel and Mr. Oliver Reynolds, both formerly of USASRDL, and Mr. Orville Harper, currently associated with this experiment. The assistance of Westinghouse Electric Corporation in providing the Permachons used in the system is further acknowledged.

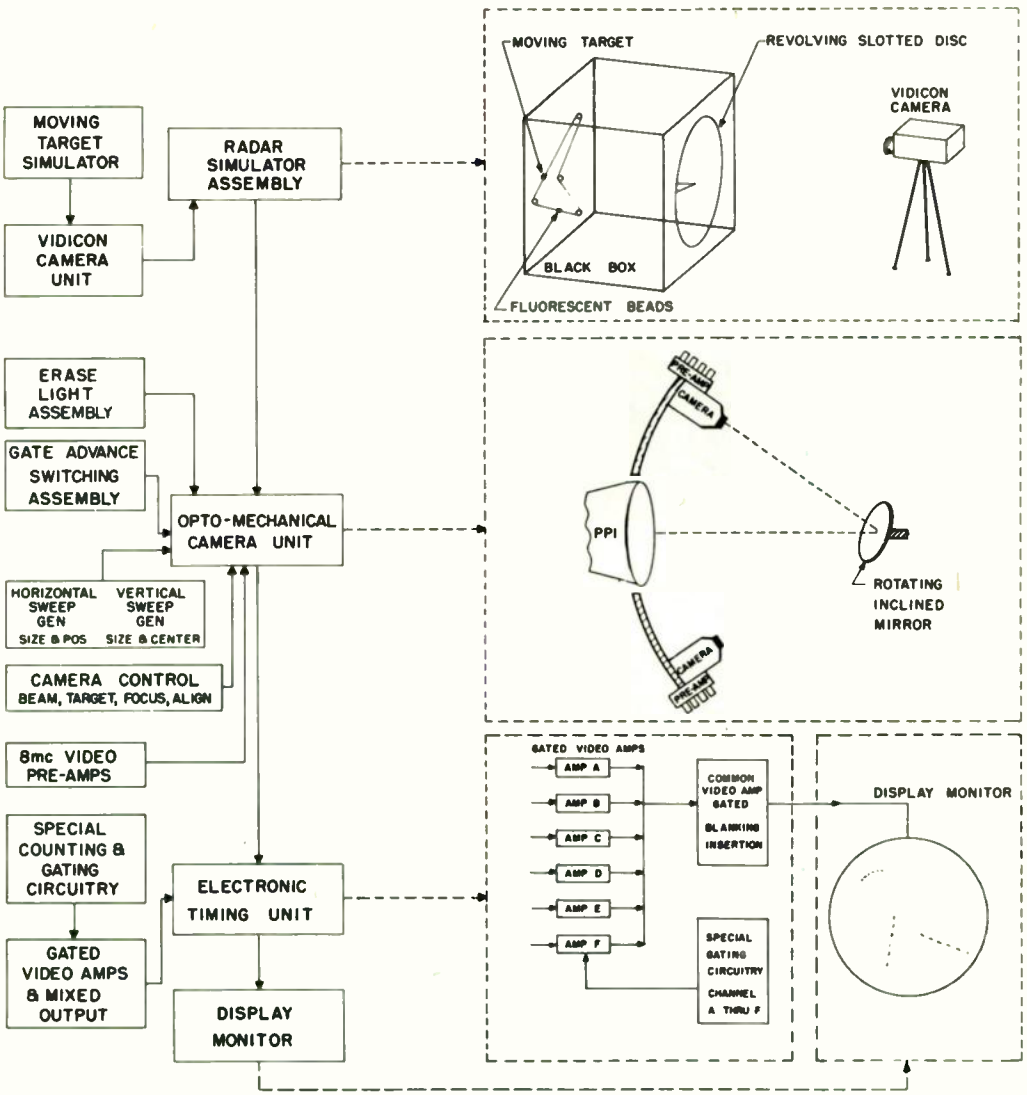


Fig. 1. System block diagram of time-compression display.

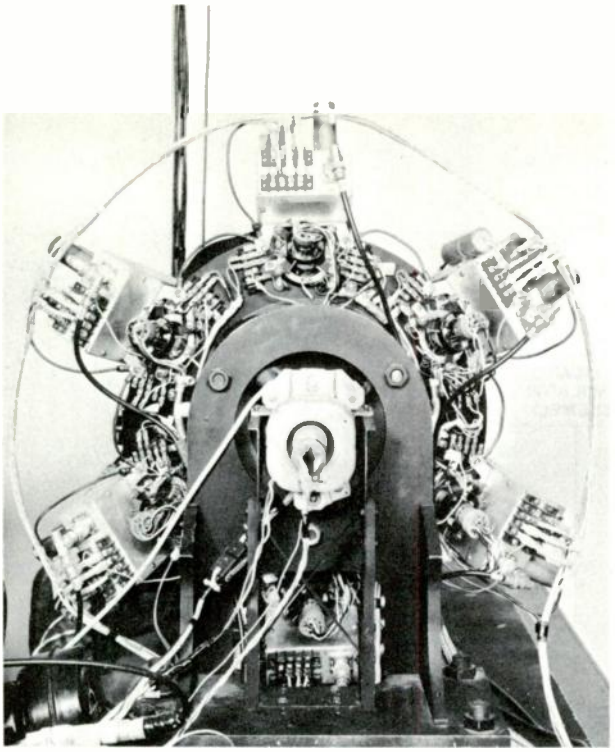


Fig. 2. Rear view of opto-mechanical camera assembly.

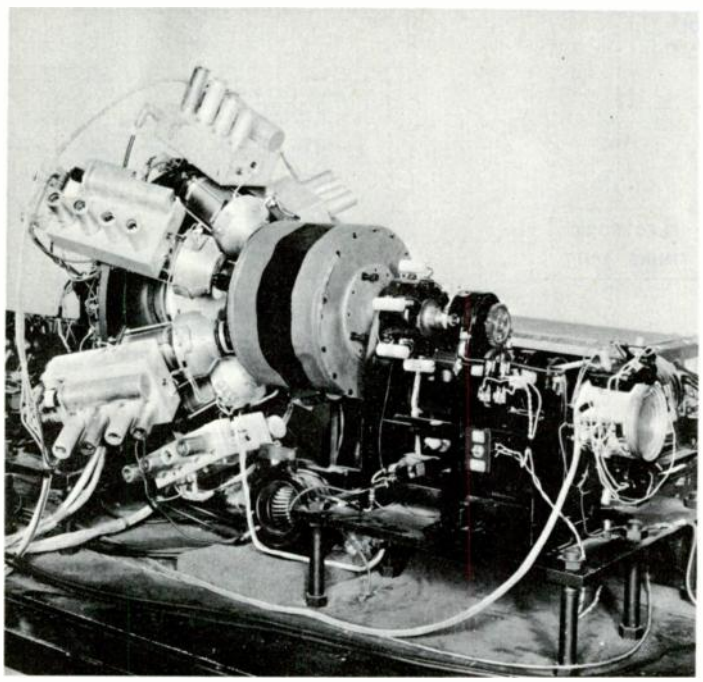


Fig. 3. Oblique side view of opto-mechanical camera assembly.

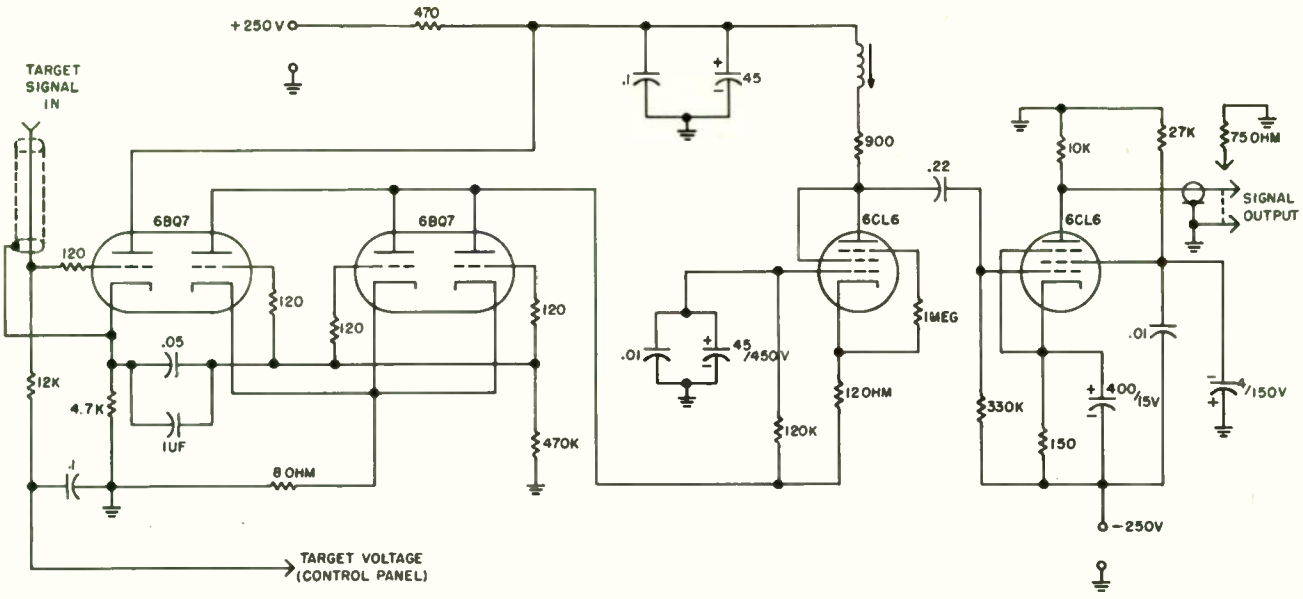


Fig. 4. Video preamplifier schematic diagram.

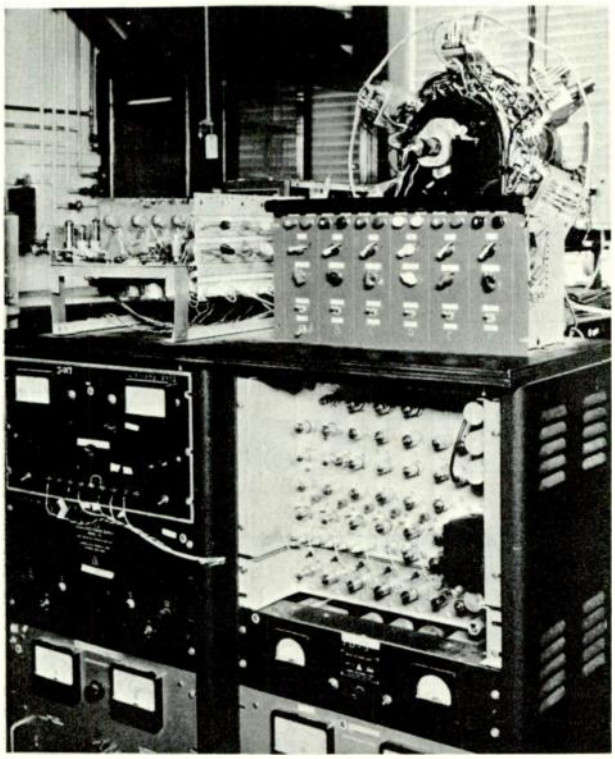


Fig. 5. Camera controls and gated video amplifiers.

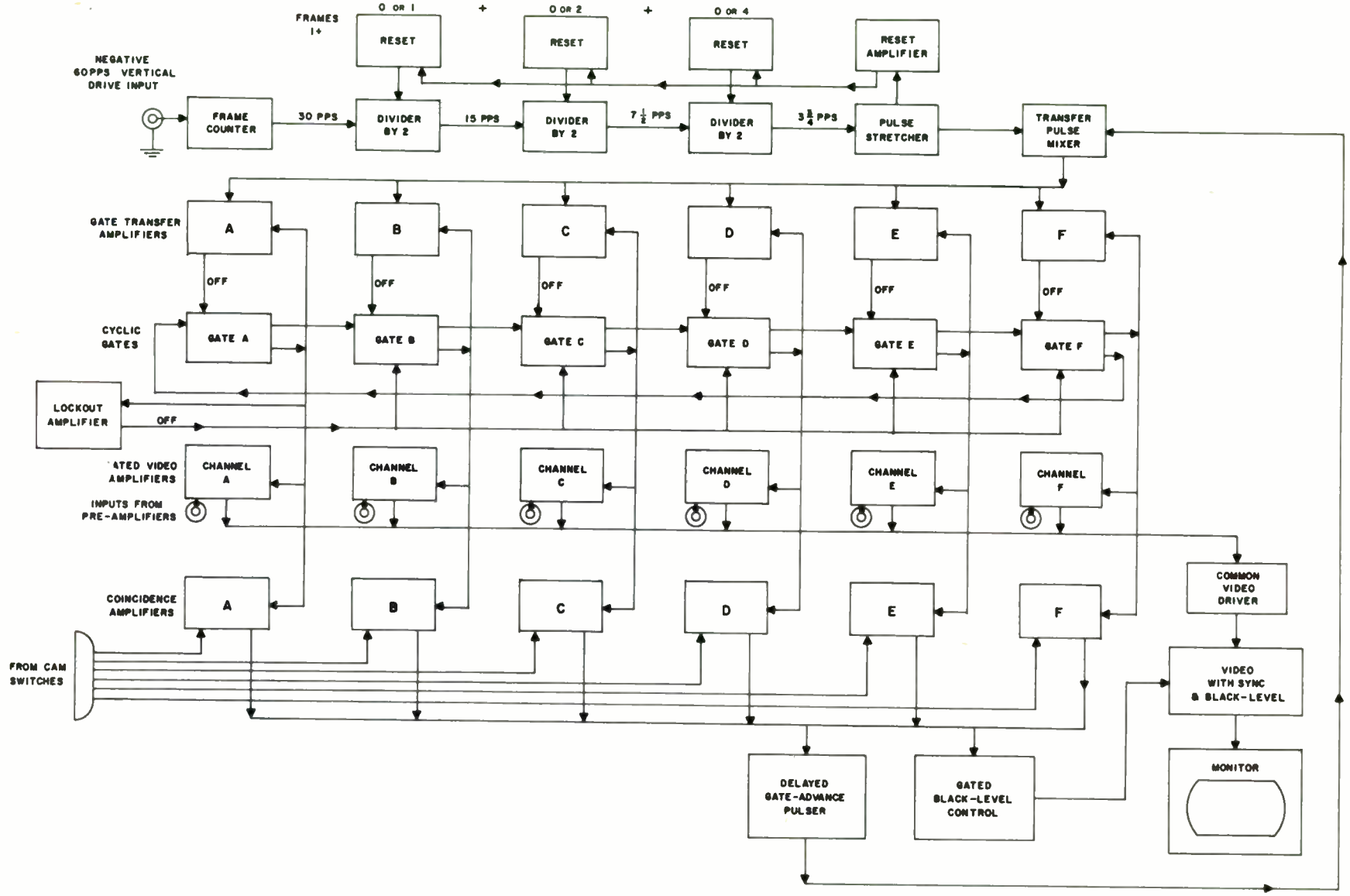


Fig. 6. Cyclic gated video amplifier.

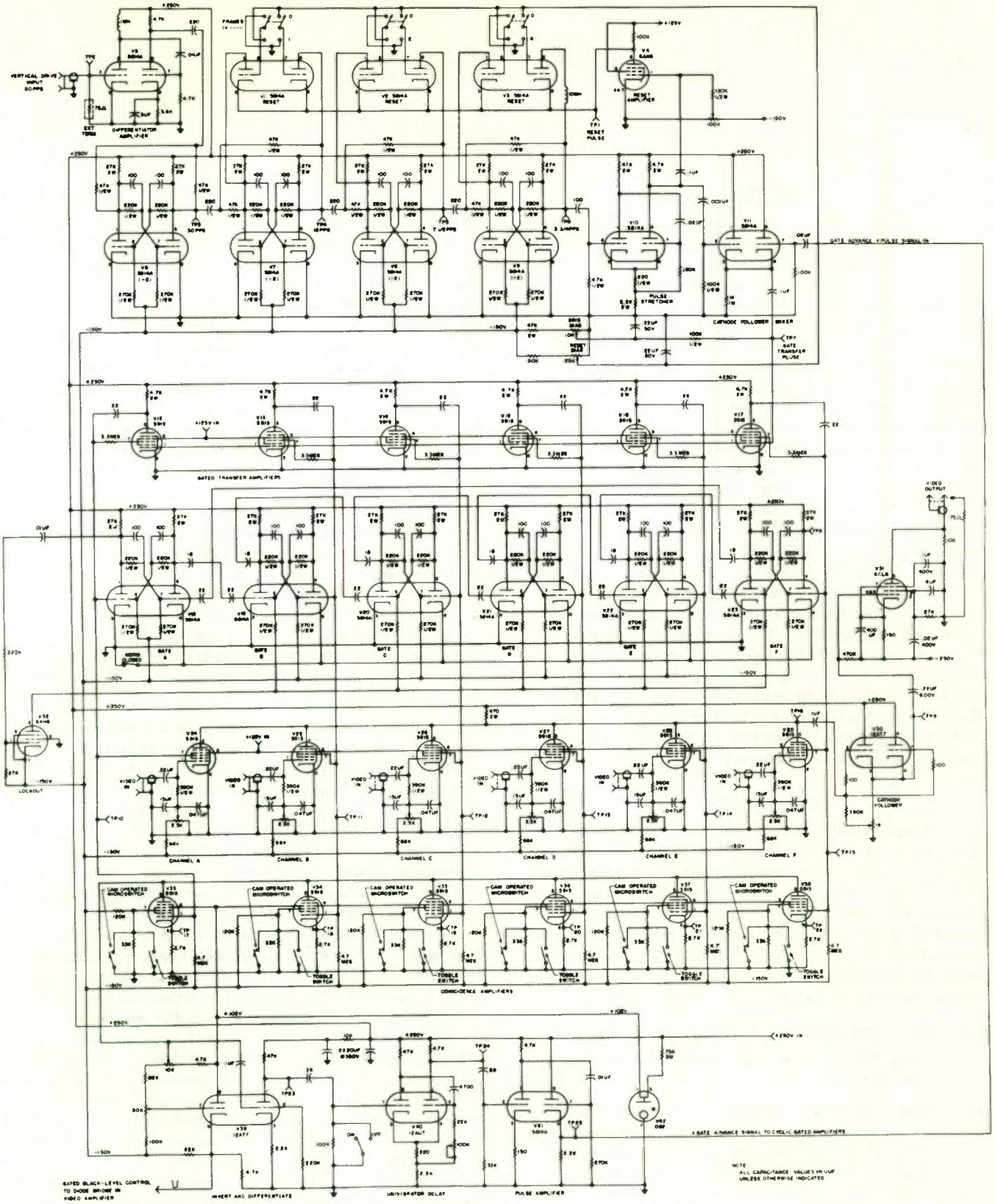


Fig. 7. Cyclic gated amplifier schematic diagram.

FREQUENCY MODULATION TECHNIQUES AS APPLIED TO PULSE DOPPLER RADAR

George R. Hetrich*
Electronic Communications, Inc.
Timonium, Maryland

Summary

Two methods of frequency modulating the transmitted carrier for ranging purposes are considered for a typical Pulse Doppler radar. The two ranging methods considered are linear FM and sinusoidal FM. An airborne environment is assumed, and therefore, the choice of FM parameters is an all important consideration due to their effects on the received ground clutter spectrum. Analogous comparisons are formulated between certain selected system characteristics for the two ranging methods. From these comparisons intelligent decisions can be rendered as to the relative merits and disadvantages of the two ranging methods for various applications. The selected system characteristics are velocity accuracy and coverage and range accuracy and coverage. In general, each particular application dictates a peculiar set of system requirements, and thus one ranging scheme may be most useful for one application and least useful for some different application. For the system characteristics selected for discussion it appears that sinusoidal FM ranging is better suited for surveillance and search applications and for utilization against high performance targets. Linear FM ranging is better suited for tracking and multiple target applications.

Introduction

Historically, Pulse Doppler radar is defined as a frequency coherent system in which a high pulse repetition frequency (prf) is employed in order to prevent velocity ambiguities and to enable enhanced range performance without the expense of transmitters with extremely high peak-to-average power ratios[1,2]. Unambiguous velocity performance is not mandatory for stationary radar platform applications, but if the radar is airborne and has a relative motion with respect to a large segment of clutter scatterers, such as the earth, then unambiguous velocity information is almost a necessity. The motion of the radar platform causes spreading of the clutter echoes over the frequency spectrum. The clutter spreading occurs around each of the transmitted sidebands causing some of the velocity coverage region to be obscured by the clutter. This predicament is most effectively overcome by operating with a high prf which enables positioning of all of the sidebands outside of the velocity coverage region of interest. See Figures 1 and 2 for clarification.

This problem of frequency-spread clutter does not exist for the stationary radar platform. The ground returns for the stationary radar are seen as discrete spectral lines and may easily be rejected with sharp cutoff filters or notches. The sharp cutoff filters minimize the loss of Doppler coverage. Thus, a stationary pulse Doppler radar can operate practically with a low prf provided no additional modulation or carrier coding is used.

It is possible to conceive of a frequency coherent pulse radar in which the prf is low enough to cause all sidebands to fall close to the carrier. This would eliminate the problem of MTAE or range ambiguity. The Doppler coverage would lie outside of the clustered sidebands as shown in Figure 3. The Doppler information would also be unambiguous. In the region of Doppler coverage the received prf sidebands would have power levels below the receiver noise level, thus preventing degradation in system performance. Under these conditions the predetection bandwidth would have a minimum bandwidth limitation of approximately one over the transmitted pulse width for optimum performance. In most practical applications this minimum filter bandwidth would be too wide to permit decent velocity resolution. The utilization of the minimum predetection bandwidth is indicative of a post detection bandwidth-to-predetection bandwidth ratio of unity thus enabling a relatively simple mechanization.

The ranging problem is greatly simplified in the low prf coherent system. Sequentially time gated receiver channels are positioned within the interpulse period which enables rapid unambiguous range interrogation with slightly less range accuracy than a similar type high

---

* Formerly with Westinghouse Electric Corporation, Air Arm Division, Baltimore, Maryland.

1. Povejsil, D.J., Raven, R.S., and Waterman, P., Airborne Radar, Volume 8 of the series, "Principles of Guided Missile Design," D. Van Nostrand Co., Inc., Princeton, N.J., 1961.

2. Goetz, L.P., and Albright, J.D., Airborne Pulse-Doppler Radar, IRE Transactions on Military Electronics, New York, N.Y., April, 1961.

prf MTAE system. Unlike the high prf systems the coherent low prf radar is mechanized such that the energy from each transmitted pulse which is reflected from targets and clutter scatterers at all conceivable ranges is received before any subsequent pulses are transmitted. Thus, none of the range ambiguity problems exist in this type system as they do in the high prf system.

If additional modulation is used such as linear FM, sinusoidal FM, or possibly pseudo-random noise modulation of the pulsed carrier, then the operating characteristics of a ground based Pulsed Doppler radar are not much better than the airborne radar with respect to clutter spreading. The additional modulation is sometimes desirable for certain applications. The prospect of using noise modulation, delaying the modulating waveform, and comparing the delayed waveform with the received echoes has become popular in recent years as an effective ECCM feature because of the high information content enherent in this type of system.

It is the purpose of this paper to briefly review the operation of a Pulse Doppler radar which has no carrier FM, to examine the effects of linear carrier FM and sinusoidal carrier FM on the certain performance characteristics of a typical high prf Pulse Doppler radar, and to make systematic comparisons of these performance characteristics for each of these two methods of modulation. An airborne environment is assumed throughout the remainder of the paper. The selected performance characteristics are velocity accuracy and coverage and range accuracy and coverage.

Pulsed Operation Only

A conventional airborne Pulse Doppler radar transmits a spectrum as partially shown in Figure 4 where f_o is the carrier frequency, and prf is the pulse repetition frequency. The received ground clutter spectrum is also shown. The envelope of the lines has approximately a sin x/x shape with the positions of the nulls depending on the duty cycle. Echoes from moving targets are shifted in frequency due to the Doppler effect and are detected in the region of Doppler coverage which is synonymous with velocity coverage. For a closing target the Doppler shift is upwards, and typically the received spectrum from such a target is shown by the dotted lines.

The frequency band of interest lies between the f_o and f_o + prf lines. A contiguous filter bank in the receiver is centered in this band. Thus, for closing targets, echoes appear in the filter bank from the f_o line and for opening targets, from the f_o + prf line. To prevent echoes from other lines from appearing in the filter bank and to separate opening and closing target returns in frequency, the lower limit on the prf is given by Equation 1.

$$prf_{min} = \frac{2v_c}{\lambda} \qquad (1)^*$$

If the radar platform is in motion, the minimum prf is further increased due to the Doppler shift of ground clutter echoes[3]. This increase is required to maintain frequency separation of target and clutter echoes.

The reciprocal prf gives the interpulse period which is related to the maximum unambiguous target range by Equation 2.

$$R_u = \frac{cT}{2} \qquad (2)$$

For rf wavelengths and targets traveling several hundred miles per hour, the unambiguous range for a single prf is only a fraction of a mile. Returns from targets at greater ranges cause ambiguities. Thus, a return appearing in the interpulse period t units of time after the transmitted pulse may actually be at any of the ranges given by Equation 3.

$$R_x = \frac{c(t+xT)}{2} \qquad (3)$$

where

x = 0, 1, 2, 3, . . .

R_x = possible target range in units of length.

t = apparent time between transmitted and received pulses in units of time.

Equation 3 reveals mathematically the existence of range ambiguities. Some conventional Pulse Doppler radars have utilized multiple prf's as the method for resolving the range ambiguities. Multiple prf ranging has been amply discussed elsewhere in the literature[4] and is not the primary subject of discussion in this paper. Other methods for performing the same task require carrier frequency modulation. Discussion of two of these other methods, linear FM and sinusoidal FM, appear in succeeding sections of this paper.

---

* See the Appendix for the definitions of symbols unless otherwise defined in the paper.

3. Coleman, S.D. and Hetrich, G.R., Ground Clutter and Its Calculation for Airborne Pulse Doppler Radar, Conference Proceedings, 5th National Convention on Military Electronics, Washington, D.C., June, 1961.

4. Skillman, W.A. and Mooney, D.H., Multiple High-PRF Ranging, Conference Proceedings, 4th National Convention on Military Electronics, Washington, D.C., June, 1960.

Sinusoidal FM Ranging

Rather than use the conventional multiple prf method to solve the problem of target range ambiguity, the pulse carrier can instead be sinusoidally modulated in frequency[5]. The transmitted waveform is then represented by Equation 4.

$$f(t) = p(t) \cos (\omega_o t + m_f \sin \omega_s t) \qquad (4)$$

where

$p(t)$ = the video pulse waveform
$= 1$ $\quad nT < t < (n+d)T$
$= 0$ $\quad (n+d)T < t < (n+1)T$
$n = 0, \pm 1, \pm 2, \pm 3, \ldots$

The transmitted spectrum consists of the carrier frequency and sidebands spaced at intervals of the prf on either side of the carrier plus FM sidebands around the carrier and all of the prf sidebands as shown in Figure 5. The spacing of the FM sidebands are at intervals of the frequency f_s as defined for Equation 4.

A typical block diagram is shown in Figure 6. A detailed explanation of Figure 6 is unimportant. Target range is a function of the phase difference between the sinusoidal waveform which frequency modulates the transmitted carrier and the sinusoidal waveform of the frequency modulated echo. These two sinusoidal waveforms appear at the input to the phase detector. The carrier and echo frequencies are different due to the Doppler effect; however, their modulating waveforms have nearly identical frequencies* but different phases. To prevent range ambiguity and to facilitate signal processing, the modulating frequency must be small, i.e., $\omega_s << c/R$ which permits the following trigonometric approximation.

$$\sin 2\pi f_s \frac{R}{c} \approx 2\pi f_s \frac{R}{c}$$

For $\omega_s << c/R$ the phase detector can be made reasonably linear, i.e., E_{out} in Figure 6 varies almost linearly with the preceding mentioned phase difference.

The energy in the central spectral line, f_o, and its FM sidebands is of most importance, because the receiver is mechanized to extract only this energy after reflection from a target. Unless the system is operating in a clutter-free environment, the reflection must occur from a moving target, and as in the conventional un-FM-modulated method of operation, the echo from the central spectral line and its FM sidebands are shifted by equal amounts in frequency due to the Doppler effect.

For the particular mechanization shown in Figure 6 the predetection filter or velocity gate must be wide enough to pass the Doppler shifted carrier and some of its FM sidebands. These FM sidebands must be above the input receiver noise power density level in order to accurately measure the phase difference between the transmitted FM waveform and the received FM waveform. The phase refers to the $m_f \sin \omega_s t$ term in Equation 4. This measurement is necessary in order to determine the target range.

It is conceivable that range information could be had by filtering only the central spectral line in the time delayed and Doppler shifted frequency modulated echo and measuring its relative phase. This mechanization, however, would require a highly selective filter since the FM sidebands would have to be rejected. The first sidebands are spaced ω_s radians from the central line. Generally, ω_s is a very small fraction of the intermediate frequency. While the highly selective filter method may possess the capability for an enhanced signal to noise ratio for a given target echo, the method proposed in Figure 6 appears as more feasible to mechanize.

The requirement that a finite number of FM sidebands be passed by the predetection filter places a minimum bandwidth limitation on the filter. This minimum bandwidth limits the accuracy of the velocity measurement. Thus, a relation exists between the velocity accuracy and the range coverage. The target velocity accuracy is given by Equation 5.

$$v_{acc} = \Delta f_d \lambda \qquad (5)$$

The accuracy of the Doppler shift measurement is approximately one quarter of a predetection filter bandwidth, BW_{PD}. The predetection filters possess finite slopes on their frequency response curves. Therefore, a single target causes a number of filters to have outputs, and by centroiding, the accuracy of the frequency measurement is about four times better than the unit filter bandwidth. In the literature[6] a criterion is established as to the minimum bandwidth requirement for an FM carrier. Using this criterion, $2n = (2\Delta F/f_s)$ sidebands must pass through the predetection filter to cause the accuracy of phase difference measurement to be limited by the phase detector. If less than 2n sidebands are passed through the filter, then the accuracy of the phase difference measurement is limited by the lack of a faithful replica of the input waveform due to too narrow a predetection filter bandwidth. The bandwidth

5. Saunder, W.K., Post-War Developments in Continuous-Wave and Frequency-Modulated Radar, IRE Transactions on Aerospace and Navigational Electronics, Volume ANE-8, Number 1, New York, N.Y., March, 1961.

* Their frequencies also differ slightly due to the Doppler effect. From the practical viewpoint this difference is negligible.

6. Goldman, S., Frequency Analysis, Modulation and Noise, McGraw-Hill Book Co., Inc., New York, N.Y., 1948, Page 150.

requirement for 2n sidebands is merely $BW_{PD} = 2n f_s = 2\Delta F$.

The range measurement is given by Equation 6.

$$R = \frac{c\phi_d}{2\omega_s} \tag{6}$$

Assume the maximum phase difference for a nominally linear phase detector is ϕ_{max} where

$$\phi_{max} << 2\pi$$

If the range coverage requirement is R_c, then the maximum permissible modulating frequency, ω_s, is given by Equation 7.

$$(\omega_s)_{max\ 1} = \frac{c\phi_{max}}{2R_c} \tag{7}$$

The maximum modulating frequency also has a limitation depending on the bandwidth of the predetection filter. Since 2n sidebands are assumed as the criterion for a faithful output replica of the input waveform, and since the sidebands are spaced f_s in frequency apart, the maximum modulating frequency is given by Equation 8.

$$(\omega_s)_{max\ 2} = \frac{2\pi BW_{PD}}{2n} = \frac{2\pi(4\Delta f_d)}{2n} = \frac{4\pi\Delta f_d}{n} \tag{8}$$

Equations 5, 7, and 8 are related in a manner which specifies the velocity accuracy for a given range coverage requirement. This interrelation is given by Equation 9 which includes the 2π conversion factor between radians per unit time and cycles per unit time.

$$v_{acc} = \frac{\phi_{max}\, n\lambda c}{8\pi R_c} \tag{9}$$

Thus, for this particular mechanization v_{acc} varies inversely with R_c; i.e., velocity accuracy is enhanced as the range coverage is increased, since a small v_{acc} is indicative of good velocity accuracy.

The received FM clutter spectrum for a moving radar platform looks like a widened and distorted replica of Figure 4. An extremely rough model of this sinusoidal FM clutter spectrum can be derived for the purpose of setting limits on the velocity coverage. For a moving radar it may be assumed that the un-FMed clutter spectrum shown in Figure 4 consists of many carriers of various amplitudes and phases depending on the antenna gain pattern, radar altitude, reflectivity coefficient, and type of scatterers. These carriers are close together in frequency and, as a composite, form the FM clutter spectrum. Further assume that the maximum amplitude of any one received clutter carrier is 60 db above the receiver noise level and, to be pessimistic, that two 60 db carriers occur one at each end of the clutter spectrum as shown by Figure 7. Actually, this condition would never occur in practice.

When sinusoidal FM occurs, these maximum extremities in frequency and power of the preceding two clutter carriers give an approximate criteria on the available unambiguous Doppler coverage. Equation 4 may be rewritten as Equation 10 showing the Bessel functions as coefficients of the sidebands of an FMed carrier.

$$\begin{aligned} f(t) = p(t)\Big\langle & J_o(m_f)\cos\omega_o t \\ & - \sum_{y=1}^{\infty} J_{2y-1}(m_f)\Big\{\cos[\omega_o - (2y-1)\omega_s]t \\ & - \cos[\omega_o + (2y-1)\omega_s]t\Big\} \\ & + \sum_{z=1}^{\infty} J_{2z}(m_f)\Big\{\cos[\omega_o - 2z\omega_s]t \\ & + \cos[\omega_o + 2z\omega_s]t\Big\}\Big\rangle \end{aligned} \tag{10}$$

The terms of interest would be those which had Bessel functions greater than 0.001 or 60 db. The sinusoidal frequency or sideband which possessed the highest order Bessel function with a value greater than 0.001 would serve as an indication of the maximum clutter spreading and thus act as a criterion for determining the available Doppler and velocity coverage. This sideband would be referenced to the extreme points on the unmodulated clutter spectrum as shown in Figure 8. Thus, the clutter spectrum would appear in the receiver to have increased in width by twice the frequency of the above defined sideband.

For the case of no frequency modulation the available Doppler coverage is given by Equation 11.

$$v'_c = (prf)(\lambda) - 2v_R \tag{11}$$

where v'_c = the available velocity coverage in units of length per unit time for the no modulation case.

For the sinusoidal FM case the available Doppler coverage is given by Equation 12.

$$v_c = (prf)(\lambda) - 2v_R - 2j\, f_s \lambda \tag{12}$$

For the Bessel function approximately equal to 0.001 and a large ratio $\Delta F/f_s$, j tends to vary linearly with $\Delta F/f_s$[7]. This leads to the conclusion that the velocity coverage tends to be somewhat independent of variations in f_s. Since range coverage varies inversely with f_s as per Equation 7, there is only a second order or negligible relation between velocity coverage and range coverage.

7. Jahnke, E. and Emde F., Tables of Functions, Dover Publications, New York, N.Y., 1945.

By combining Equations 7 and 9 and recognizing that $n = \Delta F/f_s$ the velocity accuracy is seen to vary directly with the amount of frequency deviation as shown by Equation 13.

$$v_{acc} = \frac{\Delta F \lambda}{2} \tag{13}$$

It is possible to relate velocity coverage and accuracy by comparing Equation 12 with 13; however, in the interest of uniformity this is not done. Instead velocity accuracy and coverage are related from the viewpoint of equipment with constant size and weight. This viewpoint is assumed in the linear FM discussion, and therefore, uniformity is preserved by making the same assumption for the sinusoidal FM discussion.

If a contiguous filter bank is used entirely over the region of Doppler coverage, the number of filters varies directly with the velocity accuracy and coverage requirements. If size and weight are held constant, then the velocity accuracy varies with velocity coverage as follows.

$$v_c = 4m v_{acc} \tag{14}$$

The factor m is the number of predetection filters covering the Doppler band. Equation 14 assumes the filters are spaced over the frequency band at intervals equal to their bandwidths. As the velocity accuracy improves the velocity coverage diminishes and vice versa.

The range accuracy as stated previously is ultimately limited to the accuracy of the phase detector provided the design procedure is followed as outlined; i.e., provided the FM spectrum of the target echo is reasonably well preserved when it reaches the input to the phase detector. Therefore, for a given phase detector, the range accuracy would be fixed percentage of the range coverage or range accuracy would vary directly with the range coverage. As the range coverage is increased, the absolute range accuracy will suffer and vice versa. Since the velocity accuracy varies inversely with the range coverage as per Equation 9, it is obvious that the range accuracy also varies inversely with the velocity accuracy; i.e., as the range accuracy is improved, velocity accuracy suffers and vice versa. No relationship exists between range accuracy and velocity coverage, because as indicated previously, the range coverage is practically independent of the velocity coverage, and range accuracy is dependent solely on the range coverage.

The preceding analysis has considered the relationships between range accuracy and coverage and velocity accuracy and coverage for a specific form of mechanization of sinusoidal FM ranging as shown in Figure 6. The results are conveniently tabulated in Figure 12 of the conclusion for ease of comparison with the linear FM ranging method.

Linear FM Ranging

Linear FM ranging in its simplest form consists of transmitting with a single high pulse repetition frequency and periodically ramp modulating the pulsed carrier as shown by Figure 9. The dotted lines of Figure 9 represent the target echoes. The time scales are not meant to be equal; i.e., there are many more transmitted pulses for each linear FM cycle. The receiver is interrogated at least twice during each linear FM cycle, once during the non-FM portion and once during the FM portion of the cycle. A short receiver dead time is required twice during each FM cycle because of a transient condition due to the transit time of the radiated energy. During this transient period the receiver local oscillator is synchronized with the transmitted carrier but is not synchronized with residual energy being received from previously transmitted waves. Once the maximum "out-and-back" time has elapsed, synchronization occurs, and the receiver can be energized.

The receiver local oscillator is synchronized with the transmitted carrier frequency so that after synchronous demodulation, time gating, and filtering over the region of Doppler coverage, the target signal appears as a fixed frequency tone throughout each half of the linear FM cycle. During the non-FM half the signal frequency is a function of the target velocity, while during the FM half the signal frequency is proportional to the algebraic sum of the target velocity and range. Demodulated returns from targets with high velocities are shifted higher in frequency, while returns from targets with large ranges are shifted lower in frequency during the FM mode. Thus, target velocity is obtained directly by noting the frequency of the return in the non-FM half of the cycle and inserting this value in Equation 15.

$$v = F_{d1} \lambda \tag{15}$$

Target range is obtained by Equation 16.

$$R_{FM} = \frac{c(F_{d1} - F_{d2})}{2 k_{fm}} \tag{16}$$

The linearly swept carrier causes clutter spreading of the ground echoes. This fact coupled with the frequency difference of the returns in the FM and non-FM modes seriously limits the maximum tolerable sweep rate, which in turn, limits the accuracy of the range measurement. Range accuracy is a function of the carrier sweep rate and the frequency resolution capability of the filter bank.

A typical block diagram is shown in Figure 10. Figure 10 is similar to Figure 6, and again, a detailed explanation is unimportant, since the method of extracting range has already

been discussed.

No analogous relationship between velocity accuracy and range coverage exists for linear FM as it does for sinusoidal FM. Velocity accuracy is a function of the accuracy of the Doppler shift measurement which, as indicated previously, is about one quarter of a predetection bandwidth. The range coverage is dependent on the amount of tolerable frequency shift in the target echo due to the linear FM.

A relation does exist, however, between the range coverage and the velocity coverage and/or frequency coverage. Before this relation can be realized; however, the effects of linear FM on the clutter spectrum must be examined. The effects of linear FM on the clutter spectrum influence the amount of available velocity coverage in the FM mode.

For an airborne moving radar the received clutter spectrum in the FM mode appears as a distorted replica of Figure 2. The clutter spectrum is shifted downward in frequency for an upward sweep in the transmitted carrier. The extreme downward position of the clutter spectrum would occur from clutter power received from the ground on the horizon and to the rear of the direction of travel of the radar platform. This would be maximum opening clutter power and the position of this extreme downward shift with respect to zero doppler is given by Equation 17.

$$f_{c\ max\ opening} = \frac{v_R}{\lambda} + \frac{2k_{fm}(2rh)^{1/2}}{c} \qquad (17)$$

where the first term results from the velocity of the radar platform, and the second term results from sweeping the carrier. The factor $\sqrt{2rh}$ is an approximation from the literature[8]. The clutter returns in a given doppler increment which are shifted the least result from reflections from the ground in the immediate vicinity of y axis, as shown in Figure 11.

An expression for the position for this minimum shift with respect to zero doppler is given by

$$f_{c\ min} = \frac{v_R \cos\theta}{\lambda} - \frac{2hk_{fm}}{c \sin\theta} \qquad (18)$$

where θ is the angle between the line of radar platform travel and the line of sight range to the segment of ground under consideration as shown by Figure 11. The segment of ground must lie in the immediate vicinity of the y axis. A restriction on the angle θ is as follows

$$\sin\theta > \frac{h}{R_{horizon}} \qquad (19)$$

The largest positive value of the function $f_{c\ min}$ or the extreme upward position of the clutter spectrum, is determined by setting equal to zero the derivative of $f_{c\ min}$ with respect to θ as follows

$$\frac{df_{c\ min}}{d\theta} = -\frac{v_R}{\lambda}\sin\theta_o + \frac{2hk_{fm}}{c}\csc\theta_o \operatorname{ctn}\theta_o = 0 \qquad (20)$$

or

$$\frac{\sin^3\theta_o}{\cos\theta_o} = \frac{2\lambda h k_{fm}}{c} \qquad (21)$$

It is impossible to solve for θ_o in closed form in Equation 21, and an approximating method is necessary.

The frequency coverage in the FM mode is determined by combining Equations 17, 18, and 21 resulting in Equation 22.

$$f_c = prf - \frac{v_R}{\lambda}(1+\cos\theta_o) + \frac{2k_{fm}}{c}(h\csc\theta_o - \sqrt{2rh}) \qquad (22)$$

The equivalent velocity coverage in the FM mode is given by Equation 23.

$$v_c = f_c\lambda \qquad (23)$$

Thus, the decrease in velocity coverage in the FM mode is the difference in Equations 11 and 23. This difference is given by Equation 24.

$$v_{coverage,\ no\ FM} - v_{coverage,\ FM} \approx \frac{2k_{fm}}{c}(h\csc\theta_o - \sqrt{2rh}) \qquad (24)$$

A target at maximum range has a minimum velocity limitation because of the apparent frequency shift in the FM mode due to range. If the target has a velocity less than this minimum, its echo shifts into the clutter region. The echo from a target at maximum range shifts in frequency by the amount

$$F_{d1} - F_{d2} = \frac{2k_{fm}R_c}{c} \qquad (25)$$

The minimum velocity limitation is obtained from Equations 18 and 25.

$$v_{min} = (F_{d1} - F_{d2})\lambda - \frac{2hk_{fm}\lambda}{c\sin\theta_o} \qquad (26)$$

The last term in Equation 26 accounts for the shift in the extreme upward position of the clutter spectrum. Combining Equations 11, 18, and 25 results in a relation between the velocity coverage and range coverage.

8. Reference Data for Radio Engineers, Fourth Edition, International Telephone and Telegraph Corporation, New York, N.Y., 1956, Page 742.

$$v_c = prf\lambda - 2v_R - \frac{2k_{fm}\lambda}{c}\left(R_c - \frac{h}{\sin\theta_o}\right) \qquad (27)$$

Thus, as the range coverage requirement increases, the velocity coverage decreases and vice versa.

The accuracy of the linear FM range measurement is given by Equation 28.

$$R_{acc} = \frac{c\Delta f_d}{k_{fm}} \qquad (28)$$

The accuracy of the frequency measurement, Δf_d, is constant for both portions of the FM cycle. Equation 28 takes into consideration the factors contributed due to one way transmission and due to the necessity of making two frequency measurements in order to compute the range (one in the non-FM mode and one in the FM mode). As indicated previously, the accuracy in measuring the frequency shift of target echoes is about one quarter of the predetection bandwidth. The velocity accuracy is directly proportional to the frequency measuring accuracy as given by Equation 5. Therefore, the range and velocity accuracies are related by combining Equations 5 and 28 resulting in Equation 29.

$$R_{acc} = \frac{cv_{acc}}{k_{fm}\lambda} \qquad (29)$$

Thus, the velocity accuracy and range accuracy are directly proportional.

If a contiguous filter bank is used entirely over the region of Doppler coverage, the number of filters varies directly with the velocity accuracy and coverage requirements. Therefore, the relationship between v_c and v_{acc} is oriented towards the size and weight viewpoint. As mentioned previously, if size and weight must be held constant, then the velocity accuracy would vary with velocity coverage in accordance with Equation 14.

Combining Equations 14, 27, and 29 results in relations between range accuracy, range coverage, and velocity coverage as given by Equations 30 and 31.

$$R_{acc} = \frac{\text{constant}}{m} - \frac{R_c}{2m} \qquad (30)$$

$$R_{acc} = \frac{cv_c}{4mk_{fm}\lambda} \qquad (31)$$

The preceding analysis has considered the relationships between range accuracy and coverage and velocity accuracy and coverage for a specific form of mechanization of linear FM ranging as shown in Figure 10. The results are conveniently tabulated in Figure 12 of the conclusion for ease of comparison with the sinusoidal FM ranging method.

Conclusions

The sinusoidal FM and linear FM methods for ranging have been considered seperately. Relations between range accuracy and coverage and velocity accuracy and coverage have been developed for both ranging methods. Figure 12 shows a chart which may be useful for making analogous comparisons of the respective relationships for the two ranging techniques. These comparisons can materially aid in the decision as to which ranging method is most apt to fulfill a given set of requirements.

It is difficult to make a general statement that one modulation method is better than the other especially when some of relations cannot be expressed in a closed form. In addition, there are many variations in mechanization, and the two mechanization schemes discussed herein as shown by Figure 6 and 10 may not necessarily be ideal in their ability to extract radar information. Subsequent comparisons show that linear FM is better in some respects while sinusoidal FM is better in others.

Direct comparisons of corresponding columns and rows in Figure 12 are made as follows.

1. Column 2, Row 1 to Column 1, Row 2:
   Both FM methods are equally good. In each case as velocity coverage improves, velocity accuracy becomes poorer.
2. Column 3, Row 1 to Column 1, Row 3:
   Sinusoidal FM is better because of independence between velocity coverage and range coverage. Also, range coverage degrades with improved velocity coverage for linear FM.
3. Column 4, Row 1 to Column 1, Row 4:
   Sinusoidal FM is better because of independence between range accuracy and velocity coverage. Also, range accuracy suffers with improved velocity coverage for linear FM.
4. Column 3, Row 2 to Column 2, Row 3:
   Linear FM is better because of independence between range coverage and velocity accuracy. Sinusoidal FM may be justified if necessary, since velocity accuracy improves with range coverage.
5. Column 4, Row 2 to Column 2, Row 4:
   Linear FM is better because range accuracy improves with velocity accuracy. The opposite is true for sinusoidal FM.
6. Column 4, Row 3 to Column 3, Row 4:
   Linear FM is better, because range coverage improves as range accuracy improves. The opposite is true for sinusoidal FM.

The preceding comparisons result in the following general conclusions. Sinusoidal FM ranging is best suited for surveillance and search applications or for utilization against high performance targets. This is true, because sinusoidal FM ranging produces the best performance characteristics in velocity and range coverage. These coverages are the most significant parameters to consider for the above applications.

Linear FM ranging is best suited for tracking and multiple target[9] applications. This is true, because linear FM ranging produces the best performance characteristics in velocity and range accuracy. Accuracy would tend to be the most significant parameter to consider for the above applications. High accuracy is somewhat synonymous with high resolution or at least high resolution would require high accuracy. Therefore, linear FM ranging would be better suited for multiple target applications.

The above conclusions are not necessarily paramount as criteria for selecting one modulation method over the other. For example, linear FM requires two, time discrete, frequency measurements to determine target range and velocity. Sinusoidal FM requires two measurements; one, the Doppler frequency, and the other, the relative phase of the FM on the carrier echo; however, these measurements are obtainable simultaneously. Thus, from this respect sinusoidal FM may be more attractive for certain applications where time is of the essence.

Reliability, feasibility, practicalness, and cost are just a few of the many other considerations which are not included in the above conclusions. The attempt has been to apply certain theoretical performance characteristics obtained from given modes of mechanization in arriving at the conclusions as to the better modulation method for various applications. These conclusions are reached by recognizing only certain selected considerations such as size, weight, and performance requirements. Inclusion of other considerations may justify alterations in some of these conclusions.

9. Hetrich, G.R., Linear FM Ranging and Its Application to Pulse Doppler Radar, Eighth East Coast Conference on Aerospace and Navigational Electronics, Baltimore, Maryland, October, 1961.

Appendix
Definition of Symbols

- prf - Pulse repetition frequency in units of pulses per unit time.
- f_o - Carrier frequency in units of cycles per unit time.
- v_c - Velocity coverage in units of length per unit time.
- λ - Transmitted wavelength in units of length.
- R_u - Maximum unambiguous target range in units of length.
- c - Velocity of propagation in units of length per unit time.
- T - Reciprocal of pulse repetition frequency in units of time.
- d - Duty cycle.
- m_f - Modulation index = $\Delta F/f_s$.
- ΔF - Frequency deviation in units of cycles per unit time.
- f_s - Frequency deviation rate in units of cycles per unit time.
- f_{IF} - Intermediate frequency in units of cycles per unit time.
- R - Range in units of length.
- V_{acc} - R.M.S. velocity accuracy in units of length per unit time.
- R_{acc} - R.M.S. range accuracy in units of length.
- R_c - Range coverage in units of length.
- Δf_d - R.M.S. frequency measuring accuracy in cycles per unit time ($= BW_{PD}/4$).
- BW_{PD} - Predetection bandwidth or velocity gate in units of cycles per unit time.
- n - Number of FM sidebands.
- ϕ_d - Phase difference between two FM waveforms in radians.
- ϕ_{max} - Maximum phase difference for nominally linear phase detection.
- v_R - Velocity of radar platform in units of length per unit time
- j - Unitless whole number which refers to the j^{th} sideband; i.e., the highest order, sideband with a Bessel function coefficient greater than 0.001 is defined as the j^{th} sideband.
- k_{fm} - Rate of FM sweep in cycles per unit time squared.
- F_{d1} - Frequency shift of target echo with no linear FM in cycles per unit time.
- F_{d2} - Frequency shift of target echo in the linear FM mode in cycles per unit time.
- h - Radar platform altitude in units of length.
- $f_{c\ max\ opening}$ - Frequency of maximum opening clutter power in cycles per unit time.
- v'_c - Velocity coverage for no FM in units of length per unit time.
- m - Number of predetection filters covering Doppler band.

$f_{c\ min}$ - Minimum frequency shift of a given segment of clutter spectrum due to linear FM in cycles per unit time.

θ_o - Critical angle between line of radar travel and line of sight range to the segment of ground contributing the least amount of frequency shift in clutter due to linear FM and radar motion in degrees.

$R_{horizon}$ - Line of sight range to the earth's horizon from radar in units of length.

r - Radius of the earth.

Bibliography

1. Povejsil, D.J., Raven, R.S, and Waterman, P., Airborne Radar, Volume 8 of the series, "Principles of Guided Missile Design," D. Van Nostrand Co., Inc., Princeton, N.J., 1961.
2. Goetz, L.P. and Albright, J.D., Airborne Pulse-Doppler Radar, IRE Transactions on Military Electronics, New York, N.Y., April, 1961.
3. Coleman, S.D. and Hetrich, G.R., Ground Clutter and Its Calculation for Airborne Pulse Doppler Radar, Conference Proceedings, 5th National Convention on Military Electronics, Washington, D.C., June, 1961.
4. Skillman, W.A. and Mooney, D.H., Multiple High-PRF Ranging, Conference Proceedings, 4th. National Convention on Military Electronics, Washington, D.C., June, 1960.
5. Saunders, W.K., Post-War Development in Continuous-Wave and Frequency Modulated Radar, IRE Transactions on Aerospace and Navigational Electronics, Volume ANE-8, Number 1, New York, N.Y., March, 1961.
6. Goldman, S., Frequency Analysis, Modulation and Noise, McGraw-Hill Book Co., Inc., New York, N.Y., 1948, Page 150.
7. Jahnke, E. and Emde, F., Tables of Functions, Dover Publications, New York, N.Y., 1945.
8. Reference Data for Radio Engineers, Fourth Edition, International Telephone and Telegraph Corporation, New York, N.Y., 1956.
9. Hetrich, G.R., Linear FM Ranging and Its Applications to Pulse Doppler Radar, Eighth Annual East Coast Conference on Aerospace and Navigational Electronics, Baltimore, Maryland, October, 1961.

References Not Cited

1. Woodward, P.M., Probability and Information Theory, McGraw-Hill Book Co., Inc., New York, N.Y., 1953.
2. Ore, O., Number Theory and Its History, McGraw-Hill Book Co., Inc., New York, N.Y., 1948.
3. Maguire, W.W., Application of Pulsed Doppler to Airborne Radar Systems, National Conference Proceedings on Aeronautical Electronics, Dayton, Ohio, May, 1958.
4. Bussgang, J.J., et al, Detection Range of CW and Pulse Doppler Radars, National Conference Proceedings on Aeronautical Electronics, Dayton, Ohio, May, 1958.
5. Weinberg, N.L., Walker, J.M., and Metz, R.J., Rapid Detection of Coherent Signals in Noise, Convention Record, Part 8, IRE National Convention, New York, N.Y., March, 1960.

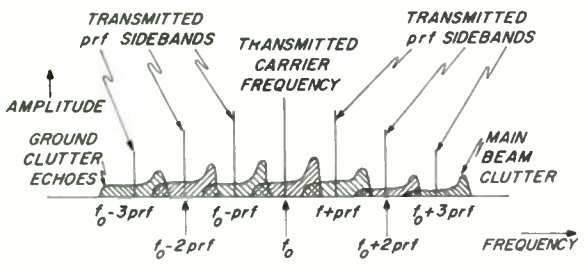


Fig. 1. Spectrum for low prf moving radar platform.

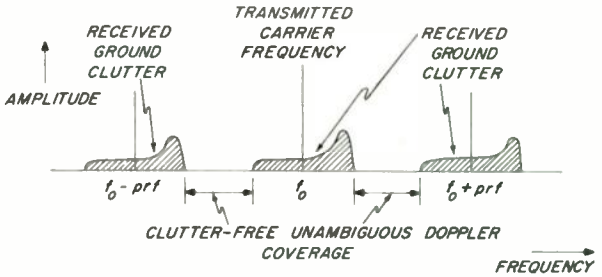


Fig. 2. Spectrum for high prf moving radar platform.

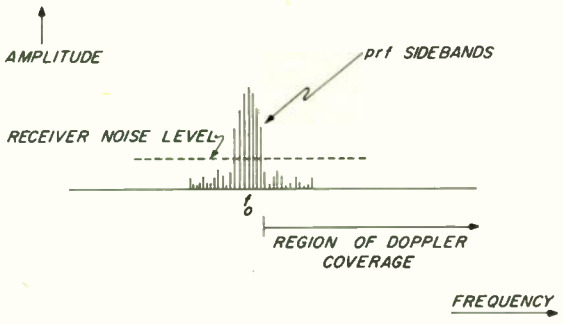


Fig. 3. Spectrum for coherent low prf operation.

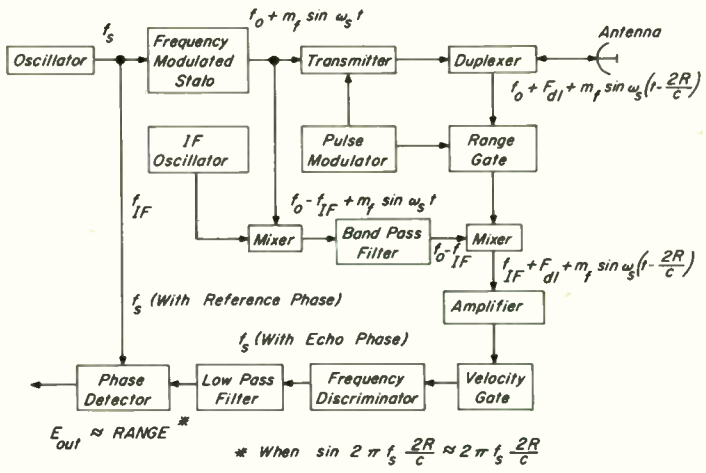


Fig. 6. Block diagram for sinusoidal FM.

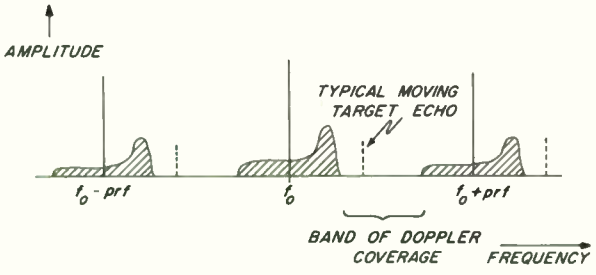


Fig. 4. Transmitted and received clutter spectrum of an airborne high prf pulse Doppler radar.

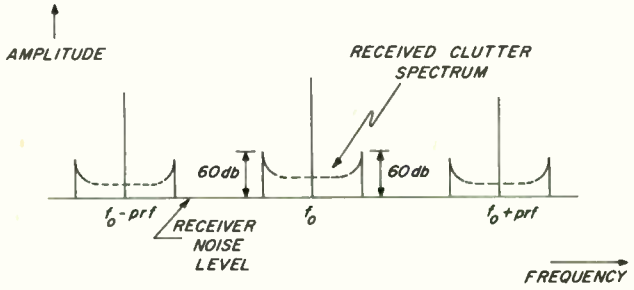


Fig. 7. Pessimistic model of received clutter spectrum for purposes of computing velocity coverage.

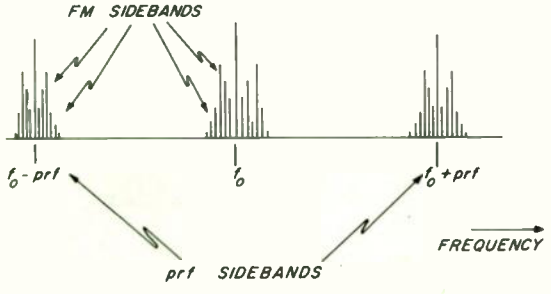


Fig. 5. Transmitted spectrum for sinusoidal FM.

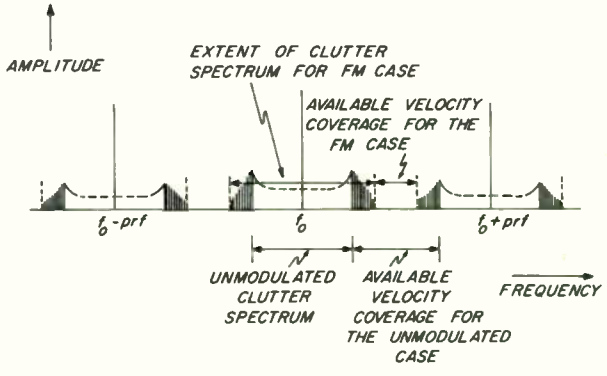


Fig. 8. Effect of sinusoidal FM on clutter spectrum of moving radar.

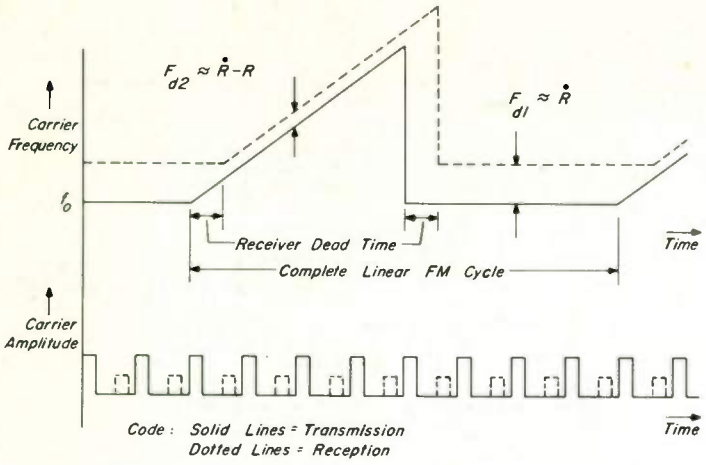


Fig. 9. Transmitted and received carrier frequency and amplitude versus time.

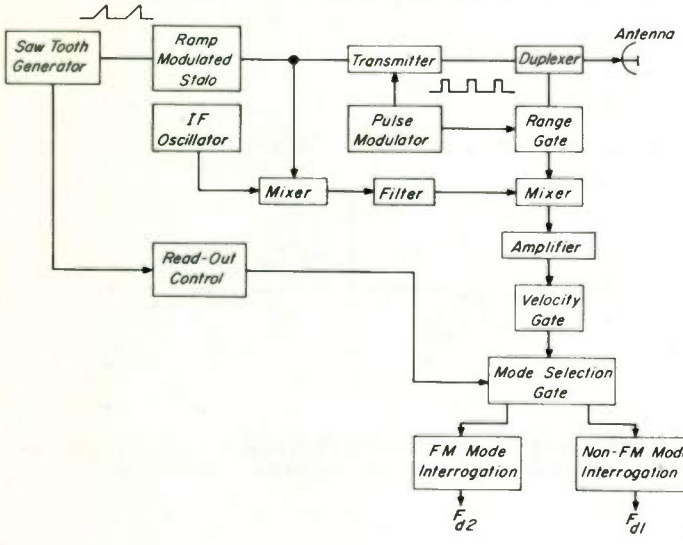
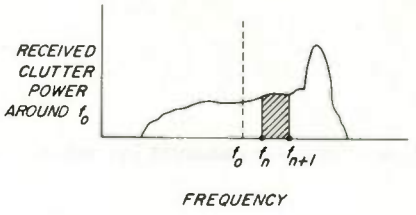


Fig. 10. Block diagram for linear FM.



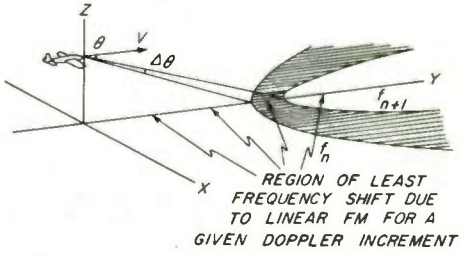


Fig. 11. Clutter scatterers reflecting energy in Doppler increment bounded by f_n and f_{n+1}.

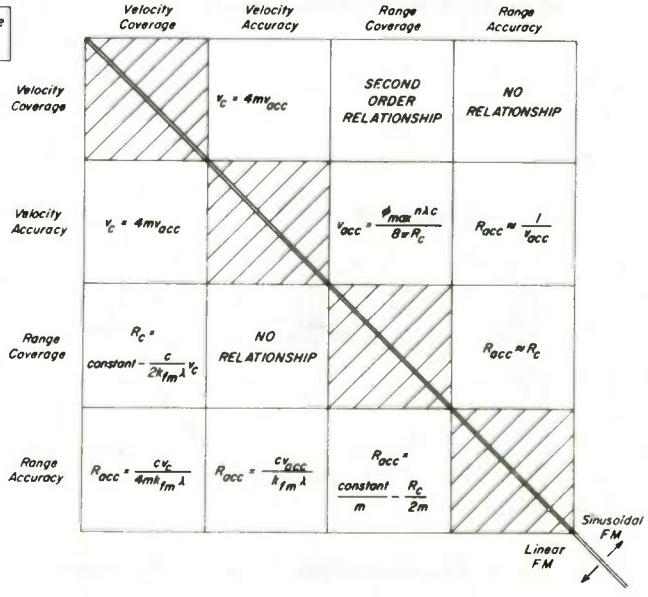


Fig. 12. Summary chart showing relationships between velocity accuracy and coverage and range accuracy and coverage for sinusoidal and linear FM.

APPLICATION OF SPACE FREQUENCY EQUIVALENCE TO RADAR

Merrill I. Skolnik

Research Division
Electronic Communications, Inc.
Timonium, Maryland

Summary

Space Frequency Equivalence (SFE) is based on the fact that the radiation pattern of a pair of array antenna elements spaced a distance d, radiating a frequency f is equivalent to the radiation pattern of a pair of elements spaced a distance d/k but radiating a frequency kf. SFE permits the trading of antenna elements for frequency. This paper describes the SFE principle and discusses some of the problems involved in applying it to radar. An example of the application of SFE to search radar is described and compared to the more conventional radar design.

Introduction

The principle of Space Frequency Equivalence (SFE) in antenna technology describes the trade than can be made between the antenna aperture size (number of radiating elements constituting an array antenna) and the frequency domain (number of spectral components constituting the radiated waveform). A pair of antenna elements spaced a distance d and radiating a frequency f, will give the same effective radiation pattern as a pair of elements spaced a distance d/2 and radiating a frequency 2f. Or, more generally, the pattern will be equivalent to a pair of elements spaced d/k radiating a frequency kf, where k is a factor that may be less than or greater than unity. This property may be used to obtain an array of a few elements simultaneously radiating many frequencies to produce a radiation pattern "equivalent" to that of a many-element conventional array radiating but a single frequency. The word "equivalent" is in quotes since the SFE array is not equivalent to the single frequency array in all respects, as will be described in this paper.

The principle of trading antenna elements for frequency spectral-lines was first described by Kock and Stone[1] for acoustical applications, but the basic principle applies equally well to electromagnetics. In addition to any practical advantages that might possibly be offered by the SFE principle, its study is of interest since it involves a dimension (frequency) not generally considered in conventional antenna theory.

Principle of the SFE Array

The radiation pattern $E_1(\theta)$ of a linear array of 2N+1 isotropic elements (2N+1 is an odd number) spaced a distance d apart may be written as[2]

$$E_1(\theta) = 1 + 2\sum_{n=1}^{N} \cos\left[(\pi/c) f n d \sin\theta\right] \quad \text{(N odd)} \qquad (1)$$

where c is the velocity of propagation, f is the radiated frequency, and θ is the angle measured with respect to the array normal. Each of the N terms in the summation of Eq. 1 may be interpreted as the pattern of a pair of elements spaced a distance 2d, 4d, 6d, . . . , 2Nd (Fig. 1a). The phase origin of the pattern is taken as the center of the array and the amplitude of the signal at each element is assumed to be unity. The total array antenna radiation pattern is that from the single center-element plus the superposition of the patterns from each pair of elements symmetrically placed about the center.

If the pattern-measuring receiver is broadband and extracts the amplitude of each frequency component and ignores the phase the same radiation pattern as given by Eq. 1 can be obtained, (with certain restrictions) from an array of three elements with the center element radiating a frequency f and the outer two elements radiating the frequencies f, 2f, 3f, . . . , Nf (Fig. 1b). Thus the index n in the factor fnd of Eq. 1 may be associated with either the distance d (as in the conventional, single-frequency array) or with the frequency f (as in the type of array described here). For example, the frequency f might be 50 Mc with N = 20, so that the signals would be radiated from the SFE array at frequencies of 50, 100, 150, . . . , and 1000 Mc. If 20 frequencies were used with the three element SFE array of Fig. 1b, the equivalent number of elements would be 41.

If the number of elements in the conventional, single-frequency array is even, instead of odd, the radiation pattern is

$$E_2(\theta) = 2\sum_{n=1}^{N} \cos\left[(\pi/c) f(2n-1) d \sin\theta\right] \quad \text{(N even)} \qquad (2)$$

where the total number of elements is 2N. The Space Frequency Equivalent array for an even number of elements consists of two elements spaced a distance d apart radiating the frequencies f, 3f, 5f, . . . , (2N-1)f. Generally an array with an odd number of elements will be assumed here since it utilizes the available frequency range to better advantage than the even-element array. If,

for example, the ratio of the upper to lower frequency limits were M, the number of frequencies that could be used with a three-element SFE array is M and would be equivalent to a single-frequency array of 2M+1 elements. In a two-element SFE array that is patterned after an even number of elements, the number of frequencies available is (M+1)/2 and is equivalent to a conventional array with but M+1 elements.

The equivalence between the frequency and the number of elements as described above is oversimplified and neglects the important fact that the pattern expression of Eq. 1 represents the amplitudes of the received signal components and does not include the carrier frequency or the phase. In a single frequency array, the carrier is the same for each term in the summation and can be taken outside as a separate factor (Eq. 3).

$$E(\theta, t) = \left\{ 1 + 2 \sum_{n=1}^{N} \cos[(\pi/c) f n d \sin\theta] \right\} \cos(2\pi f t + \phi)$$

$$= E_1(\theta) \cos(2\pi f t + \phi) \qquad (3)$$

where ϕ is the phase of the frequency f. In the equivalent expression for the SFE array, the carrier is different for each component and cannot be taken outside the summation sign. Instead, the signal received at some distant point in the far field of the array at a distance R = cT (where c is the velocity of propagation and T is the transit time) is

$$E(\theta, t) = \cos(2\pi f t + \phi) + 2 \sum_{n=1}^{N} \cos[(\pi/c) f n d \sin\theta] \cos n[2\pi f (t - T) + \phi] \qquad (4)$$

Thus the radiated energy from the transmitting array appears at the various frequencies and does not add in space as in the conventional single-frequency antenna. The measured "pattern" will depend on how the available frequency components are combined at the receiving device. Not only must the frequencies of the components constituting the SFE signal be handled properly, but the phase terms $2\pi nfT$ must also be given due consideration since they depend on the time delay T and are different for each component because of the factor n.

Another difference to be noted is that the classical reciprocity theorem of antenna theory does not hold with the SFE array; that is, the transmitting pattern and the receiving pattern are not the same. In the expressions of Eqs. 1 to 4 it was assumed that the amplitudes of the terms in the summation of the antenna pattern expression were equal. In the SFE array this means that the amplitude of each frequency component is the same and that the transmitting gain is independent of frequency. It is possible to obtain antennas based on the logarithmic periodic or the equiangular principles[3] with constant gain over wide bandwidth.[4] If the frequency components received by the SFE array are likewise to be of equal amplitude (assuming they were transmitted with equal amplitude), the effective receiving aperture should be frequency independent. It is possible, of course, to obtain an antenna with a constant receiving aperture over a broad range of frequency. (A reflector fed by a logarithmic periodic element is but one example.) However, an ordinary antenna cannot be of both constant gain G and of constant effective aperture A_e, with wavelength λ, because of the well known relationship from antenna theory that

$$G = 4\pi A_e / \lambda^2 \qquad (5)$$

If the SFE array were to utilize constant gain elements, the "effective" transmitting aperture distribution would be uniform, but the "effective" receiving aperture distribution would have an inverse taper symmetrically placed about the center of the antenna. Although the effective area varies inversely with the square of the frequency, the taper is inverse linear rather than inverse square because the aperture distribution is in terms of field strength; but the effective area relates to field intensity, the square of the field strength.

Thus the design of a radar using the SFE principle differs markedly from the usual radar design. Due care must be taken in the processing of the signal at the radar receiver so as to recover a suitable two-way "effective" antenna radiation pattern.

Application to Radar

There is more than one method for applying the SFE principle to radar. An SFE Array Radar (SFEAR) that follows from the straightforward application of the principles described in the preceding section is shown in Fig. 2. This is but one form of SFEAR. It is by no means the best, but it illustrates the basic SFE principle. Only the barest essentials are indicated in order to keep the diagram simple. Duplexers, mixers, IF and video amplifiers, power amplifiers, pulse modulators and other components needed to make a practical radar have been omitted. The three-element transmitting SFE array shown on the left of the diagram is fed by N signals of frequency f, 2f, 3f, . . . , Nf. The spacing d between the elements is assumed to be a half wavelength at the lowest frequency f. The receiving SFE array also consists of three elements with spacing d. The receiving array is shown separated from the

transmitting array for purposes of clarity. In an actual equipment a single three-element antenna with proper duplexing might serve for both transmitting and receiving.

At the receiving array the echo signal components are separated by bandpass filters whose center frequencies correspond to the components of the transmitted signal (f, 2f, etc.). The output of the filter at a frequency f corresponds to the output of the center three elements of the "equivalent" single frequency array; the output of the filter 2f corresponds to the next adjacent pair of elements; the output of filter 3f corresponds to the next pair; and the output of the filter Nf corresponds to the outer pair of elements of the equivalent single-frequency array. If each frequency is assumed to be radiated from the transmitting antenna with the same power and with equal gain, the (voltage) outputs of the various filters may be written

$$\left.\begin{aligned}
\text{Filter No. 1} &= C_1\left\{1+2\cos\left[\pi d(f/c)\sin\theta\right]\right\}^2 \times \\
&\qquad \cos\left[2\pi ft - 2\pi fT_r + \phi\right] \\
\text{Filter No. 2} &= (C_1/2)\cos^2\left[\pi d(2f/c)\sin\theta\right] \times \\
&\qquad \cos\left[2\pi 2ft - 2\pi 2fT_r + 2\phi\right] \\
&\;\;\vdots \\
\text{Filter No. N} &= (C_1/N)\cos^2\left[\pi d(Nf/c)\sin\theta\right] \times \\
&\qquad \cos\left[2\pi Nft - 2\pi NfT_r + N\phi\right]
\end{aligned}\right\} \quad (6)$$

where C_1 is a constant determined by the factors of the radar equation and T_r is the two-way transit time. The amplitudes of each frequency component are proportional to 1/n because of the inverse-square relationship between receiving aperture (or signal power) and frequency as described by Eq. 5.

There are several methods for processing the received signal. A matched filter could be designed by properly weighting the outputs at each frequency, summing and taking the envelope. An approximation to the matched filter might also be had by first summing the outputs from all elements and passing them through properly designed recirculating or tapped delay line filters that approximate the matched filter. Alternatively, the output of each filter could be converted to a common frequency and phase before summing. Still another technique is the incoherent addition of the filter outputs as shown in Fig. 2. This is by no means an optimum technique, but it is simple in concept and if only a few components are added, the loss in signal-to-noise ratio is small. For example, it can be shown from the considerations of integration loss derived by Marcum[5] that the signal-to-noise ratio is reduced by only 1.2 db when 10 signals (equivalent to 21 elements) are added, if the probability of detection is 0.99 and the false alarm probability 10^{-11}.

From Eq. 6 the effective two-way (voltage) antenna pattern of the SFE array radiating and receiving N frequencies as in Fig. 2, is given by the amplitude of the resultant video signal, or

$$E_3(\theta) = 1 + 4\cos\left[\pi d(f/c)\sin\theta\right] + 4\sum_{n=1}^{N}\frac{\cos^2\left[\pi d(nf/c)\sin\theta\right]}{n} \quad (7)$$

The two-way voltage pattern of a normal single frequency array of 2N+1 elements is

$$\left|E_1(\theta)\right|^2 = \left[1+2\sum_{n=1}^{N}\cos\left[\pi d(nf/c)\sin\theta\right]\right]^2 \quad (8)$$

The two-way patterns of the conventional radar array antenna (Eq. 8) and the SFEAR antenna (Eq. 7) are not equivalent. They differ considerably as shown by the plots in Fig. 3 of the radiation pattern of a 9-element conventional array and an SFE array radiating 4 frequencies (which is supposedly "equivalent" to the 9-element conventional array). The two-way pattern of this SFE array radar antenna leaves much to be desired and emphasizes the fact that the Space Frequency Equivalence principle is more involved than the simple statement of trading elements for frequency. Although it may be possible to synthesize better patterns than that shown in Fig. 3, the simple SFE array is basically more limited than its single-frequency counterpart. Therefore, if advantage is to be taken of the Space Frequency Equivalence principle in practical radar application, modifications of the basic concept must be made. These modifications tend to make the receiving array more complicated than shown in Fig. 2, but the flexibility of radiating many frequencies is maintained along with more acceptable two-way patterns.

Instead of receiving with the three element SFE array of Fig. 2, consider a receiving array similar to that of the conventional array except that broadband elements are used. The spacing of the elements is a half-wavelength at the lowest frequency so that at the higher frequencies the effective spacing between elements is accordingly greater. A 9-element linear array might appear as in Fig. 4. The crosses indicate broadband receiving elements. The center three elements with squares circumscribed about the crosses are both transmitting and receiving elements. The three transmitting elements are the same as the SFE transmitting array of Fig. 2, but the SFE receiving antenna is bigger than shown in Fig. 2.

The two-way (voltage) radiation pattern from an array of this type can be shown to be

$$E_4(\theta) = \frac{\sin (2N+1)\pi u/2}{\sin \pi u/2} + 2\sum_{n=1}^{N} \frac{\cos (nu) \sin [n(2N+1)\pi u/2]}{n \sin (n\pi u/2)} \quad (9)$$

where $u = d(f/c) \sin \theta$. The number of elements of the receiving array is 2N+1 and there are N frequencies radiated by the SFE transmitting array. This pattern (Eq. 9) is plotted in Fig. 5. Also shown for comparison are the normal array pattern and the three-element SFE array pattern. These assume the radiating elements are isotropic. The principle of pattern multiplication applies to the SFE array as to the conventional array. Therefore, the patterns shown in Fig. 5 must be multiplied by the element pattern to obtain the total array pattern. This will reduce the radiation at large angles and lower the relatively high lobe at 90° in the SFE array described by Eq. 9.

This is but one example of the many antenna variations that can be obtained with the SFE principle. The receiving antenna could be a separate constant-aperture antenna or a separate receiving array might be used with each frequency. Furthermore, it is not necessary to restrict the SFE array to two or three elements. A ten element SFE array, for example, radiating a total of twelve frequencies, can be made "equivalent" (in the sense described in the beginning of this paper) to a 76 element single-frequency array.

Other Radar Considerations

Beam Steering - The beam formed by the three-element SFE array of Fig. 2 can be steered in angle by using two broadband phase shifters, one in each of the outer arms, instead of the 2N phase shifters needed to steer the "equivalent" single frequency array of 2N+1 elements. The phase shifters must be broadband. Mechanically actuated coaxial line-stretchers or electronically controlled ferrite phase shifters operating on the high side of resonance[6] might be used.

Multiple Targets - The SFE array is essentially a linear device in that no multiplicative operations occur. Therefore its ability to handle more the one target is not limited as are time-averaged-product arrays or correlation arrays.[7]

Range Measurement - The measurement of range in SFEAR may be performed, as in any other radar system, with a standard pulse waveform. The use of multiple frequencies, however, has the further advantage that it may be possible to obtain a more accurate or vernier range measurement by comparing the phase differences between the various frequency components, provided they are not exactly harmonically related. This is similar to the two frequency CW measurement of range,[8,9] except that if one frequency is f the other is $2f + \Delta f$, where $\Delta f/2$ is the separation normally used between frequencies in the CW range measuring technique. Thus the use of multiple frequencies in SFEAR permits, in principle, a better range measurement than is normally extracted with a conventional single frequency pulse transmission.

MTI Capability - The SFE array radar which performs incoherent addition destroys the phase information available in the returned signal and is therefore not suited as a coherent MTI (moving target indication) radar. In other variations of the SFE, it is possible to preserve the phase and to achieve MTI performance.

Target Cross Section - In the analysis of the two-way SFE antenna patterns it was assumed that the target cross section was independent of frequency. This assumption is correct for a perfectly reflecting sphere target, large compared with the wavelength. The cross sections of complex targets are frequency dependent, but generally, the average behavior does not vary greatly with frequency.[10] However, there might be considerable fluctuation in the cross section as the frequency is changed even though the average does not change significantly. The antenna pattern as measured by the SFE receiver will certainly be affected by variations in cross section with frequency. If the variations were extreme it might be possible to employ some sort of equalizing network. Not enough study has been given this aspect of the problem to warrant further conclusions. It should be mentioned, however, that variations in cross section with frequency may not always be detrimental, even if compensation is not made. For example, Lees[11] has shown that the variation of echo amplitude with frequency may be taken as a measure of the target size.

Components - The major components which must operate over the wide frequency range needed in the SFE array are the transmission lines and the antenna elements. At VHF and UHF, coaxial transmission lines can operate with relatively high power over wide frequency ranges. Antenna elements based on the spiral, tapered helix, or the DuHamel log periodic radiator are capable of wideband operation. If beam steering is obtained by mechanical motion of the antenna, rotating joints might be necessary. Contacting joints capable of moderate power handling are commercially available from DC through X-band.

The type of broadband behavior needed in the SFEAR is different than normally required for

broadband operations. Uniform response over the entire band is not needed. Components need only be "broadband" at the N relatively narrowband frequency regions occupied by the radiated frequencies.

Frequency Diversity - The SFEAR radar is a true frequency diversity radar which operates over a wide range of frequencies. If it were not possible to operate at one of the frequencies due to malfunction, or to interference from other radiations, the effect would be similar to having two of the elements of the "equivalent" single-frequency array inoperative. The pattern would degrade, but the radar could continue to operate.

Two Dimensional SFE Arrays - The discussion in this paper has been concerned primarily with the application of the SFE principle to linear arrays. It also applies to two dimensional arrays where vertical directivity is obtained by paralleling similar linear arrays or where a linear array acts as a feed for a reflector antenna. In these cases, the SFE principle is applied in but one dimension. Little consideration has been given to applying the SFE principle in two dimensions, although there seems to be no reason why it should not be possible to do so.

Example of Radar Design

It is of interest to compare the range of a Space Frequency Equivalent Array Radar with that of a hypothetical radar of conventional design. The conventional radar is assumed to possess the following characteristics:

Frequency: 200 Mc
Transmitted energy per pulse: 40 watt-sec. (200 kw peak power, 200 μsec pulse)
PRF: 250 cps
Antenna: 9 by 4 array of dipoles
Beamwidth: 30° in elevation
14° in azimuth
Antenna gain: 70
Rotation rate: 5 rpm

The range R_m of the normal radar is given by the the radar equation

$$R_m^4 = \frac{E_t G_t^2 \lambda^2 \sigma n E_i(n)}{(4\pi)^3 k T_o B\tau F (S/N)_1 L_s} \qquad (10)$$

where E_t is the radiated energy per pulse (watts-sec), G_t is the antenna gain, λ is the wavelength in meters, σ is the radar cross section in sq.m. (σ=1), k is Boltzmann's constant, T_o is the standard temperature (290°K), B is the receiver noise bandwidth, τ is the pulse width, F is the receiver noise figure, $(S/N)_1$ is the signal-to-noise ratio required to achieve the desired probability of detection and probability of false alarm if detection were on the basis of a single pulse, n is the number of hits per scan, $E_i(n)$ is the integration efficiency, and L_s is the system loss (includes the beam shape loss, field degration, RF losses, etc.) and is taken to be 6 db in this example. The product $B\tau$ is assumed to be unity, a good approximation if maximum signal-to-noise ratio is desired.

At a frequency of 200 Mc, the noise figure which appears in the radar equation must take account of sky noise. The overall noise figure F_o of a receiver whose laboratory noise figure is F_r connected to the antenna via a transmission line with loss L_t and which views a sky of noise temperature T_s is

$$F_o = T_s/T_o + F_r L_t \qquad (11)$$

At a frequency of 200 Mc, T_s is about 700°K (maximum) and F_r is 1.9 for the 416B triode. The line loss L_t is assumed to be 1.26 (one db). The noise figure is therefore

$$F_o = 700/290 + (1.9 \times 1.26) = 4.8 \ (6.8 \text{ db})$$

The number of pulses received per scan is

$$n = \frac{\theta_B}{6} \times \frac{PRF}{RPM} \qquad (12)$$

where θ_B is the azimuth beamwidth in degrees and RPM is the antenna rotation rate in revolutions per minute. Substituting the radar parameters into Eq. 12 gives n = 116 pulses. Using data derived from Marcum,[5] the factor $nE_i(n)$ turns out to be equal to about 30.

It may be shown that a signal-to-noise ratio of 14 (or 11.5 db) is required for a probability of detection of 0.50 with a false alarm probability of 10^{-6}, if detection were to take place on the basis of a single pulse.[5] These parameters when substituted into the radar equation yield

$$R_m^4 = \frac{40 \times (70)^2 \times (1.5)^2 \times 1 \times 30}{2 \times 10^3 \times 4\,10^{-21} \times 4.8 \times 14 \times 4}$$

$$= 61.5 \times 10^{20} \text{ m}^4$$

$$R_m = 150 \text{ n mi}$$

The SFE Array Radar assumed for discussion is shown schematically in Fig. 6. It is a two dimensional version of the Fig. 4 array. The SFE transmitting array (shown by the X-boxes) is assumed to be a 3 by 4 array of broadband antenna elements. The three vertical columns of boxed-X elements (that determine the azimuth beamwidth) are those of the SFE array. The horizontal rows are identical and are connected together to provide antenna pattern directivity in the elevation plane. (Note that coverage in the elevation plane will differ from the coverage obtained with a

single frequency radar.) The SFEAR radiates equal power at frequencies of 200, 400, 600, and 800 Mc in order to simulate the 9 by 4 element array. If the total power in the SFE array is to be the same as the power in the normal array, the energy per pulse per frequency component is 10 watt-sec (which might correspond to 50 kw peak, 2.4 kw average).

The receiving antenna is the 9 by 4 array of broadband elements (X's in Fig. 6) capable of receiving over the entire band of frequencies. The physical size is the same as a conventional single-frequency 9 by 4 array. The spacing between elements is a half wavelength at the lowest frequency (200 Mc). The array pattern for a SFE Array Radar with this configuration was shown in Fig. 5.

The calculation of the radar range of the SFEAR is slightly different from an ordinary radar since some of the parameters of the range equation are a function of frequency. Therefore, the signal received at each of the frequencies must be computed separately and the total signal-to-noise contribution found after addition. In the particular SFEAR configuration under consideration, the transmitted power and the gain are taken to be the same for each frequency. It is also assumed that the target cross section is independent of frequency. The remaining two factors which are frequency dependent are the receiver noise power and the effective aperture of the receiving antenna. The noise figure at each of the frequencies and the calculated receiver noise powers are shown in Table I. The sky temperature is also shown.[12] A transmission line loss of one db was assumed at each frequency. Although the receiver noise figure increases with frequency, the sky noise temperature decreases over this frequency region. Consequently, there is little variation in the receiver noise power with frequency. For purposes of calculation it will be assumed that the receiver noise is the same at each frequency and has a value of 10^{-16} watts, which is the same as was assumed for the receiver noise figure of the single-frequency radar.

Since the effective receiving aperture decreases as the square of the frequency the receiving aperture distribution is weighted by an illumination factor which varies inversely with distance along the array. The tapered distribution reduces the overall signal-to-noise ratio, compared to the uniform illumination of the conventional array, by the factor $(1 + 1/4 + 1/9 + 1/16)/4 = 1.42/4 = 0.355$. An additional one db (efficiency factor of 0.8) of combining loss will be assumed in the SFEAR because the addition of the four frequencies is performed incoherently. The elements used in the SFE array are directive, rather than the simple dipoles assumed in the conventional array. Therefore, an element factor must multiply the array factor in order to obtain the overall antenna gain and effective area. Typically, the beamwidth of a DuHamel log-periodic element that might be used in this application is 60° by 60°, giving a gain of about 10 db. At the fundamental frequency, where the spacings are a half-wavelength, it may not be possible to realize the full element factor contribution to the gain.[13] Therefore, an overall gain of 7 db for the elements operating in the antenna system shall be arbitrarily assumed. Since the half-wave dipole gain is about 2 db, the overall increase in signal-to-noise ratio because of the directive element is 5 db one-way (10 db two-way).

With these factors, the range of the SFEAR as compared with the normal array is

$$\text{SFE Range} = \begin{pmatrix}\text{Conventional}\\ \text{Radar range}\end{pmatrix} \times \left[\begin{pmatrix}\text{Aperture}\\ \text{efficiency}\end{pmatrix} \times \begin{pmatrix}\text{combining}\\ \text{efficiency}\end{pmatrix} \times \begin{pmatrix}\text{element}\\ \text{gain}\end{pmatrix}^2\right]^{1/4}$$

$$= 150 \times (0.355 \times 0.8 \times 10)^{1/4} = 195 \text{ n mi} \qquad (13)$$

As can be seen, most of this improvement comes about because of the element gain of the broadband, directive elements. If the same type of elements were used in both antennas, the range of the SFEAR would be approximately 75% that of the conventional radar.

Conclusion

It is seen that a radar can be obtained, radiating many widely spaced frequencies, with performance comparable to that of a single-frequency radar of conventional design. The Space Frequency Equivalence principle forms a basis for the design of such a radar, but because it is an oversimplification of the antenna phenomenon envolved, the concept must be extended to account for the frequency and phase, as well as the amplitude, associated with each spectral component.

References

1. W. E. Kock and J. L. Stone, "Space Frequency Equivalence," Proc. IRE, vol. 47, pp. 1199-1200, July, 1959.
2. J. D. Kraus, Antennas, McGraw-Hill Book Co, N. Y. 1950, Sect. 4.8.
3. G. A. Deschamps and R. H. DuHamel, "Frequency-Independent Antennas," Chapt. 18, Antenna Engineering Handbook, H. Jasik, Editor, McGraw-Hill Book Co, 1961.

4. R.L. Bell, "Broadband Log-Periodic Antennas," Electronics, vol. 33, No. 25, pp. 58-59, June 17, 1960.

5. J.I. Marcum, "A Statistical Theory of Target Detection by Radar," IRE Trans., vol. IT-6, pp. 59-268, April, 1960.

6. C.M. Johnson, "Bandwidth of Ferrite Phase Shifters for Phase Array and Direction-Finding Use," Proc. IRE, vol. 47, p. 1665, Sept. 1959.

7. I.W. Linder, "Resolution Characteristics of Correlation Arrays," J. of Res. NBS, vol. 65D, pp. 245-252, May-June, 1961.

8. L.N. Ridenour, Radar System Engineering, Sect. 5.7, McGraw-Hill Book Co., New York, 1947.

9. M.I. Skolnik, "An Analysis of Bistatic Radar," Appendix, IRE Trans., vol. ANE-8, pp. 19-27, March, 1961.

10. "The Radar Cross Sections of Aircraft," University of Michigan Engineering Research Institute NEWS, vol. 8, No. 4, October, 1957.

11. R.J. Lees, "A Generalized Theory of Radar Observations," Agardograph No. 40, Avionics Research; Satellites and Problems of Long Range Detection and Tracking, Pergamon Press, 1960.

12. J.C. Greene and M.T. Lebenbaum, Letter in the Microwave Journal, vol. 2, pp. 13-14, October, 1959.

13. J.N. Hines, V.H. Rumsey, and T.E. Tice, "On the Design of Arrays," Proc. IRE, vol. 42, pp. 1262-1267, August, 1954.

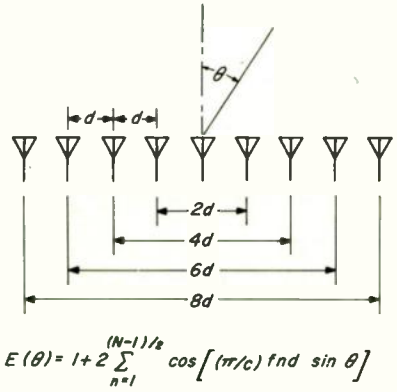


Fig. 1(a). N element linear array (radiating a frequency f) and its radiated pattern $E(\theta)$.

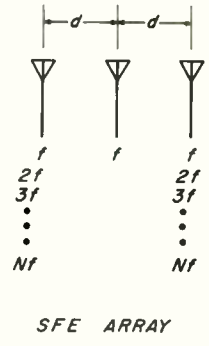


Fig. 1(b). Three-element space frequency equivalent array.

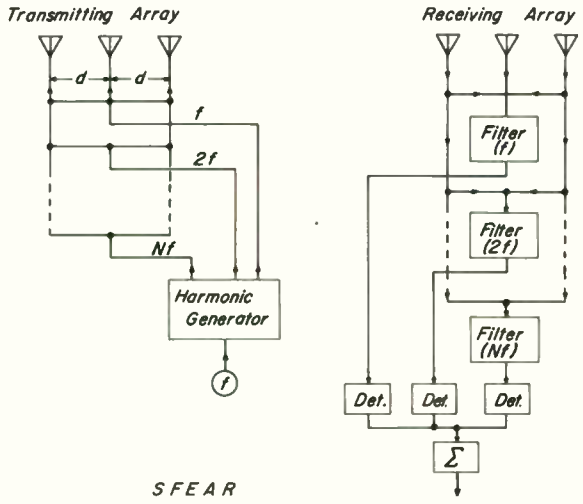


Fig. 2. One form of SFE array radar utilizing incoherent addition.

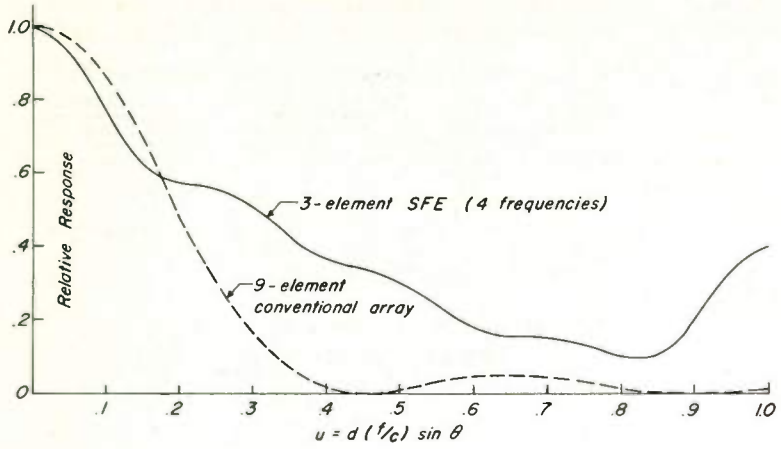
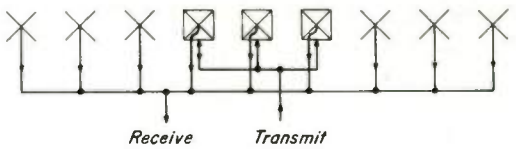


Fig. 3. Two-way (voltage) radiation patterns of 3-element SFE (Eq. 7) radiating 4 frequencies (solid curve) and 9-element "equivalent" pattern (Eq. 8) of the single frequency conventional array (dashed curve). Spacing between elements = d, frequency of 9-element conventional array and lowest frequency of SFE array = f, velocity of propagation = c, and angle measured from array normal = θ.



SFE Array Radar With "Conventional" Receiving Antenna

Fig. 4. Nine-element receiving array (X's) with 3-element transmitting SFE array (boxed X's).

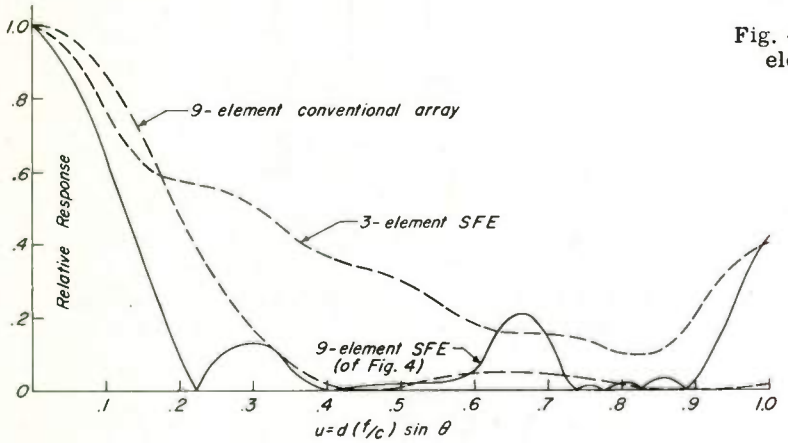


Fig. 5. Solid curve is the two-way (voltage) radiation pattern (Eq. 9) of the 9-element SFE array of Fig. 4 compared with the 9-element single frequency array (Eq. 8) and the 3-element SFE (Eq. 7).

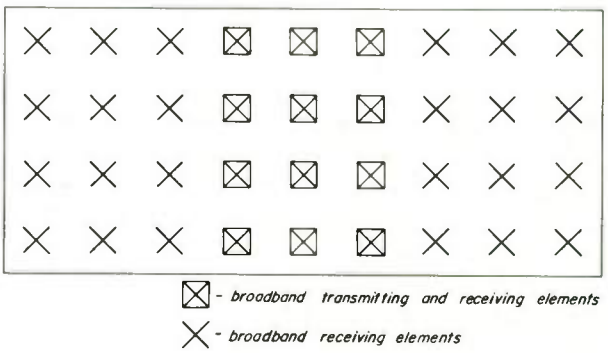


Fig. 6. Antenna configuration for SFEAR described in design example.

RADAR SYSTEM FOR UNMANNED COOPERATIVE RENDEZVOUS IN SPACE

Howard A. Reuter
Westinghouse Electric Corporation, Air Arm Division
Baltimore, Maryland

Summary

After launching a chaser vehicle to intercept a satellite target vehicle, terminal maneuvers will be required to close the gap between vehicles for docking. Pulse radar is used to guide the space crafts in this phase of the mission.

A unique radar-transponder system, now in development, will provide unmanned or manned chaser vehicles with long range acquisition, tracking and docking guidance. Special circuits are described which provide improved range accuracy at very short ranges, and two coordinate angle track without conical scan or multiple receivers. The system will acquire a target and track in range and angle over a 90-solid angle from 60 miles to docking.

The chaser radar is timed for operation when the range is about 60 miles. A series of 1 μsec interrogating pulses at a PRF of 250 pps at an L-band frequency are transmitted. These pulses are received by the target satellite and used to trigger its transponder. A 9-microsecond signal, shifted in RF from the radar frequency, is received at the chaser and used to measure range and bearing. Range rate is obtained by differentiation. The leading edge of the reply pulse is used to measure range and the remainder of the pulse to measure azimuth angle then elevation angle.

At long range the receiver bandwidth is narrow and gain high. As the range decreases, the output of the wide band amplifier becomes large enough to trigger the range multivibrator. Greater accuracy results because the long rise time, delay, and threshold jitter of the narrow band post amplifier are eliminated. Since transmission and reception are on different frequencies, and the reply pulse is always delayed two microseconds, there is no minimum range limit because of transmit pulse blockage or TR recovery as in a more conventional radar. The minimum range is limited only by the video techniques and this system will be accurate to one percent at long ranges or ± 20 ft. at short range.

Angle measurement is made by an interferometer technique not using gimballed antennas. The antenna system has four fixed plane spiral antennas facing the target, one is used for transmitting the other three for measurement of range, azimuth, and elevation angles. Angle is determined by the phase delay of the RF signal in one antenna with respect to the other; a pair of antennas is used to measure azimuth and a pair to measure elevation. (i.e., one antenna is common to both azimuth and elevation). RF phase shift is obtained by rotation of a circular polarized antenna with respect to a fixed reference circular antenna. The transponding antenna on the target satellite is a circular polarized spiral.

Introduction

Our nation has placed into orbit many earth satellites whose missions were principally data gathering. However, future project planners for civilian and military space operation now see a need for space craft whose mission will be a vast departure from previous goals. I am referring to a space craft which is placed in orbit only to seek out and join another object already in orbit. Such orbital operations are usually referred to as either cooperative rendezvous or on the other hand, unfriendly rendezvous.

In the future, we may see complex rendezvous operations designed for emergency rescue missions to return astronauts from disabled space ships; or construction of orbiting space platforms, either manned or unmanned to serve as outposts on the push to the planets; and the dispatch of crews or equipment to repair, recharge or recover communication, navigation or weather satellites. Such rendezvous operations are, however, largely a matter of speculation. A more immediate use of cooperative rendezvous may be in our efforts to begin manned exploration of the moon. Many space authorities are convinced the rendezvous technique offers the United States its only hope of being first to land a man on the moon. A lunar space ship, its rocket engine and fuel, may be lifted into earth orbits in separate launchings using Saturn boosters and assembled in orbit for its journey to the moon; instead of by one combined launching. A lengthly development would be required before 12-million-pound thrust Nova boosters are operational. Instead the 7.5-million-pound thrust Saturn C-5 (ready in 1965 or 1966) could lift the Apollo craft and its engine into orbit in separate shots.

NASA has recently announced plans to develop a two-man space craft capable of making a rendezvous while in earth orbit. Known as Project Gemini, this advanced Mercury capsule would be a scaled up version of the one-man craft and capable of testing Apollo rendezvous schemes.

Space rendezvous need not be cooperative. The military, concerned that someday unfriendly

space craft may patrol our skies, are reported to be developing their own rendezvous program for automatically inspecting and acting upon hostile satellites. This paper is, however, concerned only with friendly or cooperative rendezvous.

Having touched briefly upon the need for orbital operation, let us examine how it might be done. A target vehicle would be placed in orbit and its position computed by ground-based tracking stations. This data would be used to launch the so-called chaser vehicle to within radar distance of the target and in nearly the same orbit. The chaser, aided by radar, would close to within a few hundred feet of the target at a maximum rate of 1000 feed per second. Before docking, this rate would probably be reduced to no more than 10 feet per second.

The chaser vehicle must successfully perform the following terminal maneuvers: locate and acquire a tracking signal from the target; determine bearing angles between vehicles with an accuracy of a few milliradians from long range to docking; determine range and range rate from long range to docking; and it should be automatic, light weight, small, low power and reliable. Requirements of this sort are best satisfied using the technique of radar.

System Description

The requirements and the design possibilities were studied and a rather unique system developed. Conventional circuits and techniques were used throughout the design, but they were employed in an unusual manner to produce the system we believe is best for orbital operations.

Briefly, this rendezvous sensor is an interrogator-transponder system. The chaser vehicle interrogates the target transponder and in turn receives a broad reply pulse. The leading edge of the pulse is used to measure range, and the remainder is time shared between an azimuth or elevation antenna pair to determine bearing from interferometric techniques. Angle measurements at the target may be made in a similar manner.

The chaser radar is programmed for turn-on when the range is about 60 nautical miles. This range may be extended to several hundred miles with very little additional power. The target system may be enabled by ground control, directly by the interrogating radar, or manned operation if available.

The chaser transmits a series of 1-μsec interrogating pulses at a PRF of 250 pps at an L-band frequency. These pulses are received by the transponder receiver and, after fixed delay, used to trigger the transponder transmitter. The reply signal, 60 mc higher than the chaser radar frequency, is received by the chaser radar and used to measure range and bearing. Figure 1 illustrates the transmitted and local oscillator frequencies used in this system.

Angle measurement is accomplished by interferometer techniques, this method not requiring gimbaled antennas. The antenna system proposed is composed of four fixed-plane spiral antennas positioned on the skin of the vehicle (facing the target); one of the four is used for transmitting while the other three are used for measuring range and bearing angles. This system provides simultaneous coverage during initial acquisition and two-plane angle track over a 90-degree solid angle.

A block diagram of the system is shown in figure 2. Also included in this figure is an illustration of the radar antenna layout.

Antennas II and III form one pair and are depicted as those used for measuring the elevation component of the bearing angle. Antennas II and I, in turn, are used to measure the azimuth component. Antenna II serves another function other than acting as a common antenna for both azimuth and elevation measurements; it is the receiving antenna during range measurement on the leading edge of the pulse.

RF switches S_3 and S_4 are those used to select the azimuth and elevation angle measuring channels respectively during each received pulse.

Dither switching is accomplished by closing and opening ganged RF switches S_1 and S_2 on alternate pulses. The circuit then introduces a positive or negative differential phase shift into the reference antenna line to establish sense for servo control.

There is a 180-degree ($\pm$ the dither angle) difference between the RF path length from the reference antenna to the summing point and the length from either the azimuth or elevation antenna to that same point. Thus, this system is, in principle, a null tracker. Note that the circuit of figure 2 is a block diagram illustration, and the actual RF schematic need not contain critical phase line lengths in the connections of S_3, S_4 and the dither switches.

With the switches initially as shown in figure 2 (which is their normal position between transmission of the interrogating pulse and receipt of the reply except for the dither switches which alternate) the reply pulse will enter the receiving channel through Antenna II. The RF signal is mixed with the local oscillator signal to translate the pulse to a 30-mc IF.

The receiver gain is established by the noise AGC at far range and the signal is obtained through the narrow-band postamplifier. This establishes the 1.5-mc system bandwidth. The video output is peak detected and applied to the range and angle measuring channels after additional amplification. Range measurement is performed by the leading edge of each pulse triggering a linear range sweep.

As range decreases, AGC action holds the output constant by controlling only that voltage-controlled attenuator immediately preceding the narrow band post amplifier. When the signal strength has increased 10.3 db (the ratio of the receiver bandwidth), triggering out of the wide (16 mc) amplifier will begin. The shorter rise time associated with the wider band width permits more accurate range measurement.

A trigger pulse is applied to the gate generator when the range measurement threshold is exceeded. This pulse initiates a 3 μsec gate which closes the azimuth channel RF switch S_3, and also enables the azimuth angle measuring channel. The vector difference (null) between the antenna pair is now present in the receiver-video channel. The pulse amplitude resulting from this interferometer action is applied to a boxcar circuit where it is stored until the next transponder reply is received. This gating of the reply pulse is illustrated in figure 3.

At the end of the 3 μsec enabling gate, S_3 is opened, and the input to the azimuth channel is opened. After a 1 μsec delay, another 3 μsec enabling gate is initiated. This second 3 μsec gate closes switch S_4 and enables the elevation angle measuring channel and the pulse amplitude containing elevation angle information is stored.

When the next interrogating pulse is transmitted, the RF switches are again as shown in figure 2 except for the phase dither switches S_1 and S_2 are closed. When the reply pulse is received, range and both components of angle are again measured.

Angle Measurement

Angle measurement is accomplished with interferometer techniques. The interferometer principle employs the phase difference of the RF energy received by separated receiving antennas. When the target lies off the boresight axis, the difference in path length produces this difference in RF phase. The design of one channel to measure either the elevation or azimuth component of the bearing angle will be described with the aid of figure 3. As shown, two antennas are separated by a distance D. The difference in length of the two RF paths is shown as x. This results in a difference in phase ϕ, between the signals appearing at the two antennas of

$$\phi = \frac{2\pi D}{\lambda} \sin \theta$$

therefore, since D and λ are known constants, measurement of the electrical phase difference, ϕ, permits calculation of the bearing angle θ. The quantities actually measured directly are the direction cosines with respect to the pitch and yaw space craft axes.

The phase delay is a periodic function of the bearing angle, θ. This could result in ambiguous measurements of the bearing angle. However, if the antenna spacing, D, is approximately equal to one-half the RF wavelength, the ambiguity will not occur within $\pm$ 45 degrees of center.

The outputs of the two antennas are applied to a differencing junction. Were it not for the dither circuit, there would be no phase difference between the signals when the target is on boresight and the output would be zero. When the target lies off boresight the output would be:

$$E_1 \sin wt - E_2 \sin (wt + \phi)$$

This shows that the information out of the differencing junction does not contain information to establish on which side of the boresight axis the target lies. It also has the disadvantage of producing a signal in or near the null noise. To overcome this, phase dither is added. The phase shift of the reference channel is stepped between two values ($\pm$ an equal amount) at a rate which is synchronized to the PRF. A target off boresight results in an amplitude modulation of the dithered difference signal (E_1 and E_2) at a frequency which is half the PRF; comparing this modulation with a phase dither reference signal yields an error signal with the required sense information.

As shown earlier, measurement of the electrical phase shift between the two RF signals arriving at the antennas permits determining the bearing angle. This phase shift is obtained by controlling a calibrated phase shifter which is contained in one of the antenna lines to compensate for the difference in phase shift of the received signals. This control is obtained in the following manner:

The output of the boxcar demodulator, in each channel, is a d-c voltage proportional to the averaged value of the alternating pulses with a superimposed square wave of amplitude equal to the difference between alternate pulses. The d-c value is used for automatic gain control purposes. The square wave out of each channel represents the angular error from boresight. This signal is applied to the synchronous demodulator and the resulting polarized error voltage is applied to the antenna servo control. The RF is phase shifted until the error signal is equal for alternate pulses.

The target position is tracked by introducing a phase shift into one receiving antenna relative to the other. This is accomplished by merely rotating one spiral and holding the other fixed. The reference antenna is held fixed so that independent scanning in the component directions is achieved. This phase shift occurs because the spirals are circularly polarized and rotation of a circularly polarized antenna always results in a phase shift of the received fields equal to the angle of rotation. The amount of rotation of each antenna is then used to compute the azimuth and elevation components of the bearing angle.

No duplexer is required since sufficient

isolation can be obtained by transmitting on a separate antenna. Both transmitting antennas are duplicates of the spiral receiving antennas and radiate a circularly polarized wave from the interrogator and transponder.

A detailed analysis of bias errors in the measured angle of the target shows that angular accuracy is dependent on the excellence of the circularity of the radar receiving antennas and to a lesser extent the transponder transmitting antenna.

In an antenna having circular physical symmetry, such as a spiral, the on-axis polarization rotates as the antenna rotates. If the antenna is rotated in an imperfect system, the phase difference will also depend on the polarization characteristics of the transponder antenna at the other end of the link.

The electrical phase shift δ as a function of spiral rotation angle ϕ is given for some ellipticity combinations.

a. Transponder and receiver are perfectly polarized

$$\delta = \phi$$

b. Transponder is perfectly circularly polarized, receiver on-axis voltage ellipticity ratio (ϵ), a gain characteristic, not on the beam axis, such that the vertical and horizontal components differ by a factor k.

$$\delta = \tan^{-1}\left(\frac{k - \epsilon}{1 - \epsilon k}\right) \tan \phi$$

c. Transponder is elliptically polarized with ellipticity (m) in one direction of the chaser, receiver ellipticity (ϵ), gain characteristic (k)

(1) Major Axes Parallel

$$\delta = \tan^{-1}\left(\frac{m\,k - \epsilon}{1 - m\,\epsilon\,k}\right) \tan \phi$$

(2) Major Axes Perpendicular

$$\delta = \tan^{-1}\left(\frac{k - \epsilon\,m}{m - \epsilon\,k}\right) \tan \phi$$

The transponder ellipticity (m) was taken to be 1 db, the receiver ellipticity (ϵ) was taken to be 0.5 db on axis and the coupling ratio coefficient (k) was assumed to vary as the cosine of the angle, $k = \cos\theta$, which is somewhat severe. Antenna spacing is $\lambda/2$. Bias errors as a function of scan angle for cases c1 and c2 indicate a maximum of 0.72 angular mils error over the region $\pm$ 30 degrees from broadside for the worst case where the major axis of the transponder is perpendicular to the receiver major axis. Bias errors increase rapidly as the scan angle increases to greater than 30 degrees.[1]

The effect of system noise on angle tracking accuracy will be shown by an analysis of one channel.

The antenna system at a given antenna spacing, the IF and video circuitry, boxcar, and demodulator combine to give one of the loop constants, K_1, in volts per degree. For alternate video pulse amplitudes of 14.0 and 16.0 volts into the boxcar the value of K_1 is one volt per degree. The rms tracking error may be found from K_1, the video signal-to-noise ratio, the characteristics of the boxcar demodulator with noise input, and the tracking loop bandwidth.

The signal-to-noise ratio out of the second detector for a range of 60 nautical miles is 17 db and increases inversely as the square of the range. A 3-microsecond interval of the second detector output is selected by a gated stage and passed through a bandwidth matched to the pulse width. The matched bandwidth of a 3-microsecond pulse is given by 1.66 microsecond^{-1} or 0.4 megacycles. The signal-to-noise ratio is therefore improved by the factor of 1.5/0.4 or 5.7 db. This gives a signal-to-noise ratio at the input to the boxcar of 17 db +5.7 db or 22.7 db.

During null tracking, the input of the boxcar will be a constant voltage with a step modulation of random amplitude superimposed upon it. The power spectrum of a boxcar circuit output with this type input has been analyzed.

The analysis is made by first considering the boxcar input as noise only and determining the spectrum of its output. Then, the suppression of this output noise by the presence of a video signal at the boxcar input is obtained. Applying this reduction factor to the value obtained when noise only is considered will give the characteristics of the boxcar with signal plus noise at the input.

With noise only at the input, the output spectrum, $G_N(f)$ is given by:[2]

$$G_N(f) = \frac{2}{PRF} \cdot \left(\frac{\sin \frac{\pi f}{PRF}}{\frac{\pi f}{PRF}}\right)^2 \cdot \left(1 - \frac{\pi}{4}\right) N^2$$

where N^2 is the average noise power at the input to the boxcar.

Since the output of the boxcar is filtered by the narrow tracking loop bandwidth, only the noise power in the small region about the signal modulation frequency is of interest. Therefore, for a PRF of 250 cps and a frequency of PRF/2, the noise density is:

$$G_N = \left(\frac{2}{PRF}\right)\left(\frac{2}{\pi}\right)^2 (0.215)\, N^2$$

$$= 6.93 \times 10^{-4}\, N^2 \text{ watts per cycle}$$

For a 2.0-cycle track bandwidth, the equivalent bandpass noise power accepted by the track bandwidth is $(4)(6.93 \times 10^{-4}\, N^2)$ or $2.77 \times 10^{-3}\, N^2$ watts and the rms noise voltage is $5.32 \times 10^{-2}\, N$, where N is the rms noise on the video pulse.

The input signal level is 15 volts. With a S/N of 22.7 db or 13.7 voltage ratio the rms noise at the boxcar input is 15/13.7 or 1.1 volts rms and the noise output is:

$$N_{rms} = (5.32 \times 10^{-2})(1.1) = 5.85 \times 10^{-2} \text{ volt rms}$$

For large signal-to-noise ratios, the amplitude of the noise spectrum is suppressed by:

$$D(Z, 1) \approx \frac{1}{2\left(\frac{4}{\pi} - 1\right) Z}\left(1 - \frac{3}{4Z} + \frac{1}{8Z^2} \ldots.\right)$$

where $Z = \dfrac{\text{unmodulated signal power}}{\text{noise power}} = S/N$

$$D(Z, 1) \approx \frac{1}{2\left(\frac{4}{\pi} - 1\right)(188)} = 9.73 \times 10^{-3}$$

Therefore the noise output from the tracking loop is

$$(5.85 \times 10^{-2})(9.73 \times 10^{-3}) = 5.7 \times 10^{-4} \text{ volts rms}$$

Since the angle sensitivity is 1 volt per degree, the resulting rms tracking error is 5.7×10^{-4} degree or 0.01 milliradian. This is a negligible value and may be ignored in angle accuracy considerations.

Range Measurement

Range tracking circuits are not required to achieve S/N enhancement because of the excellent signal power available. Range is measured by initiating a sweep with the transmitted pulse and terminating the sweep with the leading edge of the return pulse. The amplitude reached by the sweep is thus a measure of range. This amplitude is retained by a peak detector, filtered, and a d-c voltage is presented for range use. Figure 4 illustrates the range-measuring circuit.

There are three voltage-to-range scale changes in the range-measuring system, the maximum ranges in the three intervals being 60 nautical miles, 4 nautical miles, and 1/4 nautical mile. Range scales are changed automatically for either opening or closing chaser conditions.

Because of the delay in the transponder, reply does not begin until the entire width of the interrogating pulse has been received. This enables the interrogating radar to measure range down to zero. There is no interference between the interrogator's transmission and the transponder's reply, even at zero range.

The accuracy of range measurement is a function of the following factors:

1. Linearity of the range sawtooth.
2. Time-varying characteristics of the video pulse leading edge at the range threshold in the radar.
3. Time-varying characteristics of the video pulse leading edge of the transponder threshold.
4. Time jitter of the reply pulse leading edge resulting from variation in the firing time of the transponder transmitter.

The linearity of the range sawtooth along with the increased sensitivity resulting from having three voltage-to-range scale factors has given a measured range-indicating accuracy of 0.25 percent of range or 5 feet, whichever is greater.

Since the video pulse will be controlled in amplitude by the AGC, the effect of the rise time in the absence of system noise can be calibrated out of the range measurement. Therefore, only the effect of system noise on the trigger time must be calculated. The results obtained will apply to items 2 and 3. Since both the chaser and the target system have the same receivers and threshold setting, each will have the same threshold jitter.

It has been shown that when the rise time of the video pulse is determined by the system bandwidth. The rms value of the threshold crossing time jitter, δ, caused by noise perturbation of the leading edge may be given by:[3]

$$\delta = \frac{1}{B}\sqrt{\frac{1}{2\ S/N}}$$

where: B = signal bandwidth

S/N = signal to noise ratio

At the maximum range of 60 nautical miles the minimum S/N ratio of 17 db and the system is operating through its 1.5-mc bandwidth. The rms time jitter for these conditions is:

$$\delta = \frac{1}{1.5 \times 10^6}\sqrt{\frac{1}{100}}$$

$$= .0667 \ \mu\text{sec.}$$

Therefore, the rms range error resulting from the threshold crossing jitter at both the range threshold and transponder threshold would be:

$$(.0667)\ (2) = .095 \ \mu\text{sec.}$$

or 47 feet at maximum range.

The error will decrease as the range decreases because of the increase in signal-to-noise.

When the target reaches approximately 18 nautical miles, operation through the 16-mc bandwidth will begin. At this range the signal will be 10.3 db stronger, but the noise will have been increased an equal amount as a result of the increase in bandwidth. Thus, the S/N at 18 nautical miles is the same as that at 60 nautical miles. But, since the bandwidth has been increased, the rms error in the threshold crossing is reduced by

a factor 16/1.5 or 10.66.

Again, if only the threshold crossing error were considered, the range measurement error would be only 47/10.66 or 4.4 feet at 18 miles.

The transmitter in the target radar has a firing delay which is of the form of a random variation superimposed on a fixed delay. The random variation has a rms value of approximately 10 nsec and the fixed delay is approximately 30 nsec. The fixed delay may be calibrated from the system.

The modulator firing jitter is independent of range and represents a constant rms error. Assuming no filtering of the range voltage, the rms (1σ) error in range voltage, considering the firing jitter only, would be $\pm$ 5 ft. A 3σ error would be $\pm$ 15 ft. When this error is considered along with the .25 percent error of the range sweep, a 1 percent overall system accuracy is satisfied down to a range of 2000 ft.; obtained by solving the equation

$$.0025\ R + 15 = .01\ R$$

Jitter errors could be further reduced, if necessary, by filtering the range voltage. Assume a control loop bandwidth, f_s, in the vicinity of 2 cps. With this amount of filtering, the circuit jitter error would be reduced by a factor of

$$\sqrt{\frac{2\ f_s}{PRF}}\ .$$

Or

$$\pm\ 15\sqrt{\frac{4}{250}} = \pm\ 1.9 \text{ feet}$$

One percent system accuracy is now satisfied down to 254 feet (found by solving .0025 R + 1.9 = .01 R).

Range Rate Measurement

Velocity is determined from the derivative of range, by an operational amplifier used as a differentiator and filter. Time constants are adjusted in each of the three range intervals in accordance with the spread of velocities to be expected in each interval.

Receiver noise and jitter in the transponder and tracking circuits induce noise in the range signals. Filtering can be used to reduce this noise. The filter in turn introduces transient and lag errors into the tracking. A proper compromise between noise errors, transient errors, and lag errors must be made to obtain the best possible system.

As was previously shown, there is negligible error in the range measuring pulse due to the threshold crossing jitter. The range measuring system produces a "boxcar" type output which has the following spectrum:[2]

$$G_N(f) = \frac{2}{PRF}\left(\frac{\sin\frac{\pi f}{PRF}}{\frac{\pi f}{PRF}}\right)^2 \left(1 - \frac{\pi}{4}\right) N^2$$

The modulator firing jitter of $\pm$ 15 ft. would be the greatest source of noise. Therefore, with the pulse repetition frequency of 250 cps:

$$G_N(f) = \left(\frac{2}{250}\right)\left(\frac{\sin\frac{\pi f}{PRF}}{\frac{\pi f}{PRF}}\right)^2 (.215)(15)^2$$

$$G_N(f) = .387 \text{ feet}^2/\text{cycle/sec for } f \ll \frac{PRF}{\pi}$$

If a differentiator with filter is used, for example, one which goes from a +1 slope to zero slope at $\omega = 1.25$ to a -2 slope at $\omega = 10$, the noise in the range rate signal can be approximated by the following integration:

$$\frac{.387}{2\pi}\left[\int_0^{1.25} \omega^2\, d\omega + \int_{1.25}^{10} d\omega + \int_{10}^{\infty} \frac{d\omega}{\omega^4}\right]$$

$$= \frac{.387}{2\pi}\left[\frac{(1.25)^3}{3} + 10 - 1.25 + \frac{1}{3\ (10)^3}\right]$$

$$= .58\ \left(\frac{\text{feet}}{\text{sec}}\right)^2 \text{ or } .762 \text{ feet/sec.}$$

Conclusions

An experimental model of this system has been in operation for several months. The measured accuracy has been close to the calculated and confirms that with additional development effort this system should prove to be a useful instrument for cooperative rendezvous navigation. All circuits are transistorized, except for the RF oscillators, providing a light weight low power equipment. The first breadboard equipment is shown in figures 5, 6 and 7. Prototype equipment of this design has been estimated to weigh 18 pounds and would consume about 30 watts for the radar, and 14 pounds and 12 watts for the transponder.

Acknowledgements

This system was developed as a Company-sponsored project and the author wishes to acknowledge the assistance of Messrs. E. D. Green, W. W. Quigley, P. S. Hacker and M. S. Wheeler who made substantial contributions to the design and analysis of the radar.

References

1. Hacker, P. S. and Reuter, H. A., "An Interferometer Radar System for Satellite Cooperative Rendezvous," IRE (PGAP) Washington, D.C., November 21, 1961.

2. MIT Radiation Laboratory Series "Threshold Signals," McGraw-Hill, New York, 1950, Section 10.3.

3. Skolnik, M. I., "Theoretical Accuracy of Radar Measurements," IRE (PGANE), December, 1960.

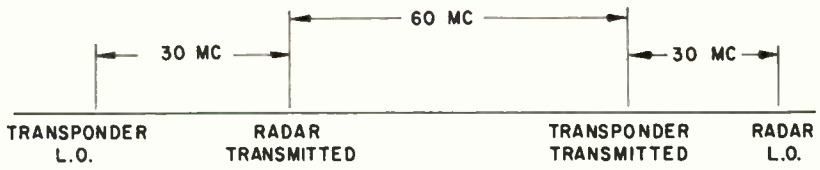


Fig. 1. RF frequencies.

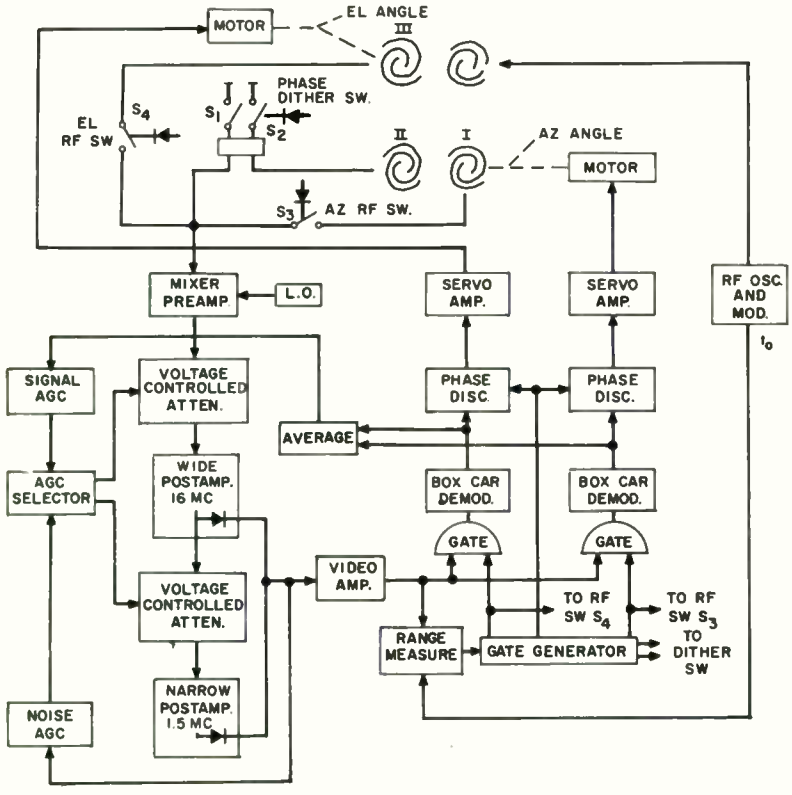


Fig. 2. Radar system block diagram.

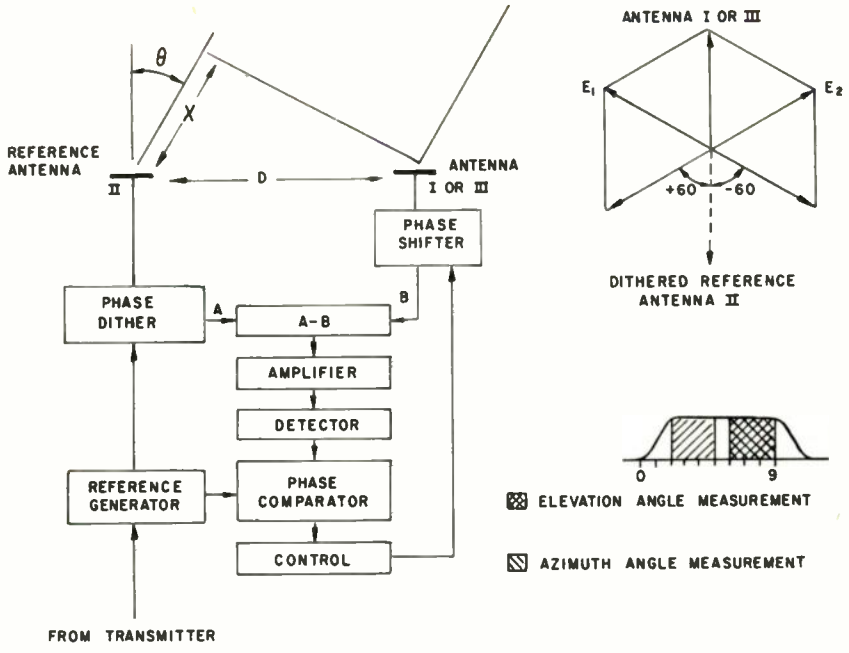


Fig. 3. Angle measuring circuit.

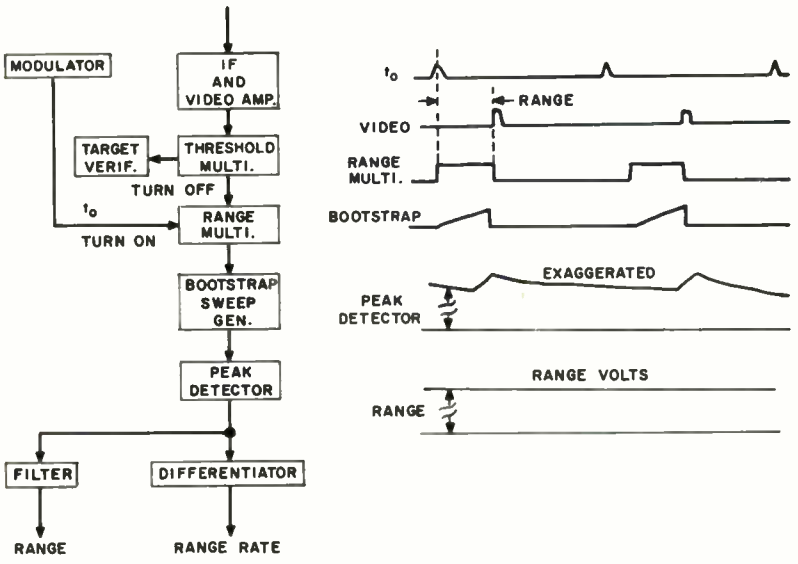


Fig. 4. Range measuring circuit.

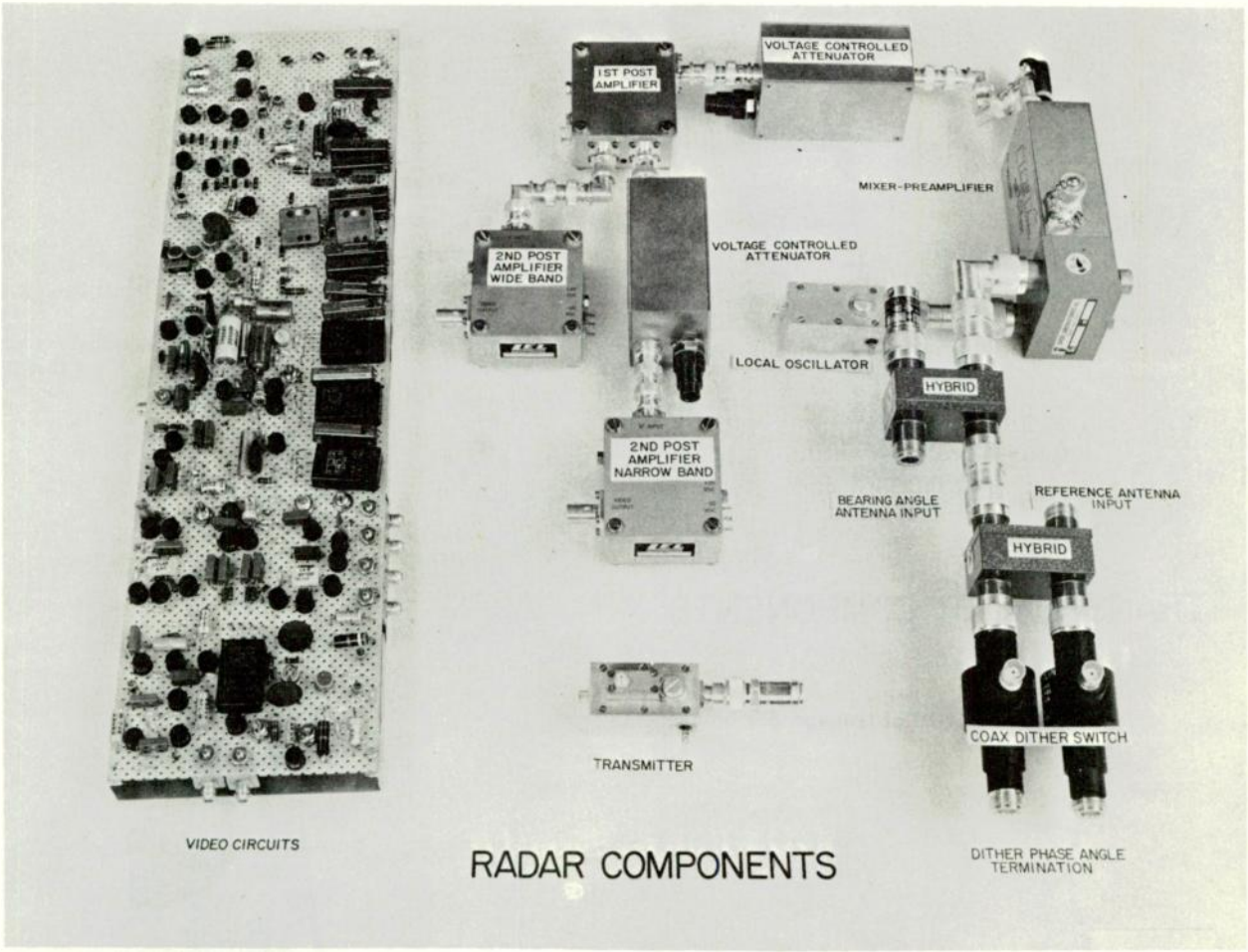


Fig. 5. Layout of radar components.

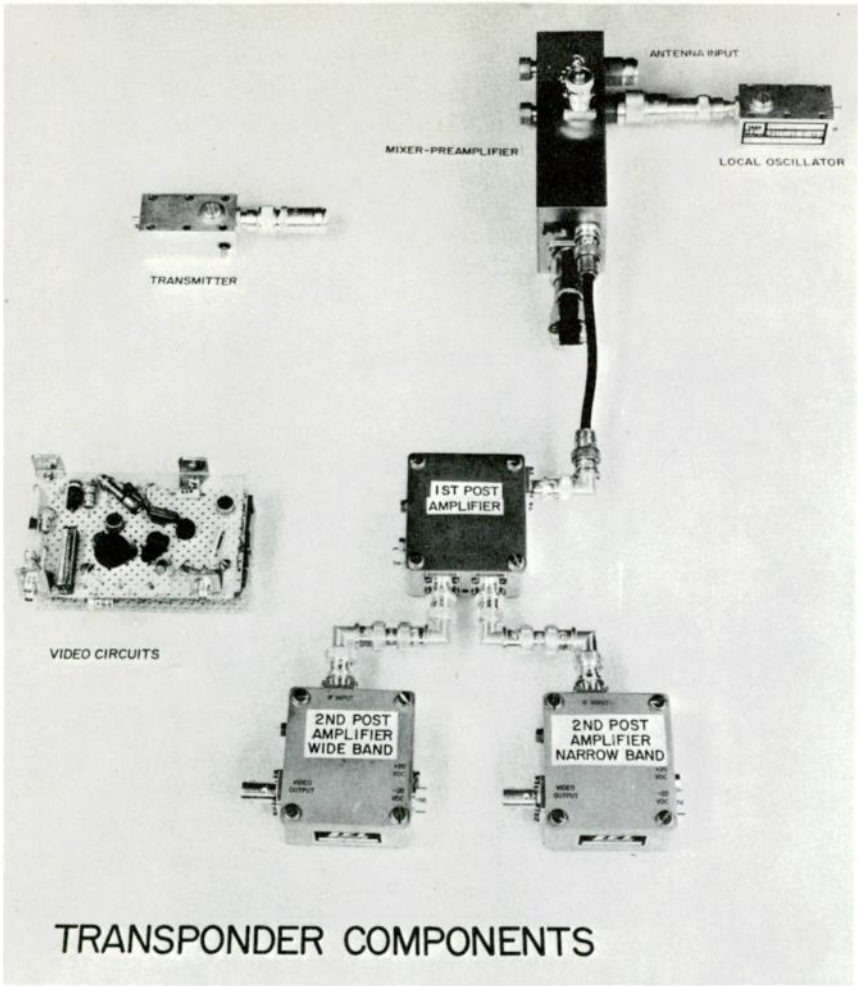


Fig. 6. Layout of transponder components.

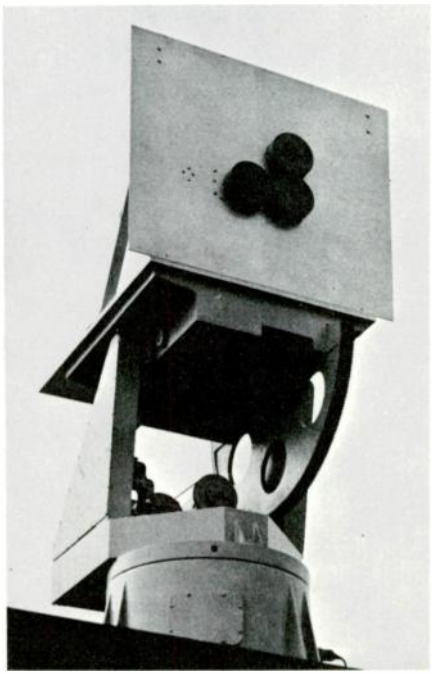


Fig. 7. Receiving antenna assembly.

A SUN-VECTOR REFERENCED EARTH-SATELLITE GUIDANCE SCHEME

Heinz Gabloffsky
Northrop Space Laboratories
Hawthorne, California

ABSTRACT

The analysis of the satellite guidance scheme which uses a sun sensor, is presented together with a numerical evaluation of a particular sun seeker system. Two sun seekers provide detection about the complete sphere. The system can be utilized for three guidance modes, yaw sensing, velocity sensing for tumble measurements, and general vehicle attitude sensing. The sun seekers may be used in conjunction with other sensors or alone. To establish vehicle orientation and position, the time related precomputed components of the sun-vector are stored in the satellite for comparison with measured values from the sun position indicators. The differences between measured and stored data are used in the control loop to initiate attitude corrections.

ACKNOWLEDGMENTS

The preparation of this paper was greatly facilitated by the availability of the IBM 7090 sun vector computer program. The Fortran program was prepared by Mr. L.S. Joyce and Mr. P.W. Soule, of the NORTHROP SPACE LABORATORIES. The writer expresses his appreciation to these gentlemen for their valuable suggestions and discussions.

INTRODUCTION

Within the last two years postulations have been made by space scientists with respect to future space programs. These programs will include space rendezvous, space station assembly in orbit, communication-, meteorological-, defense-, and reconnaissance earth satellites. All of these missions have the requirement for a long life-time and attitude control of the space vehicles within smaller or larger directional tolerances.

The particular application of the sun seeker and the sun-vector referenced guidance scheme as a yaw sensor was conceived for the semi-passive stabilization system of the ORBITAL RENDEZVOUS BASE SYSTEM (ORBS), (Ref. 1). The major goal of ORBS is the mission implementation in the years 1965 to 1967, to provide the means for manned lunar missions and possibly at a later time interplanetary missions. One major stepping stone in the establishment of feasibility of the multiple module rendezvous system was low attitude control fuel requirements during an orbital storage time of as much as 100 days. A synchronous orbit station-keeping maneuver (Ref. 1) every two weeks was another requirement for ORBS and was included in the study. The result of the studies on attitude control and guidance was the selection of a semi-passive stabilization system based on gravity gradient stabilization, which requires an initial precision ORBS-module orientation with the local vertical. The "dumbbell" effect leads to libration of the module axes about the local vertical and the vehicle's center of mass. The libration magnitudes are a function of low external forces acting upon the vehicle of initial conditions after turning off the attitude control. The magnitude and frequency of libration of each vehicle are also functions of its configuration and mass distribution. A limit switch may be operated from the horizon tracker system if the libration amplitude of the simple harmonic motion reaches plus or minus 5 degrees, and either a reactive jet thrust or an inertial reaction wheel is used for the attitude control system to reduce the oscillation.

The vehicle yaw axis is uncoupled for a vehicle configuration of symmetry about the longitudinal axis. Second-order coupling terms may make the yaw axis drift, and an independent sensor is needed to measure the displacement angle and the angular rate about the yaw axis in direct or indirect reference to the vehicle orbit plane and the local vertical. The application of a tandem sun position indicator which employs the sun vector, together with precomputed stored pointing data, is the subject of this paper. In addition to the initial yaw sensing application further studies explored the feasibility of the use of this sun vector scheme to determine the angular velocity vector of a tumbling satellite and to provide a moderate attitude alignment accuracy when applying this scheme as a sole sensor device for an earth satellite. The last two applications may be considered as emergency or redundant sensor systems.

1. NAVIGATION TECHNIQUE

1.1 GENERAL DISCUSSION

Before the coupling of ORBS modules for a space rendezvous maneuver or a space station assembly is started, many essential preparatory operations have to be completed. For example, the synchronous rendezvous-compatible orbit has to be established for each module in the injection phase, and this orbit has to be maintained through station-keeping maneuvers if the orbital storage time

exceeds a two-week period. A rendezvous transfer maneuver, terminal maneuvers, and docking or coupling maneuvers may follow the preceding operations. All maneuvers, except one -- the injection maneuver -- require the orientation of the vehicle axes with a self-contained attitude control system in respect to a selected reference frame. For a manned vehicle mission it may be permissible to eliminate some of the decision-making electronics, but certainly the vehicle will have some measuring devices such as horizon tracker, sun position indicators or other sensors.

The following chapters discuss the salient features of the attitude sensors and controls as used in conjunction with orbit-maintaining and rendezvous maneuvers which are exemplified in the ORBS concept.

1.2 ORBIT-MAINTAINING MANEUVERS

The rendezvous ascent or injection operation of each module for ORBS is similar to that of direct ascent into a parking orbit, however, the last stage of the booster is used to orient the module payload with the longitudinal axis pointing toward geocenter, and with the pitch axis of the vehicle placed normal to the orbit plane. The transfer maneuver and the lunar escape trajectory is the result of a complex flight mechanical analysis, and the details are discussed at other places (Refs. 1, 2, & 3). For the discussion of the orbit-maintaining maneuver it is only necessary to know that the selected orbit parameters will establish a synchronous rendezvous-compatible orbit of approximately 31° inclination, 263 n mi altitude, and an eccentricity less than 0.005. The selection of this rendezvous-compatible orbit permits an economic launch twice a day from the Atlantic Missile Range with a maximum payload, without any plane changes in the ascent or injection maneuvers, and it will place the added module in close to previously launched modules.

The station-keeping maneuver will restore the lost energy of angular momentum by means of a limit cycle, as shown on Fig. 1. These energy-increasing Hohman transfers are made in intervals of two weeks. The thrust level and the amount of propellant needed is extremely small. The maneuver can be accomplished with the attitude control jets. The required thrust level is approximately 1 to 2 lbs. for a $3 \cdot 10^4$ slug ft^2 pitch moment of inertia. The required two increases of orbital velocity will be 2 to 4 fps for each cycle.

To perform these station-keeping maneuvers the individual modules must have the following capabilities:

1. An attitude sensor and control system that, on command, permits orientation of the vehicle pitch and roll axes with respect to the local vertical to better than .1 degree, and the yaw axis in respect to the orbit plane within less than .6 degrees.
2. The attitude control system must be capable of thrusts on command tangent to the orbital motion.
3. A beacon or transponder should identify the module and assist in the ground-tracking. A proximity detector should provide emergency command signals and a command telemeter link should transfer the ground base commands to the attitude and station-keeping maneuver control loop.

The enclosed station-keeping control system block diagram illustrates the signal flow between the subunits (figure 2). It is of importance to the system designer to recognize the requirement for highest attitude sensor reliability. Therefore, a complete set of redundant sensors is proposed for the ORBS modules. The primary sensor set combines the sun seeker yaw sensor with four horizon trackers and the standby set consists of a two-axis yaw gyro and an integrating rate gyro.

1.3 ORBITAL RENDEZVOUS TECHNIQUE

The injection guidance together with the synchronous rendezvous-compatible orbit and the station-keeping maneuver will place all ORBS modules in a cluster of dimensions shown on Fig. (3). The approximate dimensions are 100 n mi along the orbit, and 0.4 n mi depth and height.

All impulses for module transfers should be small enough, and must not exceed the limit cycle and the space boundaries of the station-keeping scheme. The ORBS system considers thrust levels of twenty pounds (attitude jets) as adequate and the module transfer maneuver may extend in time over several orbital periods. It is assumed that during this maneuver the line-of-sight connection (radar or optical sight) is maintained continuously between the manned command module which performs or commands the complete space rendezvous and the module which will be assembled to the command module or to already previously attached modules of the space station.

It is not the intention of this paper to describe details of the guidance used for the transfer, positioning, indexing, and coupling of the modules, but rather define the attitude sensing and control methods, because they are prerequisites for the performance of these maneuvers. The paper will, in particular, emphasize the use of the sun-vector reference by means of a sun-position indicator or sensor and how it is integrated into the earth satellite guidance scheme.

The selected reference frame is shown on Fig. (4). In the right-handed coordinate system the space vehicle center of mass is the point of origin; the x axis is placed normal to the orbit plane; and the y axis is located in the orbit plane; the z axis is directed from the geocenter to the satellite center of mass along the local vertical and completes the rotating orthogonal triad of reference with the geocenter as point of

rotation. In case all vehicle axes fall in place with this reference system, the respective vehicle axes are defined as:

x^T = pitch axis x^B

y^T = roll axis y^B

z^T = yaw axis z^B

where T = Trajectory, B = Vehicle } coordinates

As stated in Chapter 1.2, the vehicle module axes will be aligned initially with the reference frame within the following tolerances:

1. Pitch and Roll axes
$$|\theta_{2_{lim}}| = |\theta_{1_{lim}}| \leqq \pm .1°$$
$$|\dot{\theta}_{2_{lim}}| = |\dot{\theta}_{1_{lim}}| \leqq 2 \cdot 10^{-4} \text{ rad/sec}$$

2. Yaw axis
$$|\theta_{3_{lim}}| \leqq \pm 0.5°$$
$$|\dot{\theta}_{3_{lim}}| \leqq 8 \cdot 10^{-3} \text{ rad/sec}$$

where θ_1, θ_2, θ_3 are angle of rotation about the x^B, y^B, z^B axes, respectively. These tolerances apply also to the self-contained attitude control system in each module.

After the initial vehicle orientation is completed the active control system is switched to a passive standby mode where all mechanical motions in the vehicle and the autopilot circuits are inactive, but the horizon tracker subsystem and the sun seeker's subsystem continue to perform their measurements.

The individual module, after being placed in orbit, is oriented in pitch and roll by the dumbbell effect. The perturbation or external forces together with residuals of initial conditions acting upon the vehicle make the pitch and roll axes librate about the center of mass of the vehicle. The libration periods are approximately 50 minutes, and they are nearly one-half of the ORBS orbital period. The libration amplitudes are a function of external forces, moment of inertia of the vehicle and the available means of damping. In case the amplitude or excursion exceeds a predetermined limit, the limit switch activates the control loop. In the meantime the yaw z^B vehicle reference may have drifted from its initial orientation and has to be reoriented with the z^T reference of the trajectory coordinate system. The sensing of the yaw displacement and displacement rate of the vehicle is performed by a sun-vector measuring subsystem. The initial orientation of the vehicle is maintained throughout all maneuvers, including the final coupling operation. Only after the ORBS vehicle is completely assembled, a new reference of orientation may be selected. This certainly is done for the orbital launch to escape from the earth orbit and to begin the journey of a lunar or interplanetary mission.

The sun will be visible 60 to 70% of each orbit. This percentage has been computed (Ref. 9) and was also established analytically by several authors (Ref. 10, 11, 12). For a maximum angular velocity of

$$\dot{\theta}_3 = 8 \cdot 10^{-3} \text{ rad/sec.}$$

the vehicle will perform without active control multiple revolutions about the yaw axis during the sun's occultation time of 1680 to 2850 seconds. To avoid tumbling of the vehicle, the attitude control together with the horizon tracker has to stay activated during the "dark" period. Any guidance maneuvers, for example orbit maintaining maneuvers, will be made in the sunny portion of the orbit and the yaw sensor orients the vehicle into the desired thrust directions.

2. ASTRODYNAMIC CONSIDERATIONS.

The exact mathematical model which describes the sun vector in respect to time and a rotating satellite coordinate system is rather complex and it is cumbersome to work with.

The particular applications of the sun position indicator for space system rendezvous operations in an orbit of low eccentricity and negligible air drag make it permissible to simplify the mathematical model when the following assumptions are made: (Ref. 9)

1. The earth moves about the sun with a constant angular velocity and at a mean distance of 80.7 million nautical miles (error $< 4 \cdot 10^{-3}$%).

2. The parallax effect between the line-of-sight from the sun to earth center and the sun to satellite center can be neglected (angular error $\sim$ 15 seconds of arc).

3. The motion of the satellite is circular and only first order secular effects of the earth's oblateness are considered in the mathematical model.

4. The vernal equinox, as reference, is fixed in space.

5. The obliquity of the ecliptic is constant, 23.45° (error $< 4 \cdot 10^{-3}$%).

The coordinate system and the orthogonal transformation of coordinates are briefly discussed in the next chapter.

2.1 COORDINATE SYSTEMS.

A right-hand orthogonal sun-fixed inertial coordinate system is used as reference frame, where the unit vector $\vec{k}^o$ is oriented along the vernal equinox, j^o is in the ecliptic plane in the direction of increasing time, and $\vec{i}^o$ is located normal to the ecliptic.

When the direction of the sunlight is used as a navigational aid, it is necessary to relate the sun vector to an orthogonal trajectory coordinate system with the center of mass of the satellite as point of origin (see Fig. 5).

Four orthogonal transformations refer the sun vector from the inertial coordinate system to the trajectory coordinate system in which the x^T axis is normal to the orbit plane, the y^T axis is in the orbit plane with the positive direction identical to the satellite motion, and the z^T axis is parallel to the local vertical (see Fig. 6). It must be recognized, however, that the (x^T, y^T, z^T) frame is a reference frame only to which the orthogonal physical body axes (x^B, y^B, z^B) may be related for the sun vector referenced attitude measurements. In case all body axes are aligned with the T-coordinate system the precomputed directional cosine values of the sun vector are identical to those values measured in the sun position indicator. Therefore, since it is desired to measure and control the displacement and rates between the T- and the B- coordinate system it was required to write a computer program to acquire precomputed data of the T-coordinate system. This 7090 IBM computer program is explained in detail in Refs. 9 and 13.

2.2 SUN VECTOR AND ORBIT PARAMETER RELATIONS.

The geometrical representation of the coordinate transformation is shown on Fig. 6. The transformation is obtained through rotation of the coordinate system about the inertially fixed unit vector of the sun-satellite line-of-sight.

The rotations are about:

$$\begin{aligned} z &\rightarrow +\beta \\ x &\rightarrow -\Omega \\ z &\rightarrow +i \\ x &\rightarrow -u \end{aligned} \qquad (1)$$

Let $C_{a\ldots z} = \cos a\ldots z$ and $S_{a\ldots z} = \sin a\ldots z$, then the sun vector can be obtained directly in reference to the T-system from the following expression:

$$\begin{bmatrix} x^T \\ y^T \\ z^T \end{bmatrix} = \begin{bmatrix} [\,C_i\,C_\beta - S_i\,C_\Omega\,C_\beta\,] & | \\ [\,S_\beta\,(S_u S_\Omega - C_u C_i C_\Omega) - S_i C_u C_\beta\,] & | \\ -[S_\beta\,(S_u C_i C_\Omega + C_u S_\Omega) + S_u S_i C_\beta\,] & | \end{bmatrix}$$

$$\begin{bmatrix} [C_i\,S_\beta + S_i\,C_\Omega C_\beta] & | \\ [C_\beta\,(C_u C_i C_\Omega - S_u S_\Omega) - C_u S_i S_\beta] & | \\ [C_\beta\,(S_u C_i C_\Omega + C_u S_\Omega) - S_u S_i S_\beta] & | \end{bmatrix}$$

$$\begin{bmatrix} [-S_i\,S_\Omega] \\ [S_u C_\Omega + C_u C_i S_\Omega] \\ [C_u C_\Omega - S_u C_i S_\Omega] \end{bmatrix} \begin{bmatrix} x^o \\ y^o \\ z^o \end{bmatrix} \qquad (2)$$

$$\begin{bmatrix} x^T \\ y^T \\ z^T \end{bmatrix} = DCBA \begin{bmatrix} x^o \\ y^o \\ z^o \end{bmatrix} \text{ and } \vec{e}^{\,T}_{ss} = E\,\vec{e}^{\,o}_{ss} \text{ where}$$

$$E = DCBA \text{ and } \vec{e}^{\,o}_{ss} = -(\cos\Lambda\vec{k}^o + \sin\Lambda\vec{j}^o)$$

The angular velocity relationships for the same rotations are given by:

$$\vec{W} = \left(\frac{\vec{du}}{dt}\right) + A\left(\frac{\vec{di}}{dt}\right) + BA\left(\frac{\vec{d\Omega}}{dt}\right) + CBA\left(\frac{\vec{d\beta}}{dt}\right) \qquad (3)$$

where:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & C_u & -S_u \\ 0 & S_u & C_u \end{bmatrix} \qquad B = \begin{bmatrix} C_i & S_i & 0 \\ -S_i & C_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (4)$$

For all practical purposes:

$$\left(\frac{\vec{d\beta}}{dt}\right) = 0; \quad \left(\frac{\vec{di}}{dt}\right) = 0 \qquad (5)$$

then

$$\vec{W} = \left(\frac{\vec{du}}{dt}\right) + BA\left(\frac{\vec{d\Omega}}{dt}\right) \qquad (6)$$

and therefore

$$\vec{W} = \begin{bmatrix} -\dot{u} \\ 0 \\ 0 \end{bmatrix} + \left[\begin{array}{c|c|c} C_i & S_i & 0 \\ -S_i C_u & C_u C_i & -S_u \\ -S_i S_u & S_u C_i & C_u \end{array}\right] \begin{bmatrix} -\dot{\Omega} \\ 0 \\ 0 \end{bmatrix} =$$

$$\begin{bmatrix} -\dot{u} - C_i \dot{\Omega} \\ -S_i C_u \dot{\Omega} \\ -S_i S_u \dot{\Omega} \end{bmatrix} = \begin{bmatrix} W_{x_T}^T \\ W_{y_T}^T \\ W_z^T \end{bmatrix} . \qquad (7)$$

or, formally expressed, one obtains:

$$\vec{W} = \begin{bmatrix} -1 & -C_i & 0 \\ 0 & -S_i C_u & 0 \\ 0 & -S_i S_u & 0 \end{bmatrix} \cdot \begin{bmatrix} \dot{u} \\ \dot{\Omega} \\ 0 \end{bmatrix} = \begin{bmatrix} W_x^T \\ W_y^T \\ W_z^T \end{bmatrix} \qquad (8)$$

Therefore, the following equations are used to determine the sun vector components, angular velocities relations, the shadow time, and whether or not the sun is visible.

$$\Lambda = \Lambda_o + Kt \qquad (9)$$

$$\text{Let } M = \left(\frac{G_2}{r^3}\right)^{\frac{1}{2}} \left[1 + \frac{J}{\hat{r}^2} (3 - 2.5 \sin^2 i)\right] \qquad (10)$$

then: (see List of Symbols p ii)

$$u = u_o + Mt \qquad (11)$$

$$\dot{u} = M$$

$$\Omega = \Omega_o - \frac{J}{\hat{r}^2} \cos i \, M \, t \qquad (12)$$

$$\dot{\Omega} = -\frac{J}{\hat{r}^2} \cos i \, M$$

and

$$\begin{bmatrix} \ell \\ m \\ n \end{bmatrix} = -E \begin{bmatrix} C_\Lambda \\ S_\Lambda \\ 0 \end{bmatrix} \qquad (13)$$

$$-\ell = C_\Lambda (C_i C_\beta - S_i C_\Omega C_\beta) + S_\Lambda (C_i S_\beta + S_i C_\Omega C_\beta)$$

$$-m = C_\Lambda [S_\beta (S_u S_\Omega - C_u C_i C_\Omega) - S_i C_u C_\beta]$$
$$+ S_\Lambda [C_\beta (C_u C_i C_\Omega - S_u S_\Omega) - C_u S_i S_\beta] \qquad (14)$$

$$-n = -C_\Lambda [S_\beta (S_u C_i C_\Omega + C_u S_\Omega) + S_u S_i C_\beta]$$
$$+ S_\Lambda [C_\beta (S_u C_i C_\Omega + C_u S_\Omega) - S_u S_i S_\beta]$$

where:

ℓ, m, n are the directional cosines of angles

$\measuredangle \alpha$ in the x-y plane.

$\measuredangle \beta$ in the x-z plane.

$\measuredangle \gamma$ in the y-z plane, respectively.

Let

$$\ell_b = -\left(1 - \frac{1}{\hat{r}^2}\right)^{\frac{1}{2}} \qquad (15)$$

then if

$\ell > \ell_b$ the sun is visible

and $\ell \leq \ell_b$ the sun is obscured

The expression for the shadow time, TS, is defined in References 9 and 13 to

$$TS = \frac{\cos^{-1}\left[\frac{1}{\hat{r}} \left(\frac{\hat{r}^2 - 1}{1 - n^2}\right)^{\frac{1}{2}}\right]}{\pi} \qquad (16)$$

where $|n| < \frac{1}{\hat{r}}$

and TS = 0 when $|n| \geq \frac{1}{\hat{r}}$.

The details of the 7090 program are obtainable from the references. They will not be discussed here. However, several boundary cases are explained which may be encountered when the sun vector referenced guidance scheme is applied.

As conceivable from figure (7) if ℓ = const, then m and n are periodical and they are $\pi/4$ periods displaced in phase. If ℓ = f(t), m and n remain periodical but the simple harmonic oscillation is distorted where the distortion is a function of ℓ = f(t). For simple harmonic oscillations the following relation yields:

$$m^2 + n^2 = 1 - \ell^2. \qquad (17)$$

A change of ℓ^2 therefore will change the magnitude on the left side of the equation without changing the periodicity and phase relation between m and n. For constant value ℓ, a point defined by the left hand side of equation (17) describes over one orbit period a circle at the z - y plane and the radius of that circle is a function of $(1 - \ell^2)$.

As shown in the following the distortion over one orbital period is small in comparison to the functional values of the simple oscillation amplitude or the related $(m^2 + n^2)$ values.

The distortion effects may be obtained from the functional relation of equation (14)

$$-\ell = C_\Lambda C_i C_\beta - C_\Lambda S_i C_\Omega C_\beta + S_\Lambda C_i S_\beta + S_\Lambda S_i C_\Omega C_\beta \qquad (18)$$

which, after several manipulations, may be written

$$-\ell = \sqrt{2}\, C_\beta S_i \left[C_{(\frac{\pi}{4}+\Lambda)} (CTN_i + C_\Omega) + \frac{1 + TAN_\beta}{\sqrt{2}} CTN_i S_\Lambda \right] \qquad (19)$$

where

$C_{(\frac{\pi}{4}+\Lambda)} = \cos(\frac{\pi}{4} + \Lambda)$

CTN_i = Cotangent of $\sphericalangle i$

$TAN_\beta = \tan \beta$

After substituting the numerical values for $\sqrt{2}$ and $\beta = 23.45°$ one obtains

$$-\ell = 1.297398\, S_i \left[C_{(\frac{\pi}{4}+\Lambda)} (CTN_i + C_\Omega) + 1.013842\, CTN_i S_\Lambda \right] \qquad (20)$$

Since i is a constant value, the equation contains three variables, $-\ell$, Ω, and Λ. For obvious reasons Ω and Λ are periodic functions where Ω will be of a much higher frequency than Λ.

The oblateness of the earth causes the ascending node of the satellite orbit to regress. The regression $\Delta\Omega$ for each orbit period T_s is a function of the satellite radius and the inclination angle, i, of the orbit plane.

$$\frac{\Delta\Omega}{REV} = \frac{-2\pi J_2}{\left(\frac{r_s}{r_e}\right)^2} \cos i \text{ (radians)} \qquad (21)$$

In the figure (8) the relation is plotted for $\Delta\Omega$/REV as a function of $\hat{r}$ and, i, the inclination angle. A reasonable range of earth satellite radii is 3700 n. miles to 6000 n. miles. Within this range of $\hat{r}$ and i = 90 to 0 degrees, the values for $\Delta\Omega$/REV vary from 0 to .6 degrees respectively. For a planned satellite mission it is assumed that the inclination angle and the satellite orbit radius is known within ±.5 degrees and ±10 n. miles respectively. For these values it is permissible to use a proportional regression rate $\Delta\Omega/T_S$ (degree/sec) in the numerical evaluation of equation (20).

The orbit period time for satellites with r = 3700 - 6000 n. miles range from approximately 5600 seconds to 9500 seconds. The corresponding values of

$$\frac{\Delta\Lambda}{T_s} = \frac{360\, T_s}{3.1558 \cdot 10^7} \text{ degree/period} \qquad (22)$$

$$\frac{\Delta\Lambda}{T_s} = \frac{3.60 \cdot .5600}{3.1558\ 10^7} = .06388 \text{ degrees/period for r = 3700 n. miles.}$$

$$= \frac{360 \cdot .9500}{3.1558 \cdot 10^7} = .10837 \text{ degrees/period for r = 6000 n. miles.}$$

The small angular rate of $\Delta\Lambda$ = 0 to .1 degrees for one orbit revolution makes it permissible again to use a proportional

rate of $\frac{\Delta\Lambda}{T_s}$ = K in the numerical evaluation of Eq. (20).

Therefore, Eq. (20) may be written in the form:

$$-\ell = -\ell(t) = 1.297398\, I_1 \left[C_{(\frac{\pi}{4}+\Lambda_o+Kt)} \left(I_2 + C_{(\Omega_o + \frac{\Delta\Omega}{T_s} \cdot t)} \right) + 1.013842 I_2 S_{(\Lambda_o + Kt)} \right] \qquad (23)$$

where I_1 = const = sin i

I_2 = const = cot i

and, further, after the summing of angles and after like terms are combined:

$$-\ell(t) = 1.013842 I_2 \left(I_7 C_{(Kt)} + I_8 S_{(Kt)} \right) + .648699 I_1 \left[2I_2 (I_3 C_{(Kt)} - I_4 S_{(Kt)}) + C_{(K + \frac{\Delta\Omega}{T_s})t} (I_3 I_5 - I_4 I_6) - S_{(K + \frac{\Delta\Omega}{T_s})t} (I_4 I_5 + I_3 I_6) + C_{(K - \frac{\Delta\Omega}{T_s})t} (I_3 I_5 + I_4 I_6) - S_{(K - \frac{\Delta\Omega}{T_s})t} (I_4 I_5 - I_3 I_6) \right] \qquad (24)$$

where:

$I_3 = \cos\left(\frac{\pi}{4} + \Lambda_o\right)$ and $K = 1.9923 \times 10^{-7}$ rads/sec

$I_4 = \sin\left(\frac{\pi}{4} + \Lambda_o\right)$ $\quad \frac{\Delta\Omega}{\text{Rev}} = f(i) = \text{const}$

$I_5 = \cos \Omega_o$

$I_6 = \sin \Omega_o$ $\quad T_s = \text{const}$

$I_7 = \sin \Lambda_o$ $\quad \frac{\Delta\Omega}{T_s} = \text{const}$

$I_8 = \cos \Lambda_o$

Values for i and T_s are selected from orbit parameters before launch.

Eq. (24) may be written:

$$-\ell(t) = f(t) \tag{25}$$

Therefore, Eq. (17) is now:

$$r^2 = m^2 + n^2 = f(t) = 1 - [\ell(t)]^2 \tag{26}$$

From geometrical observations we find the radius vector $\vec{r}$ rotating on the z-y plane of the T coordinate system with the rate $\dot{u}$. Since $u = f(t) = u_o + Mt$, we may define the position vector

$$\vec{r} = \sqrt{1 - [\ell(t)]^2} \cdot u \tag{27}$$

where $\sqrt{1 - [\ell(t)]^2}$ is the modified x^T sun vector component and u is a real variable of the position vector $\vec{r}$. Equation (27) is similar to the expression of the linear spiral

$$r = a\bullet \tag{28}$$

however, here $a = f(t) = \sqrt{1 - [\ell(t)]^2}$, and therefore a is not a proportional variable. From Equation (26)

$$r = \sqrt{m^2 + n^2} \tag{29}$$

and referring to equation (7) the components of the velocity vector in the $y^T - z^T$ plane are given by

$$W_y{}^T = - S_i C_u \dot{\Omega}$$
$$W_z{}^T = - S_i S_u \dot{\Omega} \tag{30}$$

Since S_i and $\dot{\Omega}$ are constant values the variables C_u and S_u provide simple oscillations of the vector components about the y and z axes. Here again equation (29) provides the resultant component $\sqrt{(m^2 + n^2)}$ of the sun vector and the two equations (30) provide the velocity component vectors of position vector $\vec{r}$.

In a unit sun vector diagram r can never exceed unity. This condition is confirmed in equation (27) where $\sqrt{1 - [\ell(t)]^2}$ can vary between zero and unity and r varies also within these boundaries.

2.3 APPROXIMATION INACCURACIES

The secular accelerations forced upon the earth by nearby located planets may affect the validity of the assumption that the earth moves about the sun with a constant velocity and the obliquity of the ecliptic is constant. A first order evaluation is sufficient to clarify this question. The average value in magnitude of secular perturbation is obtainable from the time difference between the tropical and the sidereal year when expressed in percent to the base of the tropical year. This average value amounts to an error of 4×10^{-3} percent.

The parallax effect between the earth-sun and satellite-sun line-of-sights amounts to an error of 15 arc seconds for the maximum orbit radius of 6000 n. miles. This error is 120 times smaller than the angular accuracy of $\pm 0.5°$ required for the yaw axis measurement. The higher order effects on the orbit caused by secular effects of the earth's oblateness are extremely small for orbit eccentricities of $<.005$ and they are also difficult to measure on satellites in orbit as has been reported in several papers. The actual effects on the orbit plane inclination angle, i, caused by perturbation can be obtained from the SRI paper by J. L. Brenner, et.al. (Ref. 14). The authors measured the variation of i on the Vanguard satellite $1958\beta_2$ over a period of eighty days. The variation from its mean value was determined to 0.006°.

For a planned mission all constant values are precomputed before launch. The equation (24) may read then

$$-\ell(t) = K_1 C_{(Kt)} \pm K_2 S_{(Kt)} + K_3 C_{(Lt)} - K_4 S_{(Lt)} + K_5 C_{(Qt)} - K_6 S_{(Qt)} \tag{31}$$

where

$$L = K + \frac{\Delta\Omega}{T_s}$$

$$Q = K - \frac{\Delta\Omega}{T_s}.$$

There is only one particular position where K, L and Q are alike and that is when $i = 90°$. In

this case $I_1 = 1$ and $I_2 = 0$ and equation (31) simplifies to

$$-\ell(t)_{i=90°} = (K_3 + K_5)C_{(Kt)} - (K_4 + K_5)S_{(Kt)} \qquad (32)$$

Therefore it is permissible to have on board for the computation of $-\ell(t)$ only one storage of sine/cosine functions which may be usable for data pick-off simultaneously for all six trigonometric functions needed for the computation.

The accuracy of computation is not reduced although the values of K, L and Q may vary slightly from their mean value, because always pairs of cosine/sine functions are used which compensates for angular errors (highest sensitivity of sin at $0 \pm k\pi$ and of cosine $\frac{\pi}{2} \pm k\pi$).

3. SUN SENSOR

The sun position indicator is an optical device which gathers the sun's radiation and detects the direction of the sunlight in reference to the optical axis of the lens system. The optical axis is identical to the x^B body axis of the previously described x^B, y^B, z^B coordinate system. The two indicators are mounted diametrically to each other, each lens pointing with its entrance aperture in the opposite direction. Both units are mounted fixed to the space vehicle and each unit searches the sky with a full field angle of 187 degrees. Thus, the two sensors have an overlapping field of view of 14 degrees and they cover the entire solid angle of 4π.

3.1 SENSOR INSTRUMENTATION

The sensor will be instrumented in accordance with the performance requirement for the particular application. There are three applications envisioned for the sun position indicator, 1) as a yaw sensor, 2) as an angular velocity sensor, and 3) as an attitude sensor of moderate accuracy. All three applications require the main or coarse sensor instrumentation which is described in this chapter. The yaw sensor and the angular velocity sensor will use the coarse and the fine or vernier sensor in tandem arrangements. The vernier sun sensor will be explained in chapter 4.

The basic components of the sun position indicators are:

1. Lens system
2. One rotating partially reflecting mirror
3. One rotating completely reflecting mirror
4. Two anamorphic lens systems
5. Two radiation tracking transducers

A typical schematic of the arrangement of components is shown in Figure 9. The required lens may have the following specifications:

Equivalent focal length	6.5 mm
Full field angle	187 degrees
Image size	.720 inch diameter
Relative Aperture	f/6.3 or less
Resolution varies between 60 and 100 lines per min.	
Distortion 67% at 75° and 23% at 45° half angles.	
Physical size:	2" length and approximately 3" diameter

The partially reflecting mirror is placed between the lens system and the first member of the anamorphic lens system. This mirror, mounted in a frame, is rotated normal to the optical axis where the rotation axis is parallel and in line with the z^B axis of the space vehicle. A digital servo motor drives the shaft of the mirror frame in such a way that the sun radiation passes through the anamorphic lens along the cylinder axis. This places the center of the focal line always in the middle of the width of the radiation tracking transducer but permits the length of the focal line, which is generated by the anamorphic lens system, to sweep freely the full height (.760 inches) of the transducer. It must be emphasized that the direction of the sunlight in the z axis is not affected by the rotation of the partially reflecting mirror and the readout accuracy is not reduced by the mirror motion. The use of the mirror eases the optical requirement of the anamorphic lens system in the direction normal to the cylinder axis because the image rays' entrance height and entrance angles normal to the cylinder axis are reduced considerably.

The same arrangement is provided for the second completely reflecting rotating mirror, only the axis of rotation is now in coincidence with the y^B axis of the space vehicle.

The block diagram figure 10 illustrates the flow of information for the sun position indicator. The indicator system with $+x^B$ generates a positive binary signal and the one with $-x^B$ generates a negative binary signal. The two way gate connects the signal generating transducer pair with the next stage of the electronic circuit. In the area of overlapping field of view, the two way gate

samples in preset sequence the signal output from each transducer pair. The rest of the diagram explains itself.

As indicated before, the anamorphic lens system is essentially an improved cylinder lens. It transforms the sun image spot of a spherical lens system into a focal line thus stretching the image spot into a long line which extends over the complete width of the .760 inches wide radiation transducer. The center portion of this focal line width can be reduced beyond the minimum angular sun size, however, toward both sides of the center the focal line "fans" out, it widens and spreads the radiation over a greater line width. The magnitude of the spread is a function of the design parameters and the correction quality of the optical system. It will be shown in section 3.3 that it is important to place the smallest grid of the transducer in the center and use the smallest line width together with the maximum binary read out.

3.2 THE ANALOGUE OPTICAL SENSOR READ OUT.

To be able to compare the analogue versus digital transducer, it is necessary to explain the application of the analogue sensor to a sun position indicator.

A simple sun position indicator is feasible when an analogue radiation transducer is used. In this case the spherical lens system alone is used and the radiation transducer is placed at the focal plane of this lens. However, the simplicity does not enhance performance, reliability, nor will it provide good rate measurements. Nevertheless, for the application of strictly position indications of celestial bodies relative to a satellite this type of sensor may be fully adequate. At the end of section 3.3 it will be substantiated why the binary read out solid state radiation tracking transducer was selected for the particular application of the sun-vector referenced guidance scheme.

The analogue radiation tracking transducer (for example: Electro-Optical Systems, Inc., Pasadena, Calif. or Micro Systems, Inc., San Gabriel, Calif) is a solid state single element silicon photovoltaic detector which determines the linear displacement of the center of an image spot in respect to a precalibrated zero - or center position. When the transducer is placed at the focal plane of a lens system, the coupled unit is capable to detect angular position of a light source relative to the optical axis and a selected two dimensional coordinate system, for example: y^B, z^B.

If a double channel amplifier and computer is added to the angular position measuring unit, the sun position indicator is complete.

The radiation transducer is a silicon p - n junction device which makes use of the flow of lateral photo currents, which is different from the conventional p-n junctions which use the transverse current flow. The lateral photo currents flow parallel to the p-n junction and the electrical read out is made by contacts placed on the periphery of the n crystal region. The off-center position of a light spot is determined from the measurement of the difference of voltage produced at the x or y terminals as shown on figure 11. A light spot centered on the calibrated zero position produces zero voltage output.

The voltage output signal is not proportional to the displacement, however, a correction factor may be obtained from the calibration grid pattern. One typical grid pattern is shown on the enclosed figure 12, Reference 15.

Since a 187 degrees full field of view of the spherical lens system must be used to have full 4π spherical coverage with the two sun seekers, the lens again has a non-linear distortion effect as illustrated in the photograph of the Los Angeles Airport which was made with the PAXAR periphoto lens (Reference 16).

The computation combining the two areas correction factors require large computer storage capacities since both corrections are nonlinear and a point to point fit to the linearization curve has to be made for each measurement in very short time sequence. (<1 millisecond). This process may lead also to decreased accuracy of read out resolution. There are several secondary analogue detector effects of unfavorable conditions when compared to the binary receptor. Therefore, the next section 3.3 will compare the direct digital detector read out with the analogue transducer when either of them is applied to the sun position indicator.

3.3 THE DIGITAL OPTICAL SENSOR READ OUT

The instrumentation which uses a digital transducer read out has been explained and illustrated in section 3.1.

The focal line generated by the anamorphic lens is placed directly on the y^B or z^B radiation transducer. The spherical lens system together with the anamorphic lens system provide image distortion and therefore the focal line will be curved at the upper and lower rim of the transducer and will be straight at the center. The curvature depends on the optical design of the complete unit. For best readout performance it is necessary to have the photo etching of the transducer grid follow the curvature of the focal line at the respective image height. This condition together with the photo etched grid combination is shown in the figure 13.

Each transducer consists of six sections and the dividing lines of each section are normal to the scan of the focal line. The y^B transducer is rotated and tilted 90 degrees from the z^B transducer (Fig. 9). There are 32 pairs of conductive and nonconductive lines located on the photo-etched grid pattern of section one. The other sections 2, 3, 4, 5 and 6 have 16, 8, 4, 2, and 1 pairs of

photo-etched surfaces respectively. The read out resolution, therefore, is

$$\frac{\pi}{2^N-1} = \frac{180}{2^6-1} = \frac{180}{63} = 2.86^\circ \qquad (33)$$

This is the resolution at the center. The resolution at the rim or at ±90° field angle depends on the contraction factor of the optical system.

A typical contraction effect may be obtained from the PAXAR lens and is shown in figure 14.

Let the contraction ratio for the optic be .384 image height for a 90 degrees half angle of the field of view, then the resolution is

$$\frac{.384}{90} = .0043\_\text{inch/degree} \qquad (34)$$

and .0128 inch for three degrees.

The minimum line width for the photo etching is approximately .006 inch. Therefore, the minimum resolution read out at the transducer may be assumed to be three degrees. The physical size of the transducer has to be .0128 x 60 = .768 inch height and the width will be selected to be the same size. This dimension provides for each section .128 width which generates an adequate solar signal strength for the switching pulse.

As explained in section 2.3, a sine/cosine storage has to be on board to compute the directional cosine values for $-\vec{\ell}(t)$. The transducers read the x^B and z^B image height which when considered together with the distortion function provides the object (sun) angles measured in the x^B and z^B respective coordinate system of the space vehicle. Let us consider a second binary read-in scale for the same cosine/sine onboard storage where the image height is correlated to the cosine function (see figure 14). Then from geometrical relation in figure 15a

$$\frac{z_1}{y_1} = P$$
$$z_1^2 + y_1^2 = a^2 \qquad (35)$$

where

z_1 is the transducer readout at the z^B coordinate - space vehicle

y_1 is the transducer read out at the y^B coordinate - space vehicle

a = image height at the focal plane of the equivalent focal length of a spherical optical system

O = Optical axis and point of reference, where the optical axis is normal to the paper surface

P = Ratio of the two transducer readouts.

Let a_{c1} be the cosine value obtained from the cosine/sine storage readout then

$$z_{c1}^2 + y_{c1}^2 = a_{c1}^2 \qquad (36)$$

and

$$y_{c1} = \sqrt{\frac{a_{c1}^2}{P+1}}$$
$$z_{c1} = \sqrt{\frac{P}{P+1} a_{c1}^2} \qquad (37)$$

The conversion of the position component vector $\vec{a}$ into the cosine value makes $a_{c1} = 1-(x^B_{c1})^2$; $x^B = f(a_{c1})$, the component vector of the unit sun vector system (see figure 15b) and also y_{c1} and z_{c1} are automatically the components of the sun vector in the x^B, y^B, z^B, satellite coordinate system.

Now, the comparison can be made between the values computed for the T and the B coordinate system. The actual procedure will be discussed under the different headings in Chapter 4.

The digital transducer read out is also used as directional signal for the digital servomotor which drives the rotating mirrors located in front of the anamorphic lens system. The y_1^B read out signal after passing through an amplifier drives the rotating mirror servo motor in the z^B axis and the z_1^B signal drives the other servo motor in the y^B axis. The servo steps of three degrees are completely adequate for the focal line alignment.

From the functional description of the analogue transducer section 3.2 and the digital transducer in this chapter, one can make the following comparison.

Although the sun position indicator which employs an analogue transducer is simpler in design and uses a lesser number of components, the inherent drawbacks of this unit reduce its reliability and stability particular in space application. For example, the transducer is very temperature sensitive. The sensitivity, dynamic resistance and noise equivalent power as function of temperature is given in figure 16. The temperature sensitivity has a great impact on the measurement accuracy when one point of the transducer surface is brought to much higher temperature than the rest of the surface. This condition is encountered for sun position indicator when the sun spot moves very slowly. The temperature gradient within the detector surface distorts the calibration grid pattern considerably and leads to incorrect measurements. The output of the radiation transducer is proportional to the incident energy

of the light spot, as well as to the displacement of this spot. In addition, any change of mean temperature of the receptor will raise or lower the output. An intensity pickoff is provided at the transducer which permits to measure the transverse transducer cell voltage (see figure 16). This voltage is used to regulate the automatic gain control in the y and z amplifiers. This arrangement, however, cannot compensate for the grid distortion and the drift of the zero point. To generate an amplitude modulated output signal, either the sun radiation itself or the electronic transducer output signal may be chopped. However, in either case temperature changes of the cell may change the impedance and will generate drift of the carrier frequency.

The digital solid state diode radiation transducer depends not on amplitude or voltage measurements, but it works on the principle of generation and counting of pulses. It is a static electro-optical device which produces directly a digital binary number where this number indicates the sun position in values of angular displacement measured in the y and z axes. Although the sensor output voltage level is sensitive to temperature and radiation incident energy, the binary pulse output for the transducer is not affected by this variation of the voltage level. However, the difference of voltage level between the sun focal line energy and the received energy from other celestial bodies or the earth or moon background is great enough to use a voltage threshold discrimination method and to assure the lock-on to the sun only.

The direct conversion of the binary transducer output to cosine values of the x^B, z^B coordinate system is of great advantage. All components involved in the measurements of the sun vector require no adjustment or calibration after launch and the drift of all instruments is minimized to a level which does not reduce the subsystems accuracy. Obviously, as shown in the previous discussion, the quantized digital indication of the sun vector has great advantages over the analogue transducer read out.

4. THE TANDEM SUN POSITION-INDICATOR

The limitation of the three degrees angular read out resolution in the coarse sun position indicator requires the simultaneous use of the coarse and the fine sun position indicators in a tandem arrangement for the yaw sensor and velocity sensor application. The coarse indicator initially acquires the sun vector and provides a direction read out of the sun position relative to the satellite coordinate system within a solid angle of three degrees. This information of y^B and z^B is used in a digital servo loop which directs the position of the optical axis of the fine sun position indicator to the center of the three degree solid angle identified by the coarse sun indicator (figure 17). To achieve a pointing error of <.08 degrees it is proposed to use an indexing pin and a registering cage. The cage should have a hyperhemispherical configuration to accommodate indexing of the fine sun position indicator for every three degrees angular distance over a field of view angle of 187 degrees. The inner radius of this cage may have a dimension of twenty inches and the 1/8 inch diameter indexing holes should be located within a tolerance of .010 inch.

The fine indicator would use the same type of digital read out as described in section 3.3. However, the much longer equivalent focal length of 7.3 inches of the anamorphic lens (only one lens is used) permits the designer to place the y^B and z^B transducer at the tangentially and sagittal focal lines of the same lens system. The focal line of greatest distance from the lens surface is rotated 90 degrees from the optical axis by a fixed partially reflecting mirror and the y^B transducer is placed at this focal line. The other z^B transducer is placed at the second focal line. The small field of view of 6 degrees of this optical system requires no centering of the sun's image lines.

The read out resolution of the transducer is

$$\frac{6}{63} = .0952 \text{ degrees per bit} = 1.662 \times 10^{-3} \text{ rad} \qquad (38)$$

The focal line curvature will be <.001 inch therefore straight line photo etched grids at the transducer may be used. The minimum time available for the determination of the sun vector is .780 minutes for 3 degree of orbital travel and a 3700 n. miles orbit radius. For a 6000 n. mile radius the time will be 1.3 minutes. If .17 minute is needed to reposition the telescope axis 3 degrees away from its previous position, a sampling time of approximately .5 to .6 minutes is available which is adequate for accurate displacement rate measurements and data smoothing.

It is assumed that the image line can be reduced to an angular width of one half the apparent angular image of the sun,

$$\frac{32}{2} = 16 \text{ arc minutes} = .266 \text{ degree} = 4.65 \cdot 10^{-3} \text{ rad}$$

Adding the indexing error, we will have an absolute read out

$$\text{accuracy of } .346 \text{ degree} = 6 \times 10^{-3} \text{ rad.} \qquad (39)$$

4.1 THE SUN POSITION INDICATOR AS YAW SENSOR

When the sun position indicator is used as a yaw sensor exclusively, it must be assumed that pitch and roll positions and rates are determined by other sensors. For example, for the semi-passive stabilization system of ORBS the pitch and roll data are acquired from the horizon tracker system. The alignment and rate tolerances are stated in section 1.3. The angular displacement

in yaw, therefore, should be considered simultaneously with pitch and/or roll displacements of less than 5 degrees from the local vertical.

At first the ideal case is analyzed where yaw drift θ_3 is observed in the $x^T - y^T$ plane.

Let $(A_1{}^B - A_1{}^T) = \theta_3$ as shown in figure (18) and consider as precomputed data

$$\begin{aligned} & x_1{}^T \\ & y_1{}^T \\ & z_1{}^T = z_1{}^B \\ & \bar{S} = \text{Unity}, \end{aligned} \qquad (40)$$

therefore

$$(y_1{}^T)^2 + (Z_1{}^T)^2 = 1 - (x_1{}^T)^2 \quad , \qquad (41)$$

$$A_1{}^T = \arccos x_1{}^T \quad . \qquad (42)$$

The measured values are

$$\begin{aligned} & z_1{}^B = z_1{}^T \quad , \\ & y_1{}^B \quad , \end{aligned} \qquad (43)$$

therefore

$$x_1{}^B = \left[1 - (y_1{}^B)^2 - (Z_1{}^T)^2\right]^{1\,2} \qquad (44)$$

and

$$A_1{}^B = \arccos x_1{}^B \qquad (45)$$

$$\theta_3 = \arccos x_1{}^B - \arccos x_1{}^T \qquad (46)$$

then

$$\begin{bmatrix} x^B \\ y^B \\ z^B \end{bmatrix} = \begin{bmatrix} C_{\theta_3} & S_{\theta_3} & 0 \\ -S_{\theta_3} & C_{\theta_3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x^T \\ y^T \\ z^T \end{bmatrix} \qquad (47)$$

$$\begin{bmatrix} x^T \\ y^T \\ z^T \end{bmatrix} = \begin{bmatrix} C_{\theta_3} & -S_{\theta_3} & 0 \\ S_{\theta_3} & C_{\theta_3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x^B \\ y^B \\ z^B \end{bmatrix} \qquad (48)$$

And the angular velocity about the $z^T = z^B$ axis is obtained from

$$\begin{bmatrix} \dot{\theta}_3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ -S_{\theta_3} & C_{\theta_3} & 0 \\ C_{\theta_3} & S_{\theta_3} & 0 \end{bmatrix} \cdot \begin{bmatrix} W_x{}^T \\ W_y{}^T \\ W_z{}^T \end{bmatrix} \qquad (49)$$

$$\begin{bmatrix} W_x{}^T \\ W_y{}^T \\ W_z{}^T \end{bmatrix} = \begin{bmatrix} 0 & -S_{\theta_3} & C_{\theta_3} \\ 0 & C_{\theta_3} & S_{\theta_3} \\ 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \dot{\theta}_3 \\ 0 \\ 0 \end{bmatrix} \qquad (50)$$

θ_3 and $\dot{\theta}_3$ can be obtained easily as shown in equations (40 through 46). The displacement and angular velocity relationships between the T and B coordinate frame in respect to the sun vector are readily obtainable from equations (47-50).

The vehicle's B-coordinate frame will not be perfectly aligned during the passive attitude control made in roll and pitch to the local vertical, which is identical to the z^T axis in the trajectory T-coordinate system. A misalignment of maximum 5 degrees is permissible before the horizon tracker initiates an alignment signal to the autopilot.

It is therefore of greatest interest to establish the measurement error of the sun position indicator as a function of limited attitude misalignment about the local vertical. The description and error analysis of the yaw error read out is presented in appendix A. The results show that for predominantly used earth satellite orbits which are located in space in such a way that the maximum elevation angle measured in the $x^T - z^T$ plane ranges between 65 and 37 degrees, the rms-yaw-error value for the sun position indicator is approximately 78 percent of the vertical attitude error, or for maximum 5° pitch or roll error the yaw measurement will be in error < 4.0 degrees.

To reduce this error considerably the following attitude control sequence is envisioned:

1. Rotate yaw angle to zero, y^B axis pointing parallel and with identical sign to the orbit velocity vector within error boundary of ±4 degrees maximum.

2. Use horizon tracker measurements to reduce alignment error in pitch and roll to less than one degree.

3. Correct yaw direction to < 3/4 degree error.

4. Correct roll and pitch error to

$$|\theta_{2_{lim}}| = |\theta_{1_{lim}}| \leq \pm .1°$$

$$|\dot{\theta}_{2_{lim}}| = |\dot{\theta}_{1_{lim}}| = 2 \cdot 10^{-4} \text{ rad/sec}$$

5. Bring yaw axis orientation and angular rate to

$$|\theta_{3_{lim}}| \leq \pm 0.5°$$

$$|\dot{\theta}_{3_{lim}}| \leq 8 \cdot 10^{-3} \text{ rad/sec}$$

The measurements and capabilities of the horizon tracker together with gravity gradient stabilization has been explained elsewhere (Reference 17).

As stated in equation (39) the fine sun position indicator resolution is less than six milliradians, when used without electronic smoothing, that is 1.3 times better than the specified rate accuracy. Improvements of this resolution may be achieved without advancements of the state-of-the-art in the optical and electronic designs and the integration of the components into the subsystem.

4.2 THE SUN POSITION INDICATOR AS AN ANGULAR VELOCITY SENSOR

As shown in the previous section, the relationship of the angular velocities between the T and B coordinate frame may be acquired also in the yaw sensor mode. However, a distinction is made in this paper between an angular velocity sensor which determines angular drift and the one which determines the angular velocity of the tumbling space vehicle B-coordinate system in respect to the sun vector. The latter sensor system is the subject of the discussion in this section and it is assumed the vehicle tumble is an emergency condition which has to be corrected. The boundary of angular drift is defined here as the angular motion of the positive x^B axis which will not exceed beyond the hemisphere of the $y^T - z^T$ plane and which contains the positive x^T axis.

Tumble of the space vehicle therefore will have angular motions of greater than 2π and the angular rates are in the order of .5 rad/sec., or greater.

To envision the correct condition of a tumbling space vehicle containing two sun position indicators in opposed diametrical orientation and in relation to the space fixed sun vector (the geocenter is assumed to be fixed also in relation to the sun for this short time of correction) it is helpful to think of a ball with two hemispheres of two different colors, one may be white, the other red (Figure 19). Gravity gradient effects are small in comparison to the kinetic energy of the tumble effect and can be neglected.

There is only one region in which the two sun seekers will see the sun simultaneously and that is in the ring-cone of ± 7 degrees from the $y^B - z^B$ plane. Otherwise each sun seeker will share the time of sun visibility equally and the time required for one half revolution of tumble may be used for rough computation of the angular tumble velocity.

The rotating or tumbling vehicle in space will have a spin or inertial stabilization and therefore will maintain its plane of rotation relative to the sun vector until any outside force generates a precession and gyroscopic reaction. The arrow of the sun vector scribes a great circle on the rotating ball with the axis of rotation normal to the plane of the circle, and the sun vector in this plane (Figure 19). The gyroscopic reaction caused by an outside force may be used to align the angular velocity vector of the tumbling vehicle with the coordinate plane of the vehicle which has the largest thrust capacity to stop the angular velocity of the tumble. Only the coarse sun position indicator is needed for the measurement of the orientation and rates of the angular velocity during the time of tumble. Since the minimum velocity is assumed to be .5 rad/sec, there will be available more than 10 bits of read out information for a three degree of transducer read out information. No comparison is made between the stored T-coordinate data and the transducer readout. The only information obtained during this emergency operation is the orientation and rates of the angular velocity relative to the B-coordinate frame (see Figure 20).

The analytical description of precession is implicit in Euler's dynamical equations (Reference 18):

$$\begin{aligned} \tau^1_{UB} &= J_1 \dot{\theta}_1 - (J_2 - J_3) \theta_2 \theta_3 \\ \tau^2_{UB} &= J_2 \dot{\theta}_2 - (J_3 - J_1) \theta_1 \theta_3 \\ \tau^3_{UB} &= J_3 \dot{\theta}_3 - (J_1 - J_2) \theta_1 \theta_2 \end{aligned} \qquad (51)$$

where τ^i_{UB} is the applied torque about the i-th principal axis and J_1, J_2, and J_3 are the principal moments of inertia of the space vehicle. If, for example, the space vehicle acquires an additional inertial angular velocity $\dot{\theta}_3$, measured about the z^B axis, which is generated by the application of a torque τ^3_{UB} about the z^B axis, then the gyroscopic reaction generates a precession torque τ_{θ_B} which rotates the vehicle $y^B - z^B$ plane about the x^B axis. Very small $\dot{\theta}_3$ values are desirable to avoid large coupling effects. Similar procedures may be

performed with the other principal axes. After the alignment of the tumble velocity vector and therefore also the sunvector with one coordinate plane normal to the spin axis, an opposed thrust is initiated to reduce the spin rate to $< 5.5 \times 10^{-2}$ rad/sec.

For example, the reoriented spin axis may be located normal to the $x^B - z^B$ plane (Figure 21). When the minimum rate of 5.5×10^{-2} rad/sec has been reached, the fine sun position indicator is activated and placed into the maximum $+z^T(t) = z^B(t)$ position of the present satellite orbit. The $z^T(t)$ value was computed before launch and stored on board. The sun sensor system commands the attitude control to bring the sun vector to this vehicle and transducer position at this precomputed time. It must be recognized that this is also the instant of time and orbit motion when the inertially stabilized vehicle with its axis normal to the sun vector may be stopped without giving rise to motion about other principal axes. Since the z^B axis was brought in coincidence with the local vertical the horizon tracker may acquire the geocenter and correct on its own the pitch and roll error. If needed the yaw mode may be used as described in the preceding section.

4.3 THE SUN POSITION INDICATOR AS A GENERAL SATELLITE ATTITUDE SENSOR OF MODERATE ACCURACY

It is a well established fact that one planet, star or sun sensor can not orient a reference frame, but only one reference axis. This paper makes no different claims. There is no unique solution to the problem of attitude reference for a vehicle tracker pointing to one celestial body only. For example, if the B coordinate frame rotates exactly about the sun vector, there is no change of values z^B and y^B as function of time. This then is not providing a workable and desired attitude orientation. It can be confirmed also when comparing the precomputed data. A small thrust along the $x^B - z^B$ plane will change this rotation about the sun vector immediately. Now, after equalization of the angular momentum the x^B axis rotates about the sun vector in a cone or a distorted configuration of a conical section if there is residual drift. As shown in section 2.2 equation (28) the correct path of the sun image at the focal plane of the sun position indicator would describe a linear spiral of extremely small gang width when a spherical lens system is used. Therefore, for a modified accuracy this radius may be taken constant (Figure 22). This means all sun image positions have to lie on a circle with radius "r" and $\ell(t)$ is constant. The first step of orientation therefore is to tilt the x^B axis to such an angle with the sun vector that the distance between the sun image and the point of origin or the optical axis stays continuously in the distance of "r". After this has been accomplished, one initiates rotation about the x^B axis, which is now also the x^T axis, until the correct position of $y^B = y^T$ and $z^B = z^T$ is established. From then on continuous active attitude control should be applied to hold the angle of deviation small, so that the attitude adjustments of the vehicle are simplified in regard to coupling effects of motions. This is particular the case since this mode must apply the constant value "a" in equation (28) and change this value in steps according to a time schedule from precomputed stored data. The complexity of this system has not yet allowed to establish an error analysis. However, it is conceivable that the accuracy is only moderate and must be assumed greater than five degrees unless at a later time an analysis will show different values.

5. CONCLUSIONS

Sun trackers and solar sensors have been described before by several authors (Ref. 19, 20, 21). Most of the previously proposed space guidance schemes which use sun position indicators were based on the inertial stabilization of the vehicle or reference platform. Other papers describe mainly the sun seeker component and its tracking function.

The new guidance scheme presented in this paper employs the tandem sun position indicators with digital two-axes directional readouts and permits the vehicle to stabilize in reference to a rotating vehicle centered coordinate system. This coordinate system is oriented in reference to the satellite orbit plane and the local vertical.

Although further development work should be performed to assure simplicity of implementation, the content of this paper shows the feasibility of the sun vector referenced guidance scheme to earth satellites of a long life time. This sun position indicator scheme is a suitable subsystem for the semi-passive attitude control for pre-space rendezvous operations, such as required for ORBS, and is capable of assisting in minor orbital correction maneuvers. The quantized readouts together with:

1. The tandem indicator arrangement.
2. The characteristic features of no mechanical or electrical drift.
3. No calibration requirements in space.
4. Threshold discriminations between the sun and radiation emitted from other celestial bodies.

make this subsystem one of greater reliability and accuracy. None of the proposed electronic instruments in the sun position indicator, the onboard storage, or the computer exceeds the state-of-the-art of today's guidance equipment.

REFERENCES

1. Herbert Reich, "Summary Report, an Orbital Rendezvous Base System," ASG TM-61-50 NORTHROP CORP., HAWTHORNE, CALIF. December 1961.

2. D. Pierce, "Velocity and Time Requirements for Satellite Rendezvous Trajectory," ASG TM-61-12 NORTHROP CORP., Hawthorne, Calif.

3. F. P. Carstens and T. N. Edelbaum, "Optimum Maneuvers for Launching Satellites into Circular Orbits of Arbitrary Radius and Inclination," American Rocket Society Preprint No. 1450-60.

4. R. A. Hord, "Relative Motion in the Terminal Phase of Interception of a Satellite or a Ballistic Missile," NACA Techn. Note 4399, Sept., 1958.

5. W. H. Clohessy and R. S. Wiltshir, "Terminal Guidance for Satellite Rendezvous," IAS Paper No. 59-93.

6. Louis W. Spradlin, "The Long-Time Satellite Rendezvous Trajectory," Aero/Space Engineering June 1960.

7. Howard Seifert, editor, "Space Technology," John Wiley and Sons, Inc., 1959. (See ch. 26 by Albert D. Wheelon.)

8. Robert S. Swanson, N. V. Petersen, and Leroy Hoover, "An Astro-Vehicle Rendezvous-Guidance Concept," NORTHROP CORP., presented to the American Astron. Society, January 1960.

9. Lawrence S. Joyce, "Computation of Sun Direction, Visibility, and Eclipse Time for Circular Earth Orbits," ASG TM-61-2 NORTHROP CORP., HAWTHORNE, CALIF., Nov. 1961.

10. David A. Pierce, "A Rapid Method for Determining the Percentage of a Circular Orbit in the Shadow of the Earth," ASG TM-61-66 NORTHROP CORP., HAWTHORNE, CALIF. 1961

11. Lawrence G. Stoddard, "Prediction of Eclipses of Earth Satellites," Amer. Astronautical Society, August 1961, Preprint 61-79.

12. George B. Patterson, "Graphical Method for Prediction of Time in Sunlight for a Circular Orbit," ARS Journal, 31(1961) p. 441.

13. Lawrence S. Joyce, Heinz Gabloffsky, "Computation and Automatic Plotting of Sun Direction, Visibility, and Eclipse Time for Circular Earth Orbits," NSL 62-25 March, 1962, NORTHROP CORPORATION, HAWTHORNE, CALIF.

14. J. L. Brenner, R. Fulton, N. Sherman, "The Determination of the Earth's Potential Field by Observations of Satellite Orbits, with Special Reference to the Determination of the Third Harmonic," AFMDC TR59-29, 2587-ITR-3, Stanford Research Institute, Menlo Park, Calif. July 1959

15. Courtesy of Micro Systems, Inc., Pasadena, Calif. "Radiation Tracking Transducer XY-20" Specification and Sales Bulletin

16. Courtesy of Pacific Optical Corp., A Division of Chicago Aerial Industries

17. Heinz Gabloffsky, "Preliminary Guidance Considerations for ORBS," Vol. II., Attitude Control Functions, ASG TM61-94, NORTHROP CORPORATION., HAWTHORNE, CALIF.

18. Connie L. McClure, "Theory of Inertial Guidance," Prentice-Hall, Inc., 1960 p 94.

19. Donald K. Wilson, Richard Wuensch, "Sun Tracker with Moon and Planet Modifications." American Astronautical Society, August 1960, Preprint 60-77

20. Paul R. Spencer, "Study of a Solar Sensor for use in Space-Vehicle Orientation Control Systems," NASA Technical Note D-885 Washington June 1961

21. James S. Albus, "A Digital Solar Aspect Sensor," NASA Technical Note D-1062 Washington September 1961.

APPENDIX A

YAW ERROR OF THE SUN POSITION INDICATOR AS A FUNCTION OF ATTITUDE MISALIGNMENT

This appendix is the descriptive and error analysis of the sun position indicator used as yaw sensor where the presumed attitude misalignment occurs only in vertical, such as in the ORBS semi-passive attitude control system.

The two frames of references, the computed and stored T-coordinate system and the space vehicle B-coordinate system, which represent the two attitude orientations are shown in figure (A-1).

One frame is rotated upon an axis $\vec{U}_\theta$ in the $x^T - y^T$ plane of the trajectory coordinate system. The angle of rotation is θ_1 in a righthanded coordinate system. The objective of this analysis is to obtain a relationship which describes the error generated

1. at the sun position indicator of system B of figure (A-1) after it had been rotated and
2. as a measure of the true angle of the same sun position indicator denoted as system T of the same figure.

The functional relation is

$$\theta_3 = (A_2 - A_1) = (A_4 - A_5) \tag{1}$$

There are two alternate expressions for the sun position indicator with the vector $\vec{S}$ which may be employed to obtain the expression for θ_3 in terms of E_1, E_2, A_2 and the rotation angle θ_1.

$$\vec{S} = \cos E_1 \cos A_1 \vec{U}_\theta + \cos E_1 \sin A_1 \vec{U}_1 + \sin E_1 \vec{K}_1$$

$$\vec{S} = \cos E_2 \cos A_2 \vec{U}_\theta + \cos E_2 \sin A_2 \vec{U}_2 + \sin E_2 K_2 \tag{2}$$

Multiplying each equation with vector $\vec{U}_\theta$ we obtain the dot product

$$\cos E_1 \cos A_1 = \cos E_2 \cos A_2 \tag{3}$$

And the same equations multiplied with vector $\vec{U}_1$ obtain

$$\cos E_1 \sin A_1 = \cos E_2 \sin A_2 \cos \theta_1 - \sin E_2 \sin \theta_1 \tag{4}$$

And since

$$\sin (A_2 - A_1) = \sin A_2 \cos A_1 - \cos A_2 \sin A_1 \tag{5}$$

We obtain from the above equations when multiplying with vector $\vec{K}_1$ the dot product

$$\sin E_1 = \cos E_2 \sin A_2 \sin \theta_1 + \sin E_2 \cos \theta_1 \tag{6}$$

The elevation angle E_2 is a function of z^B which is observed and E_1 as a function of z^T which is stored onboard as a reference function of real time.

It is desired to find $\sin A_1$ and $\cos A_1$ in terms of A_2, E_1, E_2 and θ_1. From equations (3) and (4)

$$\cos A_1 = \frac{\cos E_2 \cos A_2}{\cos E_1}$$

$$\sin A_1 = \frac{\cos E_2 \sin A_2 \cos \theta_1 - \sin E_2 \sin \theta_1}{\cos E_1} \tag{7}$$

and substituted into (5) obtains

$$\begin{aligned} \sin (A_2 - A_1) &= \sin A_2 \cos A_2 \frac{\cos E_2}{\cos E_1} \\ &\quad - \sin A_2 \cos A_2 \frac{\cos E_2}{\cos E_1} \cos \theta_1 \\ &\quad + \frac{\sin E_2 \cos A_2 \sin \theta_1}{\cos E_1} \\ &= \sin A_2 \cos A_2 \frac{\cos E_2}{\cos E_1} (1 - \cos \theta_1) \\ &\quad + \frac{\sin E_2}{\cos E_1} \cos A_2 \sin \theta_1 \end{aligned} \tag{8}$$

Limiting the angle of rotation θ_1 to small values

$$\theta_1 \ll 1$$

the term $\sin \theta_1 \rightarrow \theta_1$ to the second order accuracy.

Also $(1 - \cos \theta_1) \rightarrow 0$.

Then

$$\sin (A_2 - A_1) \approx \frac{\sin E_2}{\cos E_1} \theta_1 \cos A_2 \tag{9}$$

This is the expression for the error of measurement which is obtained in the tilted or rotated space vehicle B-coordinate system.

Since

$$\theta_3 = (A_2 - A_1)$$

and

$$\sin \theta_3 = \frac{\sin E_2}{\cos E_1} \theta_1 \cos A_2 \tag{10}$$

The rms value of $\sin \theta_3$ is then

$$\sqrt{E^2 (\sin^2 \theta_3)} = \sqrt{E(\theta_1^2)} \sqrt{E(\cos^2 A_2)} \frac{\sin E_2}{\cos E_1} \tag{11}$$

if it is assumed that the axis of rotation, $\bar{U}_\theta$, which determines the distribution of A_2 is not affected by the magnitude of the frame rotation. Presuming that A_2 is uniformly distributed around the x^T - y^T plane of the trajectory coordinate system, we have

$$E(\cos^2 A_2) = \int_0^{2\pi} \frac{\cos^2 A_2 dA_2}{2\pi} = \frac{1}{2} \tag{12}$$

and from equation (11) the rms value of $\sin \theta_3$ designated as δ_P is given by

$$\delta_P = .707 \frac{\sin E_2}{\cos E_1} \delta_{\theta_1} \tag{13}$$

For the largest possible angle of $E_1 = 90°$ the value δ_P is infinite. This condition develops when z^T and z^B reach a maximum and y = 0, the sun vector is in the z-x plane.

The predominant region of z^T max will range between E_1 = 65 to 37 degrees, that is cos z^T = .806 to .602.

The largest value of sin E_2 is given by sin $(E_1 + \max \theta_1)$. Bounding max θ_1 by 5 degrees (14) obtains the range of error in the two boundaries.

$$\sin E_{2_1} = \sin(65 + 5) = \sin 70° = .93969 \tag{15}$$

$$\sin E_{2_2} = \sin (37 + 5) = \sin 42° = .669$$

$$\delta_{P_1} = \frac{.93969}{.906} \; 0.707 \; \delta_{\theta_1} = .733 \; \delta_{\theta_1}$$

$$\delta_{P_2} = \frac{.669}{.602} \; .707 = .785 \; \delta_{\theta_1}$$

For the predominantly used satellite orbits the analysis indicates that a good compromise value for the rms value of the sun position indicator as a yaw sensor with limited attitude misalignment is approximately 78 percent of the vertical attitude error. This error amounts to less than four degrees difference in yaw angle read out if the maximum attitude error is 5 degrees and to 3/4 degrees if the misalignment is one degree.

LIST OF SYMBOLS

SYMBOLS

h	Satellite altitude (ft.)
e_{ss}	Unit vector directed from satellite to sun
i	Inclination of orbit plane with respect to the equatorial plane (deg.)
l, m, n	Direction cosines associated with α, δ, γ
l_b	Direction cosine ϕ_b
r_s	Radius of orbit = $h + r_e$ (ft.)
r_e	Radius of earth = 20.926428×10^6 ft.
$\hat{r}$	Non-dimensional radius of orbit = $\frac{r_s}{r_e}$
t	Time in seconds
u	Argument of latitude (Fig. 6)
B	Superscript for space vehicle reference frame
G	Earth's gravitational constant = 1.40771×10^6 ft^3/sec^2
J	Second harmonic constant in Legendre series expansion for the potential function of the oblate earth, .001638
K	Mean angular rate of earth's motion about sun = 1.9923×10^{-7} rads/sec
T	Superscript for trajectory coordinate system
TS	Shadowtime in percent of orbit unit
$\vec{W}$	Angular velocity of the orthogonal transformation of coordinates
α, δ, γ	Direction angles of $\vec{e_{ss}}$ when referred to trajectory coordinate system
β	Obliquity of the ecliptic = 23.45° (Fig. 6)
θ_1	Angle of rotation about pitch axis
θ_2	Angle of rotation about roll axis
θ_3	Angle of rotation about yaw axis
Λ	Angle relating the position of the earth in the ecliptic (Fig. 5)
ϕ_b	Angle between $\hat{r}$ and $\vec{e_{ss}}$
☊	Ascending node of orbit (Fig. 6)

FOR A TYPICAL ORBIT $N/m = 15/1 \quad n = 1-4$

$h = 263$ N MI $\quad i \sim 30^\circ - 40^\circ$

2 WEEK PERIOD $\Delta S = \pm 45$ N MI $\quad \Delta h = \pm .2$ N MI

$\Delta V/\text{YEAR} = 33$ FT/SEC $\left[W/S = 100 \frac{\text{lbm}}{\text{FT}^2} \text{ OR } \frac{C_D S}{m} = .7 \text{ FT}^2/\text{SLUG}\right]$

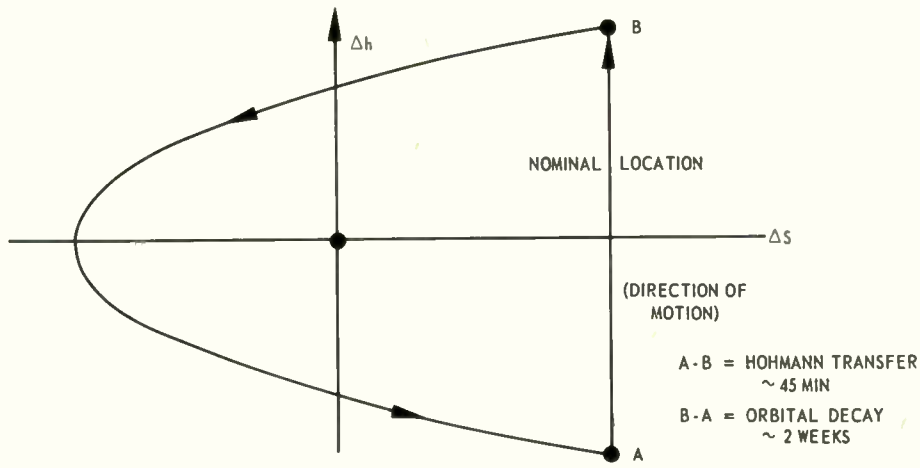


Fig. 1. Recommended station-keeping limit cycle characteristics.

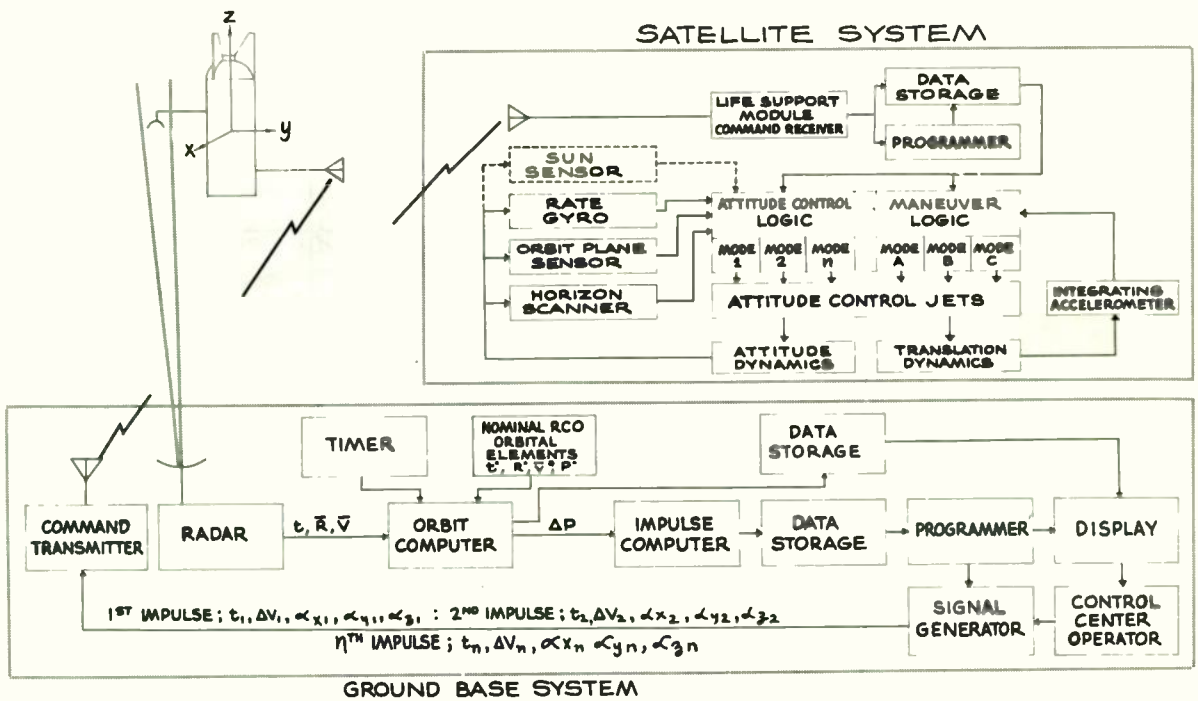


Fig. 2. Station-keeping control system block diagram.

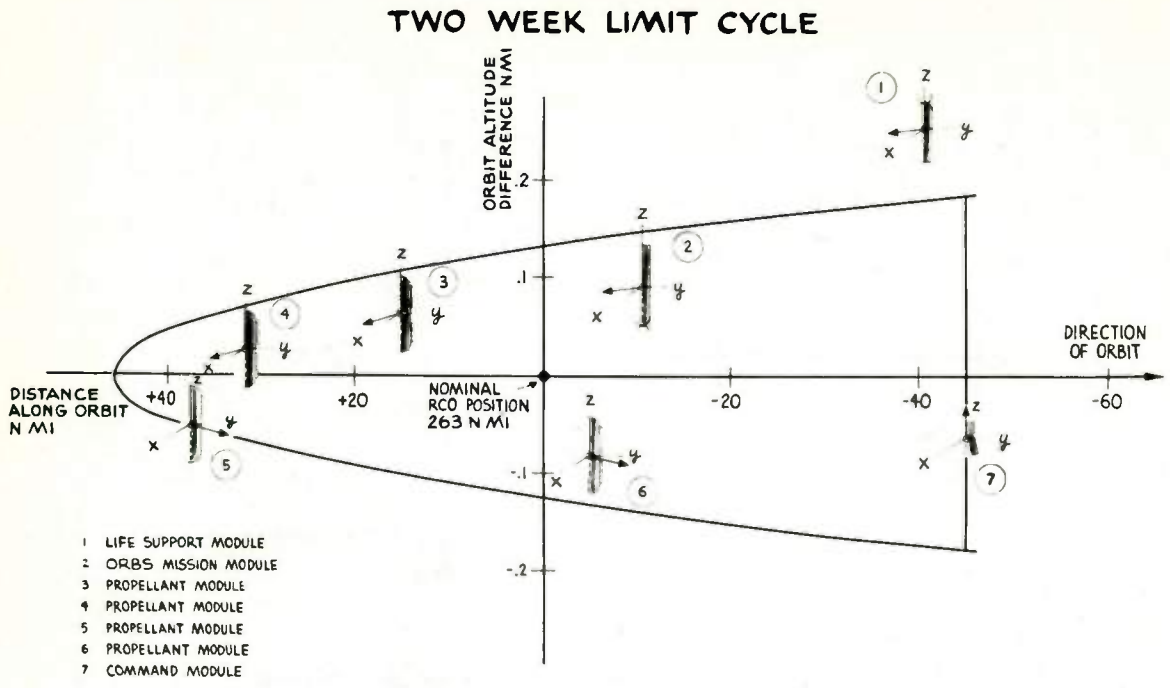


Fig. 3. Integral station-keeping two week limit cycle.

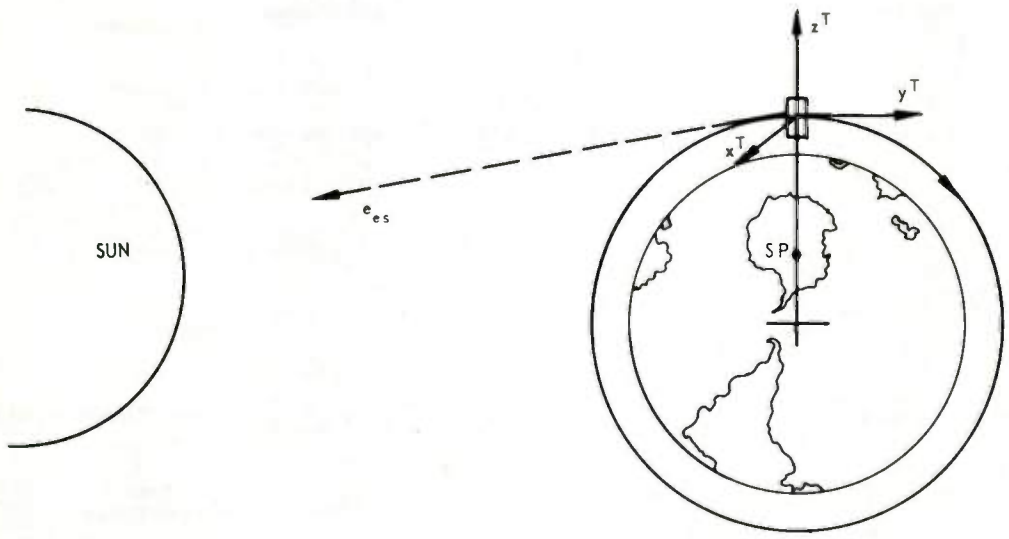


Fig. 4. Trajectory reference frame.

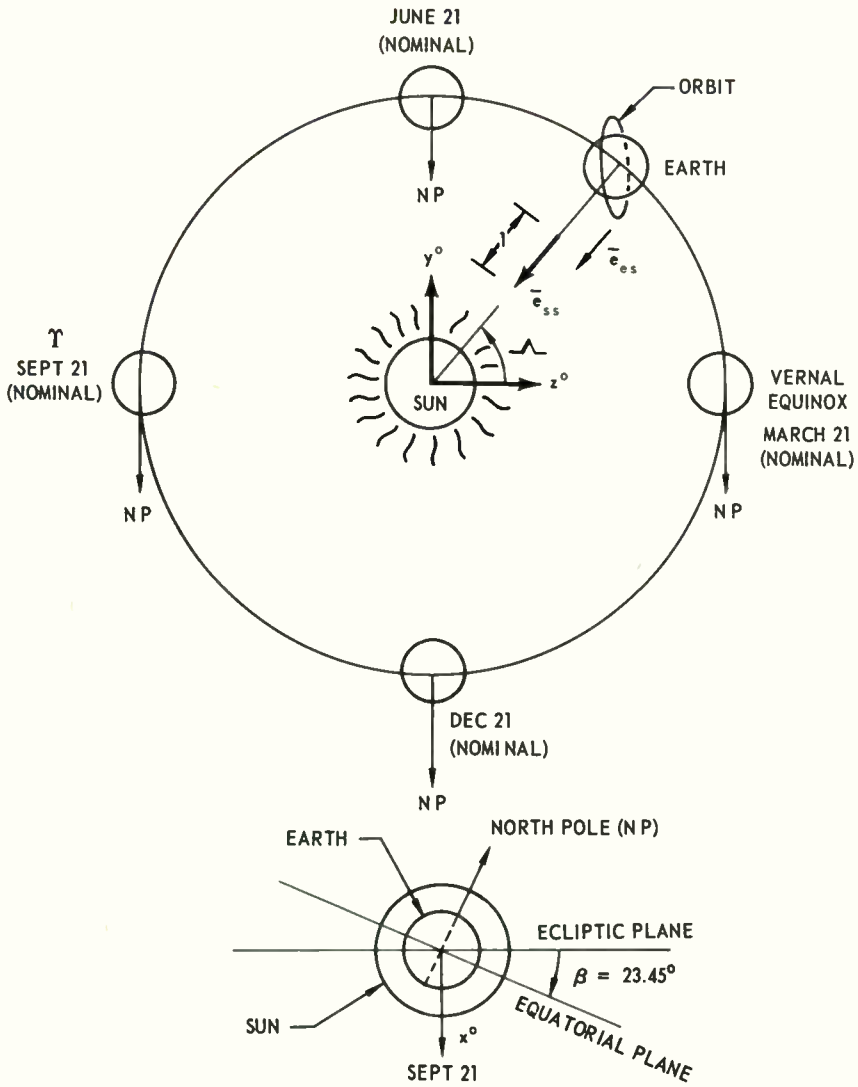
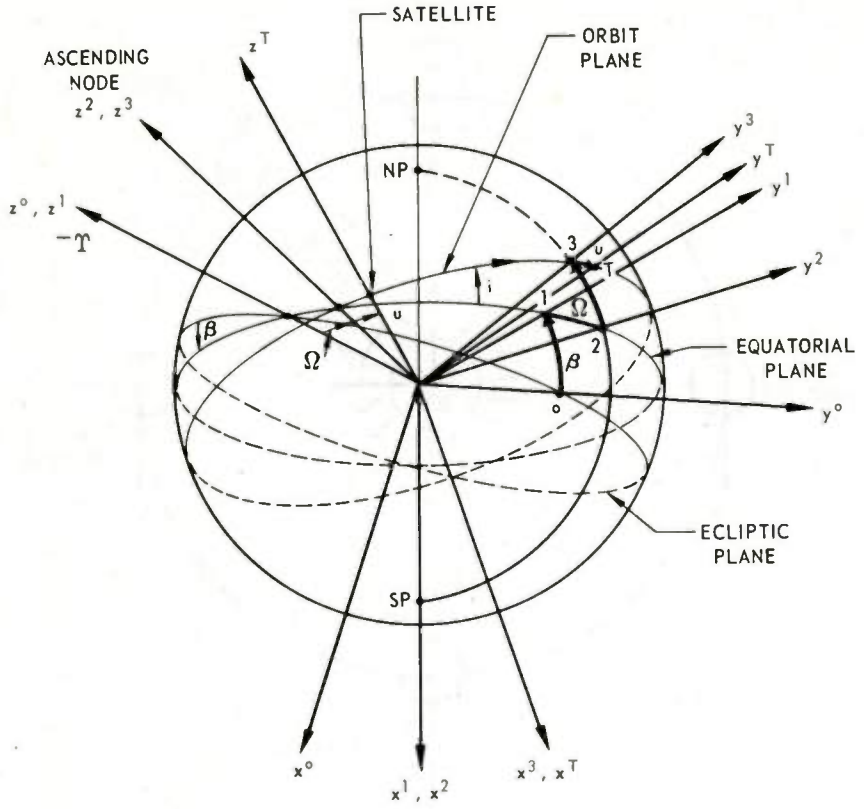


Fig. 5. Sun-fixed inertial coordinate system.



SUPERSCRIPT	COORDINATE DIRECTIONS		
	x	y	z
0	NORMAL TO ECLIPTIC	BI-NORMAL	ALONG VERNAL EQUINOX
1	ALONG POLAR AXIS (SOUTH POLE)	BI-NORMAL	ALONG VERNAL EQUINOX
2	ALONG POLAR AXIS (SOUTH POLE)	BI-NORMAL	ALONG ASCENDING NODE
3	NORMAL TO ORBIT PLANE	BI-NORMAL	ALONG ASCENDING NODE
T	NORMAL TO ORBIT PLANE	ALONG LOCAL HORIZONTAL	THRU SATELLITE

Fig. 6. Coordinate systems and transformations.

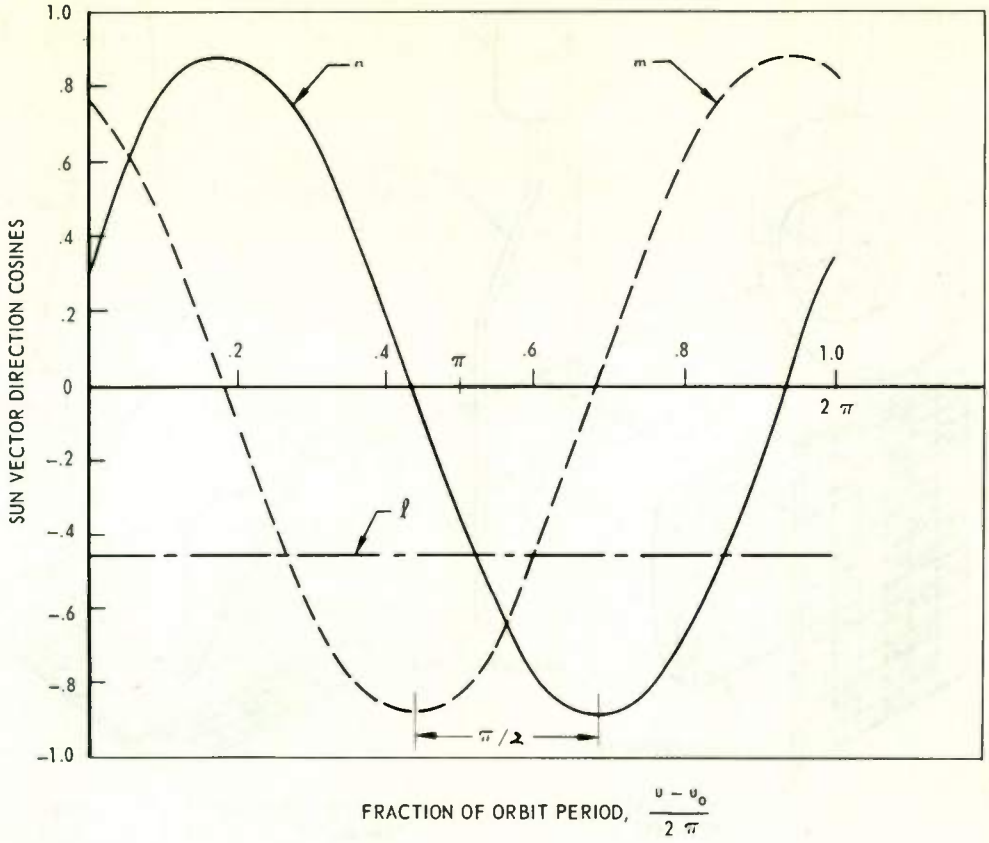


Fig. 7. Phase relation of sun vector components.

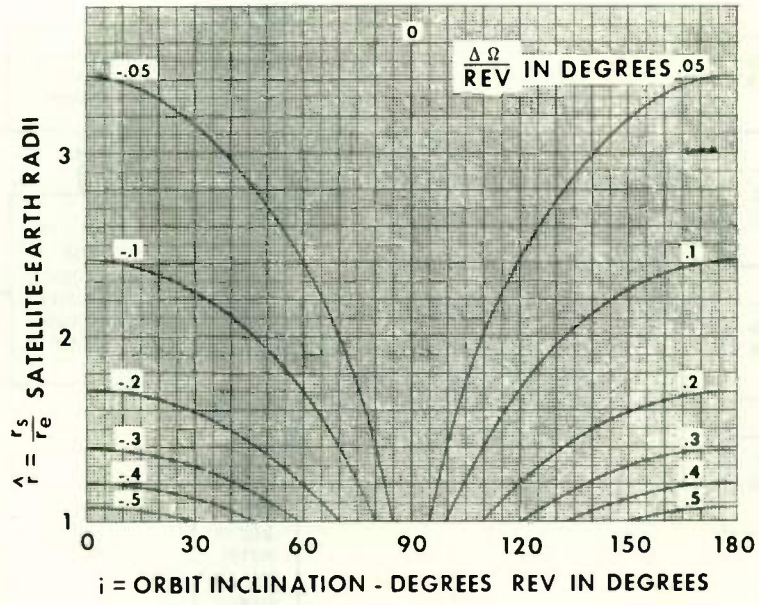


Fig. 8. $\frac{\Delta\Omega}{Rev}$ as function of $\hat{r}$ and i.

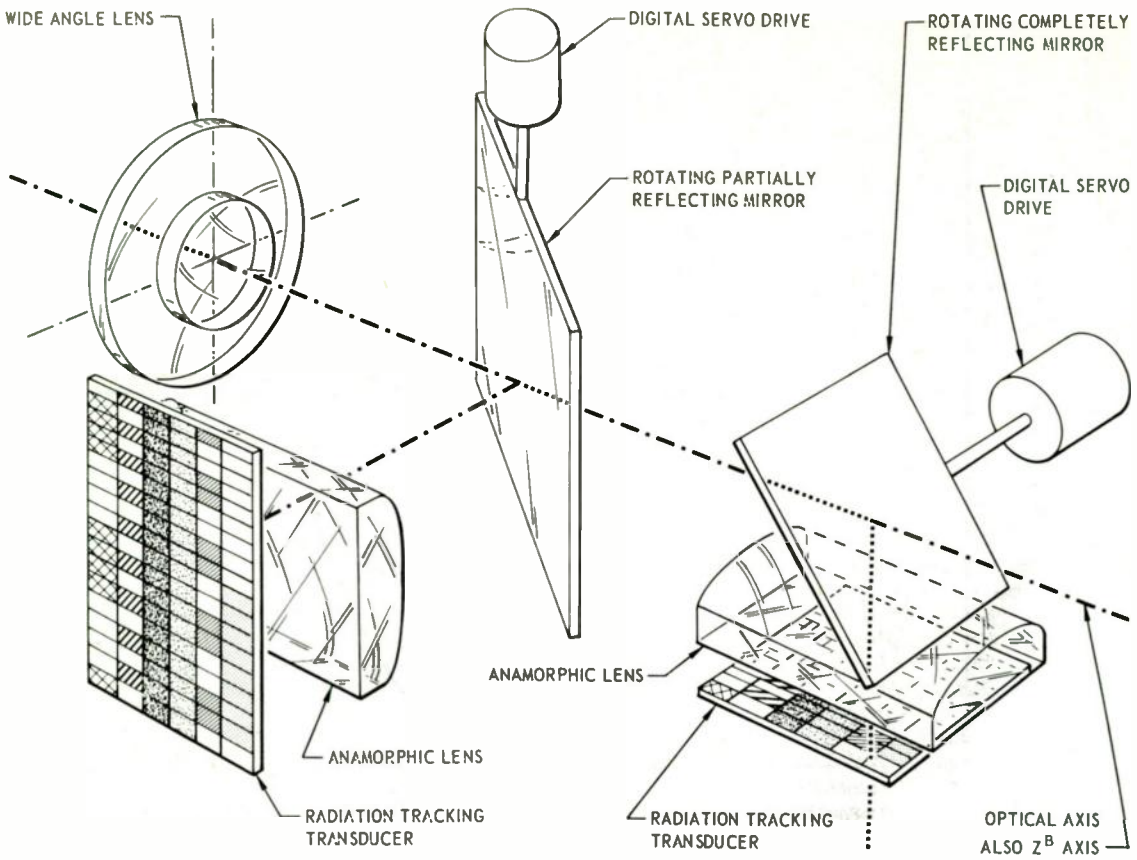


Fig. 9. Sun position indicator (coarse) schematic.

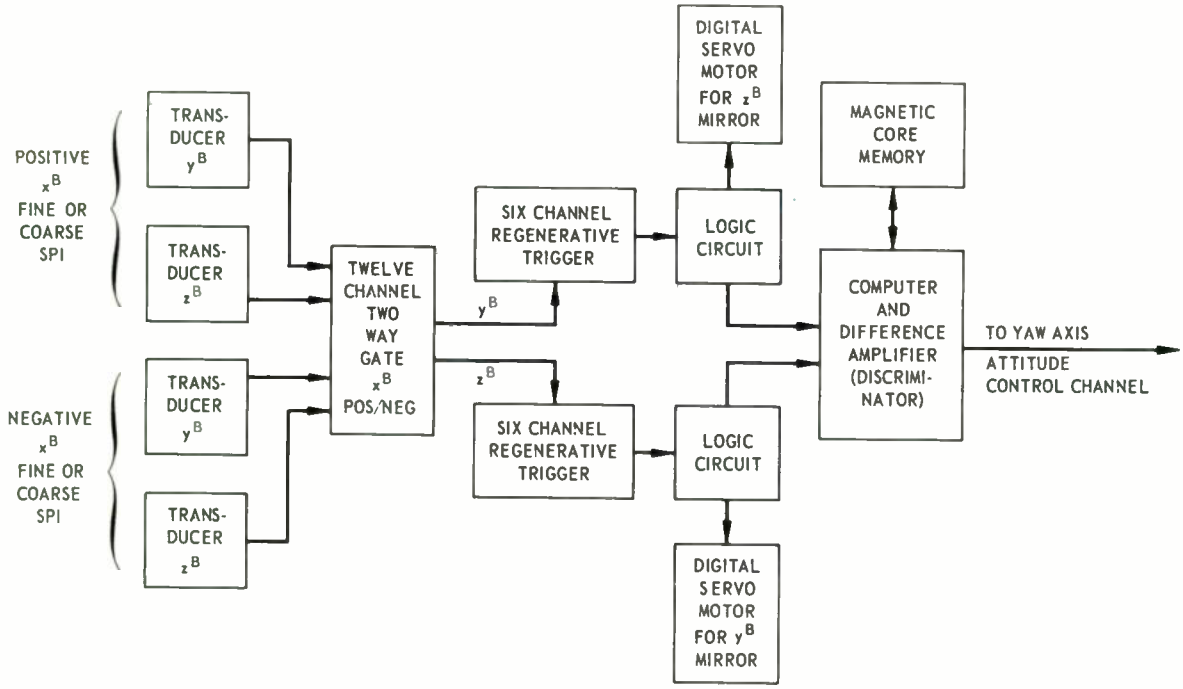


Fig. 10. Block diagram sun position indicator.

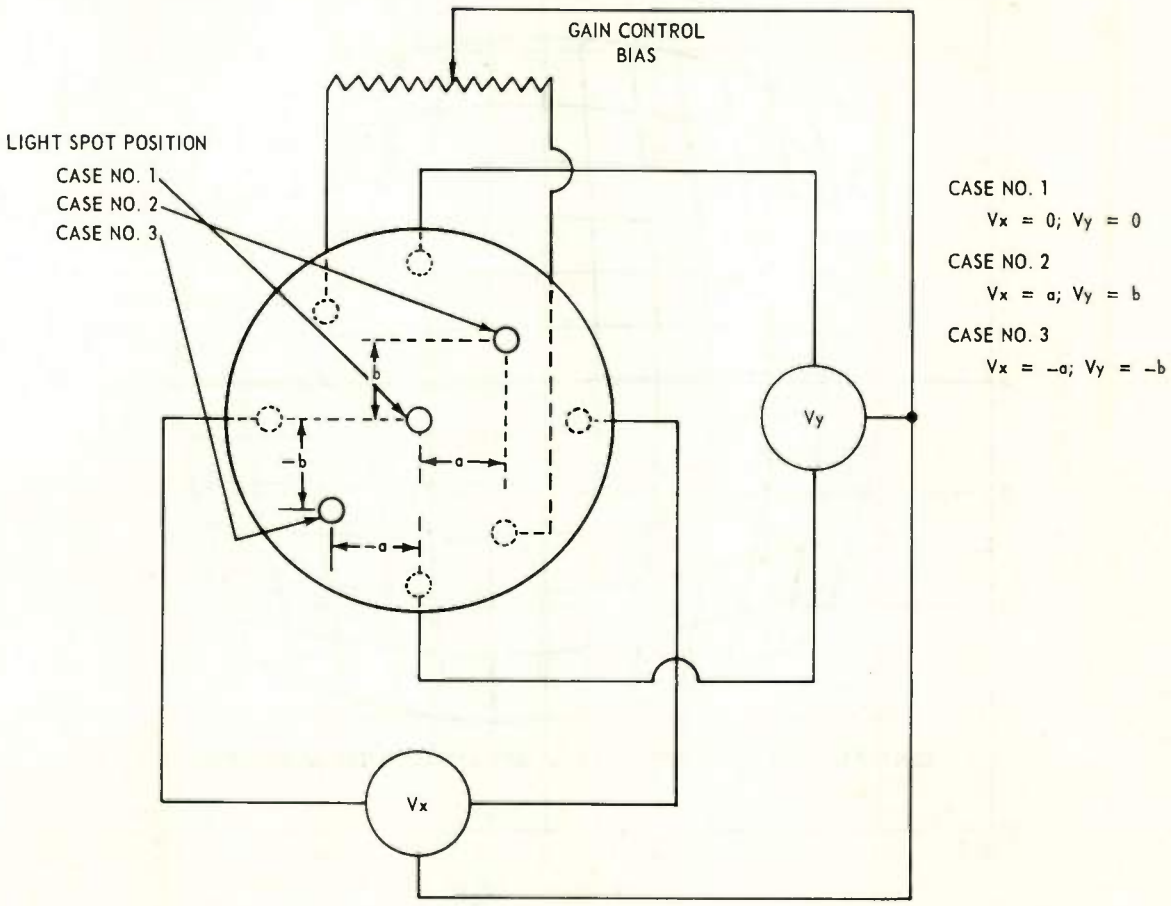


Fig. 11. Transducer operation.

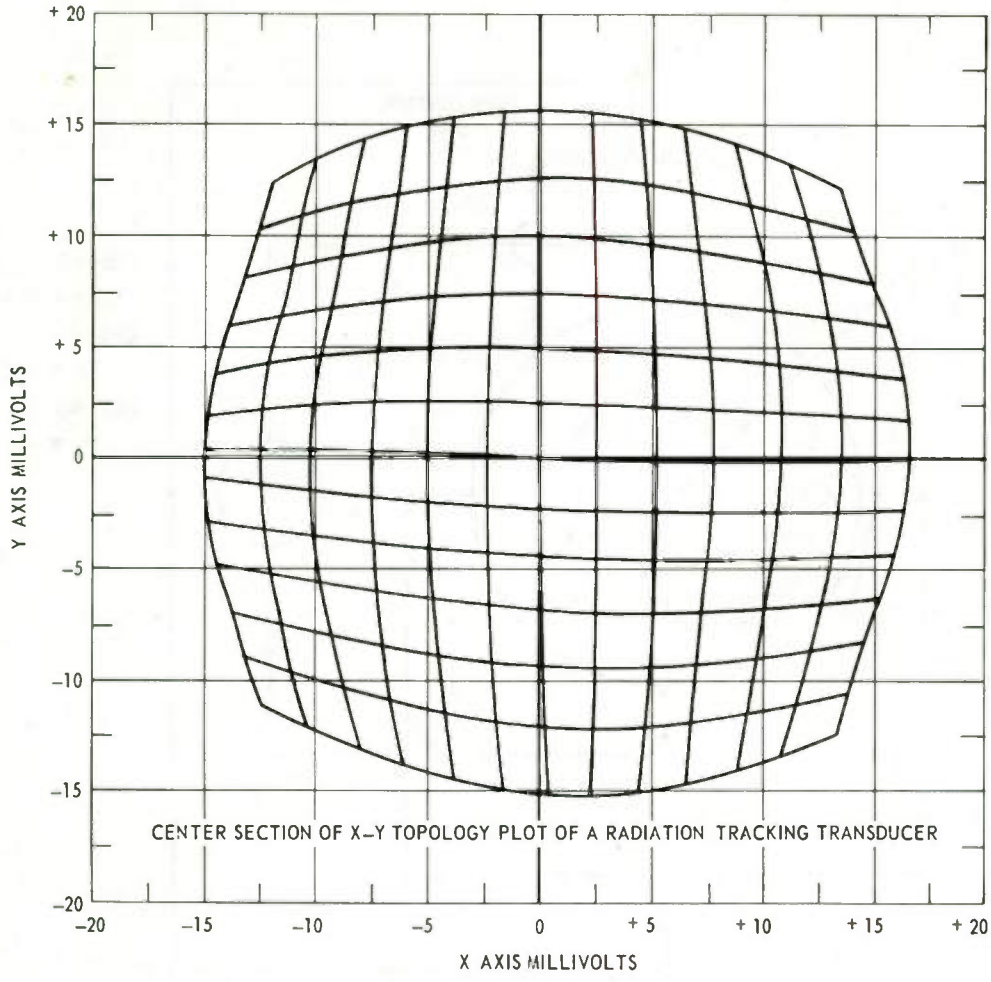


Fig. 12. Transducer XY-20 calibration grid.

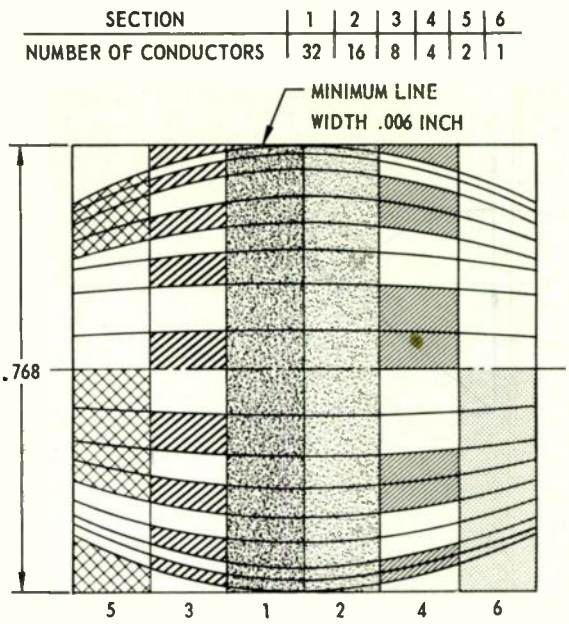


Fig. 13. Photo etched binary radiation transducer.

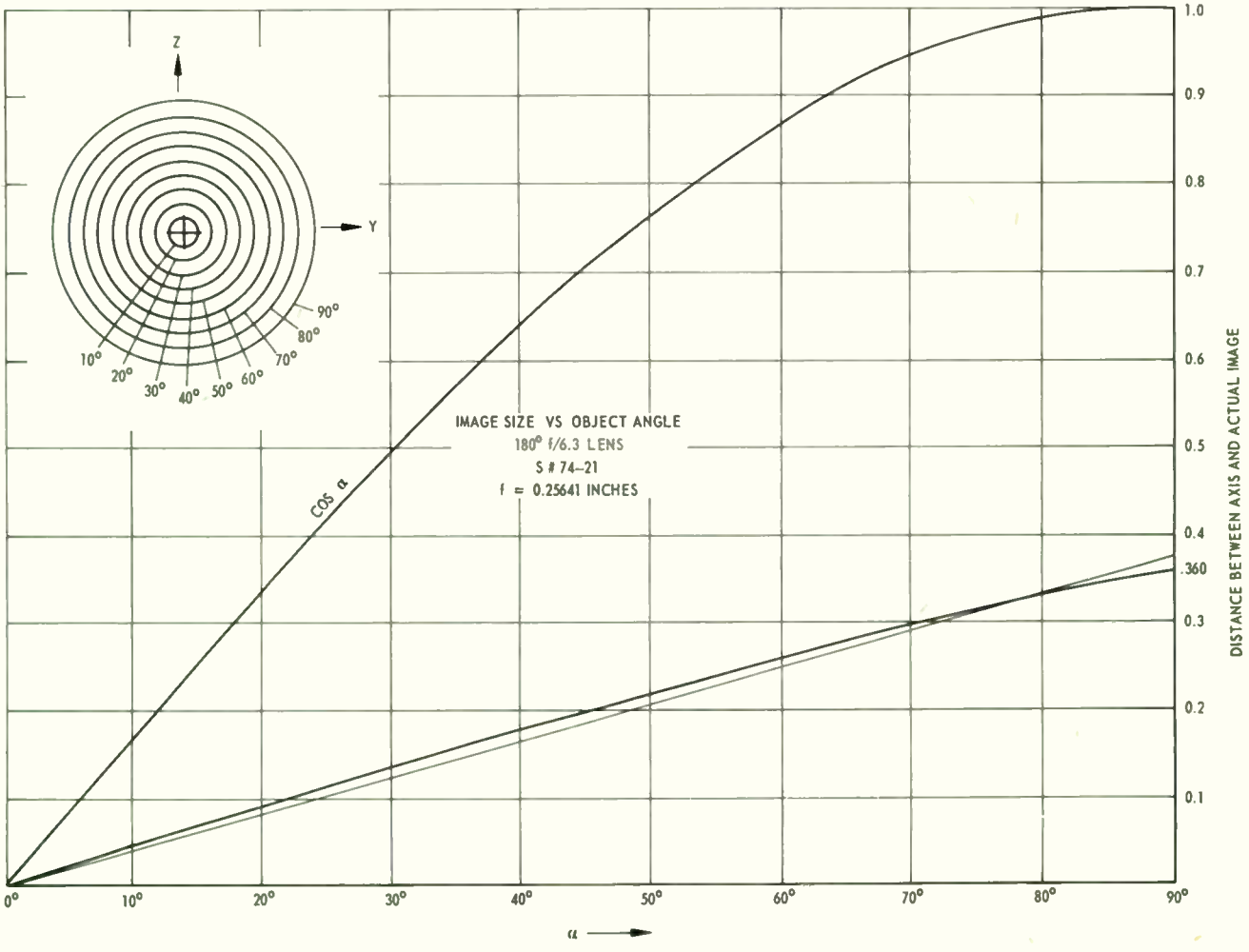


Fig. 14. Paxar contraction effect.

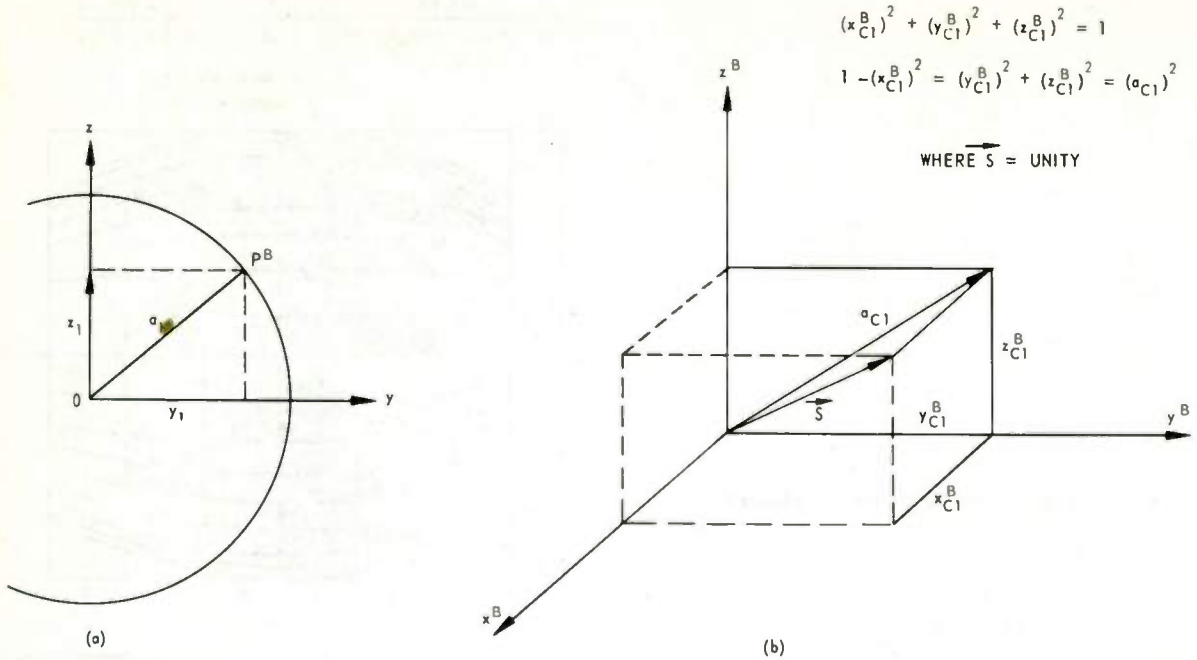


Fig. 15. Geometrical relations of position vector.

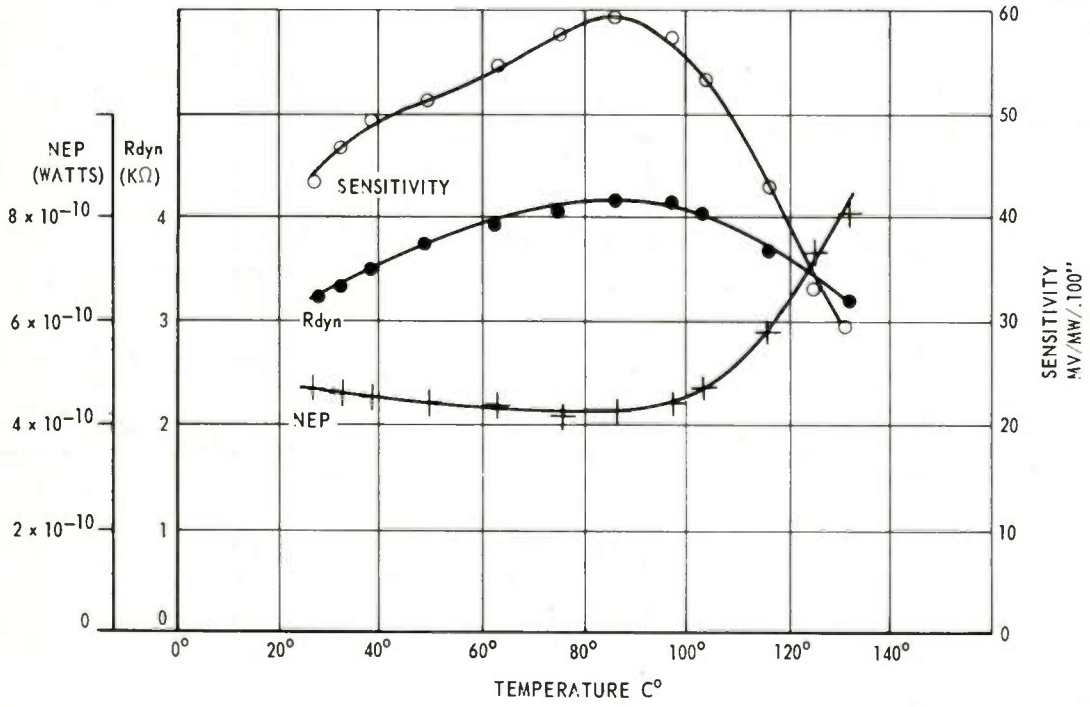


Fig. 16. Sensitivity, dynamic resistance noise equivalent power vs temperature.

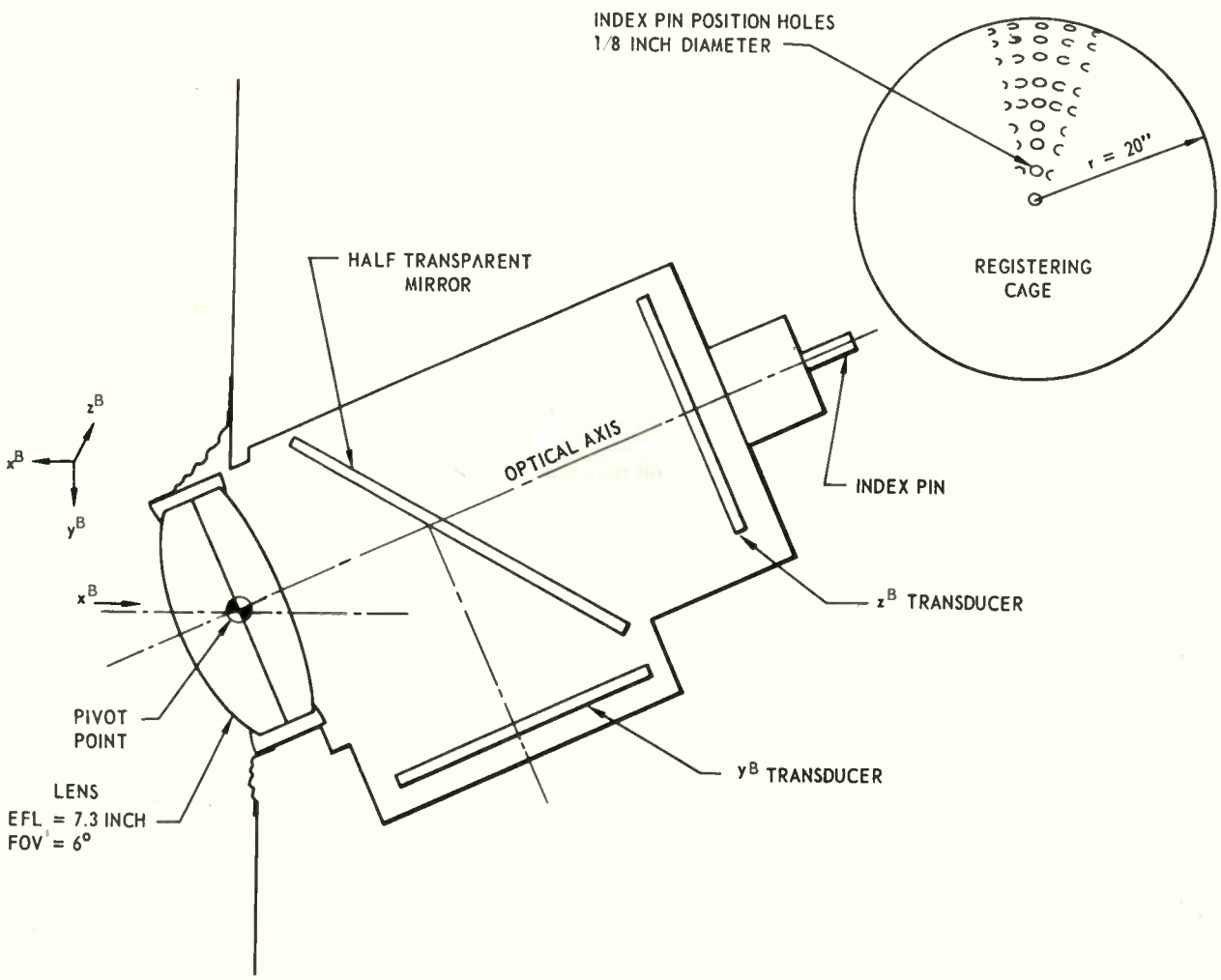


Fig. 17. Fine sun position indicator.

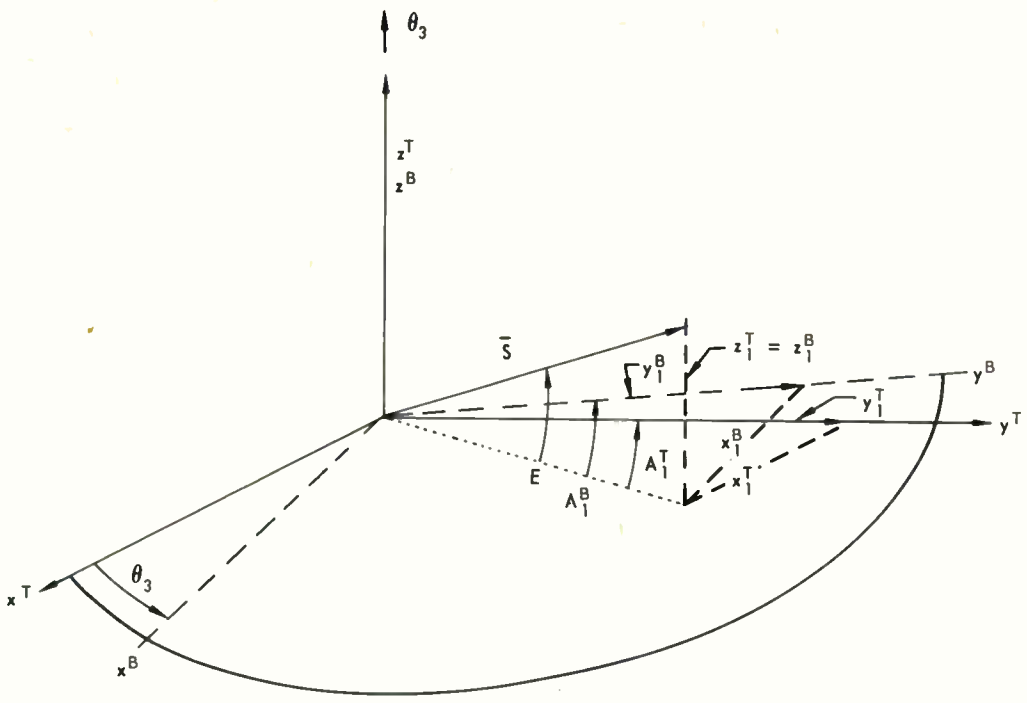
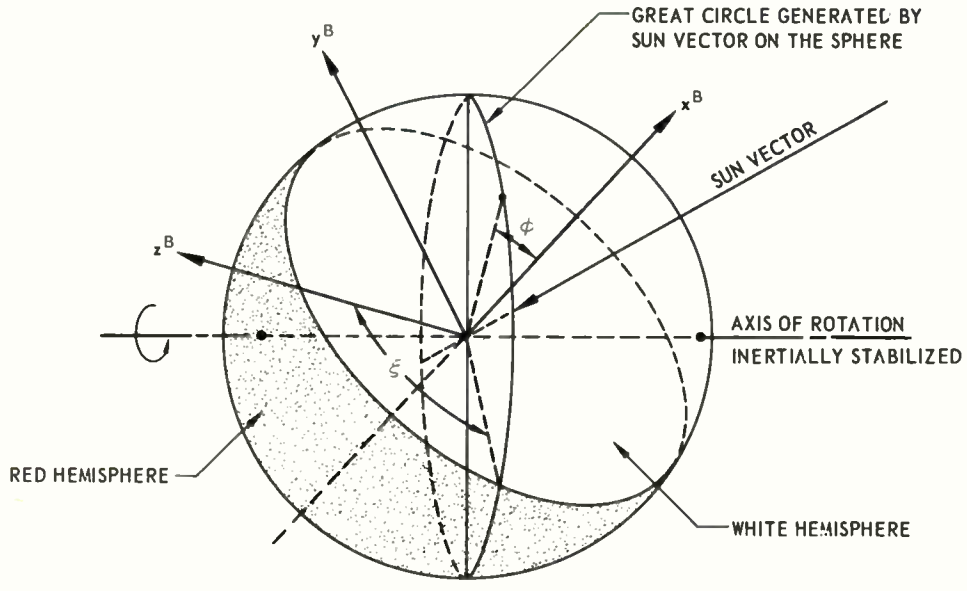


Fig. 18. Sun position indicator as yaw sensor.



ϕ = ANGLE OF ELEVATION BETWEEN PLANE OF GREAT CIRCLE AND x^B MEASURED IN COORDINATE PLANE $x^B - z^B$.

ξ = ANGLE OF AZIMUTH BETWEEN PLANE OF GREAT CIRCLE AND z^B MEASURED IN COORDINATE PLANE $y^B - z^B$.

Fig. 19. Sun position indicator as angular velocity sensor.

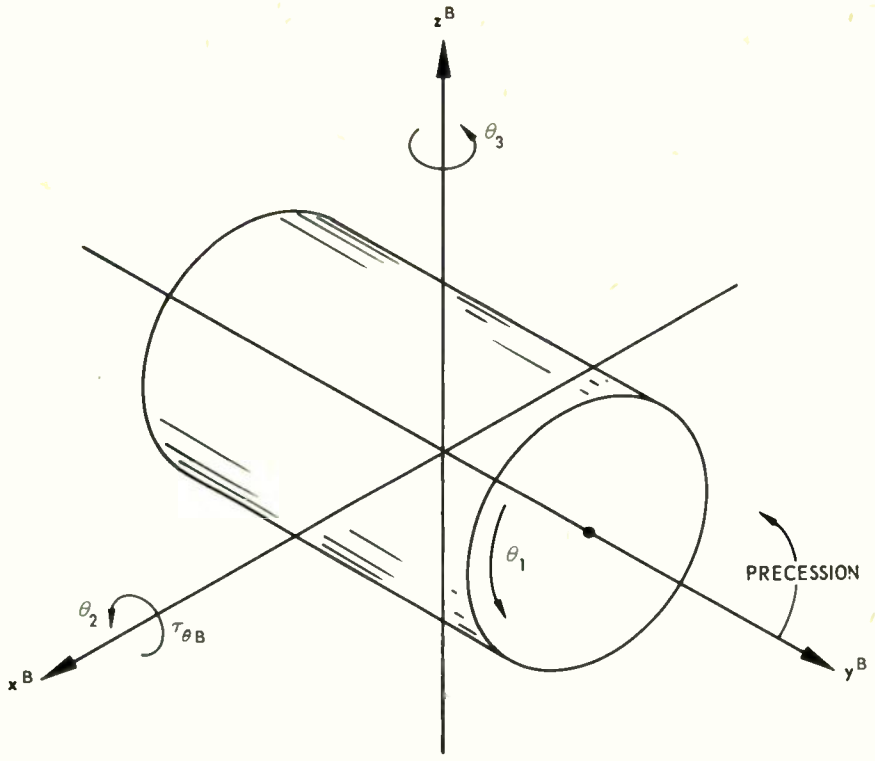


Fig. 20. Precession of rotating body under external torque $\theta_3 = \left(\tau_{UB}^{3}\right)$.

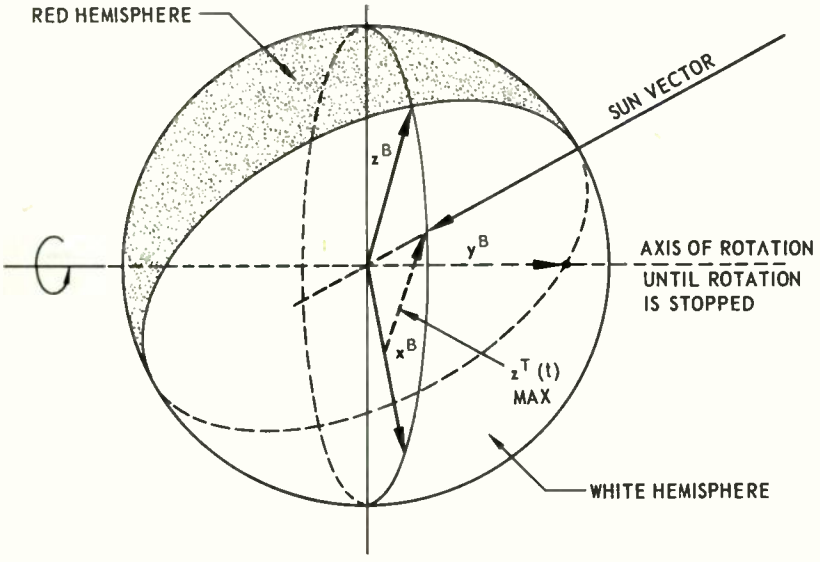


Fig. 21. Elimination of vehicle tumble.

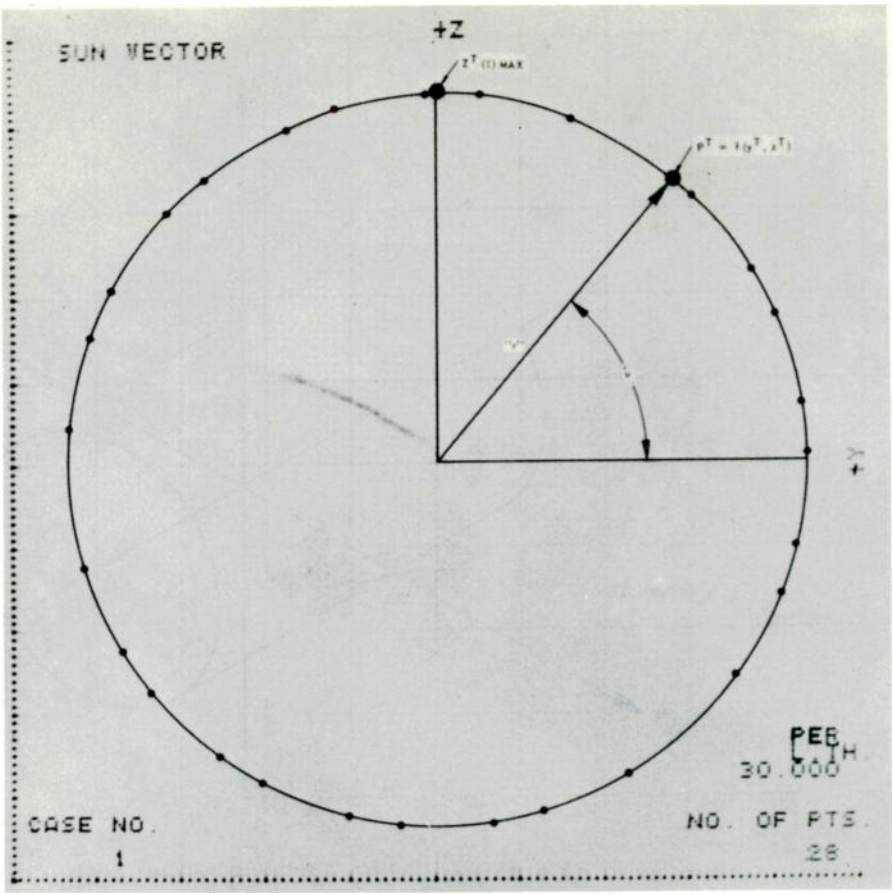


Fig. 22. Path of sun image at focal plane.

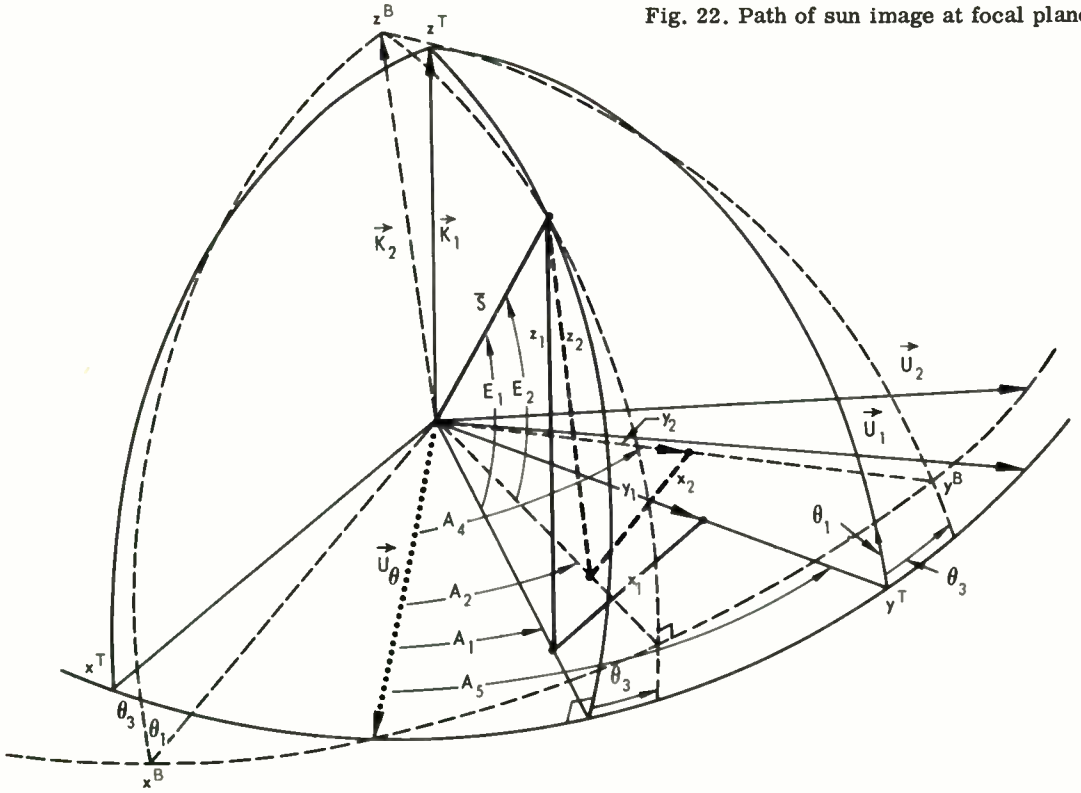


Fig. A-1. Sun position indicator error as function of attitude misalignment.

THE ASTROGUIDE--A SPACE-VEHICLE NAVIGATION SYSTEM

Morris M. Birnbaum
Program Manager, Photogrammetric Equipment Projects

Phil M. Salomon
Engineer, Circuits Design Group

Data Processing Equipment Section
Glendale Branch
General Precision, Inc., Librascope Division
Glendale 1, California

Summary

The Astroguide, a space-vehicle navigation system, comprises an array of image sensors, an electronics package, and a computer-controller. Each image sensor is composed of a wide-angle lens and a television-camera tube. By deflecting the camera-tube electron beam with direct-positioning currents on which an ac search pattern is superimposed, any small portion of the total field of view can be examined in detail.

Direct-positioning currents deflect the electron beam to "point at" the portion of the field of view being scanned; need for a gimballed platform or mechanical motion is thus eliminated. Therefore, each sensor can be rigidly attached to the vehicle shell. The azimuth-elevation angles, representing the direction of the image portion being examined relative to the optical axis of the sensor, are read by measuring the X and Y deflection currents in the camera-tube deflection coils.

Use of special scan patterns and electron-beam servoing circuitry permits a variety of navigational and control functions to be performed. The Astroguide system can be used for on-axis star tracking, off-axis star tracking, star-field matching, star-angle measuring, horizon scanning, command pointing, and field-of-view display. With field-of-view scanning, planetary mapping and image matching with stored images can be performed.

Introduction

To perform the guidance and navigation functions for space vehicles, a host of specialized instruments have been designed. These have included sun seekers, star trackers, star-field matchers, star-angle comparators, and horizon sensors. Most of these instruments use infrared detectors or multiplier phototubes and an optical system with a very small field of view. To scan a useable wide-angle field of view, rotating or wobbling reflecting optics are used; or the whole instrument is moved about on an azimuth-altitude servoed platform. Television-camera tubes have also been investigated for use in space-vehicle navigation systems. The lack of simple star tracking and accurate X-Y position-measuring circuits has limited their use to field-of-view displays.

Each of the aforementioned instruments adds its individual weight and complexity to the vehicle-navigation system. The need for (1) mechanical motions to scan a wide field of view and (2) a mechanical platform for pointing the instrument reduces the whole system's reliability and service life and introduces undesired shifts of mass in the space vehicle.

To devise a system which did not require a mechanical platform with its shaft encoder or optical pattern disk for pointing and readout of the pointing direction, Librascope engineers began in 1960 to investigate a space-navigation system using camera-tube sensors. Because a camera tube can scan a focused image electronically--with no mechanical motion--its use in space-vehicle instrumentation is attractive. Further, the imaging properties of the camera tube allow it to perform many of the functions handled by the aforementioned specialized and complex unifunction instruments. The investigation involved the search for image-processing concepts and systems suitable for space-vehicle guidance and navigation. Using a technique called "electron-beam servoing," a developmental system named the Astroguide was devised. Most of the circuits were breadboarded and tested. Results of the tests and investigation proved the versatility of the camera tube for various image-processing operations and for use as a highly accurate guidance and navigation sensor. This paper discusses the potential use of camera-tube systems to perform many tasks normally performed by several separate space-vehicle instruments.

General Description

The operation of the Astroguide is based on a sensor-system configuration composed of the following building blocks:

1. A lens with a wide-angle field of view.

2. An image-dissector camera tube with a 2-3/4-inch diameter photocathode.

3. A 0.0015-inch diameter aperture in the image-dissector aperture plate, allowing for a resolution of 1800 TV lines.

4. A dc-coupled deflection system for generating an X-Y positional deflection of the electron beam.

5. An ac-coupled deflection system for scanning and inspecting a small field of view around the point from which the electron beam has been dc-deflected.

6. Control and measuring systems.

7. A computer-controller.

If an optical image is focused on the photocathode surface of the image dissector, the photoelectrons leaving the photocathode will form an electron image. The electron image will be refocused by the dissector electron optics at the aperture plate of the image dissector. If time-varying scanning currents are sent through the deflection coils surrounding the image dissector, the electron image can be inspected by passing it through the aperture--one element at a time--and amplifying and displaying the output signal.

The Astroguide system permits inspection of a small part of the total field of view focused on the image-dissector photocathode. To do this, direct currents of known magnitudes are sent through the X and Y deflection coils to deflect a desired part of the total field of view to the aperture. Then, with high-frequency ac-scanning signals superimposed on the direct-positioning currents in the X and Y deflection coils, a small field of view around the desired portion of the total field of view can be inspected. Figure 1 illustrates the operation described. The sensor element doing the viewing has been diverted (or turned) electronically to some direction--at an angle θ with the zero reference direction--without any mechanical movement. With the physical geometry of the image-dissector tube and its deflection coils known and kept absolutely constant, and with the currents through the X and Y deflection coils measured with analog-to-digital circuits, the Astroguide can be pointed in any direction within its total field of view to an accuracy of one part in 20,000.

Circuits and systems were developed for use with a camera tube to perform the following operations without any mechanical motion of the tube:

1. On-Axis Star Tracking

   Track a star on the optical axis of the lens-camera tube. Use the error signals to change the vehicle's attitude so that the star remains on the optical axis.

2. Off-Axis Star Tracking

   Continuously scan a star image and servo the star electron beam's X-Y position to keep it always directed toward the aperture within the image-dissector tube. By keeping track of the X-Y position of the photocathode end of the electron beam, the star's position relative to the optical axis of the lens-camera tube is known.

3. Star-Angle Measurements

   Scan a field of view and read out the X-Y position of each star in a star field relative to the lens/camera-tube axis. From the data, the computer-controller can calculate the angles between stars.

4. Star-Field Matching

   Perform simple star-field matching with or without stored images of star-field patterns.

5. Horizon Sensing

   Sense the horizon and, aided by the computer, find earth vertical.

6. Image Matching and Image Viewing

   View an image and transmit it to earth, or match the viewed image with a stored image to permit the landing at a selected location on the moon or on a planet.

Obviously, a system which can perform all these operations must have certain trade-offs, as to limitations and accuracy, such as the following:

1. The system requires a computer-controller. To perform fewer functions, a smaller and simpler digital programmer can be used.

2. For infrared work above 1-micron wave length, camera tubes are not yet suitable. Although far infrared-sensitive surfaces are under development, sensitivity up to 1 micron is presently available in camera tubes.

3. For very high accuracy star tracking or horizon sensing, special instruments designed to perform a single function are more accurate than the Astroguide.

The camera-tube system fares surprisingly well when compared to available hardware. Mechanical complexity is completely avoided; the system is composed of components solidly tied to the vehicle frame. No large subassemblies move. Electronically, the special circuits needed to control the sensor and obtain intelligence from it are very simple. In all but one of the circuits, no high degree of accuracy is needed. Only in the analog-to-digital circuits, used to measure the currents in the X and Y deflection coils and the voltage to the camera tube, is a very high degree of measuring accuracy needed--one part in 20,000.

For the computer-controller, a Librascope AN/ASN-24 space computer was selected. The

computer has undergone extensive field testing and is now in quantity production.

A high degree of redundancy is present; several sensors are used and several sweep-pattern generators and analog-to-digital converter circuits are available. Therefore, the reliability of production-model Astroguide systems can be very high. The computer can be programmed to test for inoperative sensors, sweep circuits, or converter circuits, and to use the remaining ones which are still functioning.

A parts count of the Astroguide, exclusive of the computer and the image matcher, reveals that 225 transistors, 800 resistors, 150 capacitors, and 200 diodes are used. Average power of the Astroguide, excluding the computer, was estimated at 30 watts.

Sensor

The Astroguide sensor consists of a lens, a diaphragm, a camera tube, and a deflection yoke. An emitter follower, added behind the camera tube, permits transmission of the video output over several feet of coaxial cable.

Wide-angle lenses, such as are used in aerial cameras for mapping and photogrammetry, are suitable. For high off-axis pointing accuracy, resolution must be the best obtainable to minimize the star image's circle of confusion so its position can be measured accurately.

Using a 90-degree field-of-view mapping lens, a focal length of 1.5 inches, and with the diaphragm set at f/4.5, a plot of lens resolution over its field of view looks like this:

Degrees Off Axis	Visual Resolution in Line-Pairs per mm
0	219
5	217
10	212
15	172
20	170
25	165
30	109
35	81
40	62
45	46

From this plot, it can be shown that the lens can focus a star at the edge of its field of view as a circle of approximately 120 arc-seconds diameter.

For viewing terrain, rather than a star field, the diaphragm opening can be enlarged. More light will be available from the field of view, and the signal-to-noise ratio will increase. This can be done with a lens designed to function over a range of diaphragm openings.

The camera tube can be a vidicon, image orthicon, or image dissector. An image dissector was chosen for the following reasons:

1. Image dissectors with resolutions of up to 1000 TV lines per inch and 2-3/4-inch diameter screens are presently available. The Astroguide-system investigation was based on a tube resolution of 665 lines/inch to minimize focusing problems.

2. The screen (photocathode) is not damaged if one small element or area of the screen is continuously scanned for hours or days. With vidicons, the electron-scanning beam will burn a hole in the screen if left on one spot for any length of time.

3. The image dissector does not employ a thermionic electron gun. Therefore, power consumption is low, the tube is simple to operate, and its reliability is high.

4. The signal-to-noise ratio is adequate for the requirements of tracking and image displays.

The image-dissector tube has the following mechanical and electrical specifications:

Spectral Response	S1, S11, S20 typical; special responses possible, i.e., ultraviolet sensitive only
Photocathode Diameter	2-3/4 inches
Resolution	Up to 1000 TV lines per inch possible; 665 lines/inch used
Aperture Diameter	0.0015 inch
Tube Length	14 inches
Tube Diameter	4-1/2 inches

To investigate the adequacy of the image dissector for star sensing and field-of-view displays, the following equations are pertinent.

For an image dissector, the S/N ratio (noise in signal) is given by:

$$\frac{S}{N} = 6.62 \times 10^{3} D \sqrt{\frac{q N T}{2.512^{M}}}$$

where

D is the effective diameter of the lens in inches

q is the conversion efficiency of the photocathode in electrons per quanta

N is the per unit efficiency of the optics

T is the time of observation of an image element in seconds

M is the visual magnitude of the celestial body being scanned

For an f/4.5 diaphragm setting, an S11 photocathode, an observation time of 0.000125 second (f = 4kc), and a +3 magnitude star:

$$\frac{S}{N} = 6.62 \times 10^{3} \times 0.333 \sqrt{\frac{0.1 \times 0.95 \times 0.000125}{2.512^{3}}} = 1.9$$

Figure 2 shows the variation of S/N ratio with other observation times. Figure 3 shows S/N ratio versus effective lens diameter. Figure 4 shows S/N ratio for various detector quantum efficiencies. Figure 5 is a plot of visual magnitude versus star power densities. It can be seen that the dissector has more than enough sensitivity to sense over 150 guide stars with a minimum S/N ratio of 1.9 at a bandwidth of 4kc and an f/4.5 diaphragm setting. For dimmer stars, the diaphragm opening can be enlarged. However, a larger circle of confusion may result and, hence, lower pointing accuracy.

The ability of the image dissector to scan a field of view is shown by the following equations. Given a scene which illuminates the dissector photocathode with an E of 0.5 foot candle, the average instantaneous signal current, i_s, reaching the first dynode of the dissector tube will be:

$$i_s = sEa$$

where

s is the cathode sensitivity (60 microamperes per lumen)

a is the aperture area

$$a = \frac{\pi (1.5)^{2} \times 10^{-6}}{4 \times 1.44 \times 10^{2}} = 1.227 \times 10^{-8} \text{ ft.}^{2}$$

$$i_s = 60 \times 10^{-6} \times 0.5 \times 1.227 \times 10^{-8} = 0.368 \times 10^{-12} \text{ amps.}$$

$$\frac{S}{N} = \sqrt{\frac{i_s}{2ef}} = \sqrt{\frac{0.368 \times 10^{-12}}{2 \times 1.6 \times 10^{-19} \times 4 \times 10^{3}}} = \sqrt{288} = 16.9$$

If a square aspect ratio is used for displaying the field of view, a 1.9- × 1.9-inch area of the dissector photocathode will be scanned. This area contains approximately 1,660,000 image elements. With a 4kc video bandwidth, it will take about 3-1/2 minutes to scan the 1.9- × 1.9-inch area. For most space applications, this is adequate. With more light, smaller fields of view, and higher-performance photocathodes, faster scanning is possible.

Using the full 2.75-inch diameter of the photocathode for image scanning, approximately 2,740,000 resolvable elements are available for display or for data processing purposes. The camera tube, with a 0.0015-inch diameter aperture, resolves the field of view into three arc-minute elements. This is sufficient not only for high information content displays but for very high-resolution star tracking and star-field matching when special scan patterns are used.

Since the final pointing accuracy of the Astroguide depends upon measuring the values of the X and Y currents in the deflection coils, all of the parameters which affect the electron-stream deflection must be held constant. The deflection coil must be mechanically rigid, and its characteristics should not change with temperature. Lens and yoke distortions are not critical or disastrous in themselves as long as these distortions are known and remain constant. The computer-controller can be programmed to calculate the correct X-Y position of a point, once the deflection current X-Y position is known and the distortions for that X-Y point have been programmed into the computer.

The sun can damage the camera tube if it is within the field of view of the tubes's optical system. To prevent this, a silicon solar cell, within an identical field of view, is mounted parallel to the optical axis of the Astroguide tube sensor. When the sun is imaged on the solar cell, the cell's output signal is used to close the diaphragm of the sensor lens system to a small opening. The Astroguide can track the sun if desired.

Electronics Package

The electronics package contains the circuitry required for scan-pattern generation, sensor-output processing, electron-beam servoing, and X-Y measuring of the position of the field of view being scanned. The stored computer program, or commands received from earth, determine the manner for interconnecting and controlling the various subsystems within the electronics package.

Figure 6 is a block diagram of the Astroguide system. The computer connects any one sensor to the electronics package; the remaining sensors are inoperative. While the circuits and systems of the electronics package were designed to be controlled by the Librascope AN/ASN-24 space computer, other computers could also perform the control functions.

Pointing System

The computer controls the pointing to a discrete part of the celestial sphere. First, the desired sensor is chosen and turned on; then the portion of the field of view to be inspected is deflected to the aperture of the dissector tube.

A block diagram of the pointing system appears in Figure 7. As shown, the pointing system is a closed-loop servo system in which the computer serves as the error detector. The X and Y shift registers, into which the computer has placed binary commands, control the pointing of the electron beam within the sensor field of view. The binary commands are proportional to the coordinates of the desired target, as referred to the sensor optical axis.

Each shift register contains a Kirchhoff ladder network which converts the binary command into an analog signal. The analog pointing signal, received at a yoke driver stage, deflects the electron beam. Polarity-inverting circuits enable the driver stages to deflect in both the positive and negative directions.

A series resistor in each yoke coil monitors the resultant X and Y deflection currents. Each current generates a voltage drop across the resistor. A voltage-digitizing system converts this voltage into a binary number for input to the computer. The pointing coordinates are verified by comparing the binary number from the voltage-digitizing system with the original command. Any difference is algebraically added to the binary command in the shift register, and the new resultant current is monitored again. In this way, the exact pointing coordinates desired are achieved; and the feedback loop continuously samples and corrects for any drift or gain variations within the deflection driver system.

If lens and yoke distortions are large enough to be considered, the lens-yoke combination can be calibrated. It is assumed that the lens-yoke distortions will remain constant. By observing the X-Y position of the element of the field of view being pointed at versus the X-Y currents in the deflection yoke, the computer can be programmed to make the proper corrections. Thus the exact element being pointed at always can be known.

Scanning Systems

Two scan generators are used to inspect the images and the field of view. A small rosette pattern scan is used for tracking. A television scan is used for scanning and displaying a portion of or the entire sensor field of view. These scan patterns are shown in Figure 8.

The defining polar equation for a rosette is

$$r = a \sin nwt.$$

When transformed to a Cartesian system, this equation yields

$$x = \cos wt \sin nwt$$

and

$$y = \sin wt \sin nwt.$$

Because the sensor deflection yokes are mutually perpendicular, the X and Y deflecting currents must be proportional to cos wt sin nwt and sin wt sin nwt, respectively.

Several types of analog multiplying circuits are available. The bandwidth requirements of the product terms suggest the use of either quarter-square or time-division multipliers. Because of their simplicity and inherent reliability, the Astroguide system uses wide-band quarter-square multipliers to generate the rosette signals. Very accurate approximations of the square law curve can be obtained by using several diodes in parallel with appropriate biasing voltages.

Both X-axis and Y-axis deflection signals are fed to the appropriate deflection drivers, resulting in the rosette scan pattern. A rosette scan for target tracking proves advantageous because this pattern is nondirectional with respect to its center of symmetry, thereby allowing a non-weighted scan to examine the area surrounding the target. Timing pulses are also developed within the rosette generator, whose polarity and phase are later used in developing tracking-error signals.

The television scan generator provides a conventional television scan, suitable for scanning the sensor's field of view. The scan generator provides the various synchronizing pulses to be added to the sensor video output for telemetering.

Electron-Beam Servoing System

In the tracking mode, a small rosette scan pattern is superimposed on the direct-positioning currents to permit scanning of the area around the photocathode element. Through a technique called electron-beam servoing, the output signals from the sensor are processed to reduce the tracking errors to zero.

Examination of the X-axis rosette deflection signal (i. e., cos wt sin nwt) reveals that the polarity of the expression indicates whether the rosette pattern is being formed in a positive or negative X-axis direction. Inspection of the Y-axis deflection signal, (i. e., sin wt sin nwt) reveals similar information about the Y-axis position of the rosette pattern. Use of the polarity of the rosette deflection signals as gate-control pulses permits the quadrant relationship of a target pulse to the rosette pattern's origin to be uniquely determined.

Figure 9 is a block diagram of the electron-beam servoing system for pointing and for tracking. Four pulse-processing gates separate the series of sensor output pulses--obtained during deflection in the X-axis direction--from those obtained during deflection in the Y-axis direction. To separate these pulses, a logic matrix composed of four "and" gates compares the sensor output pulses with the rosette polarity pulses. From this matrix, four discrete pulse sets are possible: (1) pulses obtained with positive X deflections, (2) pulses obtained with negative X deflections, (3) pulses obtained with positive Y

deflections, and (4) pulses obtained with negative Y deflections. When an X gate and a Y gate outputs a pulse--indicating that a star has been sensed--the quadrant in which the star is located, relative to the rosette's origin, is immediately determined by noting the deflection polarity associated with the gates that conducted.

The X-axis gate outputs and Y-axis gate outputs go directly to the computer-controller. Each set of gate outputs is entered into a novel digital integrator called a "Sigmator," which is part of the computer. In the Sigmator, the pulses from the minus-direction deflection gates are reversed before being summed. As the Sigmator algebraically adds the pulses it receives, its resultant binary value including sign indicates on which sides of the rosette axes the star being sensed was positioned.

The integration of any periodic wave form is an evaluation of the dc component-time product for long integrating times. Thus the integration will yield zero if--and only if--the image is exactly centered at the origin of the rosette scan pattern. When this occurs, pulses from the plus and minus X and Y deflections will be symmetrical about a zero-voltage reference axis.

For any off-rosette-center target, the rosette X-axis and Y-axis Sigmator integrations will build up error voltages whose polarity relates functionally to the pointing error in azimuth and elevation. By integrating over many cycles of information, a very small angular error may be sensed.

The outputs of the Sigmator can be sent to one of two places. If the off-axis pointing direction of the electron beam is exactly as desired, the Sigmator outputs can be sent to reaction motors. The change in the vehicle's attitude, induced by these motors, will reduce the Sigmator outputs to zero. When this occurs, the rosette scan pattern is centered exactly on the star; the pointing direction relative to the optical axis of the sensor has not been changed; and for the X and Y axes, the plus and minus pulses from the "and" gates to the Sigmator are equal in quantity and pulse width.

To track the star without changing the vehicle's attitude, the Sigmator outputs are sent to the positioning shift registers, which control the pointing direction of the electron beam. The shift registers servo the beam to center the rosette on the star. As the vehicle moves or turns, the electron beam, with its ac rosette scan pattern, will follow the star to the edge of the sensor's field of view.

If the computer is not to be used for the aforementioned electron-beam servoing, outputs of the X and Y deflection polarity gates can be entered, respectively, into an X and a Y operational integrator. As the integrator charges, its voltage magnitude and polarity modify the direct-positioning currents in the yoke to keep the rosette centered on the star being tracked.

In servoing on a star which is imaged off axis, it is important to note that the lens will distort the star image. As focused on the dissector-tube photocathode, the star image will be a nonsymmetrical figure with the region of greatest light intensity not necessarily at the figure's center. A typical star image being scanned with a rosette pattern is shown in Figure 10.

When scanning this star image, the electron-beam servoing system will center the rosette on the geometric center of the figure, rather than on the area of greatest brightness. An error will therefore result in the exact X-Y position of the star image.

In a typical case, the star image will subtend approximately 120 arc-seconds; the smallest unit image element the dissector beam can inspect is 180 arc-seconds; and the voltage digitizers can resolve the position of the image element being scanned to approximately 8-arc-second increments. Though the dissector beam is larger than the diameter of the image, it can resolve the center of the image because it is dissecting the image from all directions; and the signals causing a rosette pattern to be traced can define the center of the rosette to within 1 arc-second.

The rosette scanning pattern will locate the star-image center to within 1 arc-second. Because the principal ray will not be at the center of the image, an error of up to 10 arc-seconds may occur between the position of the principal ray and the star-image center. If the voltage digitizer errs by one bit in resolving the X-Y position of the star image center, due to the image center falling on a boundary between two digital positions of the electron beam, another 4-arc-second error may occur in the X and Y axes, respectively. Depending on the star's location in the field of view, a pointing error of up to 11 arc-seconds may occur along an axis. This represents a positional root-mean-square error of plus or minus 16 arc-seconds, or approximately one part in 20, 000 of the total field of view.

If the angular errors between the star-image center and the principal ray can be predicted over the field of view, corrections can be stored in the computer's memory to increase the pointing accuracy of the Astroguide to a root-mean-square error approaching plus or minus 10 arc-seconds.

Computer-Controller

The Astroguide was designed for use with the Librascope AN/ASN-24 computer. However, other computers can be used by modifying the design of the electronics package which mates with the computer and the sensors. The basic requirements of any computer to be used with the Astroguide are: (1) ability to operate in a space environment, (2) a memory of more than 1000 words, and (3) ability to operate with input-output equipment consisting of analog and digital devices.

The AN/ASN-24 computer has been used successfully in space applications. The characteristics

which make this computer especially applicable for the Astroguide system are as follows:

Weight	32 pounds
Size	0.55 cubic foot
Operating temperature	-55° C to +100° C
Power requirements	120 watts
Memory type	magnetic drum
Clock frequency	160kc
Word length	25 bits
Memory capacity	2560 words
Mode of operation	serial
Number base	binary
Sigmator unit	
Track select register for controlling up to 63 relays	

For missions requiring smaller weights and power requirements, completely solid-state computers can be designed. A proposed solid-state version of the AN/ASN-24 has the following physical characteristics:

Size	0.37 cubic foot
Weight	19 pounds
Power requirements	70 watts on an intermittent basis (less than 5 watts when measuring elapsed time)

Tasks Performed By Astroguide

As described in this paper, the various Astroguide subsystems can perform the following tasks quite easily: (1) on-axis star tracking, (2) off-axis star tracking, (3) command pointing, and (4) field-of-view display.

To measure star angles, the computer controls the deflection coils so that a coarse TV raster is scanned on the face of the dissector tube. The small rosette pattern is superimposed on the TV raster. As a star is intercepted, the rosette servoes to the star-image center; and the star's X-Y coordinates are entered into the computer. The TV-raster scan is then continued, and the remaining stars in the star field are intercepted. With the X-Y coordinates of each star known, the angles between the stars can be calculated and the star field identified.

To perform star-field matching, stored images of the star fields being used as the reference, or the computer memory, can be used. If a stored image is used, a TV raster pattern or a rosette pattern can scan both the actual field of view and the stored image. This requires a special image-matching unit with its own scanning tube and file of stored images. If the computer memory is used, the pulse pattern resulting from scanning the field of view is entered into the computer and compared with pulse patterns stored in the computer. This type of star-field matching has been done very successfully at Librascope.

To perform horizon sensing, the computer controls the TV scan pattern. A coarse TV raster scan pattern with a superimposed rosette is traced on the face of the dissector tube. A high-pass filter network in the video amplifiers senses boundary crossings as pulses. The rosette servoes on the boundary, and the X-Y position of a point on the boundary is entered into the computer. Many boundary points are sensed and entered into the computer. The computer calculates the chord of the boundary arc and along which axis the vertical to earth or to a planet lies. By using two or more sensors at 90 degrees to each other, the vertical direction can be computed. Figure 11 shows the scan pattern used for horizon sensing, star-angle measuring, and image matching.

Image matching is done with the same pattern used for horizon sensing or star-angle measuring. In this case, the diameter of the small rosette pattern defines the area of the image to be matched. Using stored images and a scanning unit, the signals from the stored image are compared with the signals obtained by scanning the field of view, a small area at a time. When the rosette pattern is within pull-in range of the image to be matched, the electron beams servoes itself on the image. The matched image in the field of view can be tracked in the same fashion as a star was tracked.

Status of Astroguide

The Astroguide investigation was carried far enough to evolve the circuits and systems needed for a successful model. The following statements summarize the present status of the Astroguide.

1. The computer-controller is in production. Newer models are being designed.

2. The image-dissector tube is available commercially as a custom tube.

3. The pointing system, including the deflection circuits, and using digital voltmeters for analog-to-digital conversion, was designed and breadboarded.

4. The following systems, including all circuits, were designed and breadboarded: (a) scanning system, (b) electron-beam servoing system, and (c) an image-matching system.

Conclusions

The development of the Astroguide system proved the practicality of using camera tubes as sensor elements of a space-vehicle navigation system. Three important goals were achieved: (1) a nonmoving star-tracking system was devised which eliminated mechanical platforms, gimbals, and mechanical-optical readouts of azimuth and elevation pointing angles; (2) a relatively simple electronic system was developed which could

perform many guidance and navigation functions; (3) electron-beam servoing techniques were developed which showed great promise for use in future tracking systems of all kinds.

The development of the Astroguide showed that future work should center around development of (1) infrared-sensitive surfaces which can be scanned, and (2) smaller, lighter, computer. Both of these developments are underway at Librascope.

Acknowledgement

Many of the concepts used in the Astroguide system evolved from earlier star trackers built by Librascope for Dr. John D. Strong of Johns Hopkins University. Our contact with Dr. Strong prompted Librascope's initial interest in producing instruments for astrophysical research.

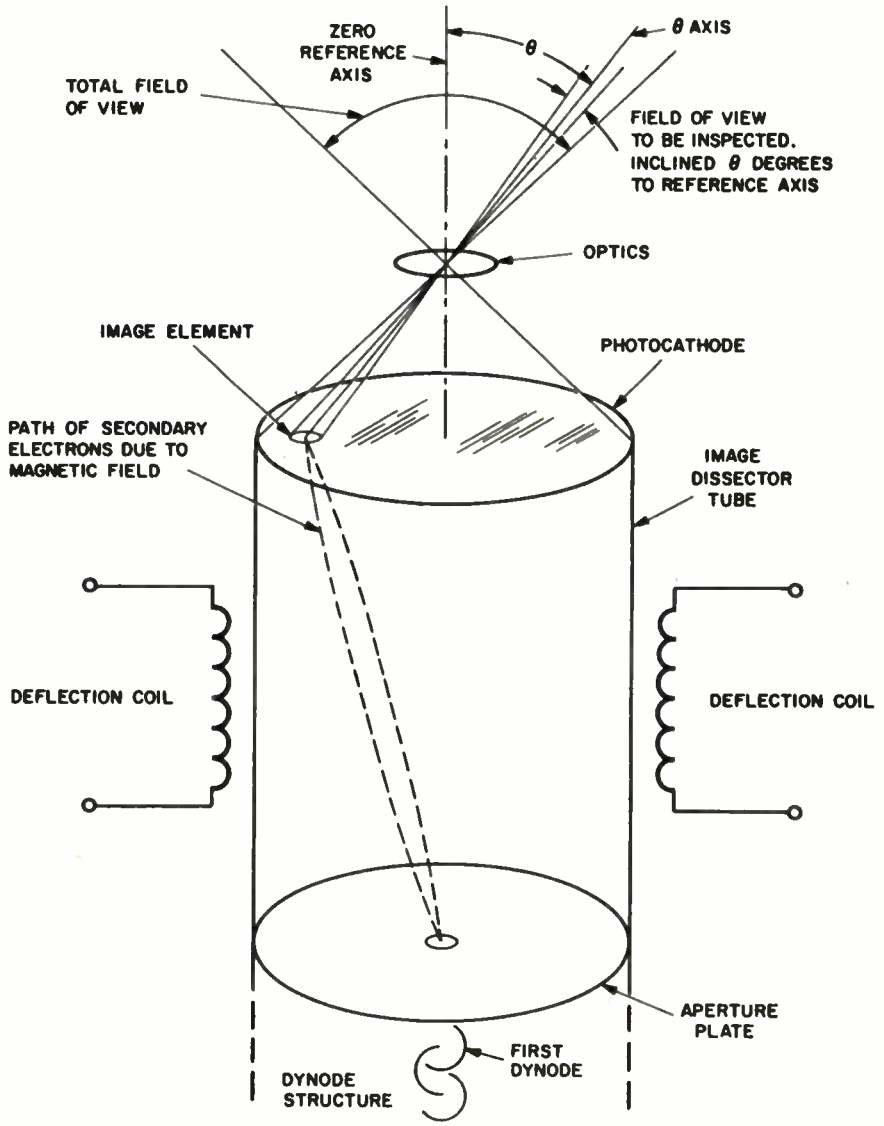


Figure 1 Electro-Optical Schematic

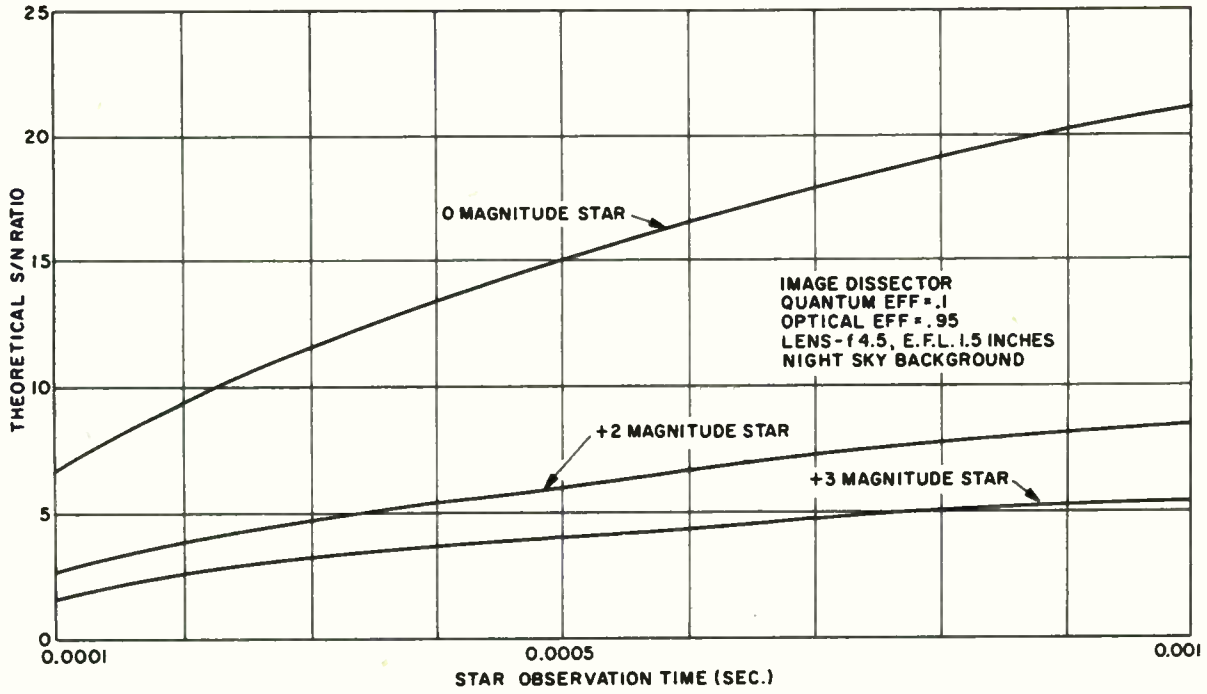


Figure 2 Theoretical S/N Ratio vs Star Observation Time

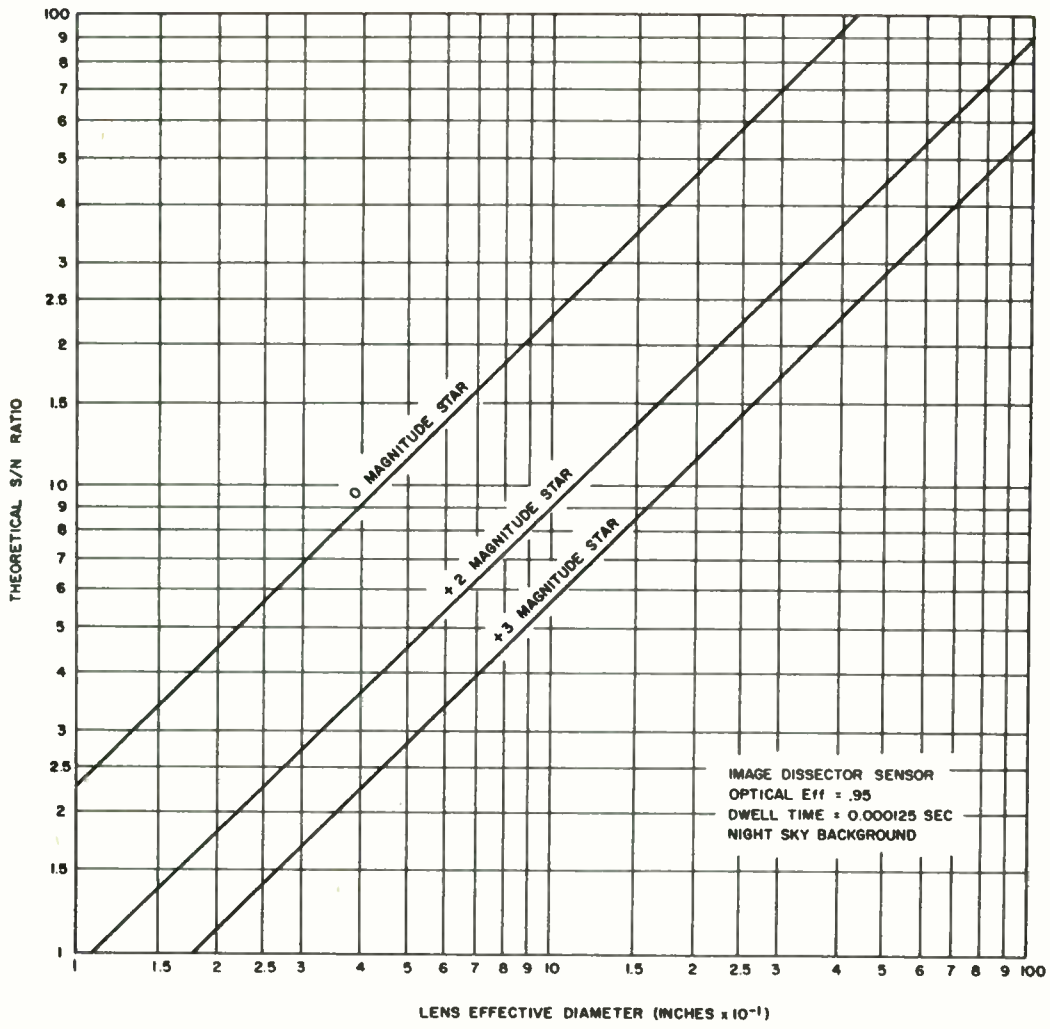


Figure 3 Theoretical S/N Ratio VS Telescope Diameter

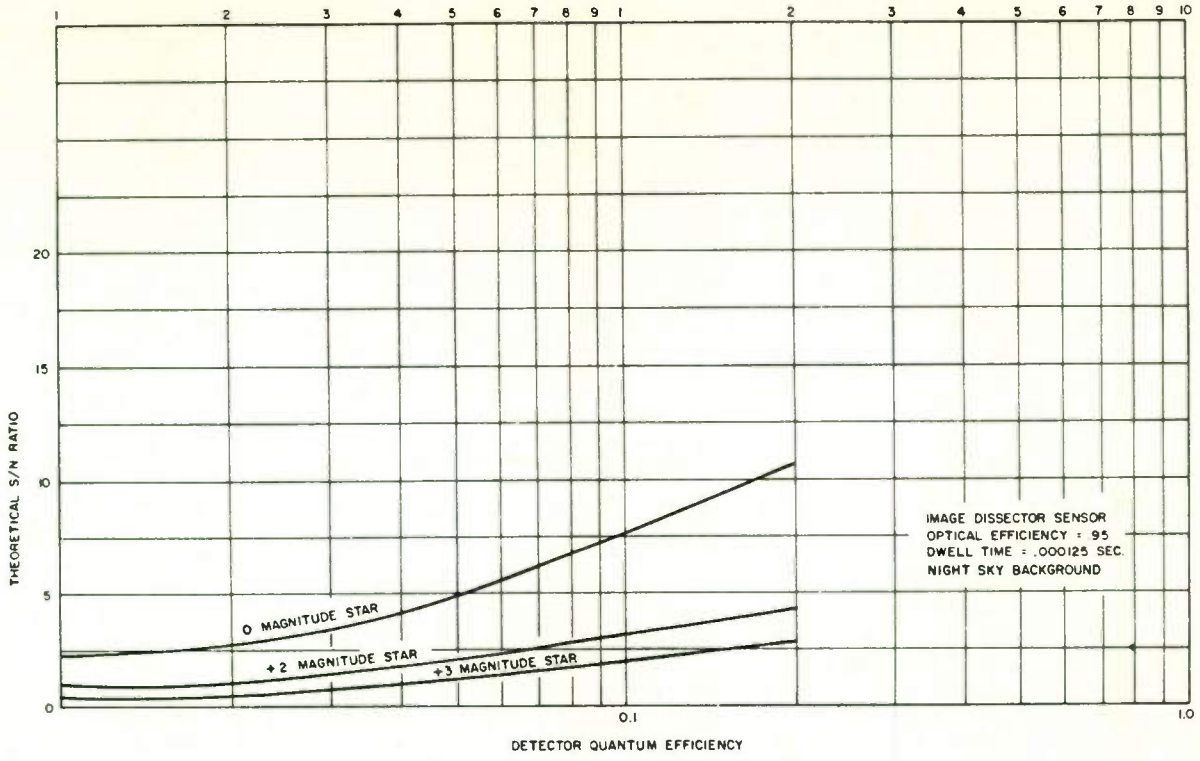


Figure 4 Theoretical S/N Ratio VS Detector Quantum Efficiency

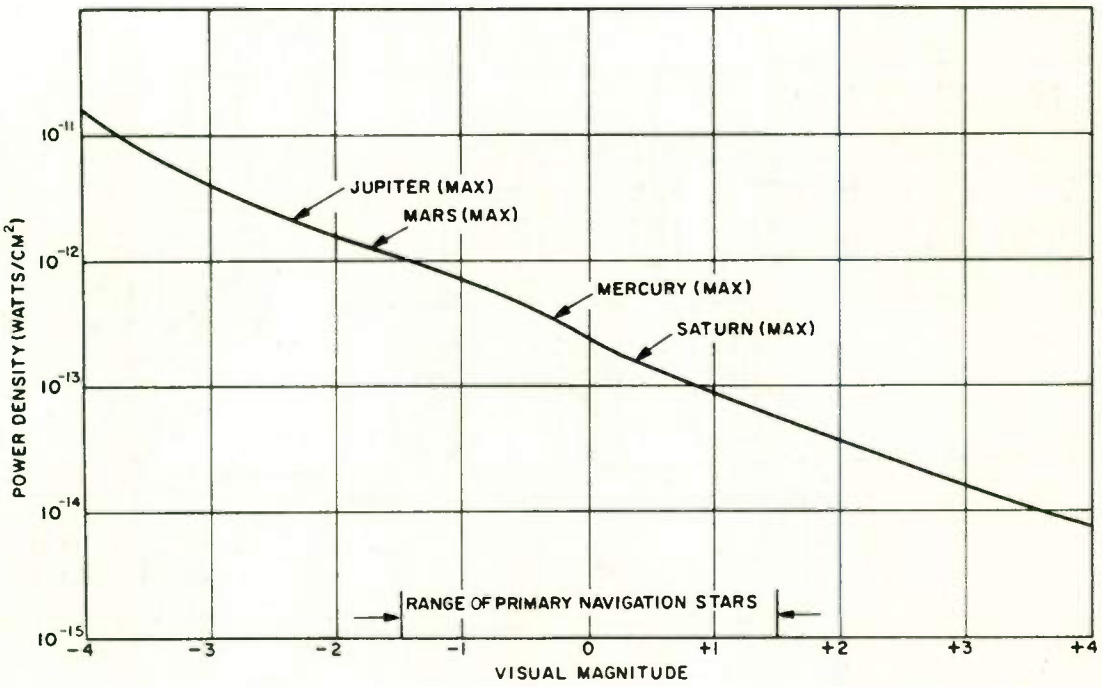


Figure 5 Power Density vs Visual Magnitude

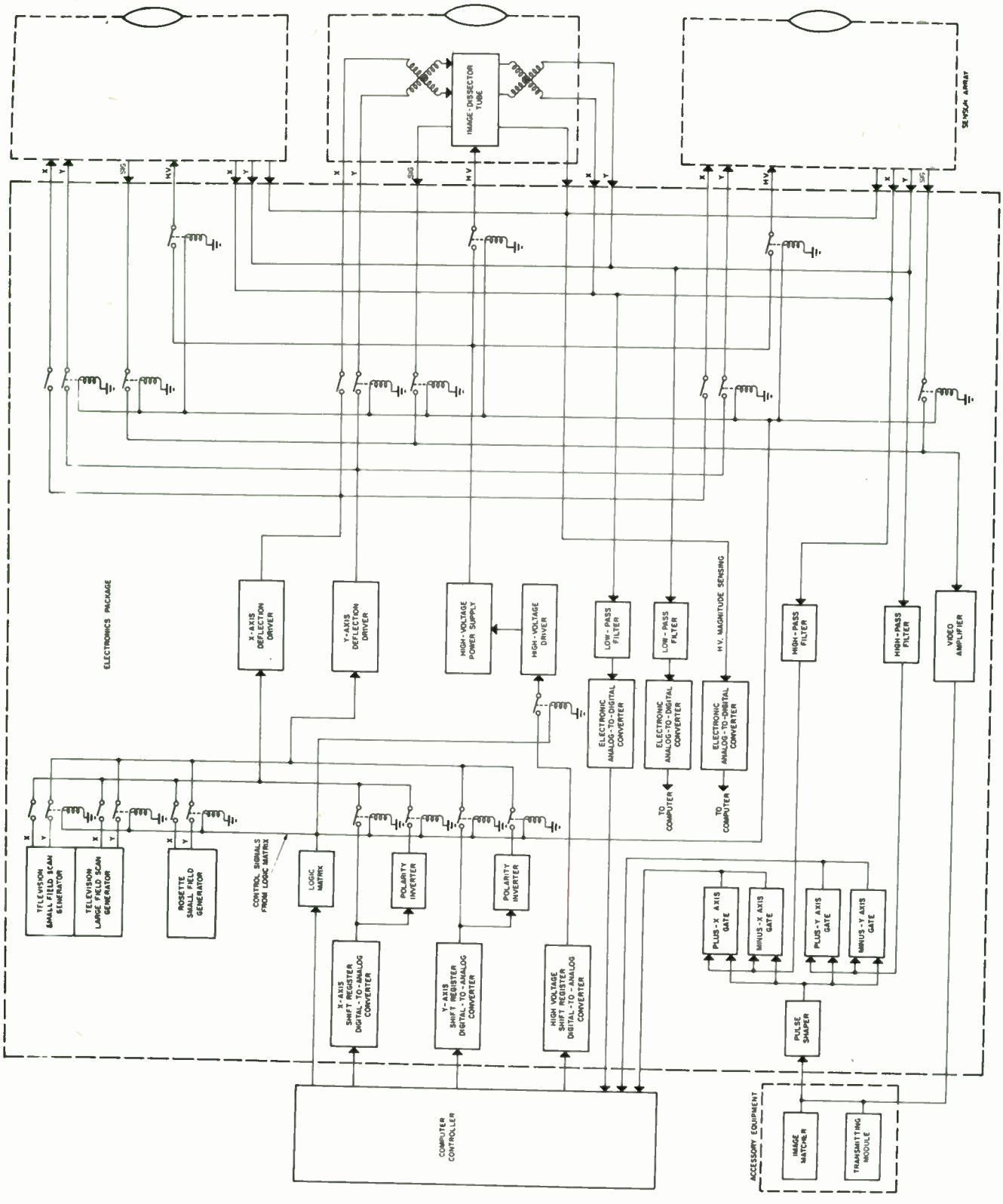


Figure 6 Astroguide System Block Diagram

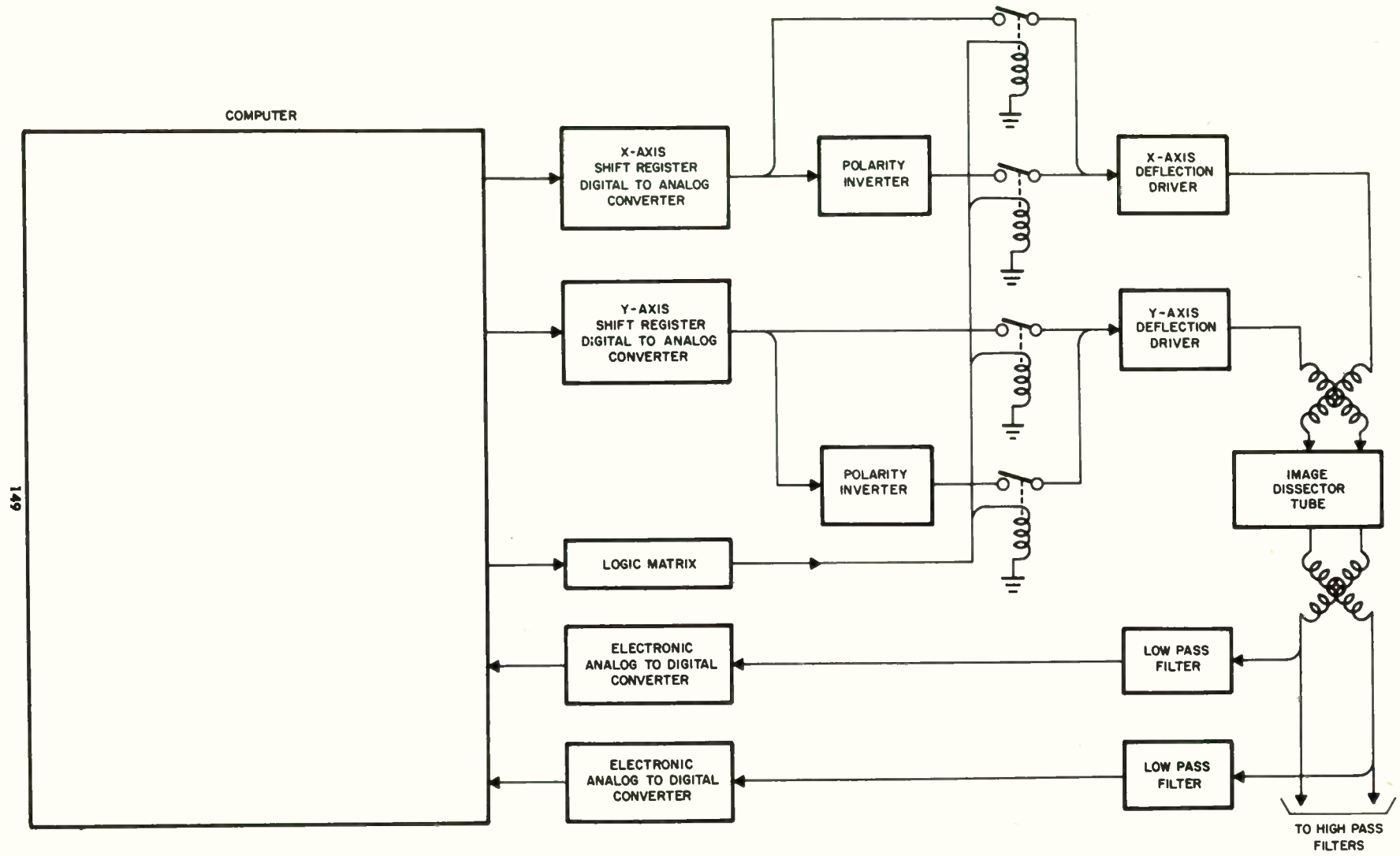


Figure 7 Astroguide Pointing System Block Diagram

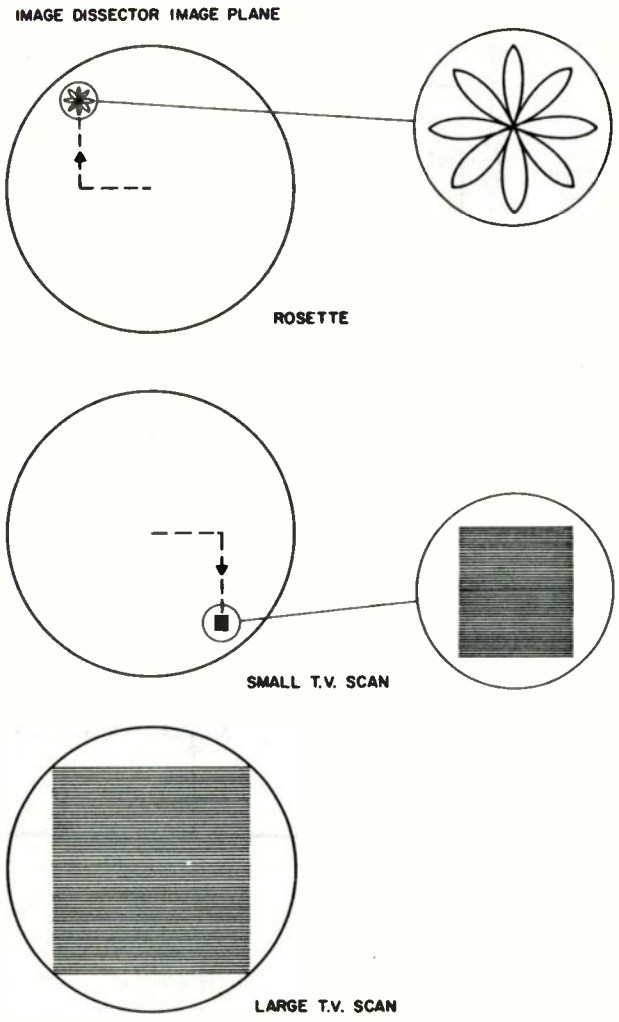


Figure 8 Astroguide Scan Patterns

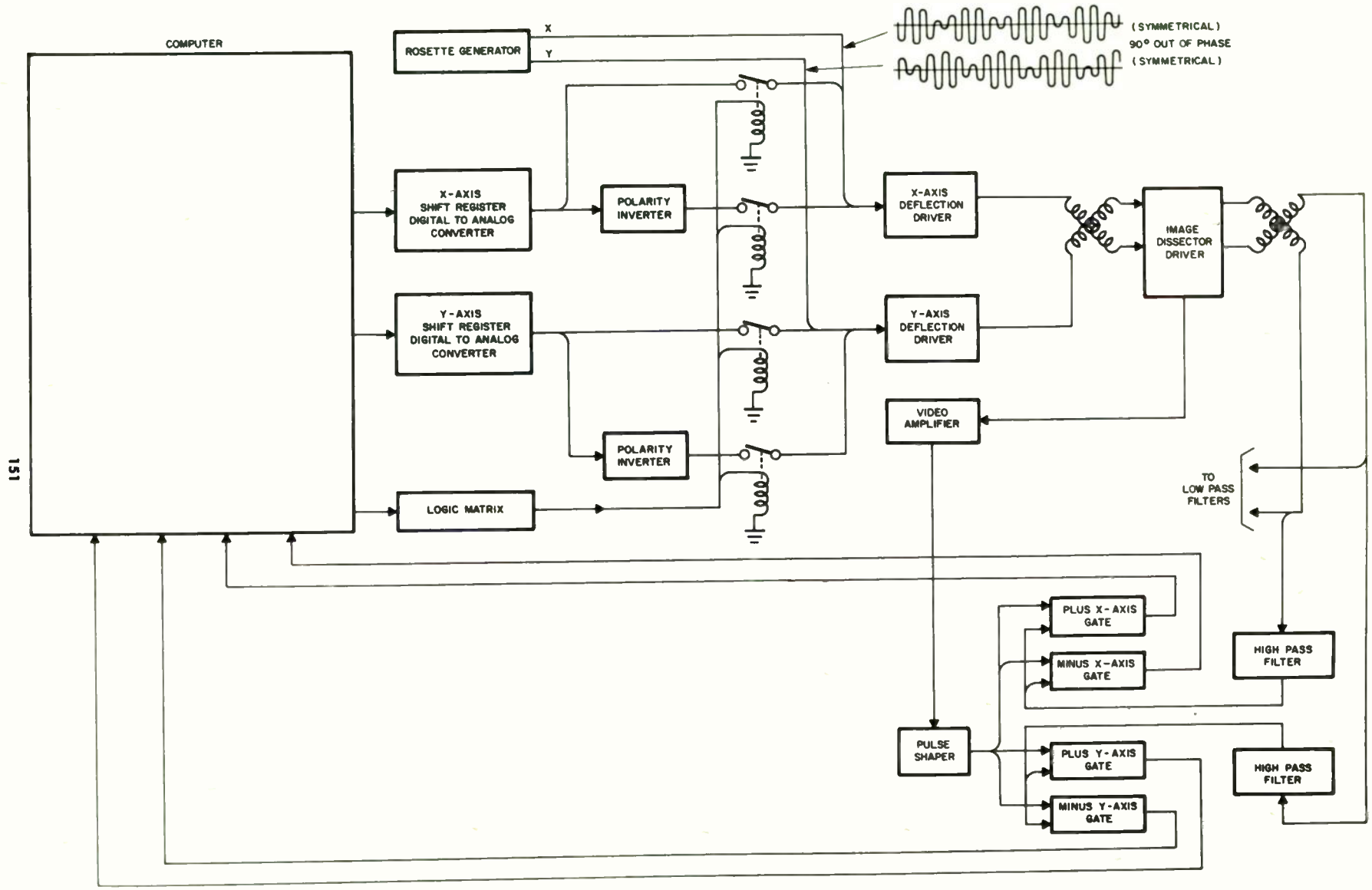


Figure 9 Astroguide Electron Beam Servoing System Block Diagram

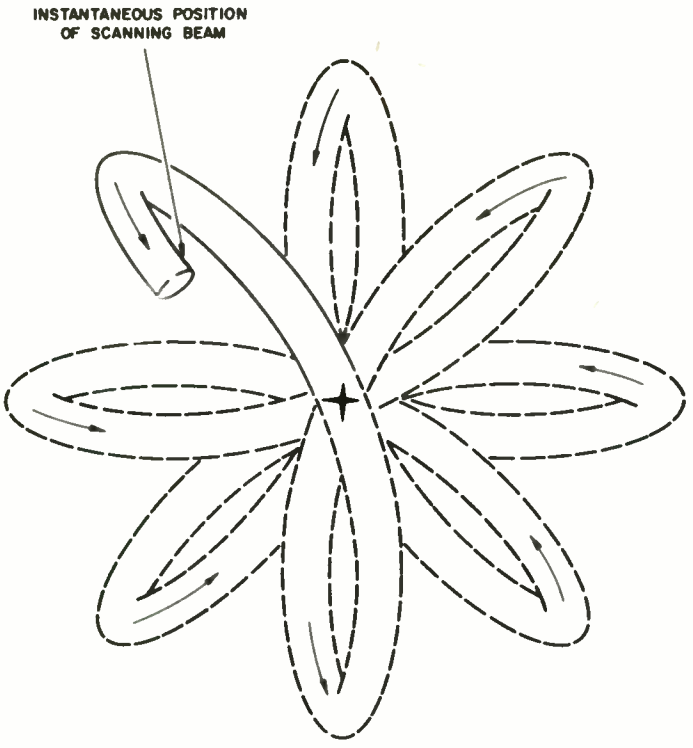


Figure 10. Typical Star Image and Rosette Scan Pattern

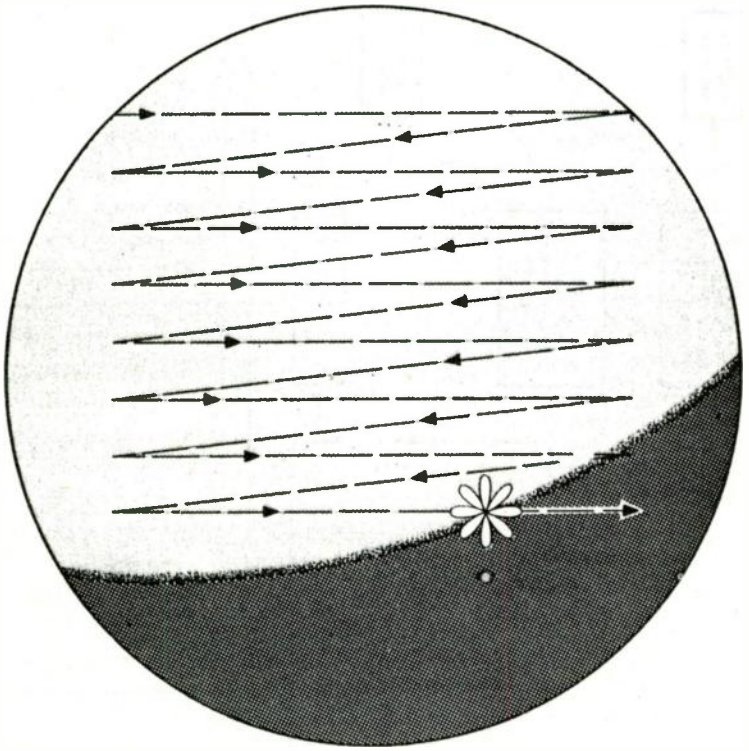


Figure 11. Scan Pattern for Horizon Sensing, Star Angle Measuring and Image Matching

EMERGENCY AND ROUTINE SPACE VEHICLE RECOVERY

J. B. Meyer and B. R. Mayo
Defense Systems Department
General Electric Company
Syracuse, New York

SUMMARY

The returning vehicle must be controlled so that it does not exceed temperature or load limits or skip out of the atmosphere. The vehicle energy must be managed so that the landing sites is reached with the proper residual energy for landing.

It is assumed that the pilot provides primary control for the craft. However, mission reliability is increased by the availability of an earth-based all-weather control and landing system. The cost of fixed ground sites to cover all possible trajectories is prohibitive. A mobile system that would be deployed in an optimum fashion for particular missions would be more economical. A mobile x-band C.W. beacon-tracking radar with integral voice, command, and telemetry channels meets the necessary requirements. Associated with the tracker is a digital data processor especially designed for real time control applications.

This terminal support system is described and various flight requirements for support are reviewed for successful re-entry, landing or recovery of the spacecraft.

Introduction

A most critical flight regime for Manned Space Vehicles is that of safe re-entry, landing and/or recovery on the earths surface. These craft approach the atmosphere at high velocities and can operate safely only under the very narrow load/thermal constraints for flight safety of the crew and of the vehicle - yet must retain sufficient energy to approach and land on prescribed landing areas.

This is a report of findings from several studies which have been underway concerning the critical re-entry and recovery phase of both manned and unmanned space flight and includes consideration of re-entry landing/recovery support for the several known space missions.

Geographical Requirements

(Figure 1) This map of the world shows typical ground tracks for orbital and lunar missions. The following assumptions can be made from these ground tracks:

1. Lunar insertion trajectories, being further south than orbital, may require ground stations in new locations.
2. Routine recovery for spacecraft may be expected at Edwards or in the Southwest, but emergency areas may be planned in the Atlantic and Pacific Oceans.
3. Ground tracks of returning spacecraft may be within an envelope extending from the South Pacific to Japan and Alaska.
4. Recovery of boosters or payloads due to an abort during launch may occur anyplace from the Cape to Africa or beyond, depending on the point on the boost rajectory where the abort occurs.
5. Recovery of spacecraft due to an abort after insertion may occur anyplace on the globe.

The orbits shown are but a few of those possible. One can readily see that fixed ground stations which can provide re-entry/landing/recovery coverage for all possible orbits would be cost prohibitive. A mobile system that could be deployed in an optimum fashion for particular missions would be more economical.

To achieve maximum utilization in view of the time required to traverse the large distances involved, mobility should be by air, using standard aircraft which are available in quantity, such as the C-119.

Flight Path Control Requirements

Re-entry

(Figure 2) Vehicle recovery includes manned and unmanned types, but the problems that will be encountered are similar. The aerodynamic coefficients, the temperature and the load limits of the vehicle define an acceptable glide corridor for vehicle entry. The upper boundary of the corridor corresponds to flying the vehicle $C_{L_{max}}$. A large lift coefficient and the resulting large lift force tend to keep the rate of change of altitude initially low, but kinetic energy is rapidly dissipated as air density increases.

Maximum range flight is obtained when an equilibrium glide is established at L/D max, approximately down the middle of the corridor.

If the velocity and altitude of the vehicle are such as to place it within the recovery ceiling the vehicle will be unable to enter without exceeding either its temperature or load limits.

This, then, provides the background for a lifting reentry. For superorbital velocities entry may begin at 400,000 feet and velocities of 36,000 feet per second. At $C_{L_{max}}$ the lift will tend to lessen the penetration into the atmosphere. If the entry is shallow the vehicle may skip out of the atmosphere, possibly into the Van Allen belt. If reentry is too steep it may exceed the temperature limit or, if pullout is not accurately controlled, may bounce out of the atmosphere.

(Figure 3) The maneuvers required to remain within the entry corridor have a significant effect on the ability to reach the final landing site. A four-phase control program could look like this. Phase I, the reentry phase, would be a programmed angle of attack, possibly constant, which would be a function of position and velocity, acceleration, structural and heating loads, and range to the landing site.

Transition to Equilibrium Glide

In order to prevent an excessive skipping trajectory a transition to an equilibrium glide is introduced. This might be at the bottom of the first skip and could include a roll to direct the lift vector downward.

Homing

The homing phase begins when a combination of altitude and velocity are met which corresponds to an arbitrarily selected equilibrium glide path. When $V = V_e$ the vehicle is controlled to fly at a constant bank angle and constant angle of attack until it is homed in on a heading toward the recovery site.

Range Control

Range control begins when the vehicle is headed for the recovery site. This includes any energy dissipation that is required to bring the vehicle to the proper position and velocity window for the final touchdown.

Need for Ground System

Pilot Backup

The premise of this discussion is that fairly broad limits in altitude and velocity will exist during phases II, III and IV when maneuvering will be performed. On board computer capacity may be limited and simple programming will be desireable. On the other hand a large ground computer capacity would allow fuller exploitation of the maneuvering capacities of the entry vehicle.

One of the primary functions of an astronaut will be to control his vehicle. This means that all sensing, stabilization and decision will be performed on-board the vehicle.

In order for him to accomplish these functions, it will be necessary to integrate the earth based tracking and guidance system with on-board inertial-stellar capability.

This will enhance the accuracy of the data for decision making, make available real-time ground-based computational support and/in event of malfunction provide an auto-pilot flight control backup for terminal phase recovery and landing.

Booster Recovery

(Figure 4) In forecasting the need for recovery we have considered unmanned boosters, unmanned payloads, and manned payloads. A typical booster mission might be the recovery of SATURN with a parawing to achieve range extension and maneuverability. This booster fits on the lower end of the altitude-velocity curves of Figure 2.

For early C-1 boosters, recovery could be made at the Cape. Higher burnout velocity and altitude of the C-3 indicate downrange recovery. Logistic and immersion complications make water recovery undesireable, and therefore, studies have been made on problems such as landing on an island and snatching by a tow plane for return to the skid strip at the Cape. It appears that the job can be done adequately with existing ground command equipment. Computer simulations indicate that a nominal 1.5 feet per second sinking speed can be achieved at touchdown. Control to touchdown requires that the ground radar see both the vehicle and the runway. Therefore, a system would be required at the landing site.

Abort Payload Recovery

During Boost. The same system could be used as a backup for payload recovery in the event of an abort during boost. For this application support gear could be spotted judiciously along the range compatible with possible abort conditions. In the event of an abort in the booster, maneuvering information could be furnished to the pilot or, for an unmanned payload, commands could be furnished to a closed loop control system. Handover capability could be provided for midcourse tracking and to provide backup to the vehicle.

After Insertion. In the event of abort after insertion into an orbital or interplanetary mission ground information is very important. A

manned vehicle might require immediate return in event of serious failure. A ground system could rapidly compute a safe trajectory, furnish information to the crew for their action, or, if necessary, take over a closed loop control system. Such a system could also track the vehicle, and, if it could not reach home base, predict its impact area as an aid to other search and retrieval forces.

Mission Example

(Figure 5) Let us now consider an example of an abort during boost of a manned vehicle. The conditions at radar acquisition are shown in the upper figure. The shaded area represents the vehicles maneuvering capability and A, B, and C represent island landing fields. The vehicle has too much energy to land at Field C but it can reach both A and B. Assuming that A is the desired site, the computer establishes a flight path that will bring the pilot to his final approach.

If for some reason the vehicle does not follow the recommended flight path its new position and heading are shown in the middle figure. We can see that its possible landing area is reduced because of energy dissipation and has shifted because its heading has changed. It can barely reach field A. The computer would then select a maximum L/D path with minimum bank for a straight in approach.

If continual tracking and computation indicates the vehicle cannot reach this field but can reach field B, the computer derives a new flight path for the pilot.

The system can be made fully automatic by inserting an autopilot in the loop.

Recovery System Design

General

(Figure 6) Work on this critical terminal phase of space flight indicates that a single type of mobile system, air transportable, could perform normal and emergency functions for manned and unmanned vehicles. A short summary of requirements is presented here to establish a basis for the recovery system design.

(Figure 7) This is a simplified block diagram of a system designed to meet the recovery requirements of Figure 6.

It consists of a ground based C.W. amplitude/amplitude comparison 10.5 KMC radar tracker with voice and command communication channels modulated onto the tracker carrier; a c.w. phase-locked frequency-offset vehicle beacon with provisions for receiving the voice and command channels from the ground and for modulating voice and diagnostic telemetry onto the beacon transmitter; and a ground based digital data processor which receives vehicle position and telemetry information from the tracker receiver, supplies vehicle commands to the tracker transmitter, and provides information for the displays.

The choice of a 10.5 KMC C.W. tracker was made from the following reasons:

1. 10.5 KMC is a good compromise between weather attenuation, antenna size required for narrow pencil beam, and ability to see through an ionized sheath.
2. 10.5 KMC is the frequency of the long baseline precision trajectory measurement systems being installed by the Air Force. The beacon is compatible with these systems and thus a single beacon could function with either the fixed long line precision systems or with the single tracker air transportable systems.
3. The use of a c.w. system permits the tracker and beacon carriers to serve as the r.f. source for command, voice and telemetry communications channels. This results in an equipment simplification with correspondingly greater reliability.
4. The use of c.w. system permits greater freedom in the location of the tracker with respect to a landing strip since there is no duplexer recovery time. Thus the tracker may be placed close to the landing strip for optimum look angle during landings in order to avoid low angle multipath.

Tracker Exciter and Power Amplifier

(Figure 8) This shows the recovery system tracker exciter and power amplifier. There are two carriers of slightly different frequencies feeding the diplexer on the antenna. One of these is fixed in frequency while the other is swept in frequency periodically in order to obtain a range calibration. The command information is impressed on the fixed frequency channel. The ground-to-vehicle voice is impressed on the swept frequency channel. Provisions are made for either 10 watt or 1KW per channel, the choice depending on whether all weather performance is desired or not.

Beacon

(Figure 9) This is the beacon. Each channel from the tracker transmitter is offset in frequency and retransmitted to the ground. The command and voice signals are extracted from the signals from the ground and the telemetry and voice are impressed on the signals to be transmitted to the ground. The beacon output signals are phase locked to the input signals.

Tracker Receiver

(Figure 10) A wide beam (low gain) comparison antenna is provided to insure that the tracker locks only on its main lobe. Vehicle azimuth and elevation information are extracted in a 4 horn amplitude-amplitude comparison circuit with an accuracy of 1/2 mil. The swept channel and fixed frequency channel signals are fed to a data-extraction unit for obtaining vehicle range to an accuracy of 2 ft. With more extensive data-extraction circuit a range

accuracy of 0.4 ft. is possible. Voice from the vehicle is extracted from the swept channel while telemetry from the vehicle is obtained from the fixed channel. Provisions are made for either a 6 ft. or 12 ft. antenna, the choice depending primarily upon the amount of weather one wishes to work through and how close to the horizon one wishes to track. The latter depends upon the landing characteristics of particular vehicles.

Digital Data Processor

(Figure 11) The digital data processor is a multi-purpose solid-state digital computer designed for military real time control applications. The basic computer contains approximately 3000 transistors on 500 cards. Its basic word length is 36 bits with word lengths of 24 bits and 30 bits available. The core memory stores 4,096 words and can be increased up to 262,144 words. Up to 20 million words of disc file storage may be added. Any number of input/ output subchannels from 3 to 256 can be provided for a wide variety of input/output and storage devices. One index register is standard but up to 32 index registers may be utilized.

The 109 decoded instructions with 9 methods of address modification were designed with emphasis on real time control and data processing applications. Frequently encountered computations, such as search within limits, coordinate transformation, data packing and unpacking, and gray to binary conversion, can be performed faster with this M-236 computer than with computers having shorter memory access times because the programmer can

1. Multiply and divide in 37.9 microseconds.
2. Use the quotient and index registers as accumulators.
3. Treat any memory location as an accumulator to which the contents of the accumulator, quotient or index registers may be added, subtracted or logically combined.
4. Use the address portion of instruction words as operands.
5. Use one command to repeat instructions a specified number of times while incrementing the address.
6. Compare within limits.
7. Read, alter and rewrite a memory location in one memory cycle.
8. Perform address modification using the accumulator, quotient or index registers (this in addition to six other modifiers; the four indirect addressing, direct operand, and increment address).

These features provide the capability for accomplishing real time computations with specified time limits.

CONCLUSIONS

(Figure 12) In summary, a single, mobile, air transportable system capable of land or shipboard operation appears quite attractive as multipurpose equipment.

For manned flight it could furnish course information to the pilot in the event of an abort during boost. It could update his inertial system, select downrange landing sights within the vehicles range, or assume automatic control if necessary.

During reentry, flight paths may be at altitudes up to 60 miles. Maximum line of sight at this altitude results in a tracker range of about 500 miles. This range capability would minimize the number of trackers required to cover a reentry corridor 6000 miles long.

Under normal conditions it could update the inertial system periodically and could furnish accurate backup information. It could be used as a gap-filler between permanent land stations for monitoring and as a data link.

Its all weather capability would permit tracking to touchdown during the landing. Deploying parawings or separating them from the vehicle after touchdown would be accomplished by discrete commands.

The same system could be employed for recovery of unmanned payloads. It offers these advantages plus vehicle recovery on test flights. There is no substitute for laboratory examination and automatic control and landing would permit this.

The transportable-equipment can also be used for experiments in communication through the plasma sheath during re-entry. This could be done with available re-entry vehicles on a non-interference basis.

An attractive possibility arises when we consider placing the airborne equipment in the payload and use it for both boost guidance and for payload recovery.

Our work indicates that all this can be obtained with todays equipment, using a single set of airborne and ground equipment. Continued effort will enable us to reduce the size and weight of these components while maintaining the accuracy and reliability we have demonstrated in the past.

ACKNOWLEDGEMENT

The design of the space craft recovery system presented in this paper is the work of many engineers. Of particular note is the work of Karl Kipp on the tracker, Cal Woods on the communications system, John Couleur on the digital data processor, and Joe Cady on aerodynamics.

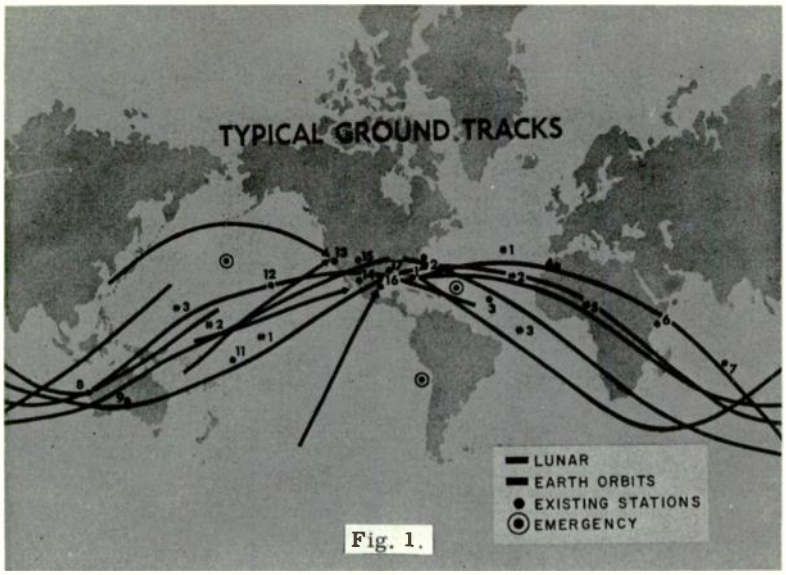


Fig. 1.

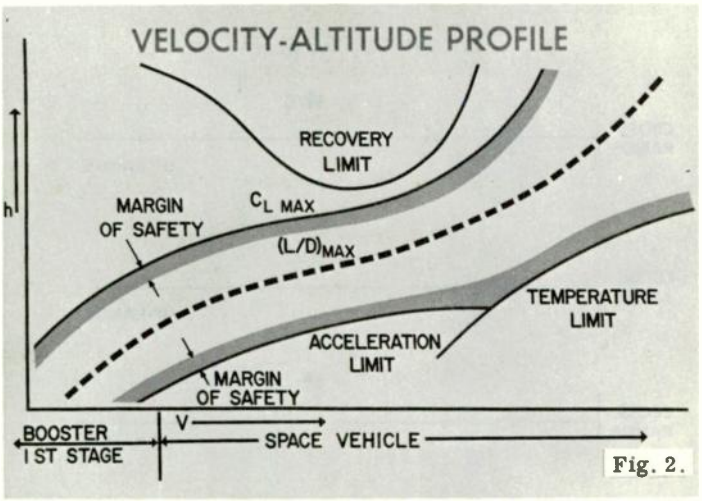


Fig. 2.

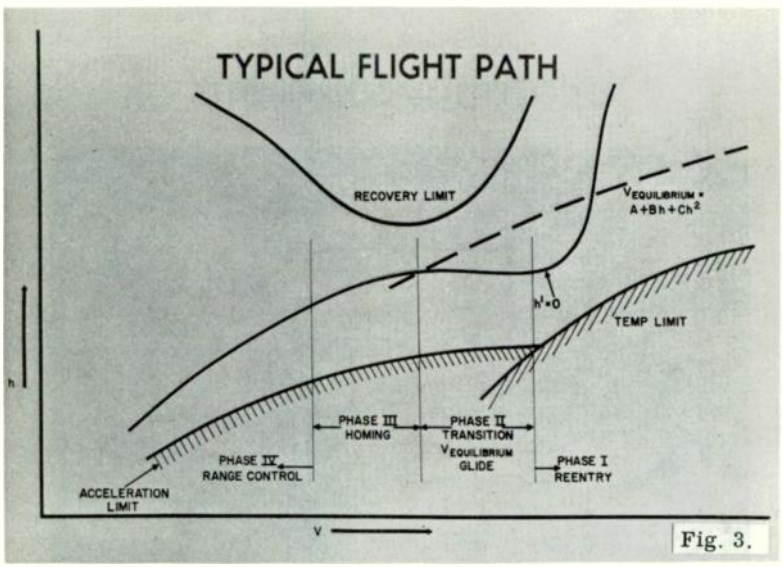


Fig. 3.

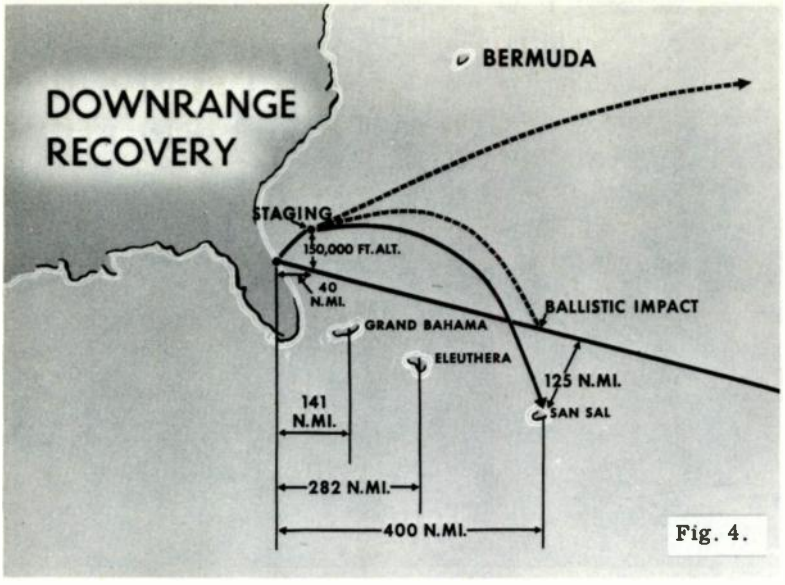


Fig. 4.

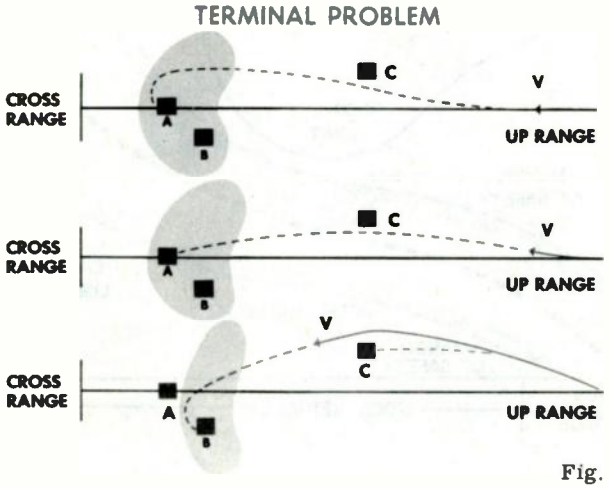


Fig. 5.

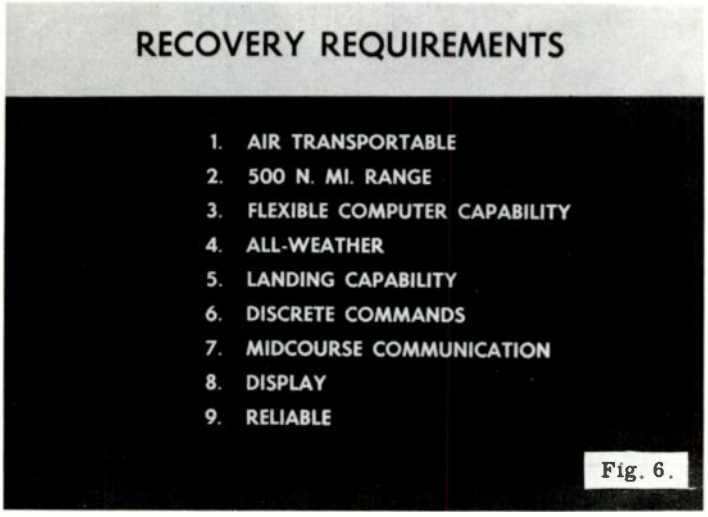


Fig. 6.

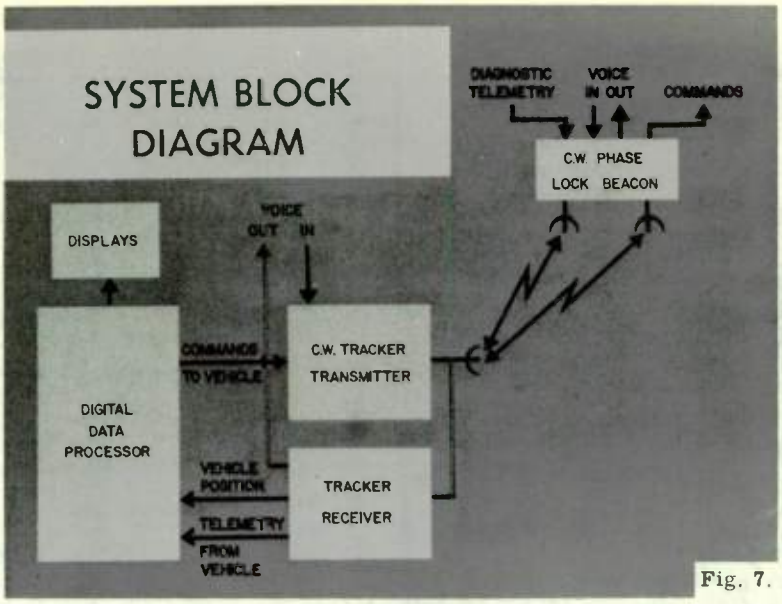

SYSTEM BLOCK DIAGRAM
DIAGNOSTIC TELEMETRY
VOICE IN OUT
COMMANDS
C.W. PHASE LOCK BEACON
DISPLAYS
VOICE OUT IN
DIGITAL DATA PROCESSOR
COMMANDS TO VEHICLE
C.W. TRACKER TRANSMITTER
VEHICLE POSITION
TELEMETRY FROM VEHICLE
TRACKER RECEIVER

Fig. 7.

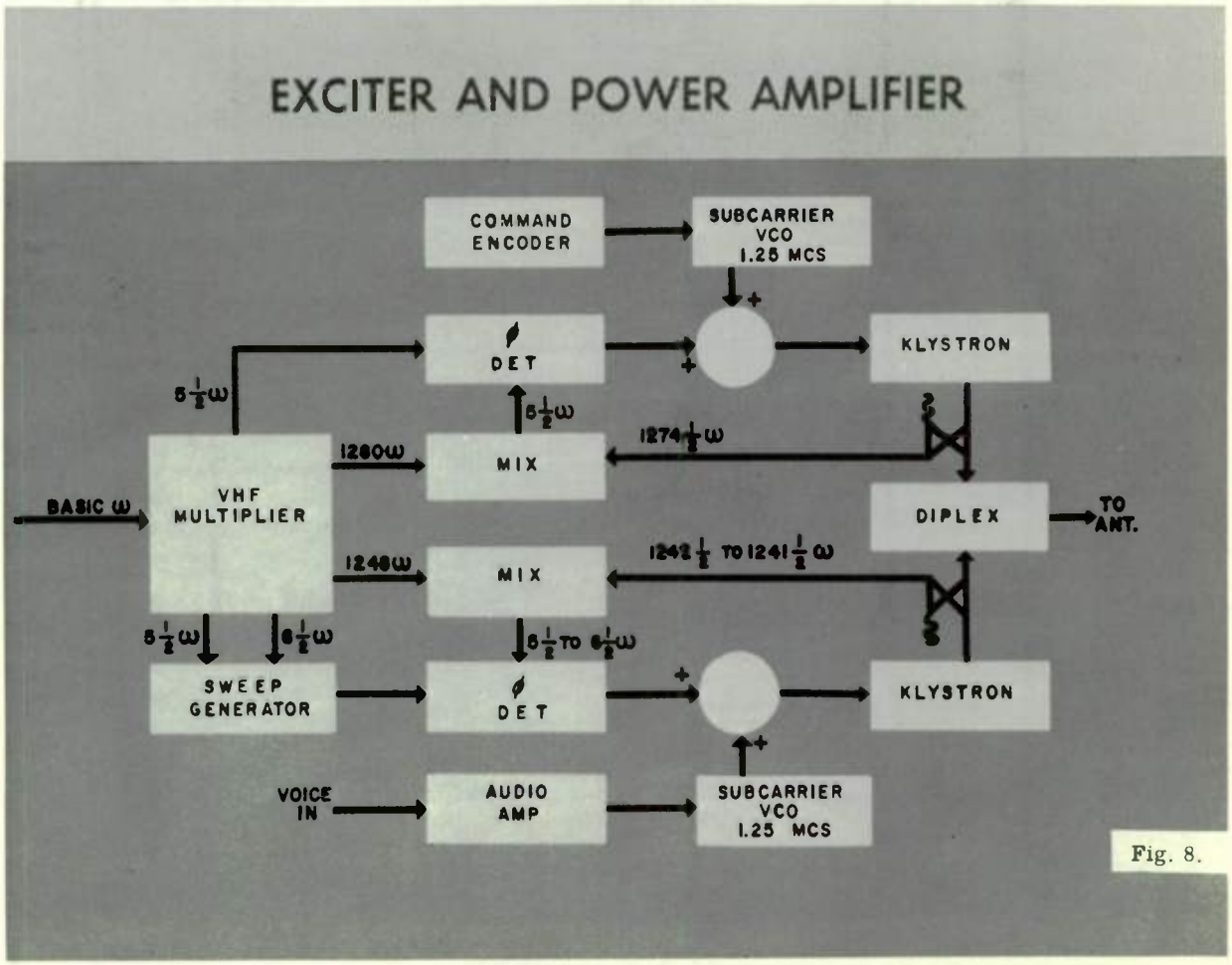

EXCITER AND POWER AMPLIFIER
COMMAND ENCODER
SUBCARRIER VCO 1.25 MCS
DET
KLYSTRON
5½ω
1280ω
MIX
1274½ω
BASIC ω
VHF MULTIPLIER
DIPLEX
TO ANT.
1248ω
MIX
1242½ TO 1241½ ω
5½ω
5½ω
5½ TO 6½ω
SWEEP GENERATOR
DET
KLYSTRON
VOICE IN
AUDIO AMP
SUBCARRIER VCO 1.25 MCS

Fig. 8.

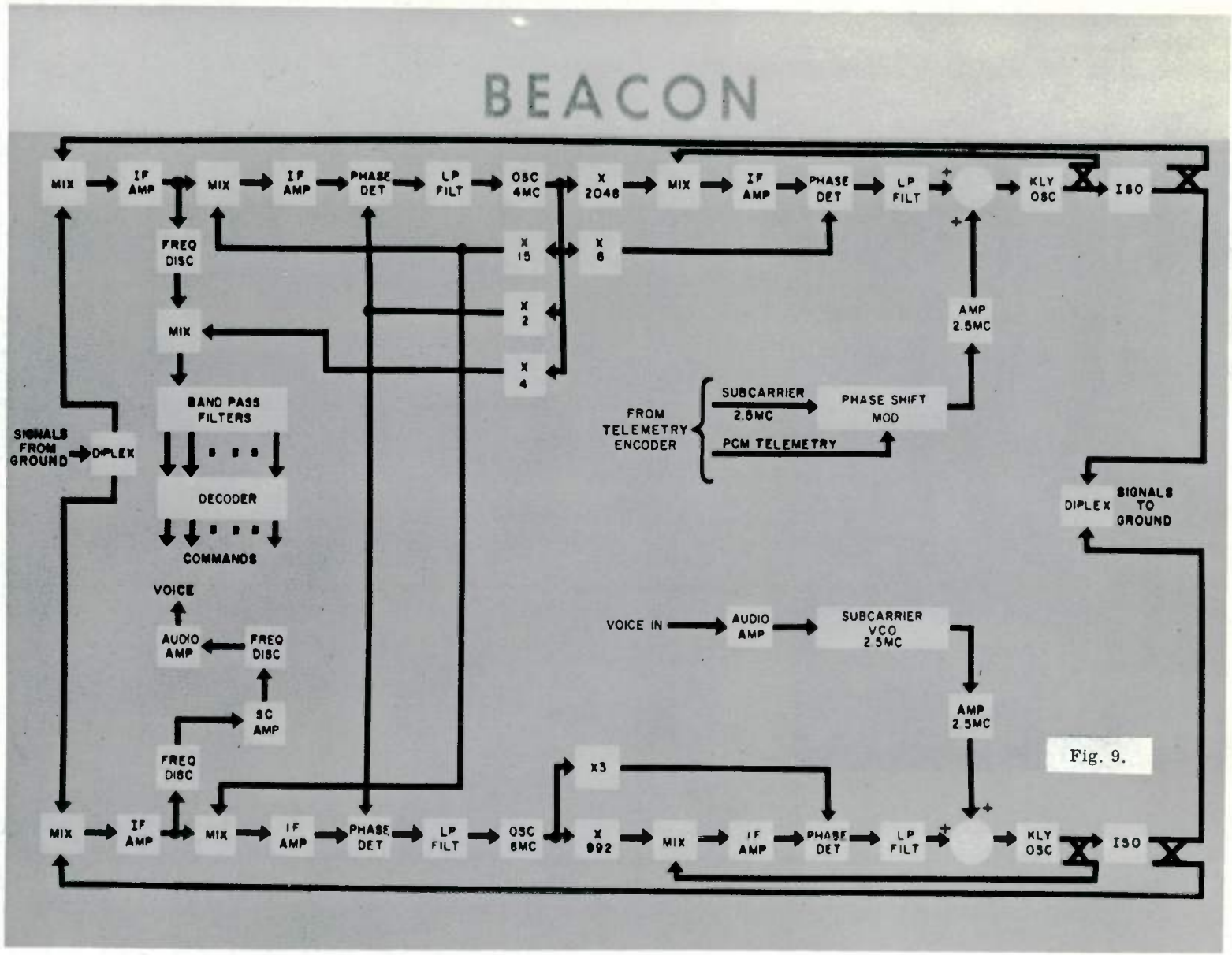

BEACON
MIX
IF AMP
MIX
IF AMP
PHASE DET
LP FILT
OSC 4MC
X 2048
MIX
IF AMP
PHASE DET
LP FILT
KLY OSC
ISO
FREQ DISC
X 15
X 6
X 2
MIX
X 4
AMP 2.8MC
BAND PASS FILTERS
SIGNALS FROM GROUND
DIPLEX
DECODER
COMMANDS
FROM TELEMETRY ENCODER
SUBCARRIER 2.8MC
PHASE SHIFT MOD
PCM TELEMETRY
DIPLEX
SIGNALS TO GROUND
VOICE
AUDIO AMP
FREQ DISC
SC AMP
VOICE IN
AUDIO AMP
SUBCARRIER VCO 2.5MC
AMP 2.5MC
FREQ DISC
X3
MIX
IF AMP
MIX
IF AMP
PHASE DET
LP FILT
OSC 8MC
X 992
MIX
IF AMP
PHASE DET
LP FILT
KLY OSC
ISO

Fig. 9.

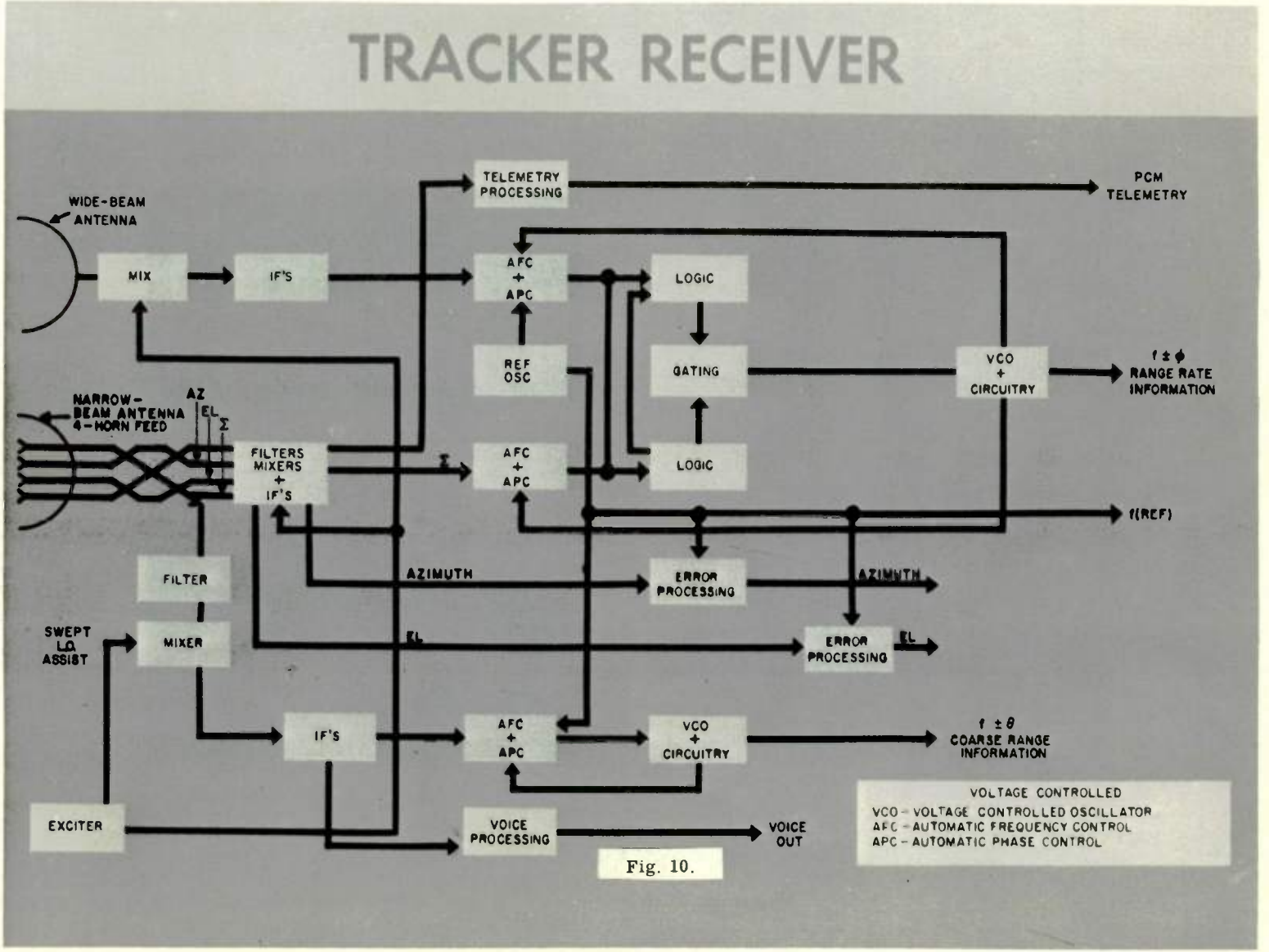

TRACKER RECEIVER
WIDE-BEAM ANTENNA
MIX
IF'S
TELEMETRY PROCESSING
PCM TELEMETRY
AFC + APC
LOGIC
REF OSC
GATING
VCO + CIRCUITRY
f ± ϕ RANGE RATE INFORMATION
NARROW-BEAM ANTENNA 4-HORN FEED
AZ
EL
Σ
FILTERS MIXERS + IF'S
Σ
AFC + APC
LOGIC
f(REF)
FILTER
AZIMUTH
ERROR PROCESSING
AZIMUTH
SWEPT LO ASSIST
MIXER
EL
ERROR PROCESSING
EL
IF'S
AFC + APC
VCO + CIRCUITRY
f ± θ COARSE RANGE INFORMATION
EXCITER
VOICE PROCESSING
VOICE OUT
VOLTAGE CONTROLLED
VCO - VOLTAGE CONTROLLED OSCILLATOR
AFC - AUTOMATIC FREQUENCY CONTROL
APC - AUTOMATIC PHASE CONTROL

Fig. 10.

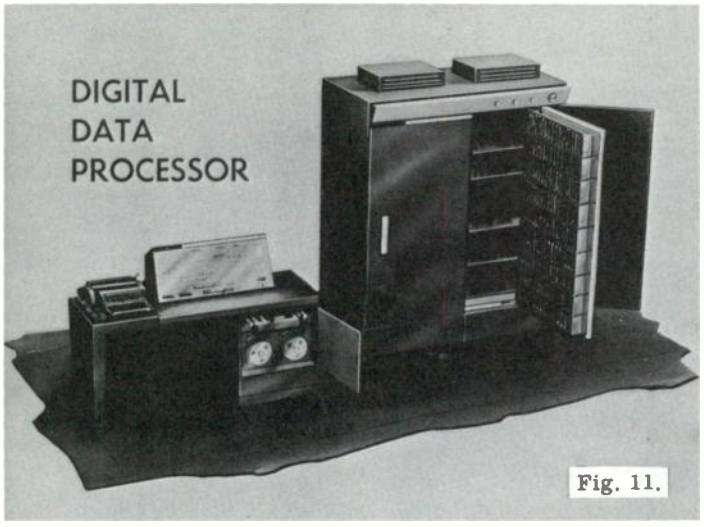


Fig. 11.

PAYLOAD APPLICATION
SUMMARY-SYSTEM CAPABILITY
ONE TYPE OF AIRBORNE EQUIPMENT

MANNED BOOST	EMERGENCY ABORT ALTERNATE FIELDS AUTOMATIC BACKUP UPDATE INERTIAL SYSTEM
MANNED RECOVERY REENTRY & LANDING	MOBILE SYSTEM MONITORING FLIGHT PATH DATA LINK AUTOMATIC BACKUP
UNMANNED	THE ABOVE PLUS RECOVERY ON TEST FLIGHTS RECOVERY FOR LABORATORY INSPECTION BEACON PAYLOAD

Fig. 12.

GIMBALLESS INERTIAL REFERENCE SYSTEM

B. Byrne, W. Murphy and R. W. Lanzkron

Martin Marietta Corporation
Baltimore 3, Maryland

SUMMARY

This paper outlines the requirements for a vehicle which uses as its inertial reference a gimballess electronic inertial package. The system consists of three strap-down (body mounted) rate gyros, three accelerometers and a digital differential analyzer. The output of this package is equivalent to the outputs of a conventional gimbaled inertial platform.

The design considerations and requirements of the gyro loop, the analog-to-digital conversion and the digital differential analyzer are discussed at length. Finally, an example is worked out in detail showing the system components combined into an inertial reference package suitable for a re-entry mission.

Much of the current emphasis on strap-down systems is directed toward making them as good as high grade inertial gimbaled platforms. Components developed for platforms are often used in these designs.

At present, new classes of problems are arising in which errors due to initial condition uncertainties greatly outnumber the errors of high grade platforms, and in which times of operation and operating environments are significantly different from those obtaining in booster rocket or manned aircraft applications.

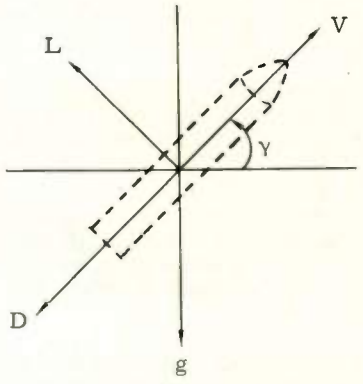
For example, requirements exist to furnish short-term reference for space vehicles to allow time for position fixing, decision and vehicle control actuation. In this environment, control rates may be reduced and structural dynamic problems held to a minimum. Digital solutions of gimbal angles are more accurate if one rate is applied at a time. There is time enough for such sequential attitude maneuvering in space.

Another requirement is for the return of drag-controlled orbital vehicles through the atmosphere. Here, the initial attitude errors resulting from sensing of the vertical preclude high accuracy in maintaining position; hence, requirements for airborne platforms may be relaxed without detriment to performance. It is instructive to examine the vehicle dynamics for this case.

In drag-controlled vehicles variations in the cg are minimal. Hence, little correction is required to body mounted accelerometers. In addition, g loadings are small. When we examine the gyro picture, we find that relaxation of long-term accuracy requirements permits us to use gyros at lower wheelspeeds and higher reliabilities.

Guidance Loop Requirements

For a near-orbital vehicle, if longitudinal vehicle equations are written in terms of velocity and flight path angle, holding angle of attack constant, it is found that the incremental response is strongly influenced by the rate of change of atmospheric density with altitude--and by the vehicle lift.



$$-\dot{V} = K_1 e^{K_\rho (R-R_o)} V^2 \delta + g_o \frac{R_o^2}{R^2} \sin\gamma$$

$$V\dot{\gamma} = K_2 e^{K_\rho (R-R_o)} V^2 - g_o \frac{R_o^2}{R^2} \cos\gamma$$

$$\ddot{R} = \dot{V} \sin\gamma + V\dot{\gamma} \cos\gamma$$

Where
- V = longitudinal velocity
- L = lift
- D = drag
- K_ρ = constant, variation in atmospheric density
- K_2, K_1 = nominal density factors
- δ = drag brake deflection
- g_o = nominal gravity
- R = geocentric radius

Using
- W_S = Schuler angular frequency
- W_o = vehicle angular frequency

After some approximation, this results in the following array of equations.

V(S)	γ(S)	R(S)	δ(S)	
$s+\frac{2D_o}{V_o}$	g_o	$-2W_S^2\gamma_o + K_\rho D_o$		$= \frac{-D_o}{\delta_o}$
$-(W_o+2\frac{L_o}{V_o})$	$V_o(S-\frac{g_o\gamma_o}{V_o})$	$-2W_S^2 - K_\rho L_o$		$= 0$
$-(W_o+\gamma_o S)$	$V_o(S+\frac{D_o}{V_o})$	$-S^2$		$= 0$

For vehicles developing lift accelerations in the range 1 to 4 g, this frequency lies in the range 0.03 to 0.06 rad/sec. This determines the longitudinal guidance loop frequency.

In view of these requirements, control loop frequencies of 0.3 to 0.6 rad/sec will adequately separate the two loops.

Let us look now at the details of the design of this system. It consists of the following components, shown in Fig. 1: three-body mounted gyros, three-body mounted accelerometers, analog-to-digital conversion system, digital differential analyzer and general purpose computer.

The gyro loop consists of the elements shown in Fig. 2. This is a rate-integrating gyro whose torquing current is measured to ascertain the change in respect to missile attitude.

Additional integrators can be used in the forward loop to reduce the hang-off angle of the gyro and so increase the accuracy of the system, Fig. 3.

The basic requirements for the gyro are: (1) Very high torquing rates, so as to prevent large transistents in the hang-off angle; and (2) Frequency response fast enough to follow the guidance and control loop.

We note here that rate stabilization is obtained with the aid of a separate set of rate gyros--as is common practice in inertial systems.

If the system in Fig. 2 is used, we can write the equation defining the system as

$$\dot{\theta}\ \text{out} = \frac{KH\ \dot{\theta}}{JS^2 + BS + K_t K}$$

or for sinusoidal input $\dot{\theta} = AW \sin wt$

$$\dot{\theta}\ \text{out} = \frac{\frac{H\,A}{K_t}}{\left(1 + \frac{S^2}{W^2}\right)\left(\frac{JS^2}{K_t K} + \frac{BS}{K_t K} + 1\right)}$$

The equation for Fig. 3 is

$$\dot{\theta}\ \text{out} = \frac{\frac{HA}{K_t}\left(\frac{KS}{K_2} + 1\right)}{\frac{JS^3}{K_2 K_t} + \frac{BS^2}{K_t K_2} + \frac{KS}{K_t K_2} + 1}$$

For the space vehicle we have discussed, since the response in the control loop will be between 0.3 and 0.6 rad/sec, let us assume 0.4 radians.

To obtain a good response of the gyro we could pick a system with a double break at 40 radians and critically damped. If we do this the first system reduces to

$$\dot{\theta}\ \text{out} = \frac{\frac{HA}{K_t}}{\left(1 + \frac{S^2}{W^2}\right)(0.025\,S + 1)^2}$$

and the second, to

$$\dot{\theta}\ \text{out} = \frac{\frac{HA}{K_t}\left(\frac{K}{K_2}\,S + 1\right)}{\left(1 + \frac{S^2}{W^2}\right)(1 + TS)(1 + 0.025\,S)^2}$$

Computing the hang-off angle for the first system, one obtains

$$\theta_{H.O.} = \frac{HAW}{KK_t}$$

while the System 2 hang-off angle is zero. Since the linearity of the torquer is a function of the hang-off angle, System 2 will be more accurate.

The total error can be defined as the difference between the integral of the rate input and the actual output of the system. For System 1,

$$E = \frac{A\,W^2\,T^2\,(W^2 + 3)}{T^2\,W^2 + 1}$$

For the frequencies discussed, this reduces to

$$E = \frac{3\,W^2 T^2\,A}{1 + T^2\,W^2}$$

If we assume

$$W = 0,4\ T = \frac{1}{40}$$

then

$$E = \frac{3\,(0.4)^2\,(0.025)^2 A}{1 + (0.4)^2\,(0.025)^2} = 2 \cdot 10^{-4}\ A$$

For a vehicle which oscillates with a limit cycle of 33° magnitude at the frequency of 0.4 radian, this is a steady state error of a hundredth of a degree.

With System 2, this error could be reduced even more.

Next, is the selection of the gyro that fits the characteristic of high torque and can be used in the above designed loop. If we pick a Kearfott Gyro 2514003, we find it has torquing rates up to 28 deg/sec--which seems adequate.

To obtain a frequency double break at 40 radians, System 1 requires.

$$\frac{J}{K_t K} = \frac{1}{W^2} = \frac{1}{40^2}$$

since $J = 30 \text{ gm cm}^2$

$$K_t K = 3.2 \cdot 10^5$$

which is a feasible design.

The damping B is obtained by recognizing

$$\frac{B}{K_t K} = \frac{2}{W} \quad \text{and} \quad B = 1.6 \cdot 10^4$$

A reasonable number too.

If we again assume an oscillation of magnitude of 33° at a frequency of 0.4 radian, the hang-off angle will be

$$\theta_{H.O.} = \frac{HAW}{K_t K} = \frac{10^5 \ 33 \text{ O.H.}}{3.2 \cdot 10^5} = 4.3° \ (H = 10^5)$$

The hang-off angle allowed is 10°.

The hang-off angle in System 1 will not be constant, so the estimate is pessimistic. The effect due to the nonlinearity of the torquer can be estimated for the particular gyro as

$$E = (A_o + A_1 \sin wt) K_t \sin \frac{C_1}{10^3} (A_1 + A \sin wt) + K_2 \cos \frac{C_1}{10^3} (A_o + A_1 \sin wt) - K_2$$

Here, we have assumed the output error curve to be a cosine and sine curve combination. A_o is the constant rate. $A_1 \sin wt$ is the oscillation on top of the constant input. Integrating this error over periods of time and setting the non-secular terms to zero, we have

$$\int \text{Error} = K_1 C_1 A_1^2 \left(\frac{1}{2 \cdot 10^5} - \frac{C_1^2 A_o^2}{4 \cdot 10^{11}} - \frac{C_1^2 A_1^2}{2.4 \cdot 10^{10}} \right) t$$

$$-t \ K_2 C_1^2 A_1^2 A_o \left(\frac{1}{1.33 \cdot 10^8} - \frac{C_1^2 A_1^2}{24 \cdot 10^{14}} \right) \quad t$$

If we assume that $A_o = 0$, $K_2 = 0$, $C_1 = 5$ millirad/deg/sec, $A_1 = 2$ deg/sec $K_1 = 2.5$ (0.025% error for 10 millirad). Error = $(2.5 \cdot 10^{-4})$ deg/sec.

This is about 0.75° per hour--the largest error in the system.

The gyro errors due to errors coupling deadspot, bearing side load, etc., were investigated and found to be of the order of 1/4° to 1/2° per hour if the frequency in the x and y axes of the package were kept apart. Should these frequencies coincide, the error will obviously be large. Therefore, in the design of the system, the pitch, yaw and roll frequencies will have to be kept apart.

The accelerometer selected for the system must match requirements of the thrust device used for mid-course correction. For lunar trajectories, chemical rocket fuels seem most probable and require accelerometers of the order of $1 \cdot 10^{-5}$ g accuracy for cutoff. Control of small thrusts, such as nuclear engines or other vernier types, may require $1 \cdot 10^{-8}$ g to utilize their small thrust most efficiently.

These, however, seem far in the future--and for the moment, a pendulous force rebalance type of instrument seems adequate. Analog force rebalance currents will be quantized in the same fashion as the gyro torquer currents--hence, should be held to about 0.01% with overall linearity of 10^{-4} g. This will provide velocity accuracy of 0.1 foot per second over thrust application periods of 10^3 seconds.

Location of the package must be very close to the vehicle center of gravity. Otherwise, corrections are required for accelerations due to angular rates about the center of gravity. If it is necessary that the displacement be significant, rates for correction are readily available from the gyros and the computation can be accomplished in the computer. An instrument with characteristics satisfying the criteria would be the Kearfott F2401 Mod II Accelerometer.

Direct and Indirect Integration

The output of the body fixed gyros and acceleration sensors consists of a time-varying signed voltage whose time integral is quantized and pulse-rate coded. A signal accuracy of 0.01 to 0.5% is required to match the gyro torquer linearity. Conversion accuracy should not exceed 0.1%. Input requirements of the coordinate conversion computer (CCC) lend themselves to dual integration; that is, the integration of some nominal value for zero and the use of the difference between this integral and the sensor integral as input. The use of dual integration is anticipated, because it avoids sign ambiguity and provides temperature compensation.

The analog signal may be integrated directly, with pulse and reset occurring when

$$K \int V_a dt = V_R$$

where K is the gain of the integrator in appropriate units, V_R the reference voltage and V_a the analog input signal. The gain K and the reference voltage V_R together determine the quantization level. Alternatively, integration may be indirect, with pulse and reset occurring when

$$K \int V_R dt = V$$

where $V_R(t)$ is usually a ramp function and where frequency division may be used as an intermediate step. If the integration is direct, the conversion is said to be in the voltage domain; if indirect, in the time domain.

The time constants are chosen to prevent large variations in V_a during the integration time. The time constants can be chosen without altering scale or quantization level, by sampling several times and averaging for each output. Space domain and other conversions shown in Ref. 1 have disqualifying features for this application.

In practice, both direct and indirect methods utilize the threshold change of a loop gain to provide regenerative pulsing and resetting. A voltage domain converter and a time domain converter are shown in Figs. 4 and 5, respectively. In some mechanizations the elements are easily identifiable; in others, notably the phantastron, they represent different aspects of the same circuit device.

Let δ_x represent the open loop error of the x element in the system. Let ϵ_{xy} be the error contribution of the element due to its interaction with the y element. Let Δx represent the total error of the x element. Then, for the error of either system, we have

$$S = \Delta I + \Delta R + \Delta C$$

$$= \delta_I + \epsilon_{IC} + \epsilon_{IR}\delta_R + \epsilon_{RI} + \epsilon_{RC} + \epsilon_{CI} + \epsilon_{CR} + \delta_C$$

where

ΔI = error contribution of integrator

ΔR = error contribution of resetter

ΔC = error contribution of comparator.

The basic features of each element and some of the techniques available for reducing error, amplification, drift compensation and time-shared calibration are discussed in the following.

Basic Features of Standard Elements

Integrators

For the simple RC integrator shown in Fig. 6

$$V_o = V_a \left(\frac{t}{K} - \frac{t^2}{2K^2} + \frac{t^3}{3K^3} \cdots \right)$$

where K = RC. The departure from linearity may be made arbitrarily small by adjustment of the ratio V_i/V_a. However, fixed error in the comparator will require either large V_a or amplification of V_o with attendant drift. Alternate approaches to the problem are shown in Fig. 7 (bootstrap integrator), and Fig. 8 (Miller integrator). Examining Fig. 7,

$$V_o = V_a \frac{A}{1 - A} \left(1 - \epsilon^{t\left(\frac{1 - A}{K}\right)} \right)$$

$$= V_a \frac{tA}{K} \left(1 - \frac{t(1 - A)}{2K} \cdots \right).$$

The quadratic error arising from a nonunity gain in the bootstrap integrator is

$$\delta_I = \frac{1 - A}{2K} .$$

From Fig. 8,

$$V_o = V_a \frac{A}{1 - A} \frac{t}{K} \left(1 - \frac{1}{2K(1 - A)} \cdots \right)$$

$$\delta_I = \frac{t}{2KA} .$$

The Miller integrator is used most often because of the relatively greater ease of achieving large gains, as compared with the holding of a near-unity gain. Other approaches, (Ref 2), including the use of a series inductance or of a transistor as a constant current device, promise little improvement.

Voltage comparators

Existing solid state comparators employ one of the following basic circuits:

A diode with a nonregenerative (saturating) amplifier

An amplifier operating at cutoff or saturation mode

A regenerative amplifier with a diode in the feedback loop

A regenerative amplifier with loop-gain switching

Solid state switches, including zener diode silicon controlled rectifiers.

The most promising is the regenerative amplifier with or without a diode in the feedback loop. These circuits, possessing enough accuracy and the required resetting features, merit consideration at the switching point. They require large d-c amplification of the signal for any accuracy (see Ref 3).

Tunnel diode. This diode is an exception. It is considered here because it promises accuracy by the quantum nature of its tunnel effect and the simplicity offered by its two-part nature. If the circuit parameters of Fig. 9 are chosen so that there are two stable states, switching will occur at some $V_o = KV_a$. Stability and load requirements can be improved by inserting a second diode--as in the circuit of Fig. 10. An ordinary diode may be used in series with the input resistor of Fig. 10 to reduce the voltage range of the comparator. Variation in the forward resistance of the diode can be made negligibly small.

Schmidt circuit. The circuit shown in Fig. 11 can be used as a voltage comparator. Regenerative feedback occurs at some voltage, V_r. For a detailed analysis, see Ref 4. Accuracy of the Schmidt circuit can be improved considerably by incorporating a diode in the feedback loops as shown in Fig. 12. Modifications of the circuit serve as a basis for nearly all existing solid state comparators. Great accuracy can be achieved by adding a chopping amplifier and phase detection circuit as shown in Fig. 13. Solid state choppers such as the 2N200I-7 are currently available. Accuracy is basically determined by the gating error, since drift in the amplifier has a minor effect.

Resetters

Nearly all existing converters achieve resetting by a switched amplifying feedback loop. From Fig. 14, we see that reset can be made independent of the loop gain by inserting clamping diodes in the circuit. An emitter follower may be used to reduce flyback time and permit d-c coupling. Final flyback time error can be partially compensated by the insertion of a resistor between the integrator and the comparator.

Error Reducing Techniques

Amplifiers are used in all accurate integrators to improve linearity. They also permit resetting without loading the comparator. Insertion of an amplifier in a circuit requiring d-c coupling calls for complex stabilization nets, if the introduced drift is not to exceed the switching ambiguity which the amplifier overcomes. Some comparators permit a-c coupling (see Ref. 5), but they have low, long-term stability due to impedance changes. The introduction of a chopper amplifier-phase comparator into the system, after the comparator, can reduce errors arising from comparator loading. However, this cannot affect the basic accuracy.

Drift compensation in the form of temperature compensation must be included in all units operating in a noncontrolled environment. The technique of dropping both signal and reference voltage through similar diodes (using a balanced amplifier, etc.) is well known. Packaging is simplified by including the amplifier and a temperature-compensating element in the same case--which also ensure similar conditions.

The use of several converters permits the cycling of duty between units and the calibration of converters during the in-flight phase. With this arrangement, accuracy essentially becomes equal to the differential error in the gating circuit, since the measurement error becomes equivalent to that of the more stable calibration standard. This improvement in accuracy is achieved at a cost in reliability and simplicity which does not seem warranted in this application.

Application

Analysis. Existing electron tube and solid state converters approach 0.01% accuracy (Ref. 6) in the comparison of an analog voltage with a gated reference voltage. Our objective is to compare an analog voltage with a nongated reference voltage to a design accuracy of approximately 0.1%. Such a goal involves the design of a voltage-controlled oscillator whose control voltage is derived from the gyro torquing loop. The problem may be stated as follows:

Quantize the current to the gyro torquer at a sufficiently fast rate, so that error in the navigational problem will be of the order of 0.25 degree.

The output of the torquer amplifier is 0 to 135 milliamperes across a torquer coil of 230 ohms resistance. The torquer is of the permanent magnet type.

The coordinate conversion computer will sample the output of the converter, synchronously, at rates up to 10,000 samples per second. However, the conversion is not required to be synchronous. It must indicate the presence of a unit incremental angle as soon as possible after the angle has been generated--and it must be capable of reset soon enough so that no appreciable increment of current is lost.

We assume the navigational problem to be that the system sweeps out approximately 60 degrees (earth central angle) in 1200 seconds about the pitch axis with practically zero sweep about the other two axes.

The pertinent input rates to the gyro torquer and their associated currents are tabulated below. We have assumed $2.78 \cdot 10^2$ ma/rad/sec as the loop gain.

Source	Rate (rad/sec)	Current (ma)
Maximum gyro response	0.485	135
Gyro drift rate	$4.85 \cdot 10^{-6}$	0.00135
Navigational input rate	$8.73 \cdot 10^{-4}$	0.243
Typical control rate	0.1	27.8

To establish basic parameters for the conversion, we assign a 4-rad/sec equivalent to the converter, which is a decade above the control frequency. Then we calculate a first approximation to the sampling rate of the coordinate conversion computer by first estimating an incremental angle as a function of the navigational problem.

Assume an incremental angle of 0.025°.

Then

$$S \frac{\text{samples}}{\text{sec}} = \frac{\frac{4 \text{ rad}}{\text{sec}} \cdot \frac{57.3 \text{ deg}}{\text{rad}}}{0.025 \text{ deg/sample}} =$$

$$9200 \frac{\text{samples}}{\text{sec}} \approx 10,000 \frac{\text{samples}}{\text{sec}} .$$

To establish the accuracy of the conversion, we investigate the effect on the navigational problem. There are $\frac{60}{0.025} = 2400$ samples. We assign a 0.5% conversion accuracy to the integrated total and--assuming, conservatively, a unidirectional error, obtain the following as the error.

$$\epsilon_{cu} = 0.005 \cdot 60 = 0.3 \text{ deg}$$

Investigating the effect of average control rates of 0.1 rad/sec by random walk, since this is an oscillatory error.

$$\epsilon_{1\sigma} = \sqrt{n1^2}$$

$$S = \frac{\frac{0.1 \text{ rad}}{\text{sec}} \cdot 57.3 \frac{\text{deg}}{\text{rad}} \cdot 1200 \text{ sec}}{0.025 \text{ deg/sample}} =$$

$$\frac{2.75 \cdot 10^5 \text{ samples}}{0.025 \text{ deg/sample}}$$

We assume each sample to be plus or minus, with random distribution, and calculate the standard deviation due to oscillation as

$$\sigma = \sqrt{2.75 \cdot 10^5 \cdot 0.03^2} = 15.7 \text{ deg.}$$

Sigma represents the probable departure from the desired orientation due to $2.75 \cdot 10^5$ steps of 0.1 rad/sec, taken in a random oscillatory fashion. Evaluating a 3 σ deviation at 0.5%, for the resulting error we have

$$\epsilon_{3\sigma} = 0.005 \cdot 3 \cdot 15.1 = 0.236 \text{ deg.}$$

Then a quantization error of 0.5% could contribute an RMS total error of 0.37 degree. Hence, as much as 0.5% may be used to satisfy the requirements--since the original knowledge of the angle is not better than 0.3°. The design goal, however, is 0.1%.

With regard to actual pick-off of the current signal, we must remember that the torquer coil will have inductance and that, even at very low frequencies, it is preferable to place a precision resistor in series with the coil. We specify R_1 = 20 ± 0.003% ± 3 ppm/deg Centigrade. We further specify that this resistor be maintained at an optimum temperature ±4° C. The voltage swing at maximum current corresponding to 0.485 rad/sec is then ±2.7 volts. The gyro drift rate equivalent is $2.7 \cdot 10^{-5}$ volts. Now we may practically compare voltages of 0.1 millivolt as the smallest bit of 0.01%--and we assign a gain which will bring to 0.1 millivolt a value of 0.1% of the smallest rate signal we find necessary to detect in a navigational interval. The navigational rate gives rise to 0.243 milliampere or 4.86 millivolts across the pick-off resistor. We specify that comparison should take place at a level near one volt and that it require no more than 0.1 second to detect the existence of the navigational rate in the integrated signal:

$$K_I (4.86 \cdot 10^{-3} \text{ volts} \cdot 0.1 \text{ second}) = 1 \text{ volt}$$

$$K_I = 2.06 \cdot 10^{+3} \frac{\text{volts}}{\text{volt second}} .$$

At the maximum rate,

$$2.06 \cdot 10^{-3} \frac{\text{volts}}{\text{volt second}} \; 2.7 \text{ volts} \cdot t = 1 \text{ volt}$$

$$t = 1.8 \cdot 10^{-4} \text{ seconds.}$$

This value is set as the nominal time constant of our integrator. The actual integrator time constant must assure buildup to within 0.1 millivolt of 1 volt in this time. A voltage gain of 3.7, together with an integrator of $t = 1.8 \cdot 10^{-4}$ seconds, will provide the required K_I. Flyback time on the comparator should be in the nanosecond region.

Basically, we must then design a monostable oscillator which has an output frequency proportional to the input voltage. The range required is 0 to 10,000 pulses per second. This corresponds to a voltage range of 0 to 2.7 volts across the input resistor. We are free to exchange voltage gain for integration time to achieve the best stability and the lowest threshold level within the limits imposed by computer sampling rate capability and practical circuitry.

Circuit design. The circuit of Fig. 15 consists of a temperature-compensated Miller integrator with a gain of approximately 100 and with a bias up to the middle of the ±30-volt range. The Miller circuit is followed by a Schmidt trigger circuit, which is used as a comparator. The comparator provides the pulse which sets the output flip-flops and resets the Miller integrator with minimum flyback time. All the components used are solid state, and the design is conservative with regard to state of the art. With a gain of 100 in the Miller circuit and a 1-millivolt capability of the Schmidt circuit, a range of $1/10^6$ is achievable without noise.

The circuit of Fig. 16 utilizes the recently developed tunnel diode. A sketch of the current-voltage characteristic of a typical unit is given in Fig. 17. The advantage of this circuit is that it may be utilized to provide integration, comparison and automatic reset with a high degree of stability at operating frequencies up to the kilomegacycle range. In the present application, we can achieve nanosecond flyback times without difficulty.

The diodes are operated in what is known as their monostable state. Thus, when a small positive increment of current appears at the node (q of Fig. 16), with the diodes biased at point A of their characteristic, TD_2 will trigger from its positive into its negative conductance state and the result is a change in the voltage drop across R_{L2}. This signal is used to set an output flip-flop.

Diode TD_1 is not affected, since the operating point is driven further down the positive conductance portion of the curve. A negative current pulse at node q will increase the potential across TD_1 and drive this diode into the negative conductance region. With the diode in this region, a change of potential results across R_{L1}. Each time the diodes switch conductance states, the potential across them returns to the stable operating point and provides automatic reset. Other configurations are possible and, in fact, the diodes may be operated in their bistable state.

The difficulties encountered with the tunnel diode circuit are associated with the stability of peak current with respect to temperature and time and the stability of bias voltage. Temperature compensation or control will enable us to stabilize peak current to about 20 milliamperes, while stability of peak current with time, for germanium diodes, is about 10 milliamperes. Stability of the operating point at the required accuracy will be achieved with bias voltages of the order of ±1 millivolt, since we may set the trigger on a differential supply. The variable load resistors enable the circuit to be balanced to achieve equal quantization increments for both positive and negative swings of the input current. The value of R_3 is selected so that a very small load is placed on the torquer amplifier.

Experiments have been conducted with tunnel diode circuitry and, as a matter of face, a tunnel diode shift register operating at 100 megacycles has been successfully demonstrated. Essentially, the same circuit as the tunnel diode oscillator in Fig. 16 has been described in Ref. 7. The frequency and amplitude stability of tunnel diode oscillators, as a function of temperature, is given in Ref. 8. Circuits similar to Fig. 16 are held to 0.1% frequency tolerance over a temperature range of -49 to ±100° C. It is expected that dynamic ranges of $1/10^5$ may be attained through relatively simple design.

Coordinate Converter Computer

The high integration rates essential to the accurate handling of vehicle accelerometer and gyro outputs are most efficiently achieved by an incremental computer, namely, a digital differential analyzer (DDA). Conversion from vehicle to inertial coordinates is achieved by using the quantized gyro outputs to update a direction cosine matrix--and the cosine matrix to update the inertial velocity or acceleration. Errors arising from mass point treatment are to be compensated for by an analog device external to the computer. Position keeping is accomplished by a general purpose computer.

The problem involved in the conversion to inertial coordinates is the solution of nine differential equations for nine direction cosines. The derivation of the equations and studies of methods for their solution are given in the following paragraphs. Preliminary results of the studies indicate the acceptability of a twenty-bit register and show little growth of the matrix for trapezoidal integration. Though results of the simpler methods for solution are not conclusive, they appear to be satisfactory for this application.

Derivation of Equations

It is necessary to transform components of acceleration or velocity from vehicle axes into a space-fixed frame. We designate our space-fixed frame as X_s, Y_s, Z_s, with the origin at target at the predicted time of impact. Axis X_s is in the direction of motion and tangent to the great circle plane containing launch and impact points. Axis Z_s is parallel to the geocentric vertical

and Y_s is defined as our cross-course vector.

The transformation from the missile axes, defined as roll (x_m), pitch (y_m), yaw (z_m), is

$$\begin{vmatrix} X_s \\ Y_s \\ Z_s \end{vmatrix} = \begin{vmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{vmatrix} \begin{vmatrix} x_m \\ y_m \\ z_m \end{vmatrix}$$

Then the derivative of a vector in the body axes frame is

$$\left[\frac{d\bar{V}_s}{dt}\right]_{body} = \left|\dot{M}\right| \bar{V}_M$$

where $\dot{M}$ refers to the derivative of the direction cosine matrix expressed in space axes,

$$\left[\frac{d\bar{V}_s}{dt}\right]_{space} = \left[\frac{d\bar{V}_s}{dt}\right]_{body} + \bar{w} \cdot \bar{V}_s$$

where the $w \cdot \bar{V}_s$ represents Coriolis accelerations and $\bar{w}$ is the rotation vector of the missile about the missile axes,

$$\bar{w} = p\,\bar{x}_{mu} + q\,\bar{y}_{mu} + r\,\bar{z}_{mu}.$$

Now we hold our space axes invariant and

$$\left[\frac{d\bar{V}_s}{dt}\right] = 0$$

$$\left[\frac{d\bar{V}_s}{dt}\right]_{body} = -\bar{w} \cdot \bar{V}_s = \left|\dot{M}\right| \bar{V}_m.$$

Equating component by component,

$$\begin{vmatrix} \dot{\ell}_1 & \dot{\ell}_2 & \dot{\ell}_3 \\ \dot{m}_1 & \dot{m}_2 & \dot{m}_3 \\ \dot{n}_1 & \dot{n}_2 & \dot{n}_3 \end{vmatrix} = \begin{vmatrix} \bar{x}_{mu} & \bar{y}_{mu} & \bar{z}_{mu} \\ p & q & r \\ V_{sx} & V_{sy} & V_{sz} \end{vmatrix}$$

and

$$\dot{\ell}_1 = r\ell_2 - q\ell_3$$

$$\dot{\ell}_2 = p\ell_3 - r\ell_1$$

$$\dot{\ell}_3 = q\ell_1 - p\ell_2$$

$$\dot{m}_1 = rm_2 - qm_3$$

$$\dot{m}_2 = pm_3 - rm_1$$

$$\dot{m}_3 = qm_1 - pm_2$$

$$\dot{n}_1 = rn_2 - qn_3$$

$$\dot{n}_2 = pn_3 - rn_1$$

$$\dot{n}_3 = qn_1 - pn_2$$

The set of nine differential equations must be solved for the nine direction cosines rapidly enough so that errors do not accumulate from extrapolation of the rates over the computation time. We note that the matrix is normal and orthogonal. Drift in the computation results in a characteristic departure from normality and orthogonality, which may be exploited to detect and redistribute errors in the computation. There are several digital methods of mechanizing the solution which involve use of the DDA or some adaptation of its logic.

Results of Methods of Solution

The following paragraphs discuss the preliminary investigation results of two methods of solution. Each method involves a simple computation scheme to determine the growth of the unit vector.

Method 1

The computation scheme is as follows:

$$(\ell_i)_n = (\ell_i)_{n-1} + (\Delta\ell_i)_n$$

$$(m_i)_n = (m_i)_{n-1} + (\Delta m_i)_n$$

$$(n_i)_n = (n_i)_{n-1} + (\Delta n_i)_n$$

$$(\Delta\ell_1)_n = r_n\Delta t\,(\ell_2)_{n-1} - q_n\Delta t\,(\ell_3)_{n-1}$$

$$(\Delta\ell_2)_n = p_n\Delta t\,(\ell_3)_{n-1} - r_n\Delta t\,(\ell_1)_{n-1}$$

$$(\Delta\ell_3)_n = q_n\Delta t\,(\ell_1)_{n-1} - p_n\Delta t\,(\ell_2)_{n-1}$$

and similarly for m and n.

Using body rates of 20 deg/sec simultaneously about the three body axes and a sampling interval of 100 microseconds, the results of Method 1 are listed below. The product of the calculated matrix by its transpose, at nine seconds, is

1.00021935	-0.00010968	-0.00010968
-0.00010968	1.00021935	-0.00010968
-0.00010968	-0.00010968	1.00021935

Method 2

The computation scheme is as follows:

$$(\ell_i)_n = (\ell_i)_{n-1} + (\Delta\ell_i)_n$$

$$(m_i)_n = (m_i)_{n-1} + (\Delta m_i)_n$$

$$(n_i)_n = (n_i)_{n-1} + (\Delta n_i)_n$$

$$(\Delta \ell_1)_n = r_n \Delta t\, (\ell_2)_n - q_n \Delta t\, (\ell_3)_n$$

$$(\Delta \ell_2)_n = p_n \Delta t\, (\ell_3)_{n-1} - r_n \Delta t\, (\ell_1)_{n-1}$$

$$(\Delta \ell_3)_n = q_n \Delta t\, (\ell_1)_{n-1} - p_n \Delta t\, (\ell_2)_{n-1}$$

$$(\Delta m_1)_n = r_n \Delta t\, (m_2)_{n-1} - q_n \Delta t\, (m_3)_{n-1}$$

$$(\Delta m_2)_n = p_n \Delta t\, (m_3)_n - r_n \Delta t\, (m_1)_n$$

$$(\Delta m_3)_n = q_n \Delta t\, (m_1)_{n-1} - p_n \Delta t\, (m_2)_{n-1}$$

$$(\Delta n_1)_n = r_n \Delta t\, (n_2)_{n-1} - q_n \Delta t\, (n_3)_{n-1}$$

$$(\Delta n_2)_n = p_n \Delta t\, (n_3)_{n-1} - r_n \Delta t\, (n_1)_{n-1}$$

$$(\Delta n_3)_n = q_n \Delta t\, (n_1)_n - p_n \Delta t\, (n_2)_n$$

Using the same body rates and sampling intervals as for Method 1, the product of the calculated matrix by its transpose, at nine seconds, is

1.00010648	-0.00002319	-0.00002319
-0.00002319	1.00010648	-0.00002319
-0.00002319	-0.00002319	1.00010648

The results of Methods 1 and 2 are plotted in Fig. 18. If the diagonal terms are represented by $1 + \epsilon_1$ and the magnitudes of the off-diagonal terms by ϵ_2, then the square of the determinant is

$$(1 + \epsilon_1)^3 - 3(1 + \epsilon_1)\, \epsilon_2^2 \, x - 2\epsilon_2^3$$

and the magnitude of the determinant is approximately $1 + 1.5\, \epsilon_1$. Hence, the percentage error after nine seconds is 0.03% for Method 1 and 0.016% for Method 2. If the results of the two methods are linearly extrapolated to 1000 seconds and the rates reduced to 6 deg/sec, then the errors for Methods 1 and 2 are 1.14% and 0.48%, respectively.

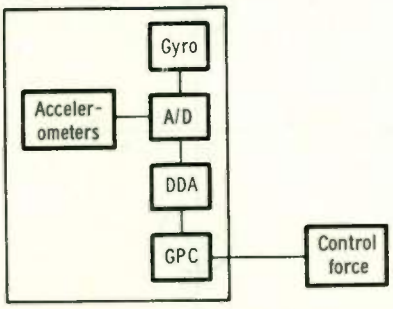


FIG. 1. GIMBALLESS INERTIAL SYSTEM

References

1. Cronin, T. R.; "An Automatic Converter to Prepare Analog Data for Digital Computation," MIT Servo Lab Report 7138-R-19, 15 January 1957.

2. Lee, F.; "Transistor Linear Sweep Generator," Electronics, 16 December 1960.

3. Enslein, F.; "Characteristics of Silicon Junction Diodes as Precision Voltage References," IRE Transactions on Instrumentation, VI-6A, No. 2, 1957.

4. "Switching Levels in Transistor Schmitt Circuits," IRE Transactions on Instrumentation, VI-9A, No. 3, 1960.

5. Millman and Taub; "Pulse and Digital Circuits," McGraw Hill, 1956.

6. Gordon, B. M., Smith, B. K.; "How Much Do Components Limit Converter Performance," Electron Design, 21 June 1961.

7. Carlson, R.; "Tunnel Diode Fast-Step Generator Produces Positive or Negative Steps." Electronics, 28 July 1961.

8. Tunnel Diode Manual, General Electric Company, 1961.

Acknowledgment

The effort of H. Wright in detail design of electronic circuitry is gratefully acknowledged.

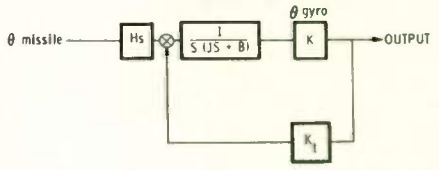


FIG. 2. GYRO LOOP

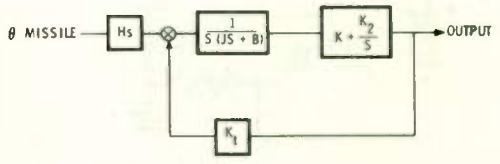


FIG. 3. GYRO LOOP WITH INTEGRATOR IN FORWARD LOOP

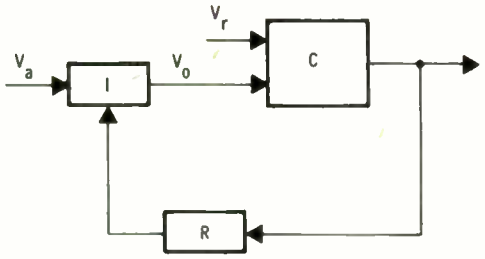


Fig. 4. Voltage Domain Converter

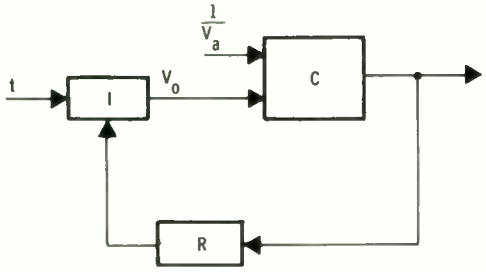


Fig. 5. Time Domain Converter

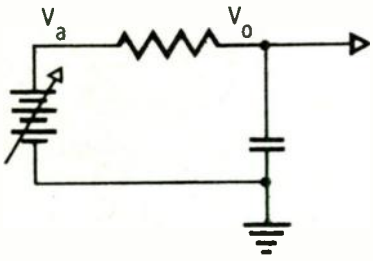


Fig. 6. RC Integrator

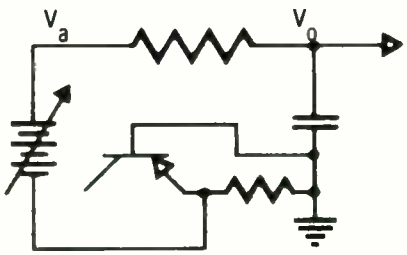


Fig. 7. Bootstrap Integrator

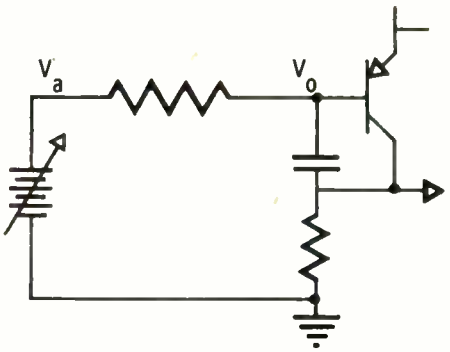


Fig. 8. Miller Integrator

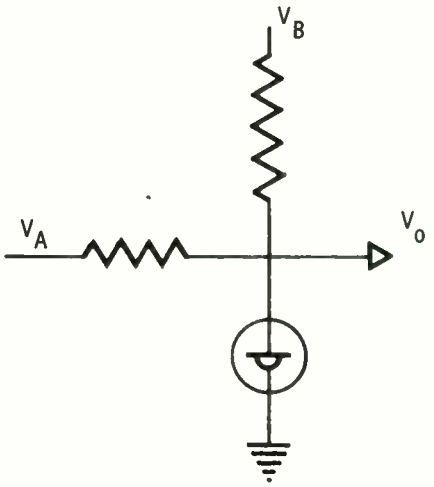


Fig. 9. Tunnel Diode Comparator

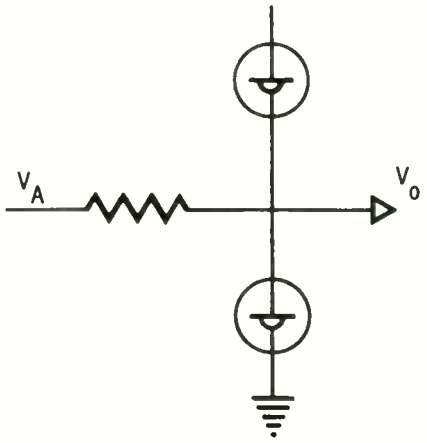


Fig. 10. Tunnel Diode Comparator

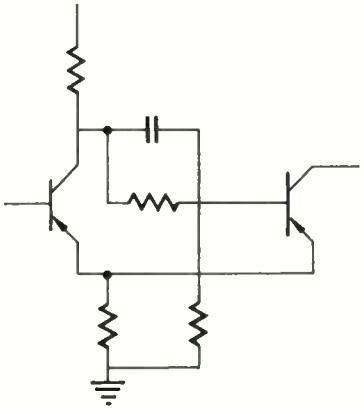


Fig. 11. Schmitt Comparator

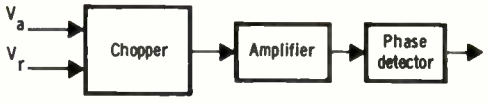


Fig. 13. Phase Detecting Comparator

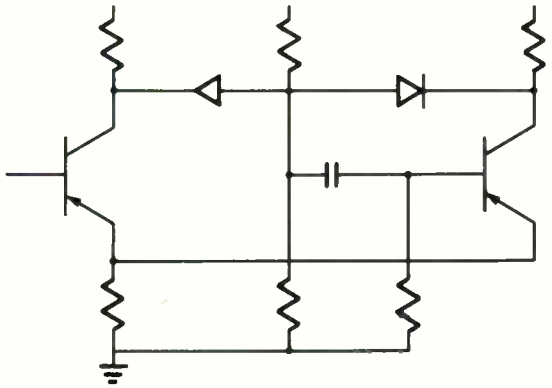


Fig. 12. Modified Schmitt Comparator

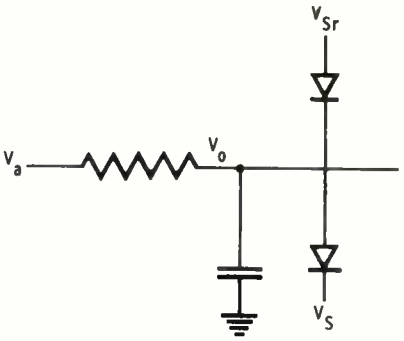


Fig. 14. Reset Clamping

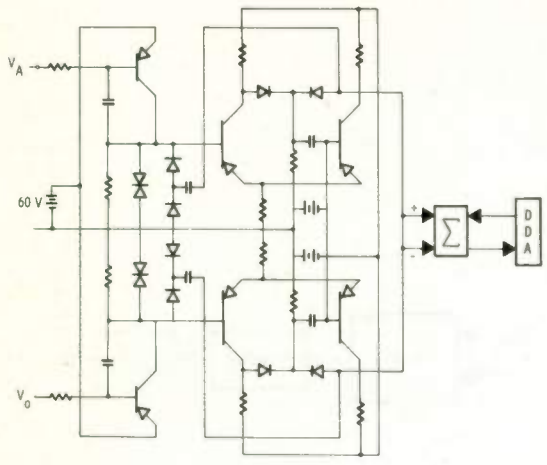


Fig. 15. Analog to Digital Conversion

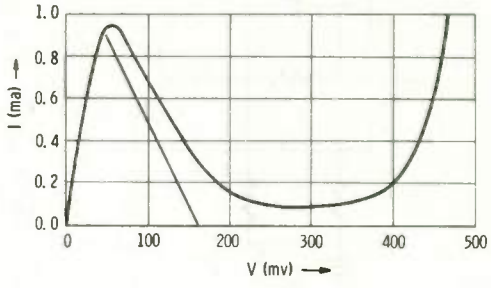


Fig. 17. Tunnel Diode Characteristic Curve

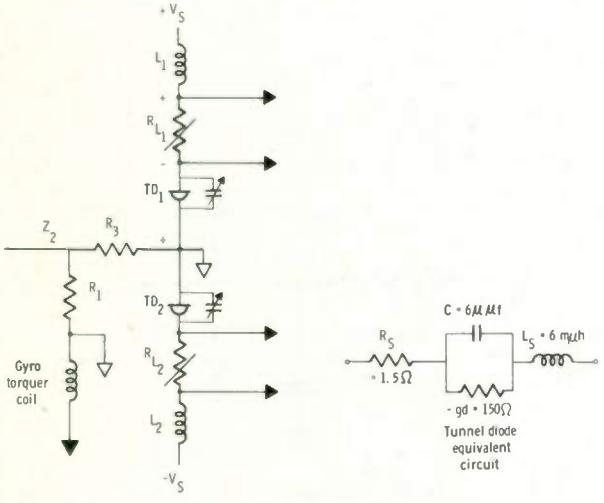


Fig. 16. Analog to Digital Converter

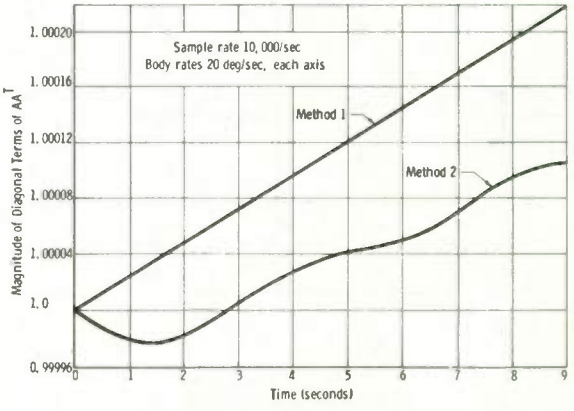


FIG. 18. COMPARISON OF METHODS OF INTEGRATION

TRANSPORT AIRCRAFT NAVIGATION WITH AIRBORNE DOPPLER RADAR

Richard N. White
Trans World Airlines
Kansas City, Missouri

Trans World Airlines will pioneer use of Doppler radar as a primary means of transport aircraft navigation over the North Atlantic. This paper summarizes flight test results using dual Doppler radar sensors and dual Doppler navigation computers on Boeing 707 long range aircraft. Test flights have been flown over TWA's North Atlantic routes for almost two years to "prove-in" the Doppler system.

An aircraft can be navigated between any two points on the earth's surface on the basis of three input parameters: heading, drift angle, and ground speed. Basically, the navigational accuracy between the two points is limited only by the accuracy of these three inputs. In the system flown by TWA, heading (direction) input is generated by electronic flux-valve compass systems. Drift angle and ground speed inputs are obtained from the Doppler radar sensor equipment. Until Doppler radar techniques made direct drift angle and ground speed readouts available, these two primary parameters could be obtained only by indirect and often inaccurate means.

The basic theory and design details of the Bendix Doppler equipment used in these flight tests is covered in three earlier IRE papers.[1,2,3]

The Air Navigation Vector Diagram

Figure 1 is the basic navigation vector diagram requiring solution. Doppler radar provides the drift angle information (angle C-A-B) and the ground speed vector quantity A-B).

In the case shown, the selected course is 90° east of magnetic north. A north wind (vector quantity C-B) causes a 10° drift to the right. The aircraft must fly a heading of 80° to fly along the preselected 90° course (along line A-B). Thus the heading correction angle must be equal and opposite to the drift angle to keep the aircraft on the selected course. This figure also shows that the actual course (sometimes called "course made good") is the algebraic sum of the heading (always a positive number) plus the drift angle (right drift being positive and left drift being negative).

The computed actual course (heading plus drift angle) is fed to the navigation computer and compared with the selected course which the pilot sets on the computer panel. Figure 2 shows the relationship of an actual course (100°) and the selected course (90°). The angular difference between the actual and selected course is 10°. The figure shows that miles traveled along the selected course is the total miles traveled times the cosine of this angular difference. The cross track departure from the selected course in miles is the total miles traveled times the sine of the difference angle. When the selected course and the actual course coincide, the miles along track is equal to the total miles traveled.

The Doppler navigation computer solves these vector diagrams and presents the "miles to go" (along track) and "offset miles" (across track) information on the computer control panel as shown in Figure 3.

The Doppler Navigation Computer - Operating Principles

Figure 4 is a block diagram of the Bendix Co. CPA-24A navigation computer used in the TWA aircraft installation.

The frequency of the ground speed oscillator in the Doppler sensor is directly proportional to aircraft ground speed. The output of this oscillator is amplified in the integrator drive circuit, then drives the synchronous motor whose rotation rate is proportional to the ground speed input from the Doppler system. This synchronous motor therefore is an integrator which provides a total shaft rotation proportional to total miles flown.

Drift angle electrical information is generated in the Doppler sensor equipment by comparing the Doppler shift in each of the four received beams reflected from the earth's surface. The drift angle electrical output drives the differential synchro in the sensor unit shown at the left. The actual course output from the differential synchro is the algebraic sum of the compass heading and the drift angle. This actual course output feeds another differential synchro which compares the course selected by the pilot with the actual course and provides an angular difference signal at the input to the control transformer of the course error servo loop which drives the mechanical ball type sine-cosine resolver.

This resolver converts the total miles traveled input into cosine (along track miles) and sine (cross track miles) outputs.

These two outputs supply the two torque motor transmitters shown. The along track miles transmitter supplies a torque motor receiver which drives the "miles to go" counter. The cross track motor transmitter feeds the torque motor receiver unit which drives the "miles offset" digital readout.

The cross track readout mechanism also drives a potentiometer which feeds the offset miles bar in the special pilot's instrument shown in Figure 5. The potentiometer supplies direct current to the offset miles meter movement directly proportional to the offset miles. This indicator is calibrated so each dot on the offset miles display represents two miles lateral deviation from the selected course.

The miles offset mechanism also drives an autopilot input potentiometer which supplies 400 cycle AC proportional to miles offset. The phase of the AC input to the autopilot reverses when the offset miles reading passes through zero. The synchro transmitter driven by the course error servo amplifier loop also supplies an autopilot input.

The ten mile warning light on the "distance-to-go" readout illuminates when the readout indicates ten miles to go along the flight segment. This gives the pilot about one minute warning, at typical jet cruising speeds, before he reaches the end of the segment, at which time the computer transfers automatically from stage A to stage B. Then stage B starts operating on the second flight segment which the pilot previously set into the stage B counters.

The selected course for the two stages can be preset into the computer panel by rotating the adjacent knobs. The along track information drives the "miles to go" digital readout. In practice the pilot sets in the length of the segment in miles with this digital counter and its knob. As the aircraft progresses along its selected course the miles to go readout "counts down" indicating the number of miles to go to the end of the selected segment. The "miles offset" counter in the center portion of the panel indicates the aircraft deviation right or left of the selected course. The miles offset readout enables the pilot, in effect, to offset the selected course to either side in the event he desires to place an offset correction into the computer.

Compass Accuracy

Since the computed actual course is the algebraic sum of the heading and drift angle the actual course flown along the earth's surface can be no more accurate than the combined accuracies of the heading and drift angle inputs. Also, since we had not been able to find measurable drift angle errors, the Doppler system cross track accuracy is tied very closely to compass input accuracy. This fact of life caused us to examine our compass systems carefully and set up a comprehensive program to increase compass accuracy.

The first objective was to determine that our Boeing 707 compass systems once accurately calibrated, would remain accurately calibrated for long periods.

The compass systems on the first 707 aircraft containing the Doppler system were very carefully calibrated at the Boeing plant prior to aircraft delivery. When the aircraft was placed on the North Atlantic routes its compass systems were observed closely and checked frequently with celestial observations. After flying this aircraft for several months over the Atlantic we concluded the compass systems on this aircraft did not drift appreciably once properly calibrated.

Meanwhile other fleet aircraft had been delivered. We soon found that their compass systems were not maintaining accuracies as good as the compass systems on the first aircraft. Some compass systems on these later aircraft were checked on compass swinging facilities at Idlewild airport and at other major maintenance bases on the TWA international system. We concluded something must be wrong with the systems on these later aircraft or with the procedures used to calibrate the compasses on the ground.

We suspected some appreciable but unknown magnetic effects might be affecting the accuracy with which we could determine magnetic North during calibration. The Idlewild airport compass rose area was surveyed with a surveyors transit, using a distant sighting point for directional reference. This survey revealed deviation errors up to 5 degrees affecting the aircraft compass transmitter at some points in the compass swinging area. These errors disappeared when the transit was moved away from the concrete ramp area. Magnetized reinforcing steel in the concrete probably caused the errors.

Meanwhile TWA constructed a 200 foot diameter nonmagnetic compass rose at Kansas City, several hundred yards from the nearest known magnetic disturbance. Reinforcing steel was not used. Carefully surveyed sighting points were located around this compass rose at considerable distance to minimize sighting parallax errors caused by the aircraft being at different points on the compass rose. But since the compass rose is 200 feet in diameter and the reference points were placed 1200 to 2500 feet away, sighting parallax corrections for position of the aircraft on the compass rose are still required.

A vertical line passing through the periscopic sextant location is very close to the 707 nose wheel vertical axis. The Kansas City compass rose was marked with reference points on the circle normally followed by the nose wheel while the compass is being swung. The compass calibration people use a set of tables showing the sighting parallax error correction to be applied when the aircraft nose wheel is over any one of these reference points.

During the calibration procedure the compass synchro output at the electrical input to the Doppler system is checked with a servo-type synchro readout device having a nominal accuracy of $\pm$ 0.1°. This method of checking compass electrical output effects a significant increase in Doppler system accuracy as compared to the earlier method using the pilots compass dial readout for

the calibration reference.

Specially designed digital type azimuth readout devices are used on the periscopic sextants used to determine angular relationship of the aircraft fuselage center line with the line between the aircraft and the reference sighting points around the compass rose. Some additional sextant readout design refinement is now in progress.

The more accurate compass systems resulting from use of the procedures and equipment described above effected a considerable increase in Doppler system accuracy. Test flight data now indicates the compass systems on our international 707 aircraft probably have typical errors appreciably less than plus or minus one degree. And continued flight testing indicates the compass systems are maintaining these low errors over relatively long time periods.

All test flights reported in this paper used the magnetic mode of the Bendix Eclipse-Pioneer Polar Path compass system. This system can be operated in the free directional gyro mode. One basic objective of these test flights was to assure that the compass system used would be sufficiently accurate and reliable so entire transatlantic crossings could be flown without requiring compass checking or resetting enroute.

Unfortunately the free gyro mode of this particular system did not meet this objective because the free gyro drift rates were either too high or too unpredictable, thus requiring resetting and/or accuracy checks enroute.

We are continuing investigations toward an improved free gyro compass system even though the magnetic system performed very satisfactorily with the required accuracy during almost two years of test flying.

Basically, free gyro compasses are very attractive for long range Doppler flight application because they permit direct great circle (minimum distance) flight between any two points and because they do not have the deficiencies a magnetic reference suffers at the higher latitudes near the earth's magnetic poles.

Aircraft Installation

Figure 6 is a composite picture showing the equipments used in the Doppler sensor system in TWA's Boeing 707 aircraft. The tracker is located in the aircraft radio rack in a compartment below the cockpit. The transmitter-receiver and the antenna assembly are located in the unpressurized radome area just forward of the nose wheel. The sensor control unit is located in the center overhead pilot's panel. The drift angle-ground speed-offset miles indicator is located on the pilot's main instrument panel.

The special drift angle-ground speed-offset miles indicator (see Figure 5) was developed by Eclipse-Pioneer to TWA requirements. We desired an "offset miles" readout on the pilot's panel which operated in the same familiar manner as the VOR "bar and bull's-eye" readout. Since each dot represents two miles lateral displacement, the offset bar in Figure 5 indicates the aircraft is a little more than two miles to the left of the course and that the course center line therefore is to the right of the aircraft. Figure 7 shows the components of the Doppler navigation computer. Two control panels are located in the pilot's center overhead control panel. The electronic unit shown to the left is located in the electronic radio rack in the compartment below the cockpit floor.

Figure 8 shows the cockpit center overhead control panel area. The Doppler sensor control panel and the associated computer control panels are grouped together in the upper center portion of this picture. The No. 1 system is on the right the No. 2 system is on the left. Figure 9 is the left main pilot's instrument panel. The ground speed-drift angle-course offset indicator is the instrument in the lower righthand corner. The switch immediately below this instrument enables the first pilot to select either Doppler computer as the input for the course offset readout.

The Figure 10 block diagram shows the interconnections used in the dual sensor-computer-compass installation. Note that both systems are supplied from the same compass system at any given time even though the aircraft contains two identical magnetic compass systems. We believe this arrangement is better than having each compass feed a navigation computer separately because with the arrangement shown, any difference in offset miles readout between the two computers must necessarily be confined to Doppler system and computer errors. Thus, compass error itself does not contribute to any difference in the offset miles readout from the two computers. The compass systems can be checked against each other procedurally by comparing the No. 1 compass readout on the pilot's panel with the No. 2 compass readout on the copilot's panel. If these compass cross checks plus any external checks indicate one of the compass systems is suspect, the crew can transfer both Doppler systems to the second compass system by actuating the switch shown at the bottom center of the Figure 10 block diagram.

Test Flight Data Summary

Figure 11 summarizes the test data for the three test flight phases. Each phase had a specific purpose.

The phase A test flights were conducted primarily to "shake down" the Doppler sensor equipment and evaluate its basic reliability in airline aircraft environment. They also permitted an initial estimate of the Doppler system's basic accuracy capability and highlighted the areas requiring further technical or operational refinement. The aircraft equipment consisted of a Doppler sensor, a drift angle-ground speed readout instrument on the pilot's panel, and a distance gone (in Miles)

counter instrument also on the pilot's panel.

The phase B flight tests were primarily a check on the basic accuracy capability of the complete Doppler-sensor-computer system supplied by compass systems which had been calibrated with newly evolved calibration procedures. Flight crews did not reset the Doppler computer enroute. Thus, along track and cross track errors at landfall after a transatlantic crossing indicated the basic system accuracy capability.

The phase C flight test procedures and test reporting methods were tailored specifically to demonstrate the Doppler system's capability as a future primary navigation aid in the North Atlantic air traffic control and navigation aid complex.

Flight crews were instructed to reset the computer enroute and adjust headings if fixes from ground based radio aids indicated lateral deviations exceeding 10 miles after flying 600 to 900 miles.

The computers were reset on 38 of the 64 phase C flights.

The basic purpose and justification for each test flight phase should be kept in mind when examining the test flight results.

Phase A Flight Test Procedures and Test Data

Prior to each flight the crews computed the magnetic courses and distances for each flight segment by reference to large Lambert Conformal North Atlantic area charts. Segment mileages were scaled directly from these charts. Magnetic course for each segment was determined on the basis of magnetic variation at the segment mid-point.

The aircraft equipment used in the phase A flights did not include the navigation computer. Thus, in effect, the pilot was his own computer. He mentally added indicated drift angle to the segment magnetic course to determine the magnetic heading required to keep the aircraft on course. This procedure required very close attention to the drift angle readout to keep the aircraft on the correct heading. Progress along the flight path was indicated by the special "distance gone" counter which indicated the total distance traveled by the aircraft.

The navigation computer described earlier computes the total distance the aircraft travels with respect to the selected course lines. The aircraft may actually fly a longer distance than the distance indicated by adding the course segment distances because aircraft in flight cannot make sharp angular turns at the course segment intersections. Instead, the aircraft must follow the large turn radius imposed by maximum bank angle limits dictated by passenger comfort requirements. At 600 miles per hour, turn radii are large. The turns did not begin until the aircraft reached the intersection of two consecutive flight segments. Thus, beginning at the flight segment intersection, the aircraft must fly a path dictated by the turn radius before it intersects the center line of the next segment. The additional flight distance in the turns probably added the extra mileage indicated by the along track error bias in the positive direction shown in the Figure 11 table for the phase A flights.

Considering the limitations imposed by the method of determining segment distances and courses, the manual method of determining correct heading enroute and the course distance errors caused by the turn radius geometry outlined above, the phase A error figures were smaller than we had expected.

Phase B Test Flight Procedures

Closely controlled test flight procedures assured uniform flight test data. The normal rules for choosing transatlantic flight routes were observed on the test flights. The flight crews, the meteorologists, and the flight dispatcher normally confer before each transatlantic flight to choose a minimum time route, after studying winds aloft data and traffic conditions. Then the chosen route is broken down into flight segments. The ends of each segment are normally defined in even degrees of latitude and multiples of 10° of longitude. (i.e., 10° west, 20° west, etc.)

Specially computed "track and distance" tables were used to convert the longitude-latitude coordinates of each flight segment into "distance to go" in nautical miles and magnetic course for each segment.

These "track and distance" tables were computed to the nearest minute of course angle and the nearest nautical mile using the Bowditch navigation tables.[4] The course angle computations were rounded off to the nearest one-half degree. Figure 12 shows a portion of the tables for typical landfall exit and entry points. Figure 13 shows a portion of the enroute tables.

The flight segment course and distance information then is transferred to the crew maps as shown in Figure 14.

The pilot sets the course and distance-to-go figures for the first two segments into stage A and stage B of the computer control panel respectively. He then activates stage A (see Figure 3), then steers the aircraft to keep the offset miles pointer centered. (see Figure 5)

When the aircraft reaches a point ten miles from the end of the first segment, a ten mile warning light on the instrument panel illuminates, indicating he has ten miles to go. Stage B is activated automatically when the stage A distance-to-go reading reaches zero miles.

The routes and flight segments for these test

flights were chosen so the last segment would pass directly over a radio facility when reaching the opposite shore of the Atlantic. Figure 15 shows the method of checking the cross track error when crossing the radio facility at the end of the transatlantic crossing.

For example, when the aircraft, approaching the eastern shore of the Atlantic from the left, reaches the range limit of the chosen ground radio facility the pilot steers the aircraft directly toward the ground radio facility while letting the Doppler system continue to compute the along track and cross track mileage. When the aircraft passes directly over the radio station the cross track "offset miles" reading in the computer panel is noted and recorded. This indicates the distance between the radio facility directly under the aircraft and the track the aircraft would have taken if the pilot had steered the aircraft to keep the offset pointer centered. Thus the offset (cross track) error is numerically the distance between the extended computed course line and the radio station.

This cross track error measurement method introduces negligible measurement error if angle β is small and negligible error in the total measurement of along track distance, shore to shore, when angle β is small.

Phase B Test Flight Results

The phase B flight test results shown in the Figure 11 table are the best data we have to date on the basic capability of our combined Doppler sensor, Doppler computer and compass system. Crews were not permitted to reset the computer enroute. The figures shown represent data from 68 navigation computer computations because the aircraft had two computers installed.

Phase C Flight Test Procedures

All of the test flight procedures applying to the phase B flights also applied to the phase C flights. But since the phase C flights permitted enroute correction of offset errors additional standard procedures were used.

About one to one and one-half hours (about 600 to 900 miles) after leaving the departing landfall point a Loran or Consol fix was obtained to check basic compass accuracies and to evaluate the accuracy of the two computers which normally operate simultaneously.

If this fix indicated more than ten miles deviation either side of the selected course and more than 0.6° of compass deviation, a second fix was taken to check the validity of the first. If the two fixes confirmed each other the offset miles error determined by the fix was placed in the computer offset miles counters of both computers. Deviations greater than 0.6° were considered to be 1° deviation, up to one degree. Any deviation computed to be above 1° was used at its full value and was applied to the remaining magnetic course segments and rounded off to the nearest one-half degree, then inserted as required into the selected course counter on the computer panels.

In establishing the amount of compass deviation, the total distance back to the landfall point from the off course position was used (even though an interim fix might have shown the aircraft to be on course).

When the ground aid fixes indicated significant deviation from course, action was taken to return the aircraft to course before the end of that flight segment was reached - but in any case within the next 200 nautical miles. The crews were instructed to question Loran fix accuracies when the fixes differed more than five to ten miles from the prescribed Doppler course.

Phase C Test Flight Results

The 95% phase C cross track error figures shown in the Figure 11 table include the effects of computer resetting on 38 of the 64 flights.

The phase B and phase C flight test procedures were identical except that phase C procedures permitted computer resetting enroute. The phase B procedures did not permit resetting.

Thus, comparison of the phase B and phase C 95% cross track error figures indicate, at least roughly, the over-all cross track error reduction that can be effected by correcting the Doppler system with fixes from present ground radio aids, (Loran and Consol).

The reduction in cross track errors effected by computer resetting and heading correction enroute was smaller than we had expected.

The phase C flight test reports included equipment outage records. Figure 16 shows the No. 1 and the No. 2 sensor equipment outage times for each of the 64 flights. The two graphs show a general resemblance, but the dissimilarities are significant. For example, Flight 12 had a four hour outage on the No. 2 sensor equipment and a 30 minute outage on the No. 1 sensor equipment. This would indicate a definite No. 2 equipment malfunction (as opposed to the lack of adequate received signal level).

On the other hand, outage times for Flights 23, 24 and 25 show very similar outage times for both sensors. In these cases we can conclude reasonably that the reflected signal return during the outage periods was probably marginal. The No. 1 and No. 2 sensors were not inoperative simultaneously on any single flight for more than 30 minutes. This speaks well for over-all dual system reliability.

When the signal drops out in flight, the ground speed and drift angle sensor outputs "freeze" at the last correct values. Thus, the computer continues to operate during the periods

of received signal dropout from these "frozen" drift angle and ground speed inputs.

Experience to date indicates insignificant errors due to dropout times of 30 to 45 minutes or less for a single Atlantic crossing. This is true probably because the Doppler system was flown in relatively constant altitude cruising flight typical for North Atlantic jet transport operation.

Figure 17 shows the computer malfunctions during the phase C flights.

Figure 18 shows the along track deviation, in miles, at landfall. Along track deviations were not appreciably affected by any lateral (cross track) correction the crew may have made enroute. These curves do not indicate percentage along track deviation because the distance from shore to shore varied depending on the particular route used. The total mileage for 59 of the flights was about 1900 nautical miles. The mileage for the remaining five flights was about 2730 nautical miles. (Nantucket Island to the Azores)

Both the along track deviations and cross track deviations at landfall were plotted in miles instead of percentage of distance flown because the main purpose of the phase C test was to demonstrate the aircraft did not exceed its time-distance estimates by more than 30 to 35 miles (about 3 minutes of time) or exceed the North Atlantic ATC maximum lateral displacement limits of 60 miles.

Figure 19 shows the cross track deviation at landfall for both computers for each of the 64 flights. Missing points in the Figure 18 and Figure 19 graphs indicate lack of valid data caused by sensor or computer failure in flight.

Most of the larger deviations were caused by computer resetting or heading correction errors.

For example, flight No. 43 shown a No. 1 computer deviation about 50 miles to the right and a No. 2 computer deviation about 56 miles to the right. In this case an error in the heading correction caused both computers to indicate larger deviations than would have been indicated if the heading correction had been applied properly.

Drift Angle Output Accuracy From Doppler Sensor

As the compass system errors were reduced, approximately proportional increases in total Doppler system accuracy were noted. This indicates drift angle errors are probably insignificant. Attempts to measure drift angle accuracy have been to no avail because available measurement methods are not accurate enough.

We can only conclude that if the antenna electrical longitudinal axis is closely aligned with the aircraft fore and aft axis, the drift angle error is probably some small fraction of one degree in level cruising flight.

Coupled Autopilot Flight Tests

In February 1961 TWA conducted flight tests with the Doppler radar computer coupled to the autopilot. The autopilot heading command input was supplied by a combination of Doppler computer cross track offset and course error outputs. These two outputs were combined in a mixing network to provide a single autopilot heading command input. The ratio of the two input signals was not critical. The flight tests were very encouraging. The autopilot steered the aircraft smoothly to maintain zero miles cross track offset with no instability. It was recognized that computer stage transfer would generate a large autopilot step input which would cause the autopilot to initiate unacceptably large aircraft roll rates while the autopilot was attempting to align the aircraft to the new selected course.

The solution to this step-input problem is a "command modifier" in the autopilot heading control circuitry. This device limits autopilot roll rate regardless of step input magnitude. Command modifiers are now planned for all TWA jet aircraft autopilots. When this installation is completed we anticipate very smooth and tightly coupled Doppler-autopilot operation. The switching functions to implement the Doppler-autopilot coupled function are planned for all TWA international jet aircraft. When this aircraft installation is completed, Doppler-autopilot coupled flights for the entire Atlantic crossing are possible. When stage transfer takes place at the end of a flight segment, the aircraft will automatically begin following the next segment which the pilot had set into the other computer stage before the stage transfer took place.

Future Test Program

The phase C flight test data supported TWA's presentation to FAA for technical approval for Doppler radar as a primary long range navigation system.

This test data, particularly the along track accuracy data, suggests the possibility of reducing longitudinal and possibly lateral aircraft separation in the North Atlantic area, thereby increasing the aircraft traffic handling capacity in this area.

The next step is a follow-on error measurement program using carefully calibrated ground radio facilities and possibly long range ground radars. This program is in the study phase now and probably will be implemented later.

Bibliography

1. Cordry, B.L., "A Doppler Navigational Radar Utilizing New Techniques," Conference Proceedings of 6th Annual IRE East Coast Conference of Aeronautical and Navigational Electronics, 1959.

2. Kirner, E.O., "A New Doppler Radar Frequency Tracker," Conference Proceedings of 6th Annual IRE East Coast Conference of Aeronautical and Navigational Electronics, 1959.

3. Willey, Robert E., "Flat Array Antenna for a Doppler Navigation System," Conference Proceedings of 6th Annual IRE East Coast Conference of Aeronautical and Navigational Electronics, 1959.

4. Bowditch, Nathaniel, "American Practical Navigator," U.S. Navy Department Hydrographic Office Publication H.O.9, 1943 Edition.

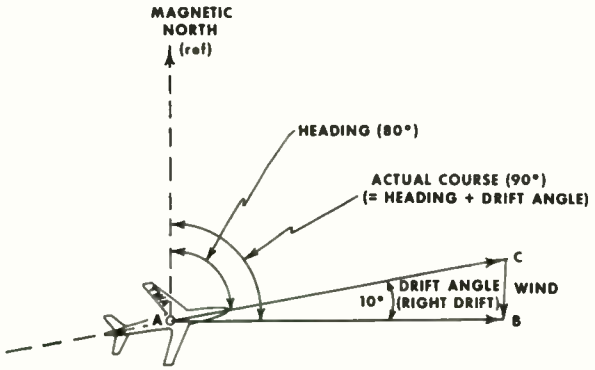


Fig. 1. Relationship of heading, drift angle, and actual course.

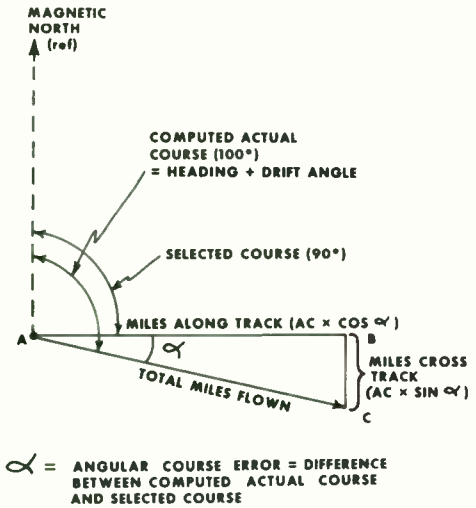


Fig. 2. Resolution of total miles flown vector into miles along track and miles cross track vector components.

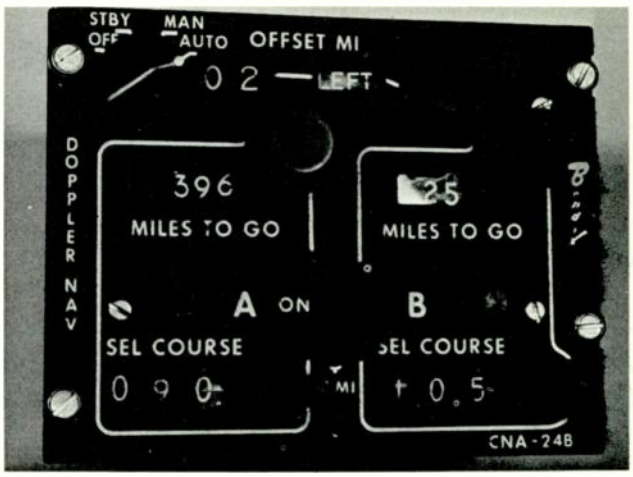


Fig. 3. Bendix radio Type CNA-24A-2 Doppler radar navigation computer control panel.

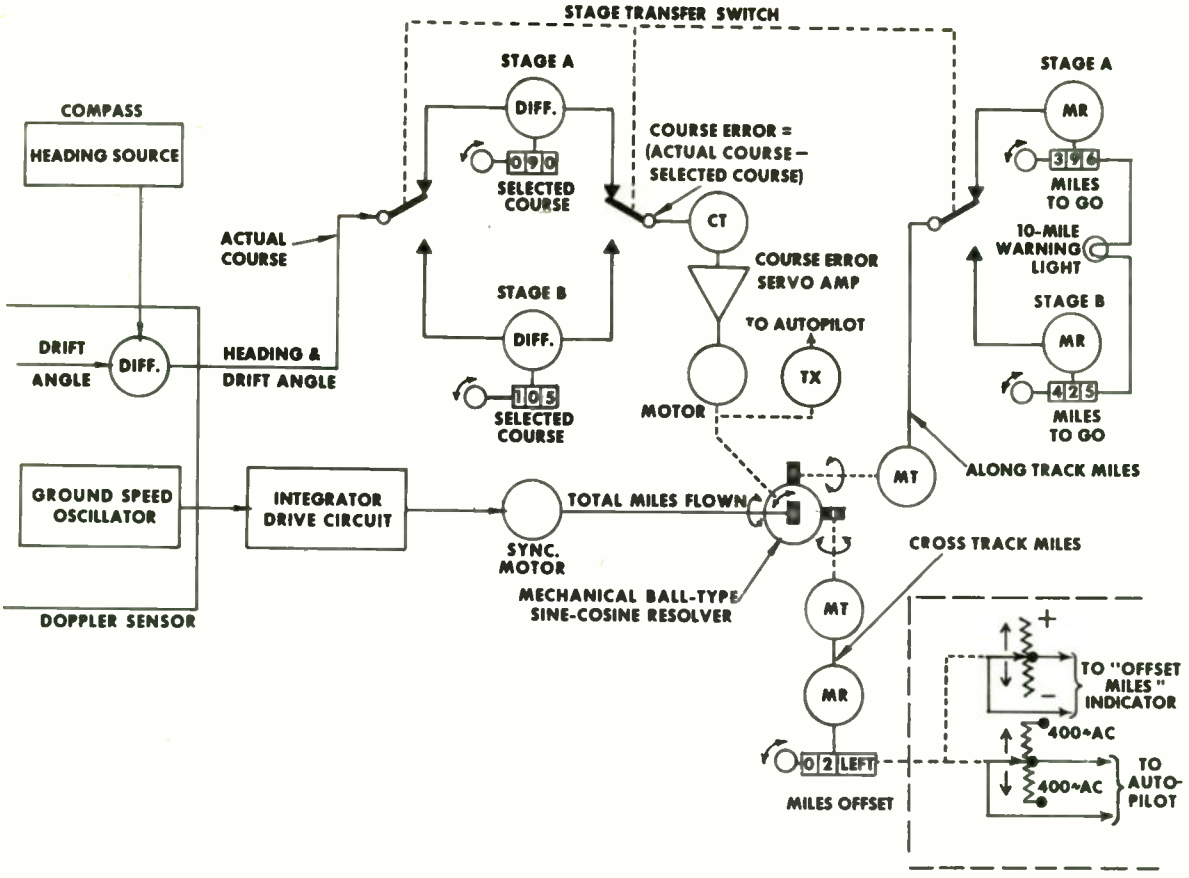


Fig. 4. Bendix Radio Company Type CPA-24A Doppler Navigation Computer - block diagram.

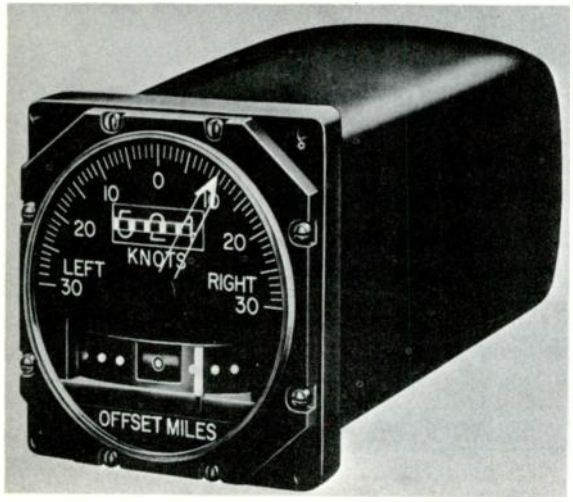


Fig. 5. Special ground speed - drift angle - offset miles indicator developed by the Eclipse-Pioneer Company to TWA requirements.

Fig. 6. Bendix Type DRA-12A Doppler sensor equipment. Two sets of this equipment were installed in each aircraft used in the Phase B and Phase C flight tests.

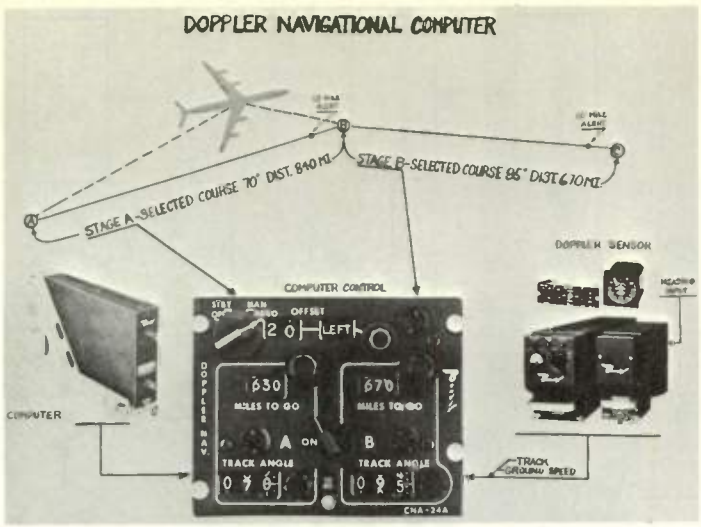


Fig. 7. Bendix Radio Type CPA-24A Doppler navigation computer equipment. The Doppler sensor equipment supplying the actual course and ground speed inputs to the computer is shown on the right.

Fig. 8. Pilot's center overhead control panel. The two Doppler computer control panels are located in the upper portion on either side of the center line. The Doppler sensor control panels are immediately above the related computers. The Number 1 system is on the left, the Number 2 system on the right.

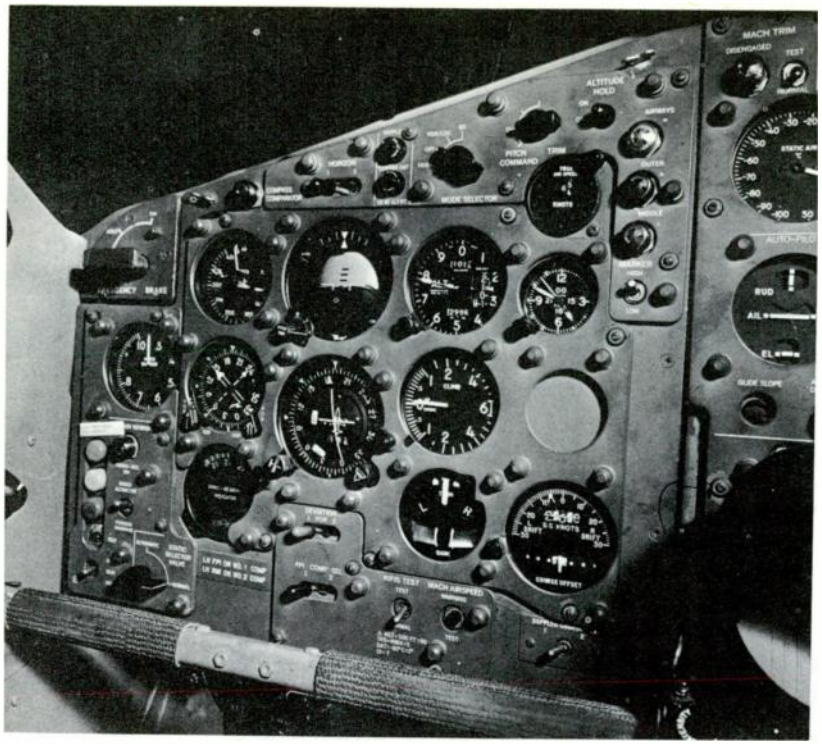


Fig. 9. Pilots left main instrument panel, TWA-Boeing 707 jet aircraft. The special ground speed - drift angle - course offset indicator is in the lower right-hand corner. The Doppler "sensor out" and "ten-mile alert" indicator lights are in the vertical row above and slightly to the right of the gyro horizon instrument.

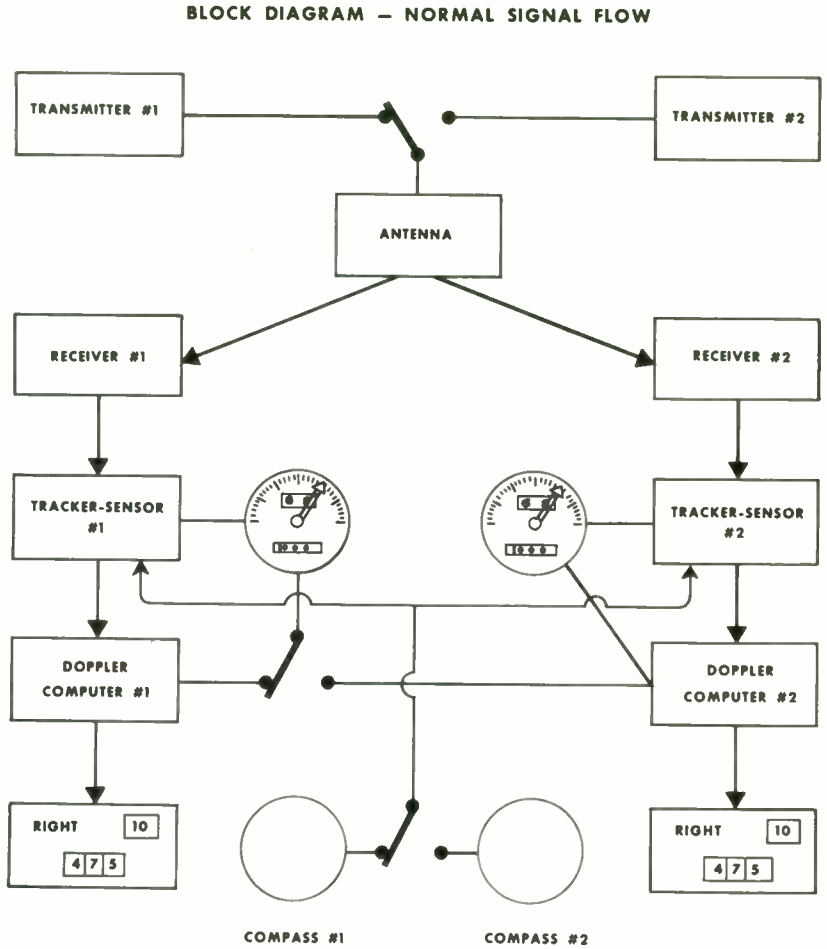


Fig. 10. Dual Doppler sensor - Dual Computer interconnection block diagram. This arrangement was used in the Phase B and Phase C test flights.

TRANS WORLD AIRLINES DOPPLER RADAR TEST FLIGHTS - SUMMARY

TEST FLIGHT PHASE	PURPOSE	FLIGHT TEST DATES	AIRCRAFT EQUIPMENT	AIRCRAFT USED	NO. OF ATLANTIC CROSSINGS	ALONG TRACK ERROR (% OF DISTANCE FLOWN)		CROSS TRACK ERROR (% OF DISTANCE FLOWN)		REMARKS
						AVERAGE	95%	AVERAGE	95%	
PHASE A	DOPPLER SENSOR EQUIPMENT RELIABILITY AND INITIAL ACCURACY TESTS.	JAN. 1, 1960 TO JULY 22, 1960	SINGLE BENDIX SENSOR AND "DISTANCE GONE" INDICATOR (NO COMPUTER)	ONE BOEING 707	110 (41 FLIGHTS ANALYZED)	-	+ 2.75% - 0.5%	1.33%	± 3 1/4%	PILOTS CORRECTED HEADINGS MANUALLY TO COMPENSATE FOR THE DRIFT ANGLE INDICATED BY THE SENSOR INDICATOR. "95%" FIGURES INDICATE ERRORS NOT EXCEEDED ON 95% OF THE FLIGHTS.
PHASE B	ACCURACY TESTS OF BASIC SENSOR-COMPUTER-COMPASS SYSTEM.	SEPT. 22, 1960 TO FEB. 15, 1961	DUAL DOPPLER SENSORS, DUAL NAVIGATION COMPUTERS AND SPECIALLY CALIBRATED COMPASSES	ONE BOEING 707	34	± 0.47%	± 1.2% *	+ 0.88% (± 0.5°)	+ 2.2% * (± 1.26°)	COMPUTERS NOT RESET AND HEADINGS NOT CORRECTED ENROUTE.
PHASE C	TO TEST TOTAL SYSTEM ACCURACY AND RELIABILITY IN ENVISIONED AIR TRAFFIC CONTROL AND GROUND BASED RADIO NAVIGATION AID ENVIRONMENT.	MAY 12, 1961 TO JUNE 12, 1961	DUAL DOPPLER SENSORS, DUAL NAVIGATION COMPUTERS AND SPECIALLY CALIBRATED COMPASSES.	NINE BOEING 707'S	64	± 0.54%	± 1.28% *	± 0.65%	+ 1.69% * (± 0.96°)	ON 38 OF THE 64 FLIGHTS THE COMPUTERS WERE RESET AND HEADINGS CORRECTED ENROUTE AS INDICATED BY CONSOL OR LORAN FIXES TAKEN APPROXIMATELY 600 TO 900 NAUTICAL MILES OFFSHORE.

* GAUSSIAN 2σ VALUES

Fig. 11. Doppler radar flight test summary.

CHANGES: COD 69°W Revised to 68°W.

ENTRY/EXIT POINTS	TO LAT	DIST	MAG TRK E / W
BELLE ISLE (52° 03'N 55° 08'W) T O 50° W AND			
	57	350	066.5/246.5
	56	303	071.5/251.5
DISTANCES	55	259	080.0/260.0
B.I. - Heath Pt. 298 NM	54	223	091.5/271.5
Heath Pt. - QM 220 NM	53	200	106.5/286.5
B.I. - QM 510 NM	52	193	123.5/303.5
B.I. - JT 241 NM	51	206	140.0/320.0
CAPE HARRISON R/R (54° 46'N 58° 26'W) T O 60° W AND	62	507	071.0/251.0
	61	458	074.0/254.0
	60	415	078.5/258.5
	59	374	085.0/265.0
	58	339	092.0/272.0
	57	312	101.0/281.0
	56	293	111.0/291.0
	55	286	122.5/302.5
	54	294	133.5/313.5

ENTRY/EXIT POINTS	TO LAT	DIST	MAG TRK E / W
FORT CHIMO "VP" (58° 02'N 68° 32'W) T O 60° W AND	65	485	073.5/253.5
	64	436	076.5/256.5
	63	392	081.5/261.5
	62	348	087.0/267.0
	61	316	095.0/275.0
	60	289	104.5/284.5
DISTANCES	59	274	116.0/296.0
VP-UL-KIDL 1088 NM	58	270	128.0/308.0
VP-PQI-KIDL 1185 NM	57	280	140.0/320.0
GANDER VOR (48° 53'N 54° 32'W) T O 50° W AND	57	513	052.0/232.0
	56	455	054.0/234.0
	55	401	057.0/237.0
	54	348	062.0/242.0
	53	298	066.5/246.5
	52	252	074.5/254.5
	51	214	085.5/265.5
	50	188	099.5/279.5
DISTANCES	49	178	118.5/298.5
CYQX-QM-KIDL 1003 NM	48	188	136.5/316.5
CYQX-QI-KIDL 990 NM	47	214	151.5/331.5

Fig. 12. A portion of the entry - exit point track and distance tables. These tables are used to determine magnetic course and distance from landfall points to adjacent latitude - longitude defined points.

45°N - 49°N

FROM	TO	DIST	60W TO 50W / 50W TO 60W	50W TO 40W / 40W TO 50W	40W TO 30W / 30W TO 40W	30W TO 20W / 20W TO 30W	20W TO 10W / 10W TO 20W
45°N	50	504	081.5/334.0	081.0/334.0	078.0/330.5	073.0/326.0	068.0/321.0
	49	474	087.5/327.5	086.5/327.5	083.5/324.0	079.0/319.5	074.0/314.5
	48	450	093.5/320.5	093.0/320.5	090.5/317.0	086.0/312.5	081.0/307.5
	47	434	100.5/312.5	100.5/312.5	097.5/309.5	093.0/305.0	088.0/300.0
	46	424	108.0/304.5	108.0/304.5	105.0/301.5	100.5/297.0	096.0/292.0
	45	424	116.0/296.0	116.0/296.0	113.0/293.0	108.5/288.5	104.0/284.0
	44	433	123.5/287.5	123.5/287.5	121.0/285.0	116.5/280.5	112.0/275.5
	43	448	130.5/279.5	131.0/279.5	128.0/277.0	124.0/272.5	119.0/268.0
	42	471	137.0/272.0	138.0/272.0	135.0/270.0	131.0/265.5	126.0/261.0
	41	500	143.5/265.5	144.0/266.0	141.0/263.5	137.0/259.0	132.5/254.5
	40	534	148.5/260.0	149.0/260.5	146.5/258.5	142.0/254.0	137.5/249.0
46°N	51	500	082.0/335.5	080.5/335.0	077.5/331.5	073.0/327.0	070.0/321.5
	50	468	087.5/329.0	086.5/329.0	083.5/325.0	079.0/320.5	074.0/315.5
	49	444	094.0/321.5	093.0/321.5	090.0/318.0	085.5/313.5	080.5/308.5
	48	426	101.5/313.5	101.5/313.5	097.5/310.0	093.0/305.5	088.0/300.5
	47	417	109.0/305.0	108.5/305.0	105.5/302.0	101.0/297.5	096.0/292.5
	46	417	116.5/296.5	116.5/296.5	113.5/293.5	109.0/289.0	104.0/284.0
	45	426	124.5/288.0	124.0/288.0	121.5/285.0	117.0/280.5	112.0/276.0
	44	440	132.0/280.0	131.5/280.0	129.0/277.0	124.5/273.0	119.5/268.0
	43	464	138.5/272.5	138.5/272.0	136.0/270.0	131.5/265.5	126.5/261.0
	42	494	144.5/266.0	144.5/266.0	142.0/263.5	137.5/259.0	133.0/254.5
	41	528	150.0/260.0	150.0/260.0	147.0/258.0	143.0/253.5	138.0/249.0

Fig. 13. A portion of the enroute track and distance tables used for determining course angles and segment distances between latitude - longitude defined points in the North Atlantic area.

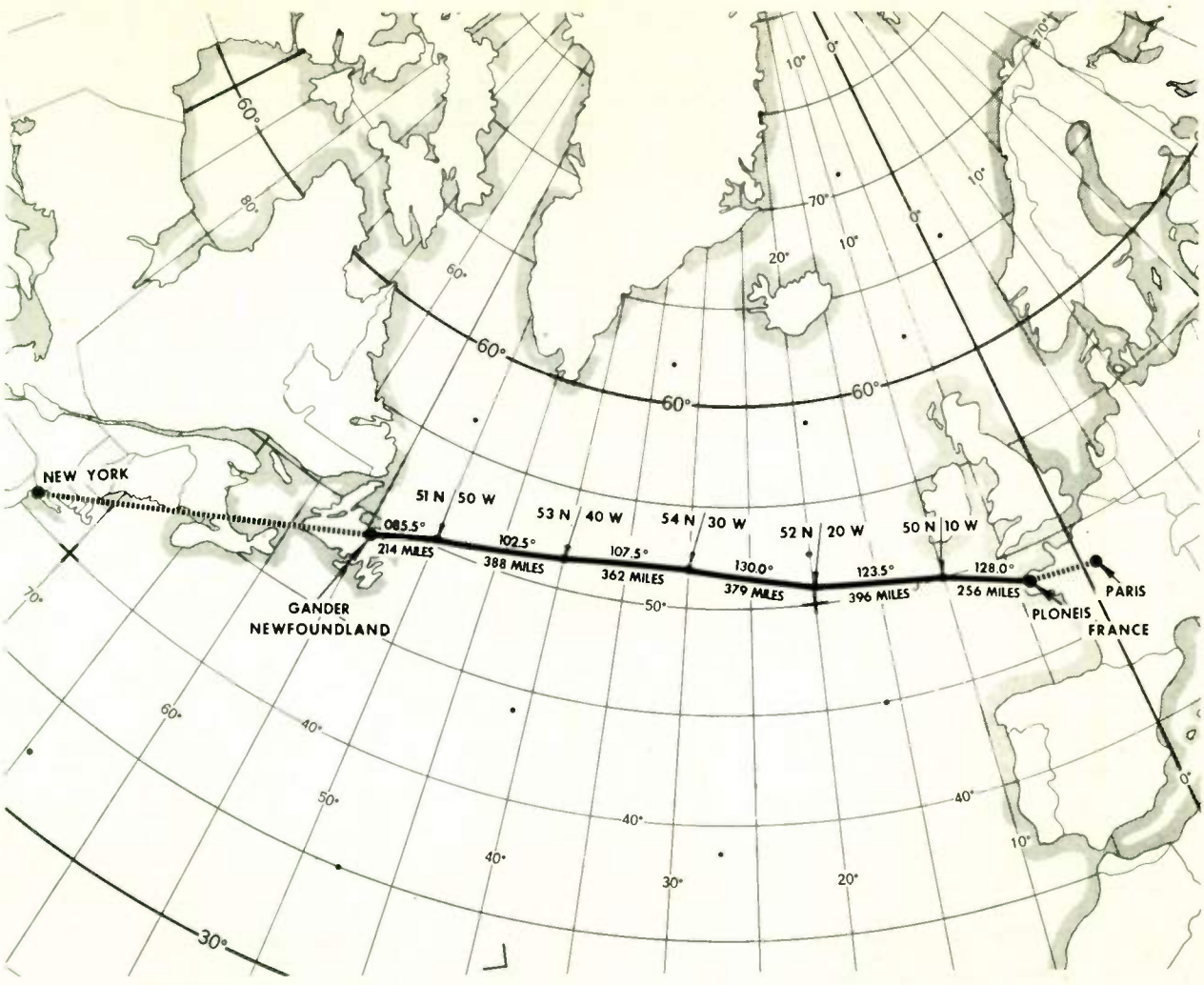


Fig. 14. Typical North Atlantic flight path showing magnetic course and distance (nautical miles) for each flight segment.

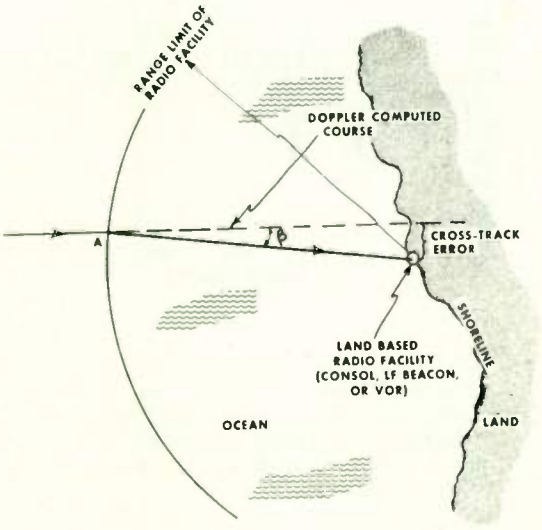


Fig. 15. Method for determining Doppler system cross track errors after a transatlantic flight.

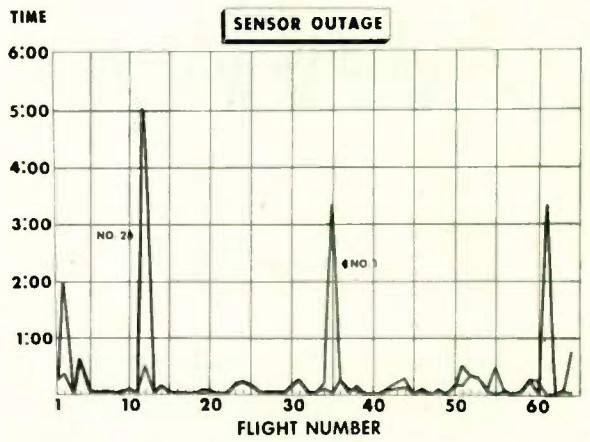


Fig. 16. Doppler Sensor Outages - Phase C flights - caused by insufficient return signal or by equipment malfunction. Dashed line shows Number 1 sensor outages, solid line shows Number 2 sensor outages.

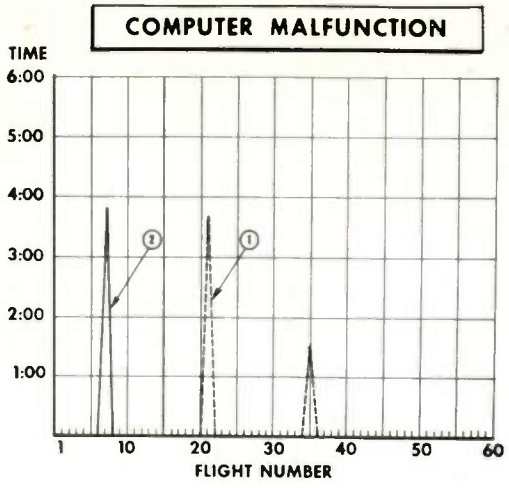


Fig. 17. Computer Malfunctions on Phase C Flights for Number 1 (dashed line) and Number 2 (solid line) computers.

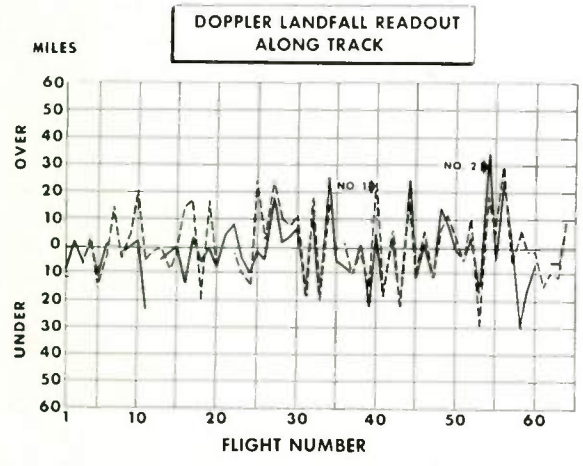


Fig. 18. Along track errors at landfall - Phase C flights for Number 1 (dashed line) and Number 2 (solid line) systems. Total flight length for 59 of the 64 flights averaged 1900 nautical miles. Total flight length for remaining five flights averaged 2730 nautical miles.

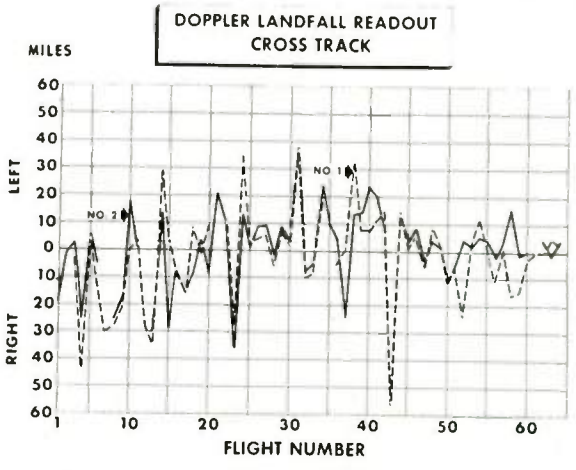


Fig. 19. Cross track errors at landfall - Phase C flights for Number 1 (dashed line) and Number 2 (solid line) systems. Total flight length for 59 of the 64 flights averaged 1900 nautical miles. Total flight length for remaining five flights averaged 2730 nautical miles.

PROPAGATION-OF-ERROR EQUATIONS FOR AIRBORNE DOPPLER NAVIGATION

Nathaniel Braverman
Federal Aviation Agency, NAFEC
Atlantic City, N. J.

Nathan Marchand
Marchand Electronics Laboratories
Greenwich, Conn.

Summary

The general equations for the propagation of error with distance traveled for the case of fix-corrected doppler navigation along a path which consists of one or more rhumb-line legs are derived and shown to consist of a combination of six types of error variances. The specific expressions when the position finding system used for fix correction is either a rho-theta (i.e. VORTAC) or a hyperbolic (i.e. Loran) type are given. The expressions for the resulting error spreads in terms of RMS or arbitrary x and y components are shown to be straightforward and easy to apply in most cases. The exact expressions for the size and spatial orientation of the elliptical equi-probability error contours are shown to be more complex and more difficult to apply practically. However, it is pointed out that the RMS error specification method is easily applicable in all of the practical cases considered. The use of the RMS error circle approximation is always safe although in some cases it may lead to "overdesign."

Introduction

The purpose of this paper is to derive the general equations for the position-output error variances and for the two dimensional distribution at any point along a rhumb line or multiple rhumb-line course when employing airborne doppler navigation. In addition to the errors accumulated because of the characteristics of the doppler sensor, the navigation computer, and the heading reference, the errors inherent in the starting fix, fix insertion, doppler computer resetting and readout will be incorporated into the mathematical model. However, it is possible to express the distribution of error at any point in a number of ways; as rectangular coordinate variances, as a root-mean-square variance or as the elliptical equi-probability error contour. In the case of a single-rhumb-line course, all three methods will be derived and discussed. Useful equations will be obtained for the multiple-rhumb-line case but the equi-probability error contour, because of its complexity and doubtful usefulness, will not be derived but its derivation will be indicated.

The results obtained in an analysis of the error standard deviation for doppler navigation[1] by Marchand will be used as a starting point in the derivation of the mathematical model. In that paper it was demonstrated that when the doppler navigation equipment is flown over a straight-line course the variance (the square of the standard deviation) of the position output error accumulated with distance consists of two parts; one dependent upon the distance traveled and the other dependent upon the square of the distance traveled.

Definition of Errors

K = the error (or error rate). All K's refer to a single flight or flight leg.

$K_{ac\alpha}$ = the along course error rate (in miles per mile flown) for the integrated error which is proportional to the distance flown.

$K_{ac\beta}$ = the along course error rate (in miles per square root of miles flown) for the integrated error which is proportional to the square root of the distance flown.

$\sigma_{ac\alpha}$ = the along course error rate standard deviation (in miles per mile flown) for those integrated errors which are proportional to the distance flown.

$\sigma_{ac\beta}$ = the along course error rate standard deviation (in miles per square root of miles flown) for those integrated errors which are proportional to the square root of the distance flown.

Similarly the cross course error rates will be designated by the subscript cc so that the two error rates will be $K_{cc\alpha}$ and $K_{cc\beta}$ and the two standard deviations will be $\sigma_{cc\alpha}$ and $\sigma_{cc\beta}$. (All variances will be designated by σ^2.)

$K_{ac\phi 1}$ = the along course error (in miles) introduced by the fix source. It is independent of distance flown, is introduced at the fix correction point, and is effective until the next fix is introduced.

$K_{ac\phi 2}$ = the along course error (in miles) introduced when the fix correction is set into the computer. It is independent of distance flown and is effective until the next fix is introduced into the computer.

$K_{ac\phi 3}$ = the along course error (in miles) introduced when setting (or resetting) the

doppler computer. It is independent of the distance flown and is introduced each time the computer is reset to a new "stage" (change in rhumb-line course). $K_{ac\phi 3}$ includes the combined effects of the error in translating a computer reading with respect to one rhumb line into a value to be used with reference to the next rhumb line and the error made in inserting this value into the computer.

$K_{ac\phi 4}$ = the along course readout error (in miles). It is independent of distance flown and is introduced at the point where the accumulated positional error is being evaluated (whenever a position reading is taken).

$K_{cc\phi 1}$, $K_{cc\phi 2}$, $K_{cc\phi 3}$, $K_{cc\phi 4}$ = the cross course fix source, fix introduction, computer resetting and readout errors respectively (in miles).

$\sigma_{ac\phi 1}$, $\sigma_{ac\phi 2}$, $\sigma_{ac\phi 3}$, $\sigma_{ac\phi 4}$, $\sigma_{cc\phi 1}$, $\sigma_{cc\phi 2}$, $\sigma_{cc\phi 3}$, $\sigma_{cc\phi 4}$ = the error standard deviations (in miles) of the $\phi 1$, $\phi 2$, $\phi 3$ and $\phi 4$ type errors in the along and cross course directions.

It will be assumed that the starting fix is in error by the amounts S_x and S_y with normal probability distributions whose standard deviations are σ_{sx} and σ_{sy}. In order to completely define the error distribution of the starting fix it will be necessary to specify the correlation function between the x and y starting fix errors and this will be noted as $\rho_{S_{x,y}}$.

Derivation of Equation for Single Rhumb Line

A diagram of the i'th flight over a rhumb-line course of length L and at an angle of θ to the x axis is shown in Figure 1. The subscript indicates that the results are for a single flight, the i'th flight of a population of n flights. The final errors for this flight, between the actual and indicated end positions, are given by Δx_i and Δy_i as shown in the figure.

The value of Δx_i is obtained by summing up all the components of error along the x direction so that

$$\Delta x_i = \left[K_{ac\alpha i}L + K_{ac\beta i}\sqrt{L} + K_{ac\phi 1i} + K_{ac\phi 2i} + K_{ac\phi 3i} + K_{ac\phi 4i}\right]\cos\theta - \left[K_{cc\alpha i}L + K_{cc\beta i}\sqrt{L} + K_{cc\phi 1i} + K_{cc\phi 2i} + K_{cc\phi 3i} + K_{cc\phi 4i}\right]\sin\theta \quad (1)$$

where the subscript i on each of the K's designates the error for the i'th flight.

Similarly

$$\Delta y_i = \left[K_{cc\alpha i}L + K_{cc\beta i}\sqrt{L} + K_{cc\phi 1i} + K_{cc\phi 2i} + K_{cc\phi 3i} + K_{cc\phi 4i}\right]\cos\theta + \left[K_{ac\alpha i}L + K_{ac\beta i}\sqrt{L} + K_{ac\phi 1i} + K_{ac\phi 2i} + K_{ac\phi 3i} + K_{ac\phi 4i}\right]\sin\theta \quad (1a)$$

If enough of a population is considered, both Δx and Δy will tend to have zero mean values. Thus zero mean values will be assumed. To assume otherwise would be to assume that a significant universal natural, manufacturing or calibration phenomenon exists which will affect different manufacturers or users in the same undetected manner. No such significant systematic phenomenon is suspected in the system under analysis. In addition it will also be assumed that there is no correlation between the errors of the starting fix and the errors encountered in the use of the doppler equipment as well as no correlation between the various types of errors as separated in equations 1 and 1a (except for the components of the starting fix). Thus the starting fix correlation coefficient, $\rho_{\phi 1}$, it will be assumed, can exist.

The variance σ_x^2 of the error at the end position is given by

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\Delta x_i^2 \quad (2)$$

Substituting for Δx_i from equation 1 into 2 and performing the summation

$$\sigma_x^2 = \left[\sigma_{ac\alpha}^2 L^2 + \sigma_{ac\beta}^2 L + \sigma_{ac\phi 1}^2 + \sigma_{ac\phi 2}^2 + \sigma_{ac\phi 3}^2 + \sigma_{ac\phi 4}^2\right]\cos^2\theta + \left[\sigma_{cc\alpha}^2 L^2 + \sigma_{cc\beta}^2 L + \sigma_{cc\phi 1}^2 + \sigma_{cc\phi 2}^2 + \sigma_{cc\phi 3}^2 + \sigma_{cc\phi 4}^2\right]\sin^2\theta - 2\rho_{\phi 1}\sigma_{ac\phi 1}\sigma_{cc\phi 1}\sin\theta\cos\theta \quad (3)$$

Similarly, solving for the variance σ_y^2 at the end position

$$\sigma_y^2 = \left[\sigma_{cc\alpha}^2 L^2 + \sigma_{cc\beta}^2 L + \sigma_{cc\phi 1}^2 + \sigma_{cc\phi 2}^2 + \sigma_{cc\phi 3}^2 + \sigma_{cc\phi 4}^2\right]\cos^2\theta + \left[\sigma_{ac\alpha}^2 L^2 + \sigma_{ac\beta}^2 L + \sigma_{ac\phi 1}^2 + \sigma_{ac\phi 2}^2 + \sigma_{ac\phi 3}^2 + \sigma_{ac\phi 4}^2\right]\sin^2\theta + 2\rho_{\phi 1}\sigma_{ac\phi 1}\sigma_{cc\phi 1}\sin\theta\cos\theta \quad (3a)$$

The starting fix errors are described by the error standard deviations σ_{sx} and σ_{sy} as well as the correlation coefficient $\rho_{S_{x_i}y}$. From the geometry of Figure 1 it will be noted that the error S_{xi} can be resolved into two components, $K_{ac\phi 1i}\cos\theta$ and $K_{cc\phi 1i}\sin\theta$. Therefore

$$\sigma^2_{sx} = \sigma^2_{ac\phi 1}\cos^2\theta + \sigma^2_{cc\phi 1}\sin^2\theta - 2\rho_{\phi 1}\sigma_{ac\phi 1}\sigma_{cc\phi 1}\sin\theta\cos\theta \quad (4)$$

and similarly

$$\sigma^2_{sy} = \sigma^2_{cc\phi 1}\cos^2\theta + \sigma^2_{ac\phi 1}\sin^2\theta + 2\rho_{\phi 1}\sigma_{ac\phi 1}\sigma_{cc\phi 1}\sin\theta\cos\theta \quad (4a)$$

Substituting 4 into 3

$$\sigma^2_x = \sigma^2_{sx} + \sigma^2_{ac\phi 2}\cos^2\theta + \sigma^2_{cc\phi 2}\sin^2\theta + \left[\sigma^2_{ac\alpha}L^2 + \sigma^2_{ac\beta}L + \sigma^2_{ac\phi 3}\right]\cos^2\theta + \left[\sigma^2_{cc\alpha}L^2 + \sigma^2_{cc\beta}L + \sigma^2_{cc\phi 3}\right]\sin^2\theta + \sigma^2_{ac\phi 4}\cos^2\theta + \sigma^2_{cc\phi 4}\sin^2\theta \quad (5)$$

and similarly

$$\sigma^2_y = \sigma^2_{sy} + \sigma^2_{cc\phi 2}\cos^2\theta + \sigma^2_{ac\phi 2}\sin^2\theta + \left[\sigma^2_{cc\alpha}L^2 + \sigma^2_{cc\beta}L + \sigma^2_{cc\phi 3}\right]\cos^2\theta + \left[\sigma^2_{ac\alpha}L^2 + \sigma^2_{ac\beta}L + \sigma^2_{ac\phi 3}\right]\sin^2\theta + \sigma^2_{cc\phi 4}\cos^2\theta + \sigma^2_{ac\phi 4}\sin^2\theta \quad (5a)$$

The first three terms of equations 5 and 5a represent the error of the starting fix and its insertion into the computer. The fourth and fifth terms represent the error accumulated in traveling a rhumb-line course of distance L and angle θ to the x axis. The sixth and seventh terms represent the errors encountered in readout of the final position.

The actual length of error for any flight i, $\Delta\tau_i$, is given by

$$\Delta\tau_i = \sqrt{\Delta x_i^2 + \Delta y_i^2} \quad (6)$$

The RMS variance is noted as σ^2_{rms} and is obtained from

$$\sigma^2_{rms} = \frac{1}{n}\sum_{i=1}^{n}\Delta\tau_i^2 \quad (7)$$

Substituting 1, 1a and 6 into 7 and simplifying

$$\sigma^2_{rms} = \sigma^2_{srms} + \sigma^2_{ac\phi 2} + \sigma^2_{cc\phi 2} + \sigma^2_{ac\alpha}L^2 + \sigma^2_{ac\beta}L + \sigma^2_{ac\phi 3} + \sigma^2_{cc\alpha}L^2 + \sigma^2_{cc\beta}L + \sigma^2_{cc\phi 3} + \sigma^2_{ac\phi 4} + \sigma^2_{cc\phi 4} \quad (8)$$

where $\sigma^2_{srms} = \sigma^2_{sx} + \sigma^2_{sy}$

In order to determine the elliptical equi-probability contour it is necessary to derive the coordinate correlation coefficient, $\rho_{x,y}$, for the end position. It is given by

$$\rho_{x,y} = \frac{1}{n\sigma_x\sigma_y}\sum_{i=1}^{n}\Delta x_i\Delta y_i \quad (9)$$

Substituting for Δx_i and Δy_i from equations 1 and 1a and simplifying, using the assumption that there is no correlation between the starting fix errors and the doppler errors as well as no correlation between the doppler errors themselves as separated into α, β and the various ϕ errors,

$$\rho_{x,y} = \frac{1}{\sigma_x\sigma_y}\left[\rho_{S_{x,y}}\sigma_{sx}\sigma_{sy} + \frac{1}{2}\left(\sigma^2_{ac\phi 2} + \sigma^2_{ac\alpha}L^2 + \sigma^2_{ac\beta}L + \sigma^2_{ac\phi 3} + \sigma^2_{ac\phi 4} - \sigma^2_{cc\phi 2} - \sigma^2_{cc\alpha}L^2 - \sigma^2_{cc\beta}L - \sigma^2_{cc\phi 3} - \sigma^2_{cc\phi 4}\right)\sin 2\theta\right] \quad (10)$$

The correlation coefficient $\rho_{x,y}$ and the variances σ^2_x and σ^2_y completely describe the error distribution at the end position of a one-rhumb-line flight.

An important basic characteristic of the error distribution is the equi-probability ellipse which is described by the angle which the axis of the ellipse makes with the x coordinate axis and by the magnitudes of its two axes. The angle ω which the axis of the ellipse makes with the x coordinate axis is obtained by rotating the coordinate axes until the correlation coefficient as given in equation 10 is zero. Doing this and solving for ω

$$\omega = \frac{1}{2}\tan^{-1}\frac{2\rho_{x,y}\sigma_x\sigma_y}{\sigma^2_x - \sigma^2_y} \quad (11)$$

Substituting 5, 5a and 10 into equation 11 will give the value of ω for this particular case. To obtain the magnitudes of the two axes of the equi-probability ellipse, the coordinate axes are rotated by the angle ω as given in equation 11 and solved for the values of the variances along the elliptical axes. Calling these new values σ^2_1 and σ^2_2

$$\sigma^2_1 = \frac{1}{2}\left[\sigma^2_{rms} - (\sigma^2_y - \sigma^2_x)\sec 2\omega\right] \quad (12)$$

$$\sigma^2_2 = \frac{1}{2}\left[\sigma^2_{rms} + (\sigma^2_y - \sigma^2_x)\sec 2\omega\right] \quad (12a)$$

where σ_1 makes an angle of ω with the x axis and σ_2 is at right angles to σ_1. Substituting from 5, 5a, 8 and 11 into equations 12 and 12a will give the required magnitudes.

Multiple Rhumb-Line Courses

Now consider the case of a flight consisting of a number of rhumb-line courses, one following the other, from departure point to the point of read-out. Calling the length of the first leg L_1 and its angle to the x axis θ_1, it will reach from the departure point to the beginning of the next leg which will have a length of L_2 and an angle to the x axis of θ_2 and so on. It will be assumed that there are m legs to the course.

For any one flight, taking into account that the $\phi 3$ error is different for each rhumb line, the x error Δx_i (for the i'th flight) will be given by

$$\Delta x_i = S_{xi} + K_{ac\phi 2i}\cos\theta_1 - K_{cc\phi 2i}\sin\theta_1 + \sum_{j=1}^{m}\left[K_{ac\alpha ji}L_j\cos\theta_j + K_{ac\beta ji}\sqrt{L_j}\cos\theta_j + K_{ac\phi 3ji}\cos\theta_j - K_{cc\alpha ji}L_j\sin\theta j - K_{cc\beta ji}\sqrt{L_j}\sin\theta_j - K_{cc\phi 3ji}\sin\theta_j\right] + K_{ac\phi 4i}\cos\theta_m - K_{cc\phi 4i}\sin\theta_m \qquad (13)$$

However, $K_{ac\alpha i}$ and $K_{cc\alpha i}$ are bias error rates and are constant for the whole trip. This means that they will be independent of the leg of the flight, or in other words, will be constant and independent of the number of the rhumb line. It should be noted that $K_{ac\beta ji}\sqrt{L}$ and $K_{cc\beta ji}\sqrt{L}$ are the errors at the end of the j'th leg for the i'th flight which are due to random occurances during that leg. The values for successive legs of the same flight will in general be different. $K_{ac\beta ji}$ and $K_{cc\beta ji}$ are each a member of a specific population, each of which may be described by a standard deviation and mean. The means for each of these populations are zero.

Thus simplifying equation 13 and substituting into equation 2 for the variance[2]

$$\sigma_x^2 = \sigma_{sx}^2 + \sigma_{ac\phi 2}^2\cos^2\theta_1 + \sigma_{cc\phi 2}^2\sin^2\theta_1 + \sigma_{ac\alpha}^2\left(\sum_{j=1}^{m}L_j\cos\theta_j\right)^2 + \sigma_{cc\alpha}^2\left(\sum_{j=1}^{m}L_j\sin\theta_j\right)^2 + \sum_{j=1}^{m}\left[\sigma_{ac\beta}^2L_j\cos^2\theta_j + \sigma_{ac\phi 3}^2\cos^2\theta_j + \sigma_{cc\beta}^2L_j\sin^2\theta_j + \sigma_{cc\phi 3}^2\sin^2\theta_j\right] + \sigma_{ac\phi 4}^2\cos^2\theta_m + \sigma_{cc\phi 4}^2\sin^2\theta_m \qquad (14)$$

Similarly, the variance of the y error, σ_y^2, may be obtained

$$\sigma_y^2 = \sigma_{sy}^2 + \sigma_{cc\phi 2}^2\cos^2\theta_1 + \sigma_{ac\phi 2}^2\sin^2\theta_1 + \sigma_{cc\alpha}^2\left(\sum_{j=1}^{m}L_j\cos\theta_j\right)^2 + \sigma_{ac\alpha}^2\left(\sum_{j=1}^{m}L_j\sin\theta_j\right)^2 + \sum_{j=1}^{m}\left[\sigma_{cc\beta}^2L_j\cos^2\theta_j + \sigma_{cc\phi 3}^2\cos^2\theta_j + \sigma_{ac\beta}^2L_j\sin^2\theta_j + \sigma_{ac\phi 3}^2\sin^2\theta_j\right] + \sigma_{cc\phi 4}^2\cos^2\theta_m + \sigma_{ac\phi 4}^2\sin^2\theta_m \qquad (14a)$$

Summing up the terms $L_j\cos\theta_j$ over the m legs of the course takes into account the sign of $\cos\theta_j$ so that the summation is actually equal to the resultant x distance between departure fix and read-out points. Calling this distance X,

$$X = \sum_{j=1}^{m}L_j\cos\theta_j \qquad (15)$$

Similarly, summing up the terms $L_j\sin\theta_j$ over the m legs of the course is actually equal to the resultant y distance between the departure point and the destination. Calling this distance Y,

$$Y = \sum_{j=1}^{m}L_j\sin\theta_j \qquad (15a)$$

Thus the fourth and fifth terms on the right hand side of equations 14 and 14a can be simplified by substituting X^2 and Y^2 for the summation portions.

(Formula continued in next column)

The terms containing the squares of sine and cosine θ_j have to be handled differently since the effect of the sign of the sine and cosine as summed up is eliminated by the squaring of the terms. Thus calling D_{xj} the square of the x component of the square root of the distance L_j on the x axis

$$\sqrt{D_{xj}} = \sqrt{L_j} \cos \theta_j \tag{16}$$

Now letting D_x be equal to the sum of the squares of these components

$$D_x = \sum_{j=1}^{m} L_j \cos^2 \theta_j \tag{17}$$

Since the effect of sign is eliminated, this term depends upon the total distance traveled. Similarly

$$D_y = \sum_{j=1}^{m} L_j \sin^2 \theta_j \tag{17a}$$

Substituting from equations 15, 15a, 17, and 17a into the equations for variance (equations 14 and 14a) we get

$$\sigma_x^2 = \sigma_{sx}^2 + \sigma_{ac\alpha}^2 X^2 + \sigma_{ac\beta}^2 D_x + \sigma_{cc\alpha}^2 Y^2 + \sigma_{cc\beta}^2 D_y + \sigma_{ac\phi 3}^2 \sum_{j=1}^{m} \cos^2 \theta_j + \sigma_{cc\phi 3}^2 \sum_{j=1}^{m} \sin^2 \theta_j + \sigma_{ac\phi 2}^2 \cos^2 \theta_1 + \sigma_{cc\phi 2}^2 \sin^2 \theta_1 + \sigma_{ac\phi 4}^2 \cos^2 \theta_m + \sigma_{cc\phi 4}^2 \sin^2 \theta_m \tag{18}$$

and

$$\sigma_y^2 = \sigma_{sy}^2 + \sigma_{cc\alpha}^2 X^2 + \sigma_{cc\beta}^2 D_x + \sigma_{ac\alpha}^2 Y^2 + \sigma_{ac\beta}^2 D_y + \sigma_{cc\phi 3}^2 \sum_{j=1}^{m} \cos^2 \theta_j + \sigma_{ac\phi 3}^2 \sum_{j=1}^{m} \sin^2 \theta_j + \sigma_{cc\phi 2}^2 \cos^2 \theta_1 + \sigma_{ac\phi 2}^2 \sin^2 \theta_1 + \sigma_{cc\phi 4}^2 \cos^2 \theta_m + \sigma_{ac\phi 4}^2 \sin^2 \theta_m \tag{18a}$$

Now letting

$$Z^2 = X^2 + Y^2 \tag{19}$$

and

$$D = D_x + D_y \tag{20}$$

so that when σ_x^2 and σ_x^2 are added together to obtain σ_{rms}^2 the following is obtained

$$\sigma_{rms}^2 = \sigma_{srms}^2 + \sigma_{ac\alpha}^2 Z^2 + \sigma_{ac\beta}^2 D + \sigma_{cc\alpha}^2 Z^2 + \sigma_{cc\beta}^2 D + m\sigma_{ac\phi 3}^2 + m\sigma_{cc\phi 3}^2 + \sigma_{cc\phi 2}^2 + \sigma_{ac\phi 2}^2 + \sigma_{ac\phi 4}^2 + \sigma_{cc\phi 4}^2 \tag{21}$$

The term Z is actually the distance between departure fix and readout points. The term D is a fictitious distance which is defined by equations 17, 17a and 20 and has the dimensions of length to the first power.

If it is desirable to obtain the elliptical equiprobability error contours it will be necessary to obtain first the correlation coefficient by substituting into equation 9 and solving for $\rho_{x,y}$. Equations 11, 12 and 12a can then be used to obtain the two dimensional characteristics of the error distribution.

Specific Equations for a Rho-Theta Fix System

Suppose now that the starting fix is obtained from a rho-theta station (such as a VORTAC or radar fix).[3] It is necessary now to obtain the equations for σ_{sx}^2, σ_{sy}^2 and $\rho_{S_x,y}$ so that the previously given equations can be used. The errors encountered in a rho-theta fix system are essentially proportional to the distance between the aircraft and the rho-theta station at the time of position determination. Thus for any one fix where the distance in miles from the aircraft to the rho-theta station is R (See Figure 2)

$$\text{Radial error} = K_{ri} R \tag{22}$$

$$\text{Angular error} = K_{\theta i} R \tag{22a}$$

Thus K_r is the radial error rate in miles per mile of radial distance to the rho-theta station. K_θ is equal to the angular error rate (of θ_s) in radians and is here expressed in miles per mile of radial distance to the station. For small errors, such as are encountered, $K_\theta R$ can be considered to be perpendicular to the radial line from the rho-theta station to the aircraft.

Calling S_{xi} the error in the x direction for the i'th flight and S_{yi} the error in the y direction for the i'th flight

$$S_{xi} = \left[K_{ri} \cos \theta_s - K_{\theta i} \sin \theta_s \right] R \tag{23}$$

$$S_{yi} = \left[K_{ri} \sin \theta_s + K_{\theta i} \cos \theta_s \right] R \tag{23a}$$

Substituting equations 23 and 23a into the equations for the variance (see Equation 2) and assuming that there is no correlation between the distance errors and the angular errors, the variances are

$$\sigma_{sx}^2 = \sigma_r^2 R^2 \cos^2\theta_s + \sigma_\theta^2 R^2 \sin^2\theta_s \qquad (24)$$

and

$$\sigma_{sy}^2 = \sigma_r^2 R^2 \sin^2\theta_s + \sigma_\theta^2 R^2 \cos^2\theta_s \qquad (24a)$$

where σ_r is the standard deviation of the radial error rate (in miles per mile of distance to the station) and σ_θ is the standard deviation of the angular error rate (in miles per mile of distance to the station).

And substituting into equation 9 for the correlation coefficient

$$\rho_{S_{x,y}} \sigma_x \sigma_y = \left[\sigma_r^2 R^2 - \sigma_\theta^2 R^2\right] \sin\theta_s \cos\theta_s \qquad (25)$$

The RMS variance of the fix error, σ_{srms}^2, is obtained by taking the sum of σ_{sx}^2 and σ_{sy}^2. Thus the RMS variance is given by

$$\sigma_{srms}^2 = \sigma_r^2 R^2 + \sigma_\theta^2 R^2 \qquad (26)$$

Specific Equations for a Hyperbolic Fix System

The error distribution around a fix obtained from the intersection of two hyperbolic lines of position obtained from two pairs of stations has been worked out in a report from the Office of the Chief Signal Officer.[4] Because of the complexity of the equations, first the RMS error will be given here. The constants are defined as follows, where the subscript 1 refers to the hyperbola from the first pair of stations and the subscript 2 refers to the hyperbola from the second pair of stations. (The two pairs may have one station in common.)

k_1 (and k_2) is a constant giving the gradient (perpendicular to the hyperbola) of the difference in the time of arrival of the signals from the two stations. It is expressed in miles per microsecond for each hyperbola at the point of intersection.

σ_1 and σ_2 refer to the standard deviations of error (in microseconds) in time difference readings for each hyperbola. However, these are usually equal and in the equations given here they will be assumed equal and will be designated as σ_t (time-difference-measurement error-standard-deviation).

θ_h is the angle of intersection of the two hyperbolas at the fix point. (See Figure 3)

$r_{1,2}$ is the correlation constant between the two readings of time differences taken to determine the fix. In standard Loran (Loran A) $r_{1,2}$ is usually assumed to be zero.

The RMS fix error variance is given by

$$\sigma_{srms}^2 = \frac{\sigma_t^2}{\sin^2\theta_h}\left[k_1^2 + k_2^2 + 2r_{1,2}k_1k_2\cos\theta_h\right] \qquad (27)$$

It is possible to obtain the value of σ_{srms}^2 by making some simple measurements on the hyperbolic chart (such as the Loran chart) at the point of fix.

Using the previous definitions of σ_t and $r_{1,2}$

$$\sigma_{srms}^2 = \frac{\sigma_t^2 \lambda_a^2 (1 - r_{1,2}) + \sigma_t^2 \lambda_b^2 (1 + r_{1,2})}{20{,}000} \qquad (28)$$

which should give the same value as equation 27. λ_a and λ_b are dimensions taken from the Loran chart as shown in Figure 3.

Aligning the x axis so that hyperbola 1 is parallel to the x axis, the rectangular coordinate variances are given by[4]

$$\sigma_{sx}^2 = \sigma_t^2 \frac{k_1^2\cos^2\theta_h + 2r_{1,2}k_1k_2\cos\theta_h + k_2^2}{\sin^2\theta_h} \qquad (29)$$

$$\sigma_{sy}^2 = \sigma_t^2 k_1^2 \qquad (30)$$

$$\rho_{S_{x,y}} = \frac{\sigma_t(k_1\cos\theta_h + r_{1,2}k_2)}{\sin\theta_h\sqrt{\sigma_{srms}^2 - k_1^2\sigma_t^2}} \qquad (31)$$

The values for any other orientation of the coordinate axis may be obtained by a rotation of axis.[5]

Thus, if the starting fix is a hyperbolic fix, the results of either equation 27 or 28 for the RMS distribution, or equations 29, 30, and 31 for the coordinate description of the error distribution of the fix can be used in the doppler-error-propagation equations and the error distribution at the readout position determined.

In the case of Loran A, charts which show contours of RMS-error standard deviation are sometimes available. If these charts include airborne Loran equipment and reading errors they may be used for directly obtaining specific values of σ_{srms}.

Discussion of Errors

The question now arises as to when the various descriptions of the error distribution should be used. Obviously, this depends upon the specific application. The question might be as follows: When two aircraft report their positions (as read from the doppler computer) to be fifty miles apart, what are the chances that they are within five miles of each other? If one of the aircraft is following the other along the same course (the course is assumed aligned along the x axis) then the values of σ_x for each may be used to estimate the probability. If the two aircraft were passing each other on parallel courses which were set up to be fifty miles apart, then the values of σ_y for each would be of interest in answering the above question. If the two aircraft were on courses which intersected each other, then the size and orientation of the equi-probability ellipse for each would be required to best answer the above question.

If the question were: Within what area would you expect to find some specified percentage of all aircraft which report themselves (as read from their doppler computers) to be over a specific reporting point along a given airway? In this case the best description of the error distribution which would answer this question would be that which gave the size and orientation of the equi-probability ellipse.

Describing the error distribution by a circle whose radius is some multiple of σ_{rms} is an approximate method which is usually much easier than using the equi-probability ellipse. The calculations are much less complicated, especially for multiple-rhumb-line cases, and there are no axis directions involved. However, the σ_{rms} circle description is not as accurate as the equi-probability ellipse description, especially when the actual two-dimensional distribution is an ellipse of high eccentricity. The σ_{rms} circle will contain a variable percentage of the errors depending upon the ratio of the major to minor axes of the actual equi-probability ellipse. For instance, a one σ_{rms} circle will contain anywhere from 63.2 to 68.3 percent of the errors. Similarly the $3\sigma_{rms}$ circle will contain anywhere from 99.7% to 99.99% of the errors. In addition, when the actual two-dimensional error distribution is an ellipse with high eccentricity the use of the approximate σ_{rms} circle gives especially pessimistic values of error spread (too high) in the direction of the minor axis of the ellipse. Thus, using the σ_{rms} circle approximation is safe in all cases, although in some cases it may lead to "overdesign." Since it is often impossible to be sure that the errors are normally distributed (especially in the "tails" of the distribution curves), use of approxmations such as the $3\sigma_{rms}$ circle to represent a two-dimensional error distribution can be justified in many cases.

The question may arise as to the applicability of the various equations for error distribution at the readout point since they were arrived at using a planar geometry and since actual flight is over a spherical earth. Because of the nature of the error accumulation process and the fact that actual distance over the earth along the intended course is usually calculated and inserted into the doppler computer for each rhumb line, the error distribution equations for a single-rhumb-line flight give values which are close approximations when rhumb-line distances ordinarily flown are of interest. This is true as long as the coordinate axes are rotated before plotting the error distribution results so that the angle between the x coordinate direction and the rhumb line as plotted on the navigator's chart at the readout point is the same as the angle between them at the start of the rhumb line. There is a second order effect inherent in this procedure, since it does not eliminate the errors of calculation (introduced by assuming a planar earth) insofar as an initial fix error is concerned. However, this error due to using a planar earth model is small because for rhumb-line lengths which are sufficiently long for the shape of the earth to be of significance the initial fix error ordinarily contributes very little to the error distribution at the readout point.

At the end of a long (where the shape of the earth becomes important) flight which was made up of a series of rhumb lines and where the angles between successive rhumb lines were large the use of the equations derived in this paper for the error distribution expressed in terms of the size and orientation of the equi-probability ellipse or in terms of σ_x and σ_y will lead to error. In this case the safe procedure would be to describe the error distribution at the readout point in terms of σ_{rms}.

References

1. Nathan Marchand, "The Evaluation of Error Standard Deviation in the Accuracy Evaluation of the Doppler Navigator," IRE Transactions on Aerospace and Navigational Electronics, vol. ANE-8, pp. 91-94; September, 1961.

2. A. Hald, "Statistical Theory with Engineering Application," John Wiley and Sons, Inc.; New York, N. Y., pp 117-118.

3. N. Braverman, "The Use of VOR/DME for Referencing Doppler Self-Contained Navigation Systems," presented at The International Aviation Research and Development Symposium, Atlantic City, N. J., in April, 1961.

4. Report #ORS-P-23, "The Range Reliability and Accuracy of a Low Frequency Loran System," Office of the Chief Signal Officer, The Pentagon, Washington, D.C., January 1946, by W. Q. Crichlow, J.W. Herbstreit, E. M. Johnson, K. A. Norton, and C. E. Smith, ASTIA No. ATI 52 265.

5. A. Hald, "Statistical Theory with Engineering Application," John Wiley and Sons, Inc., New York, N. Y., pp 596-602.

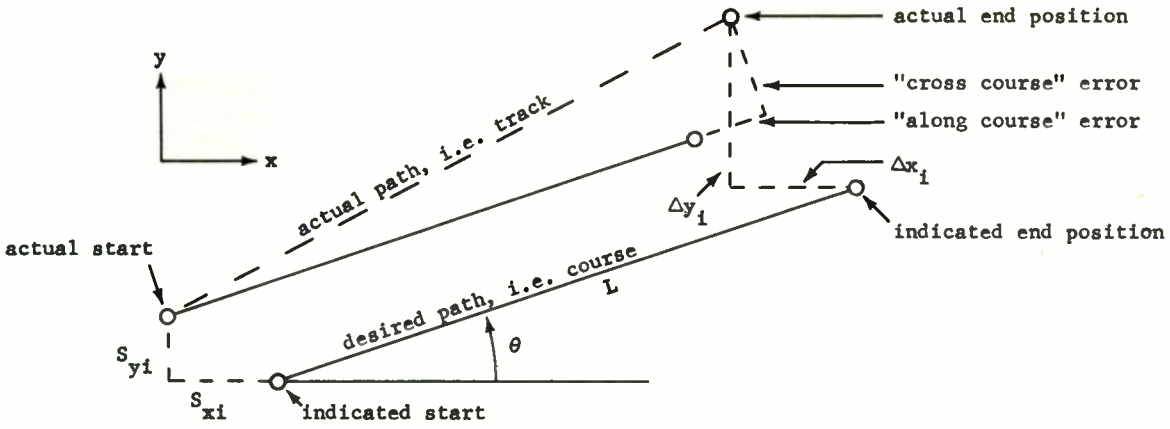


Fig. 1 - Indicated and actual positions of an aircraft for the i'th flight along a rhumb-line course of distance L at an angle θ to the x direction.

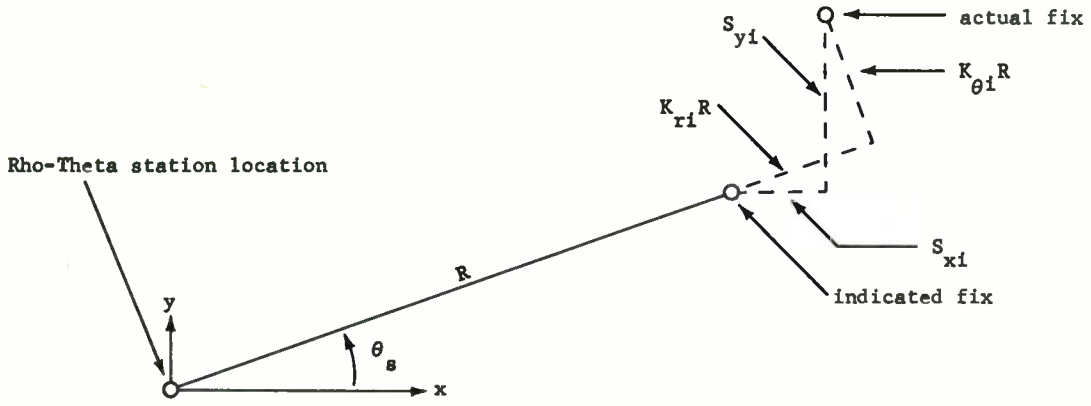


Fig. 2 - The errors in a rho-theta fix at a distance R from the station and at an angle θ_s to the x coordinate axis.

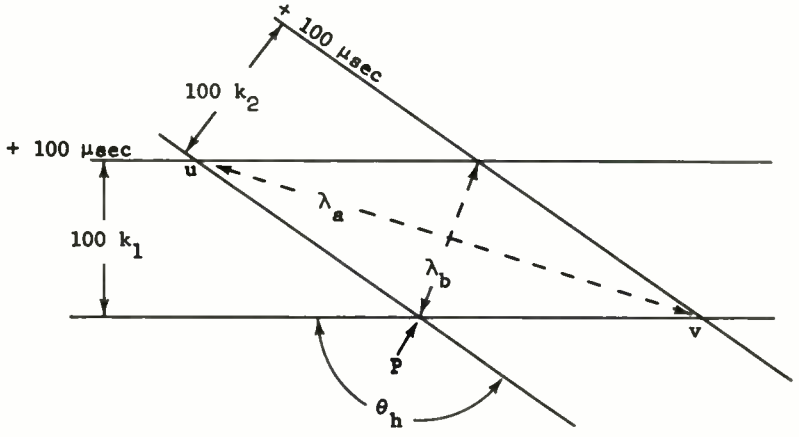


Fig. 3 - Two intersecting hyperbolas at point P and their hyperbolas 100 microseconds away. The distances λ_a and λ_b will determine σ^2_{srms}.

RADAR/TV POSITION LOCATOR

AID TO NAVIGATION

A. Roberts
General Precision, Inc.
Tarrytown, New York

This paper describes a new Radar/TV aid-to-navigation system primarily designed for all-weather marine-traffic operations in congested harbors and along shore lines. In this system a shore-based installation transmits over a TV channel a scan-converted PPI radar image of the harbor, its traffic, buoys, channels, etc., and a unique means for passive identification. Use of the system is voluntary and user-equipment aboard a vessel is nothing more than a standard TV receiver.

The novel feature of this system, which distinguishes it from previously proposed Radar/TV systems, is the means for the user to identify his vessel's "blip" among the other traffic by direct viewing of his TV display.

To the viewer, the TV picture will consist of the scan-converted radar image on which is intermittently superposed a white line of position emanating from a point on the display corresponding to the location of the TV transmitter. Identification is indicated when the white line appears on the display and intersects a blip.

This paper describes how the means for position location is generated at the shore station, the picture on the user's TV display, some considerations for a reduced bandwidth and the growth potential of this system.

Introduction

Since the earliest days of sailing history, marine authorities have been concerned with improving the safety and expediting the flow of marine traffic along coast lines and in ports. In the U.S., aids to marine navigation have come a long way since the early wooden spar-buoys used on the Delaware River, and the Act of Congress in 1789 which established the authority and funds for maintenance and erection of lighthouses. Since then, a diversity of aids have geen developed and applied. Flashing lights and distinctively shaped buoys with color-coding have been built into a comprehensive system of visual aids for the mariner. Bells and horns, provide aid in poor or zero visibility. In recent years, electronic aids, such as radio beacons, Loran, Shoran, etc., supply valuable passive tools for determining position without regard to visibility conditions. Although such electronic aids are important for position fixing, they do not provide a traffic picture.

Marine interests are increasingly concerned with traffic densities and accidents occurring in ports and along coast lines. Ship companies are motivated by economic competition to maintain schedules. There is constant pressure for traffic to continue in safety in all types of weather and visibility conditions. The number of pleasure-boat owners is continually increasing. In the future, the speedy hydrofoil will further complicate the traffic picture in a busy harbor or along a coast line.

For ships at sea, radar is an all-weather device that has been useful in providing both a traffic picture and relative-position fix. However, it has some shortcomings in the harbor environment. The minimum of space does not allow time for interpretation of the radar screen and maneuvering. The clutter-returns, in many cases, interfere with the recognition of close-range targets, and the dim display usually is blurred by ship motions and course changes.

Also, radar is an expensive device for the small-craft operator and requires thorough training in its use and maintenance. Pleasure-boat owners will only accept simple and inexpensive aids.

Some ports are using shore-based radar to survey the harbor and communicate by radio to guide individual vessels along a waterway during poor visibility. This is the case in Rotterdam[1], Southampton[2] and other ports.

For low-density traffic in single channels and at low speeds, this system appears adequate. However, it is not available to a large number of simultaneous users, because each vessel, seeking guidance, must establish and maintain radio contact with a shore-based operator. Also, the basis exists for a legal conflict involving the division of responsibility between the shore-based operator and a ship's captain if an accident occurs.

Although originally proposed for air traffic control[3], Radar/TV systems appear to suit, particularly well, the harbor geography and varied traffic demands existing in U.S. ports and along shore lines. As planned by the U.S. Coast Guard[4], the system would be available to all on a voluntary basis, (as with a lighthouse). It would provide a traffic picture, be easy to use, and inexpensive enough to include the mass of small boat users.

The following sections discuss the advantages of a basic Radar/TV system; the system for passive determination of one's own position on the display; considerations for a reduced bandwidth; and some areas of growth potential for this system.

Basic Radar/TV System

In this system a central shore-based PPI (Plan Position Indicator) radar detects the traffic situation in the harbor. The PPI picture is scan-converted into a television image and is transmitted to all craft that are equipped to pick this information out of the air. Such a system has the following advantages:

1. A TV receiver aboard a vessel is inexpensive when compared with radar (a minimum ratio of 1:10), making it financially feasible to the small-boat owner.

2. The display will be as easy to operate as an ordinary TV set.

3. The transmitted picture can be a high-resolution, bright, stable display of much better quality than that provided by the average ship's radar display. The image appearing on the screen will be a north-up, plan-position display showing all marine traffic, buoys, navigational hazards, shore lines, etc. Moving targets will show wakes. And viewing can be done in daylight.

4. The display will not be affected by a vessel's motion or course change and would be a true-motion, true-bearing presentation.

5. Additional information generated at the shore station can be inserted into the display for transmission.

6. All ships and boats equipped with TV receivers will receive the same display showing all traffic in the area.

7. This is a voluntary system not requiring government — mandatory equipment, procedures, or participation on the part of any vessel owner.

8. Permanent records of all traffic situations taken from the monitor at the shore station can be kept by the U.S. Coast Guard.

However, one important problem that arises with the basic Radar/TV system is the lack of identification of one's own vessel amidst the many radar "blips" on the TV display.

Of course, in good visibility, an observer could correlate the blip position on the display with what he sees out his window. However, with decreasing visibility, it would become more difficult to relate one's own vessel to an obscured environment.

Several methods of identification are available, but all require active cooperation on the part of the operator or vessel. These are:

1. Identification by maneuvering. This is obviously undesirable in the close quarters of a harbor.

2. The use of radar reflectors. However, there is the possibility of confusion with other vessels' reflectors.

3. The use of transponders or Radio Direction Finders. This involves time-consuming operational techniques and extra equipment aboard the vessel, over and above the TV set.

The self-identification and position-location feature described next is a simple and relatively inexpensive addition to the basic Radar/TV shore station. Using an ordinary receiver, the viewer will be able to locate or identify his vessel's blip on the received TV picture.

User's Display and System Operation

Figure 1 is an illustration schematically indicating the received TV display aboard a vessel in a harbor. The harbor outline, each vessel and the buoys are detected by the PPI radar. The PPI scan is converted to a TV scan, and a "line-of-position" cursor is superposed. The composite TV picture is transmitted by a specialized antenna having both omni- and highly directional properties.

The picture on a vessel's TV receiver would show the shore-based PPI display and a superposed recurring "line-of-position" which intersects that vessel's radar blip on the received picture. The line of position would appear as a bright line emanating from the position on the display representing the location of the shore-based TV transmitter. The intersection of the blip with the bright line fixes the position of the vessel.

The insets in Figure 2 show how the displays would look at various receiver locations in the harbor at a given instant. The line of position is superposed on the telecast PPI display and appears on Vessel A's TV picture only. At the same time Vessels B and C are receiving the telecast but not the line of position. The line is shown crossing the center of A's blip. The appearance of the line lasts a fraction of a second on A's blip and will be repeated every few seconds.

Vessel B next receives the line superposed on its display when it is illuminated by the directional section of the TV antenna array. Vessel C in turn will be illuminated by the antenna which finally returns to Vessel A. Throughout the sequence, all vessels receive the total PPI display. The picture that is telecast from the shore-based station will be generated with precision sweeps and in accordance with standard TV practice.

To overcome distortion problems in ordinary receivers, a rectangular grid-overlay can be placed on the display at the shore station. Thus, a vessel with a TV receiver that has non-linear sweeps will distort the grid as well as the picture, and the observer can judge his position relative to surroundings by observing the distorted grid pattern around his vessel's blip. Thus, an ordinary portable TV set will be adequate for receiving the telecast.

Of course, ships that are equipped with precision-sweep TV receivers will assure themselves of high-fidelity reception.

Operation of the shore-based installation can be completely automatic, although, at times, a monitor might be desirable, such as during hazardous weather conditions.

How the System Works

Figure 3 is a block diagram indicating the system configuration. The functional components indicated are integrated into a complete central-station system.

The PPI radar scans the harbor area and detects the land masses, waterways, vessels, buoys, etc. The scan-converter converts the radial scan of the PPI radar directly to a standard television sweep (based on horizontal scanning).

The sweeps from the scan-converter are controlled and the video signals mixed with standard TV synchronizing pulses. The composite signal is led to an RF modulator and prepared for TV transmission in the standard manner on an omni-directional TV antenna.

The identification or position-location feature is achieved by the incorporation of the remaining apparatus shown in Figure 3.

The line of position-display generator is essentially a white line painted on a black background. The whole display rotates, being slaved to a directional rotating antenna which is part of the complete antenna array. As the antenna array rotates, the white line display rotates in synchronism.

A television camera is positioned to view the rotating white line display. This camera is controlled by the same synchornizing circuits that control the radar to TV scan-converter.

In accordance with standard practice, the camera scans the rotating display, one horizontal line at a time. Considering one line, as the dark background is scanned, no video signal is generated. As the white line is scanned across, the video signal increases abruptly and then falls as the line is passed. Thus, it is apparent that the video output of the TV camera is a series of pulses, one for each TV line.

This series of video pulses is led to the control terminal of a gate circuit. The input of the gate is obtained from the output of the RF-modulated TV signal.

Whenever a video pulse from the white line appears, the gate is opened to the passage of the instantaneous RF signal, which is also being continually fed to the omni-directional antenna. The gated RF signal is fed to the directional antenna but before it is radiated it is phase-reversed.

Figure 3 shows one of several possible forms for the antenna system. The omni-directional section consists of two crossed dipoles arranged to form a "turnstile" antenna. These dipoles are excited in phase quadrature with each other to generate a substantially circular radiation pattern. Aligned on either side of the "turnstile" are a series of dipoles which constitute the directional portion of the antenna array. All of these dipoles are excited in phase with each other with equal currents to form two narrow beams.

Figure 4 shows the instantaneous radiation patterns for the omni-directional antenna, the directional antenna and the net radiated pattern. The current fed to the omni-directional antenna is N times the current fed to the directional section, where N is the number of dipoles.

At the instant the vidicon scans the white line of the rotating display, it can be seen that the radiation is of very low intensity, approaching zero in the direction toward which the directional antenna is pointing. Zero RF signal constitutes white on a TV receiver, and at that instant a white dot will appear on the TV receiver. This white dot occurs on a receiver each time the white line is scanned at the central station and when the directional antenna is pointing at the receiver.

Consider now the image produced by the television receiver assumed to be installed aboard a vessel. During the major portion of the revolution of the Antenna system the directional antenna will have no effect on the receiver. The signal picked up is solely that emitted by the omni-directional section of the antenna, and accordingly, the PPI presentation is reproduced on the TV receiver screen.

Now consider what happens when the directional section of the antenna is pointing at the TV receiver. When the dark background is scanned, no video signal is generated, the gate is blocked and the directional antenna is not energized. Accordingly, the PPI presentation is traced out on the TV

receiver during this time. When the white line is scanned, the directional antenna is energized and the net pattern of Figure 4 is created. The TV receiver sees no signal, which causes a white dot to appear on its screen. Successive scan lines cause successive bright spots to appear on the receiver creating a bright line. This is a line of position and appears superposed on the always – present PPI display.

During each revolution of the antenna, two nulls of radiation are directed toward the receiver, and the line of position crosses the field of view twice. If the antenna rotates at 3RPM, there are six cycles of operation per minute and the line of position appears on the receiver screen every ten seconds, persisting while the antenna is effectively pointing toward the receiver. For a 5 degree beamwidth, this will be for approximately $\frac{1}{4}$ sec.

At standard TV frame rates, $\frac{1}{4}$ sec. represents $7\frac{1}{2}$ frames -- more than enough to establish the line in the viewer's mind.

Considerations for Reduced-Bandwidth System

In the foregoing discussion the system bandwidth required was based on the use of a standard unmodified entertainment TV receiver. If one considers the fundamental factors that determine TV bandwidth and allows the flexibility of modifying the TV receiver, then it is feasible to reduce the TV channel bandwidth.

Consider the basic factors involved in reproducing a picture by means of a TV scanning system. The basic equation for the television bandwidth is:

$$f = \frac{cN}{2T_f}$$

f=olectrical frequency in cyles/sec. (bandwidth)

N=total number of black and white elements in the picture

T_f=time to scan one frame

c=a numerical constant determined by the active percentages of line and frame-scanning periods.

For standard entertainment transmission, T_f has been selected at 1/30 sec. or 30 frames/sec. and 2 fields/frame. This frame rate will allow the fast-moving objects to be reproduced and also avoid flicker problems for the viewer. The phosphor on the receiver TV screen is selected on the basis of its persistence properties -- short enough so that the image can be up-dated by the next scan line and long enough to provide some storage of the image between scans.

The number of effective scanning lines nas been selected at about 410. The product of the scanning lines and the horizontal resolution in line numbers determines the number of picture elements transmitted in one frame time and the resolution capability of the system.

Now consider the TV reproduction of the PPI radar image. Let us assume that resolution requirements for harbor traffic remain the same as in standard TV. Thus, the active number of scan lines and horizontal resolutions are retained and the factor cN remains constant.

However, ships in a harbor, and even helicopters, present targets that are moving at relatively slow speeds compared to standard TV image motions. In relation to the time between frames, blip displacements are extremely small. Thus, the frame rate can be considerably reduced without affecting the dynamics of the problem, and if the scanning lines and horizontal resolution remain the same, the resolution is not disturbed.

However, the reduction of frame rate will require a consideration of the display tube to be used.

If the standard TV tube is used, a reduction of frame rate to some lower value will produce flicker.

To an extent, flicker can be avioded by using a long-persistence phosphor on the TV tube. If, in conjunction with a long-persistence phosphor the fields per frame are increased, then further reduction of frame rate is possible without getting into flicker problems. The exact extent of these possibilities for reducing frame rates really requires detailed study and experimentation, but a factor of 4 appears practical. Thus, a 6 megacycle bandwidth would be reduced to 1.5 m.c.

In addition to changing the phosphor on the TV screen, it will also be necessary to change the time constants associated with the vertical sweep circuits in order to accommodate reduced frame rates and a new interlaced scanning.

These modifications will increase the cost of the user's TV receiver, and it will no longer be a standard device that can also be used for entertainment purposes.

In addition to using a standard CRT (Cathode Ray Tube), the use of a storage tube will allow considerable reduction of bandwidth requirements while preserving resolution. However, since storage tubes are expensive, this reduction in bandwidth will result in more expense to the user.

Nevertheless, these tradeoffs might be considered in view of the lack of channel-space in the TV section of the spectrum.

Growth Potential

Important advantages of this system are its flexibility in meeting particular coast-line or harbor requirements and its capability, after initial installation, of evolving to more sophisticated forms in step with increasing traffic densities and changing concepts or port operation.

When applied to coastal navigation and harbor approaches, this system could supplement lighthouse beacons and, in zero visibility, provide a display to ships located a considerable distance from the coast line, possibly in excess of light-beacon range. Each ship would have at its disposal a detailed display of navigation hazards, traffic in the area, and the passive means of position fixing.

A system, once installed, is capable of improvements and refinements especially in relation to central-station displays. For instance, a display could be used with a standard chart as its basic format. This chart could be quite large, detailed, and of high contrast. Superimposed on this chart would be the moving radar detected targets. Other advisory information such as weather could be added to the picture or transmitted on the audio portion of the TV channel. A shipboard computer is being developed for the U.S. Maritime Administration for use with shipboard radars. The computer, using radar data, will track other ships, compute their courses and determine potential conflicts. A shore-based installation of similar function could be used to advise all ships of dangerous traffic situations. Because it is shore-based it could be more accurate, analyze a larger number of ship's courses and have available expert operators and maintenance personnel. All this refinement, possible in a shore-based station, would be expensive if each ship had its own computer. But if shore-based, the benefits would be available to all ships via the telecast at the cost to them of a TV set.

It is conceivable that it may be desirable to interrupt the standard display being telecast to send an enlarged highly detailed picture of a section of the entire harbor picture. The ships involved in a conflict could still identify themselves with the line of position.

If desired to telecast a special display or information to a particular vessel, the rotating line of position could be stopped and held stationary on a vessel. The vessel could be previously informed that when the line is stationary on its display blip, information is for its specific use. Use of the audio channel

could alert all the viewers to check their displays.

Other potential areas of application for this system are for helicopter operations and certain military requirements.

Acknowledgements

Messrs. G. Stavis and S. Slawsky of GPL Division, General Precision, Inc., who are co-inventors of this system.

References

1. "Rotterdam's Shore Based Radar Chain" by the Pilotage, Beacons & Lighthouse Service of The Netherlands.

2. "Port R/T Communication Scheme, Harbour Surveillance Radar," Port Operation and Information Service, Southampton Harbour Board Publication.

3. "Teleran," Electronics, February 1946.

4. Brunner, L. E., Capt. USCG, "U.S. Coast Guard Views on Shore-Based Radar as an Aid to Navigation," Vol. II, Symposium Papers, RTCM April 6, 1961.

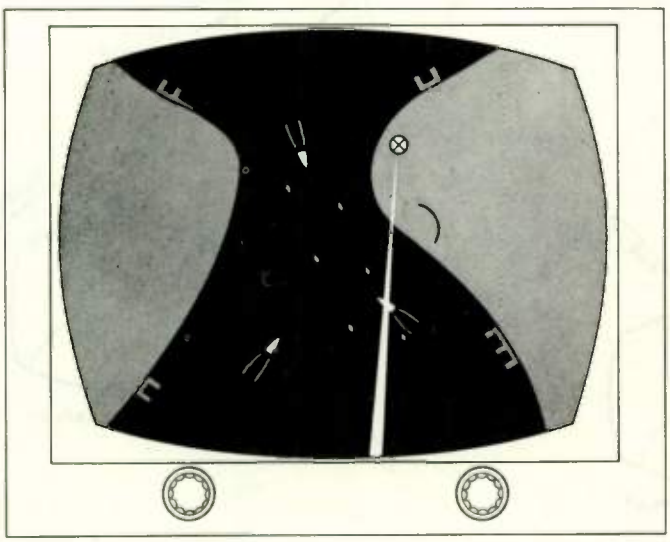


Fig. 1. Vessel's TV display.

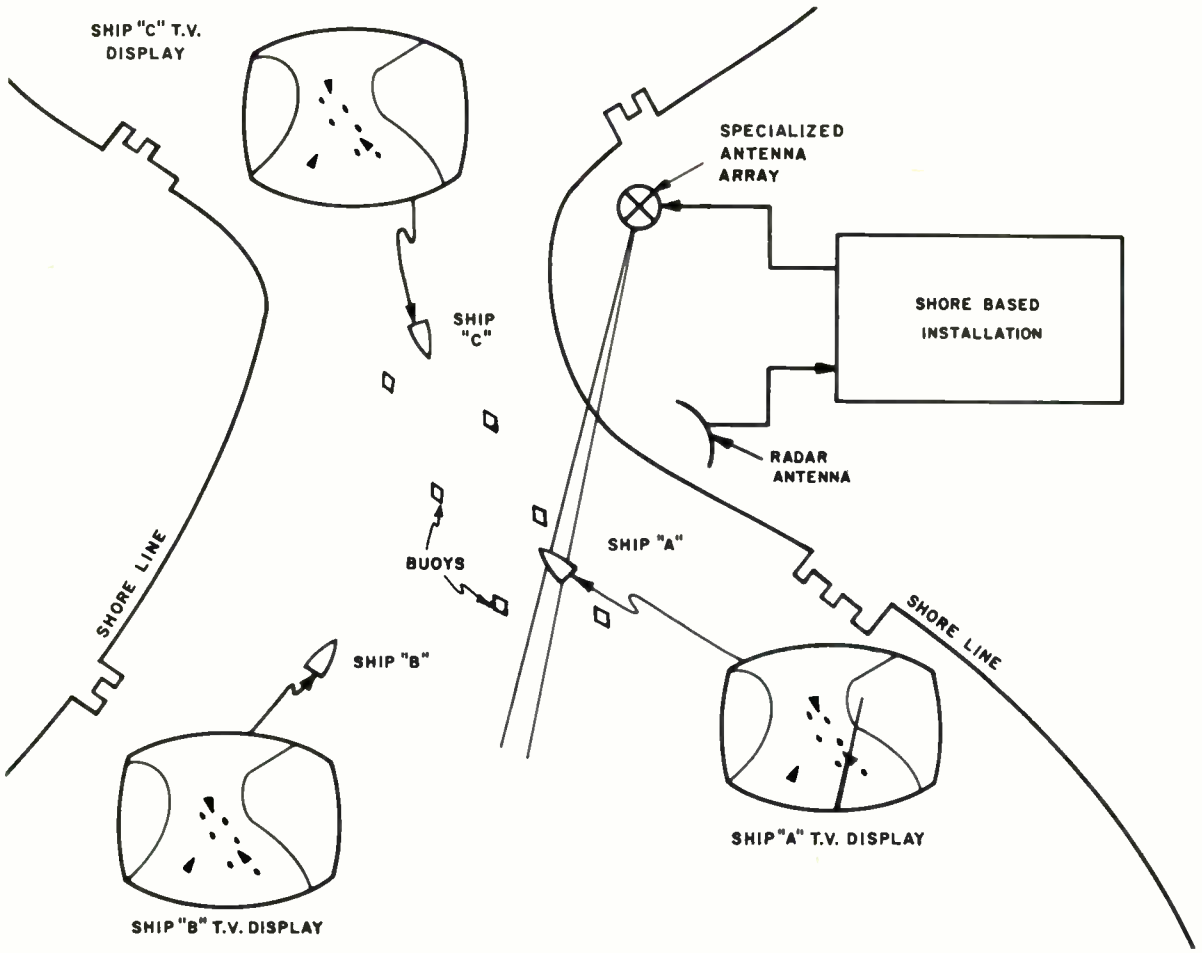


Fig. 2. System description.

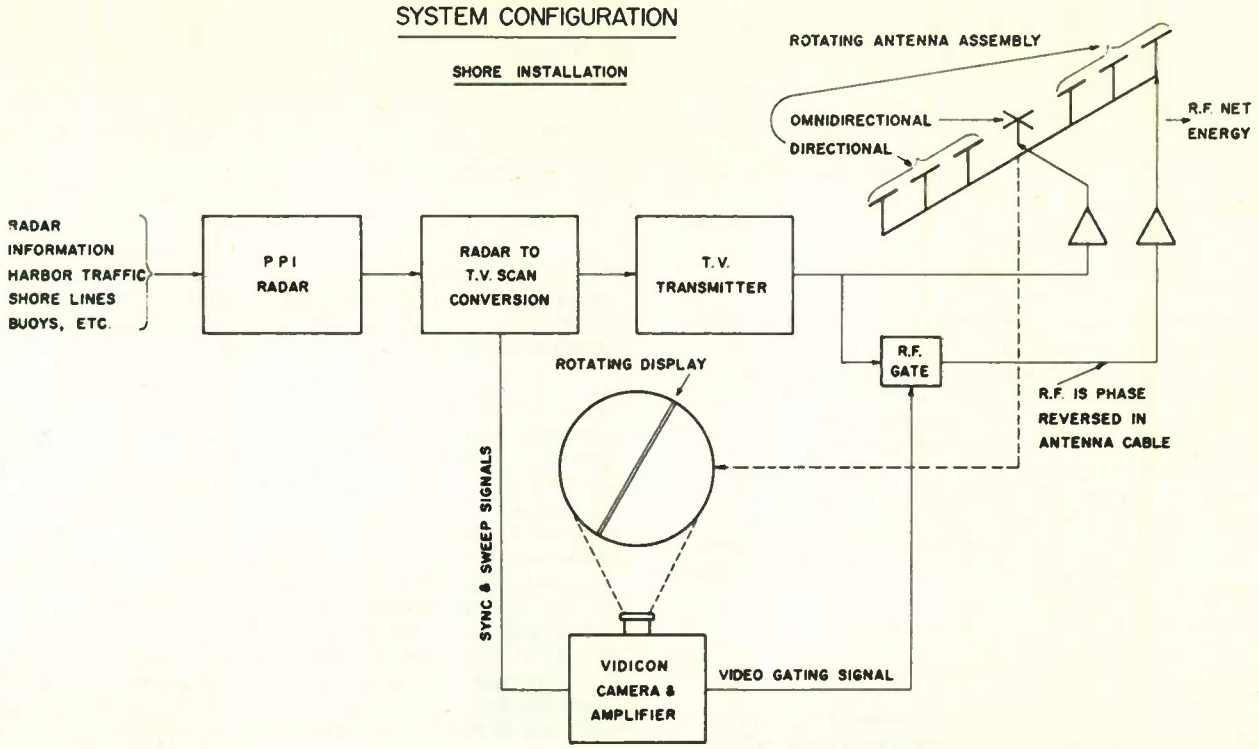
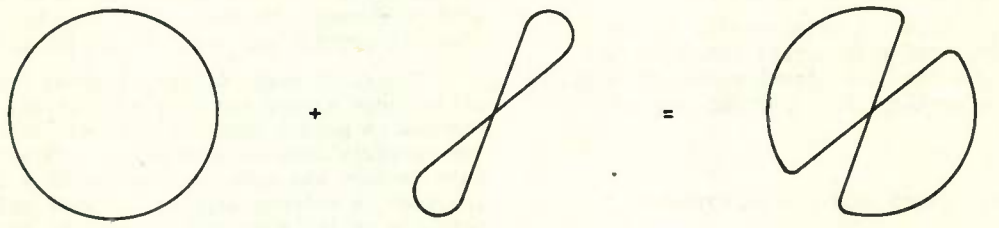


Fig. 3.



OMNI-DIRECTIONAL ANTENNA PATTERN

FIXED DIRECTIONAL ANTENNA PATTERN ✱ (PHASE REVERSED R.F.)

✱ THIS PATTERN EXISTS ONLY WHEN VIDICON IS SCANNING ACROSS WHITE DISPLAY LINE

NET RADIATED PATTERN

Fig. 4. Antenna radiation patterns.

DIRECTIONAL UHF GLIDE SLOPE SYSTEM

E. R. Hollm and J. L. O'Connor

Airborne Instruments Laboratory
A Division of Cutler-Hammer, Inc.
Deer Park, Long Island, New York

Summary

Standard equisignal and null reference glide slope systems are briefly described, with some discussion of the reasons limiting their operational usefulness. To overcome these limitations, the glide slope technique was developed further, and some of these developments are enumerated without, however, evaluating their value as operational systems. AIL has designed an experimental directional UHF glide slope system and its parameters and advantages are delineated. Some of the features to be incorporated in this system, as finally developed, are noted.

Introduction

The Instrument Landing System (ILS), consisting of a glide slope, localizer, and two markers, is based upon developments begun more than 20 years ago. It has been an international standard for 15 years, with about 500 installations presently in use around the world.

Although the ILS has not lived up to its promise as a true all-weather instrument landing system, nevertheless it has provided civil aviation with a relatively simple and reliable low approach system. The accumulated investment in ground and airborne equipment, by Governments and aircraft operators, has tended to perpetuate the ILS's existence, despite its limitations.

This paper offers a brief review of the existing glide slope system, and describes the development of a new directional system by Airborne Instruments Laboratory (AIL) for the Federal Aviation Agency.

Established Glide Slope Systems

Practically all glide slope systems presently in operation can be classified as either the equisignal or null reference type. Both systems use a two-element antenna system, but differ in the method of excitation and antenna position relative to the ground plane.

The equisignal type was the original system. For a 3-degree glide slope angle, the lower antenna is located 4 feet above ground and is excited with a carrier modulated 95 percent at 90 cps. The upper antenna is located 18 feet above ground and is modulated at 150 cps. The resulting field patterns are shown in Figure 1. The lowest intersection of the two patterns defines the desired glide angle, which can be varied by changing the antenna heights.

The null reference type is a later development, which is used in most operating glide slope systems. Its upper antenna is located at an above-ground height that will place the first null at the desired glide angle. This is normally about 29 feet for a 3-degree glide angle. The upper antenna is excited with equal 90 and 150 cps sideband energy only. The lower antenna is located at half the upper antenna height and is excited with 90 and 150 cps modulated carrier energy. The resulting field patterns are shown in Figure 2. The 90 and 150 cps sidebands are phased for an addition and cancellation in space to provide a predominance of 150 cps modulation below and 90 cps modulation above the desired glide angle.

Limitations of Existing Systems

All present-day operational glide slope systems require the ground plane to form a path in space. Irregularities in the ground plane will produce perturbation in the glide slope, since the Fresnel zone changes constantly as the aircraft approaches the runway. At most airports it has been possible to level the ground sufficiently to provide an acceptable glide slope. At some airports, however, it is impossible to achieve acceptable performance because it is economically impractical to provide the required ground plane.

The glide angle of the original equisignal glide slope system was very susceptible to periodic changes in ground plane level, particularly those due to variations in snow cover. This objectionable feature was reduced somewhat by a limited increase in antenna heights. A more effective solution to the snow problem was the development of the null reference glide slope system in which the glide angle is established exclusively by the height of the upper antenna. Since normal snow accumulation represents a relatively small percentage change in the 29-foot null reference antenna height, the resulting shift in the glide angle is correspondingly small.

It should be noted that, though the null reference glide angle is more stable with respect to snow cover, in practice it is actually more susceptible to path bends due to ground irregularities--that is, the Fresnel zone for the higher

null reference antenna extends considerably beyond the normally flat airport property. The practical solution has been to use the null reference system wherever possible, a modified equisignal system where necessary, and no facility where neither one provides a flyable path.

Glide Slope Requirements

An improved glide slope system, fully compatible with existing airborne installations, has been needed for many years. This need is becoming more pressing as new sites fail to meet the level ground requirements. The stringent requirements of modern aircraft, the desire to use approach couplers, and the trend toward lower approach limits dictate a further tightening of path tolerances.

Future requirements will be even more demanding. Three of the foremost landing system developments depend upon the glide slope for initial approach guidance to altitudes below today's limits. As runways are lengthened, less level land normally remains for the glide slope reflection area. Consequently, an increasing number of major airports find it difficult to meet today's relatively broad tolerances on path perturbations.

Other Glide Slope Systems

Many attempts have been made to improve the two existing glide slope systems. The most noteworthy of these are the so-called "M" and "B" arrays, and the Flush-Mounted Directional Glide Slope System. Without discussing the relative merits of these systems, it is, however, significant to note that they have been under development for a number of years and have not yet become generally accepted as operational.

Antenna Height Restrictions

A review of the last 15 years of glide slope development discloses no particular shortage of qualified personnel or financial support. After consideration of the limited progress made in the reduction of path perturbations, it was decided that further development, limited by existing antenna height restrictions, offered little promise of success.

Height restrictions near the edge of the runway are constantly being relaxed. This has been effected by an appreciation of the significant safety advantages provided by modern facilities, outweighing the rather arbitrary assessment of the hazards presented by these facilities when they are located within several hundred feet of the runway.

An ideal directional glide slope antenna array, capable of restricting ground illumination to low-level side lobes, would be more than 100 feet high. This would probably be unacceptable, considering the present tolerance of 50 feet at 400 feet from the runway centerline. Fortunately, an improved system can be designed with a directional array which, though somewhat higher than 50 feet, will actually present a lower obstruction angle since it can be located further from the runway.

Directional Glide Slope System

The basic difference between the directional glide slope system and such systems as the equisignal, null reference, M array, and B array is that the directional system radiates a path independent of the ground. The other four systems depend equally upon the direct and ground-reflected signals to form the path. Dependence upon the ground plane limits the capabilities of the last four systems to overcome path perturbation caused by ground irregularities.

The experimental directional glide slope system that has been fabricated, and is presently undergoing tests, consists of two arrays; one is excited with a 90-cps modulated carrier and the other with a 150-cps modulated carrier. The resulting free-space field patterns are shown in Figure 3. When the arrays are placed above a ground plane, these patterns are modified somewhat, particularly at zero degrees, where the lower lobe is equal to zero.

Basic Array Design

The electrical design of the antenna was initially accomplished at X-band. Use of scale-model antenna techniques allows the rapid evaluation of the antenna design without waiting for full-scale fabrication and testing. The major electrical design objectives of the antenna, as dictated by system requirements, can be summarized as follows:

1.	Elevation beamwidth	3.5 degrees
2.	Elevation side lobes	-20 db or better
3.	Polarization	Horizontal
4.	Cross polarization	-26 db or better
5.	Azimuth beamwidth	60 degrees
6.	Azimuth pattern ripple	1 db or less
7.	Squint angle	-1 degree to +3.5 degrees

The full-scale UHF antenna was to operate over the glide slope band of 329 to 335 Mc. Electrical characteristics and mechanical requirements set by airport location indicated that a slotted-waveguide array would be a suitable antenna type. Control of the slot excitations would allow the required low side-lobe level to be obtained.

The beam direction from a slotted-waveguide array is tilted away from the normal to the array, unless the slot spacing is half a waveguide wavelength. To retain zero and small tilt or "squint" angles over a frequency band, requires control of the phase velocity in the waveguide. This can be achieved by mechanical adjustment of the width or broad dimension of the waveguide, a technique adopted in engineering the glide slope antenna.

Initially, however, on the X-band model, standard waveguide was used and a slot configuration set up to establish the adequacy of the illumination taper in securing the required radiation pattern, squint angle, efficiency, and VSWR. These were found to be a satisfactory basis for further design. An X-band pattern is shown in Figure 4.

To obtain some indication of the effect of ground on the beam pattern, a copper mesh ground plane, 30 feet by 15 feet, was constructed, and the X-band antenna mounted vertically at one end. At the other end, a receiving antenna could be moved in the vertical plane to yield the antenna plus ground-plane pattern.

With the beam axis horizontal, the 3-db beamwidth decreased from 3.5 degrees to 2.65 degrees with some distortion of the beam shape. On elevating the beam, the beam broadened and the distortion was reduced, until, at about 2 degrees elevation, ground-plane effects were almost negligible.

Although the results of the ground-plane measurements on the scale-model antenna were an indication of the performance to be expected from the UHF antenna, they cannot be rigidly applied to accurately predict the UHF ground-plane effects. Differences in reflection of terrain at UHF, and of the copper mesh at X-band, may produce varying effects in the antenna patterns.

Because it had been proposed to provide the full-scale antenna with adjustable waveguide walls capable of being accurately positioned, X-band was regarded as too high a modeling frequency. An adjustable array was therefore constructed at S-band, and its performance investigated. The expected control of the pointing direction of the beam was established. Both side walls of the waveguide were symmetrically moved; moving only one wall produced unacceptable pattern distortion.

For the experimental system, two side-by-side antennas were required. A dual S-band antenna system was studied. This allowed the mutual coupling and shadowing effects to be investigated and minimized. Chokes fitted to the arrays were found to be advantageous in reducing both mutual coupling and cross-polarization, and they assisted in shaping the azimuth patterns. A full pattern analysis of the dual S-band antenna included frequency runs and variation of the scissor angle between the arrays. The data obtained furnished the requisite design information for the full-scale antenna.

Full-Scale Array

The full-scale antenna is constructed in seven sections from fabricated aluminum waveguide having 26 x 13 inch internal rectangle dimensions. Coaxial to waveguide transitions are fitted to the ends of the array. A coaxial load is attached to the upper transition, and the antenna is fed at the bottom transition. The antenna is 60 feet long.

Both side walls of the antenna are movable. Therefore, the antenna is tunable over the glide-slope band, and the squint angle can be adjusted to any desired value at any frequency.

Full-scale pattern measurements of the array, mounted horizontally on a turntable, give a pattern in agreement with the X-band model pattern shown in Figure 4.

Full-Scale Experimental System

The installation of the full-scale dual array system at Peconic River Airport, New York, is shown in Figure 5. This experimental facility permits independent control of the phase, amplitude, squint and tilt angles of the two beams. Tests are being conducted to determine the effect of ground illumination as these parameters are changed. Although ground and flight tests have recently commenced, sufficient data have been gathered to prove that it is feasible to construct a directional glide slope fully compatible with existing airborne equipment and capable of operation without a ground plane.

Design of Operational System

One of the dual array antennas will be modified to provide a single array system as soon as the present series of tests on the experimental system have been completed. The dual array is not being proposed as an operational system. Sufficient flexibility of control can be provided in the single array after the basic data have been gathered from the dual array tests. In addition to cost reduction, the single array will eliminate the carrier cancellation inherent in the dual array because of the 500-degree spacing between the two antennas.

The final single array design will include a modification to the lobe structure shown in Figure 3, to reduce the difference between the 90 and 150 cps radiation. This was found to be permissible, because during initial flight tests it was found that the clearance ratio between the 90 and 150 cps radiation below the path was considerably greater than required. This will result in a reduction in ground effect and will give comparable performance with reduced antenna height requirements.

An additional improvement in the final design will be achieved by maintaining an optimum phase relation between the direct signal and the low-amplitude reflected signal. In practice, it is impossible to maintain the ideal in-phase or out-of-phase condition along the entire path. A method has been devised to provide the optimum phase condition along the most critical section of the approach path. The technique has not been tested in the field, but promises to provide a very substantial reduction in path bends.

Conclusion

An experimental directional glide slope system has been designed and fabricated. The limited flight tests completed to data have proven the feasibility of this system. Effort is continuing on the design of the operational system that will be less expensive and provide superior performance in several respects.

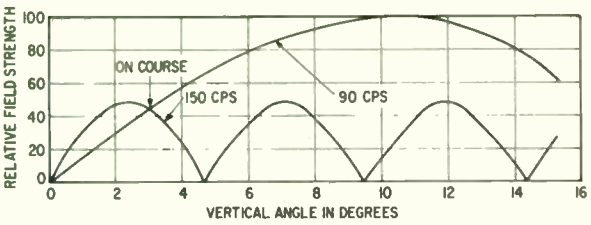


Fig. 1. Equisignal glide path.

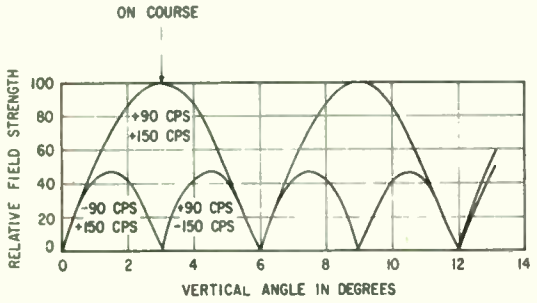


Fig. 2. Null reference glide path.

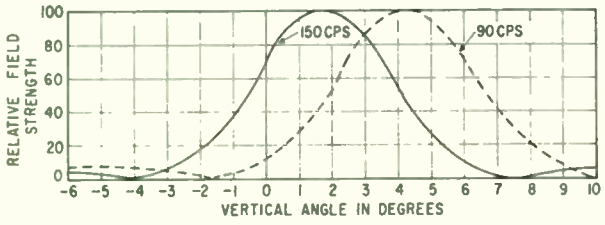


Fig. 3. Directional glide slope free space field patterns.

Acknowledgment

The authors wish to acknowledge the assistance of Messrs. Jack Ramsay, Consultant, and Boris Popovich, Engineer, of the Antenna Section of AIL's Department of Radar Systems, for the design and testing of the antennas used in this project.

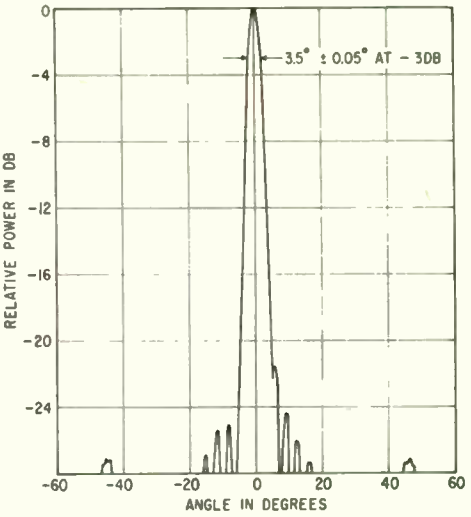


Fig. 4. X-band model elevation pattern.

Fig. 5. Directional glide slope.

DESIGN GOALS, TECHNIQUES, AND TESTS OF A NEW INSTRUMENT LANDING AID

F. H. Battle, Jr.

Airborne Instruments Laboratory
A Division of Cutler-Hammer, Inc.
Deer Park, Long Island, New York

Abstract

A scanning-beam angle measurement technique has been designed to add all-weather capabilities to the standard ILS. One-way transmission and airborne reception of a coded K_u-band fan beam will accurately define the aircraft elevation over the runway. The ground-based equipment, the encoding scheme, and novel internal functions of the receiver are described. Compatibility of the parameters and circuits with a possible system expansion are indicated. The design goals stressed universality of application, high signal quality, and system reliability. Preliminary tests show RMS angular errors less than 0.02 degree, and manually controlled landings have been accomplished.

Introduction

It is generally recognized that routine all-weather landings cannot fully depend on the standard Instrument Landing System (ILS). However, the reliability of ILS for low-approach service is well established, the system is widely installed, and thousands of pilots are proficient in its use. The evolution of all-weather capabilities should therefore depend, not on replacing ILS with new techniques, but on extending and supplementing this mode of guidance.

Since the flare-out maneuver and actual touchdown demand more precise guidance than ILS can provide, additional techniques are required. The nature and use of these techniques should be operationally compatible with ILS, but at the same time technically compatible with a future improvement of basic ILS service.

The scanning-beam technique to be discussed is philosophically similar to ILS, because one-way (ground-to-air) transmission is used rather than radar. It can be introduced as an adjunct to ILS, sharing the same cockpit instruments and requiring no change of piloting procedure. However, many of the technical characteristics were chosen to allow a future expansion of dependence on the new technique, which should prove superior in the basic ILS role as well. The configuration and operational criteria of such a future system have been described.[1] A survey of the state of the art in radio guidance techniques and their usefulness for instrument landing has also been reviewed;[2] this survey and critique led to the choice of the scanning-beam technique discussed here as the basic means of position measurement.

After a brief statement of the operating principles of the technique, its characteristics and specifications will be justified in terms of operational goals. Some novel functions of the electronic equipment will then be described. Finally, the results of preliminary tests will be discussed.

General Description

The proposed means of guiding aircraft to actual landings depends on extremely accurate measurements of angle by reference to coded transmissions on a thin microwave fan beam. Such a beam may be scanned so as to define either elevation or azimuth angles, but the initial use of this technique in instrument landing experiments provides measurements of a single elevation angle. A vertically scanning beam radiated from beside the runway, well behind the standard ILS glide slope transmitter (Figure 1), can extend the ILS mode of guidance through flare-out to the touchdown point, thus adding capabilities of all-weather operation with little additional equipment or training.

The rapidly scanning beam carries pulse-coded data representing its instantaneous scan angle. Upon interception of the beam signal, an airborne receiver decodes and memorizes the elevation angle; this measurement is updated by each scan. Essentially continuous knowledge of the aircraft position is obtained from the three angular coordinates provided by the scanning beam and the ILS, up until the flare-out maneuver begins (Figure 2). During this maneuver, with less than 30 seconds to go, progress along the runway is automatically dead-reckoned; however, ILS centerline guidance and the scanning-beam measurement of elevation above the runway continue until touchdown occurs (Figure 3).

Choice of Specifications

The design goals for the scanning-beam technique and for its possible future use as a full landing system have heavily emphasized (1) universality of application (pertaining to both airports and aircraft), (2) quality of the signals and derived data, and (3) system reliability. The influence of these objectives on the choice of

technical parameters was quite direct, and very few choices were made arbitrarily.

The microwave fan beam has half-power beamwidths of approximately 20 degrees horizontally and 0.5 degree vertically. The wide horizontal pattern allows coverage of the touchdown zone from a point alongside the runway; it also provides initial-approach service before the aircraft is aligned with the runway. The narrow vertical beamwidth avoids interference from ground reflections while low elevation angles are measured; a still narrower beam would require either a higher radio frequency (discussed below), or a greater vertical aperture at the ground-based antenna. The aperture now used is 8 feet, and the center of radiation from a larger one might be so high that service to aircraft near the runway surface requires a downward "look"--hence renewed trouble from reflections. Tests indicate that the 0.5-degree beamwidth is satisfactory.

A radio frequency of about 16,000 Mc was chosen as high enough for reasonably small aperture dimensions, and low enough to avoid excessive atmospheric attenuation. At this frequency, one-way transmission through a typical rainstorm to a superheterodyne receiver 20 miles away requires, theoretically, about 500 watts peak pulse power. (Figure 4 illustrates this case, but with a transmitter power yielding a receiver signal-to-noise ratio about 12 db higher than theoretically required.) In contrast, at 24,000 Mc the theoretical required power would be over 8 kw, and at 36,000 Mc it would be 6 megawatts. The only economical choice among these that allows an ample safety margin is 16,000 Mc.

The beam is scanned upward from slightly below the runway surface to an elevation of 20 degrees, and back down. The lower limit includes the low angles required for flare-out and touchdown guidance. The upper limit allows angle measurements (especially for monitoring in advance of flare-out) during a final-approach glide at any angle up to 20 degrees. High-performance military jets and future civil transports might use considerably steeper initial glides than the present ILS allows. Guidance at high angles could also be used by jets for let-down toward the airport from en route altitudes, prior to final approach.

The scanning rate was chosen as 10 scans per second. Computer simulation and flight tests have indicated that flare-out maneuvers entered from a 3-degree glide at 120 knots can be adequately guided by error-free determinations of position five times per second, whereas at any appreciably lower sampling rate the stability of control seems inadequate. Scanning at 10 scans per second allows a 2:1 margin for higher angles and approach speeds, as well as for finite errors of measurement. Significantly higher rates of scan would impose difficult requirements on the angle-encoding scheme and on the antenna scanning mechanism.

It is advantageous, of course, to make the basic measurements of angle as accurate as possible with uncomplicated techniques. One advantage of angular accuracy is freedom to site the ground equipment at desirable locations, as noted below; another advantage is reliability of angular rate signals derived from the measurements. The maximum allowable error was taken as 0.05 degree, and the ultimate design goal as 0.01 degree.

The proposed siting of equipment (Figure 5) offers: (1) directly useful signals, without an airborne coordinate converter; (2) a system geometry providing position fixes precise enough for flare-out guidance from angular measurements of the specified accuracy; and (3) independence of real estate remote from the airport. As indicated in Figure 5, the measurements of two elevation angles can be made from the same or opposite sides of the runway. The longitudinal distance between these two origins is not critical (though it should eventually be standardized within about 5 percent), and may range from 2,000 to 4,000 feet.

The pulse code transmitted on the scanning beam has a novel format that seems uniquely suited to this application. It is required to provide full numerical values of angular position, with 0.01-degree resolution, to aircraft at any unspecified angle (hence, at all angles) between 0 and 20 degrees--and to do this ten times per second. Furthermore, to allow for future system expansion, time is reserved for shared-channel reception of similar data from two other scanning beams. Thus, 1/30 second is allotted for encoding a 20-degree sector in increments of 0.01 degree. To accomplish this with a digital code would require a pulse data rate of about 300,000 pulses per second, and would entail considerable complexity in the airborne receiver. With conventional analog encoding, such as variation of the interpulse spacing between successive pulse pairs, the required timing resolution would be uncomfortably fine--about 5 nanoseconds. Although such conventional codes might be usable if the system objectives were curtailed, the novel pulse code that has been adopted offers satisfaction of the foreseeable future requirements without depending on future improvements in the techniques used.

The code format consists of a simple pulse train, the interval between successive pulses varying progressively (as the scanning-beam angle changes) from 16 to 98 μsec. The encoding process is independent of the rate of antenna motion; if the scan were halted at any angle, a uniform pulse train would be transmitted such as to represent that angle according to the relation

$$\phi = \left[(\text{pulse spacing}) - 18.00\ \mu\text{sec}\right] \times 0.25\ \text{degree}/\mu\text{sec}$$

Thus, the pulse spacing is a linear function of scan angle, whereas the PRF varies inversely with scan angle between 62,500 and 10,204 cps. At zero angle, the pulse data rate just matches the intended scan rate--that is, the scan angle changes by 0.01 degree every 18 μsec, which is also the interval required to establish a new encoded value. At all higher angles, the encoding is controlled so that the interval preceding each pulse represents the angle at the instant of transmission of

that pulse within ±0.01 degree, though the data are transmitted at larger increments of angle. (The potential accuracy of decoding at high angles is thus preserved, since a minimum of 15 or 20 values per scan will be received, and these can be averaged to yield a resolution better than ±0.02 degree. At low angles, the resolution obtained by averaging is theoretically much finer.)

From the equipment designers' point of view, this analog code format offers several advantages. It can be decoded by a variety of analog and digital techniques, and its simplicity should encourage continual development of better and more economical airborne receivers. The use of analog decoding circuits facilitates the enhancement of accuracy by averaging processes, which are especially effective in guarding against noise and interference. The single disadvantage that has appeared thus far is a minor one--the requirement to design electronic circuits to behave as desired with 6:1 variations in pulse recurrence frequency; this is more a matter of forethought than of expense.

Ground-Based Equipment

The main elements of ground-based equipment are a narrow-beam scanning antenna, an angle-data encoder, and a transmitter. Figure 6 illustrates the functions of these units.

The scanning antenna (Figure 7) is a parabolic pillbox array of lightweight but rigid construction. It forms part of a mechanically resonant assembly that also includes horizontal torsion bars and a counterbalancing mass. This assembly is driven through a clutch so as to scan the antenna through a 40-degree mechanical sector, starting 10 degrees below horizontal; the beam is radiated within the central 20-degree portion of this sector, which it traverses ten times per second in alternate directions. An engraved scale affixed to the side of the scanning pillbox passes beneath a light source and a photoelectric pickoff system, thereby generating a series of pulses representing successive angular increments of 0.01 degree. As the first step in generating the pulse data code, these increment pulses are counted within the angle-data encoder. After each scan through the active 20-degree sector, an upper or lower limit switch on the scanner is activated, whereupon the total count of increment pulses is automatically tested; if it does not equal exactly the expected number, an alarm circuit is triggered, and the count is then forced to the correct value in preparation for the next scan. (This is one of several automatic-alarm tests that are incorporated to guard against the transmission of spurious data. Another such test requires a pulse from an RF detector in the antenna transmission line to coincide with each modulating pulse produced at the output of the encoder.)

The transistorized angle-data encoder is shown in Figure 8, and its internal functions are illustrated by Figure 9. This device supplies the transmitter modulator with a continuous train of pulses, the intervals between which are kept proportional to the instantaneous tally of angle-increment pulses from the antenna pickoff. A coarse-fine timing technique is used, with the coarse interval controlled by counting clock pulses derived from a crystal oscillator, and the subsequent fine interval controlled by analog interpolation between clock pulses. The accuracy required of the analog interpolator is only about 1 part in 16; although it could be eliminated by using 25-Mc counters, the present arrangement (counting at 3.125 Mc) was considered more conservative and reliable.

Omitting several bookkeeping circuits that guard against undesired coincidences, the operation of the encoder circuits will be described in terms of three stages of interval generation (Figure 9). Upon the termination of each coarse interval, the next timing cycle begins by counting 50 pulses at a rate of 3.125 Mc to produce an initial interval of 16 μsec, which is the minimum specified code spacing. During this time, a preparatory transfer of the angle increment tally is made into the variable-interval counter, where (after the initial 16 μsec) the resulting number is counted down toward zero. The time required to reach zero corresponds, within one clock pulse, to the scan angle existing at the instant of transfer; if the angle changes by one or more increments during the counting process, one more or one less clock pulse may be counted. More often, the additional increments are accounted for by changing the duration of the later interpolation delay.

When a zero value is reached in the variable-interval counter, the coarse timing loop has completed one cycle, and starts to repeat; however, the resulting modulator-trigger output is delayed by some multiple of 1/8 of the clock period. This interpolation delay is established, basically, by the excess of the stored count of antenna pulses over those represented by an integral number of clock pulses. (The clock-pulse interval corresponds to 8 hundredths of a degree.) However, the delay corresponding to this excess must be added to the interpolation delay applied after the last timing cycle, since it is the interval between delayed modulator triggers that is meaningful. The results of the necessary addition are transferred once per cycle, during the initial interval, in the form of a non-zero reset to the interpolation counter; any resulting overflow modifies the variable interval as mentioned above. Any increment pulses arriving from the antenna after a new value is transferred into the interpolation counter are tallied directly in that counter, so that the delay applied to the modulator trigger is up-to-date within 0.01 degree of scan angle.

The modulator and transmitter are shown in Figure 10. The modulator design is straightforward; vacuum-tube switching is used to accommodate the high maximum PRF of the system at a pulse width of 0.2 μsec. Some special attention to power-supply regulation was necessitated by the 6:1 variation of duty factor. Although a 2-kw peak-power magnetron is used in the present experimental equipment, the duty factor could also be sustained in available tubes with peak-power ratings up to 85 kw.

Airborne Equipment

Each aircraft served by the guidance system will carry the standard ILS receivers, as well as a new receiver that derives flare-out guidance from the scanning-beam signals. A flare-out path computer might also be carried as a separate package, though in some forms the circuits are simple enough to be incorporated in the guidance receiver or the approach coupler. For example, in early flight tests successful use has been made of passive RC circuits, which can produce various modifications of paths generally similar to an exponential decrease in height.

A prototype package of the scanning-beam receiver is shown in Figure 11. This is a transistorized, superheterodyne receiver and angle tracker, designed for airline service conditions; its primary power drain is less than 50 watts. The larger unit is contained in a standard 1/2-ATR case, and weighs about 20 pounds. The smaller unit, containing the RF components and IF preamplifier, measures 3 by 4-1/2 by 7-1/2 inches, approximately, and weighs under 5 pounds. An integral package could be produced, but this separation avoids a long run of waveguide from the antenna.

Figure 12 indicates the primary internal functions of the scanning-beam receiver. Most of these functions will be described only briefly, but some details will be given on the more novel features.

The "front end" of the receiver is a conventional superheterodyne design, using a cavity-stabilized klystron local oscillator. The IF, centered at 60 Mc, has an overall bandwidth of about 20 Mc; this can be reduced to about 5 Mc, if a crystal-controlled oscillator and multiplier are later substituted for the klystron.

The discriminator functions shown in Figure 12 will be discussed at some length below.

Angle decoding is accomplished by means of a sawtooth voltage that is reset to a reference value by each received pulse, then sampled and again reset by the next pulse. As shown in Figure 13, the difference between the sampled voltage and the reference value is proportional to the pulse spacing minus 18 μsec; hence, referring to the equation previously given, this voltage difference is proportional to the instantaneous scan angle. The reference voltage is varied so as to equal the angle output voltage from the memory, and the expected value of the sawtooth voltage at the instant of sampling is near ground. Therefore, the sampled voltage (relative to ground) is proportional to the pulse-to-pulse error between the angle represented by the receiver output voltage and the angle represented by the spacing of received pulses.

Immediately after each pulse is received, the existing error voltage from the decoder is used to produce an incremental correction of the charge stored in the angle-memory integrator. The magnitude of each correction is also controlled by the amplitude of the incoming pulse, and is thus proportional to the product of amplitude and instantaneous angular deviation. The angle voltage delivered by the receiver is continually servoed, by this process, so as to represent the "center of gravity" of the many instantaneous values of scan angle that are sampled as the beam passes over the aircraft. This averaging method of angle determination minimizes the effects of noise and interference.

Another safeguard against noise and interference is provided by a timing gate--one of the "video discrimination" functions indicated in Figure 12. Prior knowledge of the approximate angle to be decoded is used to control the timing of a track-gate signal, so that only pulses that recur at the expected intervals are consistently passed to the angle tracking circuits.

The received signals are also subjected to two forms of automatic amplitude discrimination--AGC in the IF amplifier and video threshold control. In both cases, the functional design takes advantage of peculiarities of the desired signals and of other features of the receiver.

Peak detection of the scanning-beam envelope is used to establish the AGC level. The point of detection is beyond the track gate and video threshold circuits, which prevent disturbance of the AGC level by interference. The AGC response is about 5 db per second, and the dynamic range is about 65 db.

When the aircraft is about to land, the scanning beam is received at a very low angle above the ground. Reflected energy then combines with the direct signal so as to produce asymmetrical skirts on the beamshape pattern, though there is little angular displacement of the nose of the beam. Therefore, it is desirable to derive angular data only near the peak of the beam envelope, ignoring the distorted low-level skirts. When the aircraft is at a high angle, however, the angle data are represented by fewer (widely spaced) pulses per half-power beamwidth; a more reliable determination of the angle is obtained by lowering the receiver's video threshold so as to accept more of the weak pulses on the beam skirts, which are now undistorted by reflections. Hence, a video threshold level that varies as a function of angle tends to give optimum results.

Figure 14 illustrates the pulse-count threshold control circuit. The video threshold of the receiver is automatically varied, over a considerable range of aircraft angles, by a servo loop that maintains a constant number of received pulses per beam passage. Since the angular position of the aircraft cannot change rapidly, the average number of pulses during many successive scans will suffice as a threshold control. The received pulses that exceed the threshold level are regenerated to a standard amplitude, and then integrated. Any change in the voltage from the integrator is applied as a corresponding change in back-bias on the pulse regenerator; by this means, an effective beamwidth corresponding to about 42 pulses is maintained. The rate at which the threshold level can respond to variations in pulse count is limited to about 1 db per second--much slower than the AGC response--to minimize interaction.

At an elevation angle of 7 degrees, with the scanning antenna running at 10 scans per second, 42 pulses occur between points 15 db down on the beam envelope; whereas at an angle of 0 degree, 42 pulses occur at -2 db, approximately. The minimum threshold level is limited at -15 db to avoid acceptance of side lobes and other spurious low-level signals. As a result of this limit, the number of pulses per beamwidth gradually decreases after the angle exceeds 7 degrees, reaching a value of 20 pulses at 20 degrees; operation at this point is still satisfactory.

The output of the pulse-count threshold circuit consists of unmodified video pulses that are gated through the circuit by coincidence with the standardized pulses. The absence of base-clipping avoids signal drop-out due to minor amplitude fluctuations, which would cause disproportionate changes in the voltage by which the pulses exceed threshold level. Since standardized pulses are triggered only by incoming video that exceeds the threshold, the effect of the adopted scheme is to shear off the skirts of the beam, passing the central portion with its original amplitude characteristics.

The basic decoding and tracking functions of the receiver are illustrated by the heavy lines in Figure 15; the means that are used for automatic control of the calibration of these circuits are also indicated. Switching between the beam reception mode of operation (which is shown) and the autocalibration modes is accomplished by gated diode circuits. No relays are used.

For beam reception, all incoming pulses that pass the track-gate and threshold circuits are applied to the angle decoder and to the sample-gate circuit. When the decoder sawtooth voltage is halted by each pulse, its deviation from ground potential tends to produce a proportional positive or negative current flow from the error detector. The video pulse simultaneously triggers the sampling gate, and the error current flows into the angle memory (an integrator) for a time proportional to the amplitude of the video pulse. Thus, the incremental correction to the stored angle value depends on the product of amplitude and angular deviation for each received pulse in turn. After all pulses within the beamwidth cause similar incremental corrections, the net correction of the stored angle is toward the centroid of the received envelope; hence the term "center-of-gravity" tracking.

The pulses received during each beam passage are also applied to a series of pulse-integrating, triggering, and gate-generating circuits, as represented in Figure 15 by the Calibration Sequence Control block. The response of these circuits to the received beam signals is to select, alternately, the two calibration modes of operation--one mode or the other being activated for about 1 millisecond after each beam passage. During this brief autocalibration period, a series of pulses from one of two crystal-controlled oscillators is injected into the normal video channel of the receiver. The spacing of pulses in this simulated beam signal represents either 0 degree or 11.25 degrees. The reference voltage for the decoder sawtooth is switched accordingly, so that no corrective current is produced by the error detector if the decoding and tracking circuits are properly calibrated. In the zero calibration mode, any error current is integrated over several seconds and applied as a correction to the nominal ground potential against which the decoder sawtooth is compared. In the slope calibration mode, the error current is separately integrated over several seconds, and applied so as to correct the slope (in volts per μsec) of the decoder sawtooth. The result of these continual calibration processes is to prevent any slow variation in circuit parameters, such as temperature effects, from degrading the accuracy of the angular output value.

Initial Test Results

The accuracy of the basic technique of elevation angle measurement was tested extensively over airport terrain during 1960. The possibility of erroneous measurements at low angles, where reflections from the ground tend to distort the received beam envelope, was a primary concern. Accordingly, hundreds of test series were made with a receiving antenna mounted at various heights on a 40-foot pole, which was moved to different locations on the field. All data were reduced to show the angle relative to the average ground plane as the independent variable.

These ground-based tests of static performance included investigations of ground reflectivity, beamshape distortion, system stability, and receiver angle readings. The receiver used for the tests was a vacuum-tube breadboard model that incorporated essentially all of the functions that have been described above, except the track-gate circuit. Optimum conditions of measurement were found to include tall grass and low installation of the scanning antenna; both of these factors reduce the magnitude of reflections.

Under typical operating conditions, with a stationary receiver, the overall system error was generally within 0.02 degree. Figure 16 shows, for example, the mean errors and RMS dispersions in readings taken at various elevation angles, and at distances between 1500 and 3000 feet, along a single test line. The peak error at angles greater than 0.5 degree was 0.03 degree, and mean errors were within 0.01 degree down to 0.2 degree. Angles below 0.5 degree are less critical for aircraft landing guidance, because of the system geometry.

Early flight tests of this technique have used simplified instrumentation that offers less precision in the determination of errors than in the case of ground-based tests. Emphasis in flight testing, thus far, has been on the suitability of the derived guidance signals for manual and automatic control of landing maneuvers. Flare-out paths have been programmed in terms of the continuously measured angle, so as to produce an approximately exponential decrease in rate of descent. Such data as have been recorded, as well as qualitative observations, show that the deviation signals displayed or coupled to the autopilot

are exceptionally quiet, and yet rapidly responsive to maneuvers and gusts. (No damping is applied to the receiver output signal, other than the lag inherent in the ILS crosspointer--without a capacitor--and in the Glide Slope channel of the approach coupler.)

About 150 manually controlled landings and 25 fully automatic touchdowns have been achieved with this technique. In many cases, the pilots were guest observers from airlines, military agencies, and other activities; their ILS training was found sufficient for these tests. However, much remains to be done in arriving at better display instrumentation and at optimum gain adjustments in the automatic control channel. Further tests are being conducted under the auspices of the Federal Aviation Agency and other interested organizations.

Conclusions

A practicable technique of elevation angle measurement for application to the all-weather landing problem has been developed, using a rapidly scanning, pulse-coded, K_u-band radio beam and an airborne receiver. Reliability of the derived guidance signals is ensured by design features that minimize the effects of noise and interference, take advantage of redundant information, and maintain critical calibration by reference to crystal oscillators.

An exhaustive series of ground-based tests has proven the static angular accuracy of the technique, in the region of operational interest, to be within 0.02 degree RMS. Flight tests in conjunction with the standard ILS indicate that this technique can be applied successfully toward the accomplishment of routine all-weather landings.

References

1. F. H. Battle, Jr., "A Practical Approach to Automatic Instrument Landing," Interavia, May 1960.
2. G. B. Litchford, et al, "A Look at the Future of Automatic Landing Systems," IRE Transactions on Aeronautical and Navigational Electronics, Vol ANE-6, p 118-128, June 1959.

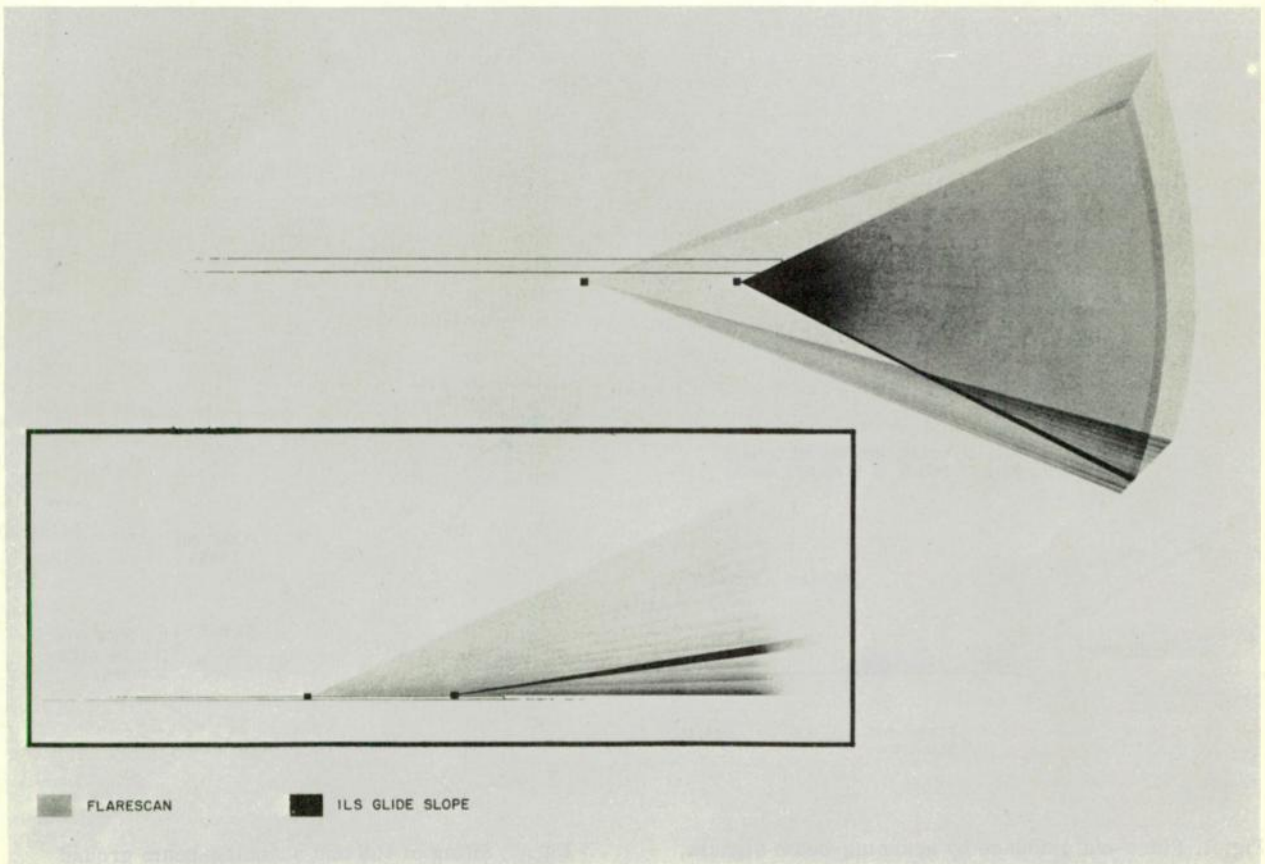


Fig. 1. Overlap of scanning beam with ILS glide path.

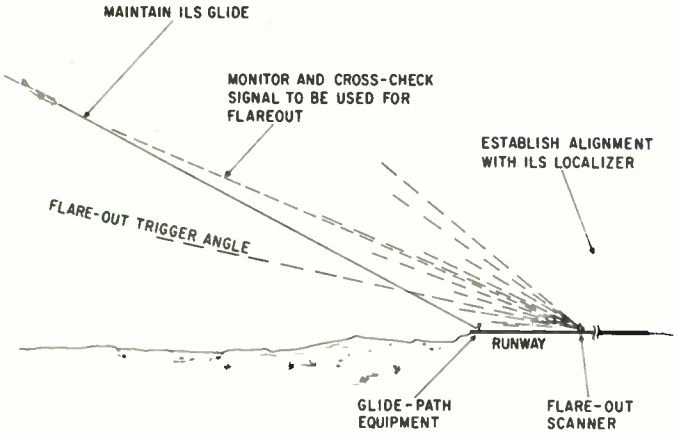


Fig. 2. ILS glide before flare-out maneuver.

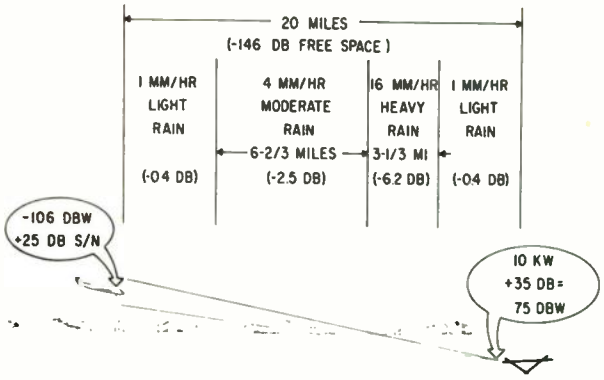


Fig. 4. Attenuation of K_u-band signal by rainstorm.

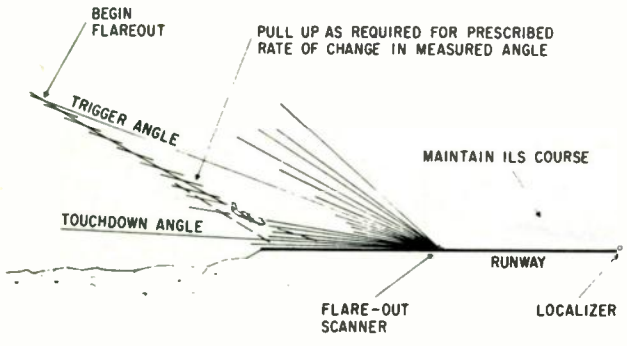


Fig. 3. Flare-out guidance by scanning-beam signals.

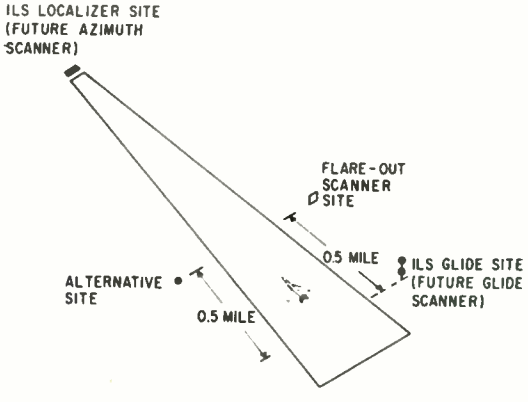


Fig. 5. Siting of ILS and scanning-beam ground equipments.

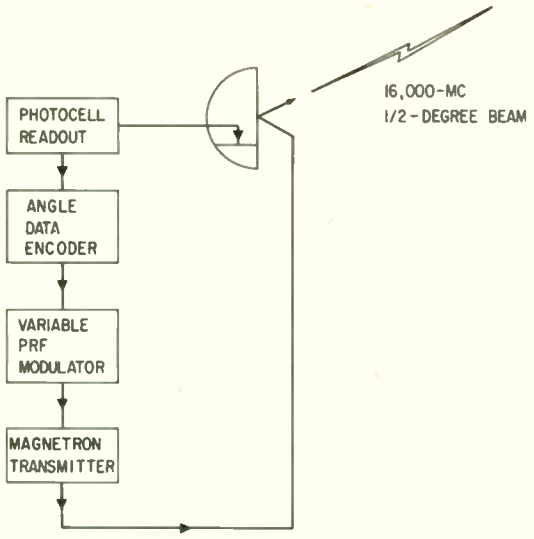


Fig. 6. Functions of ground-based equipment.

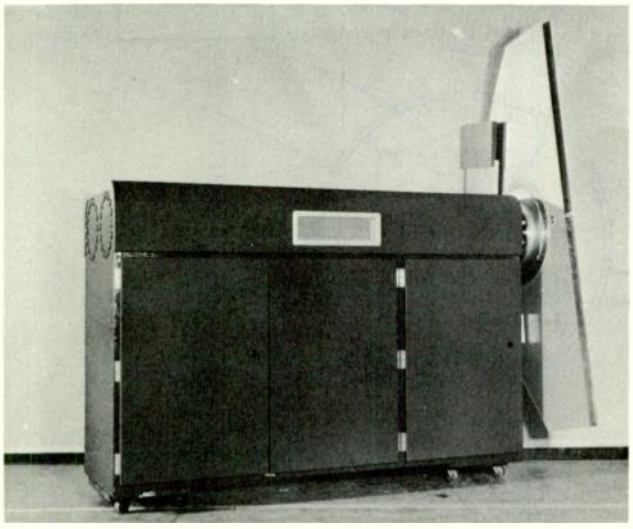


Fig. 7. Mechanically resonant scanning antenna.

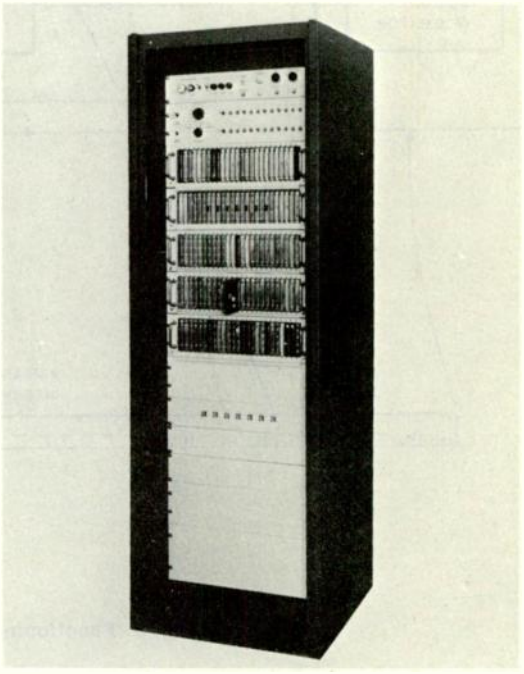


Fig. 8. Angle-data encoding equipment.

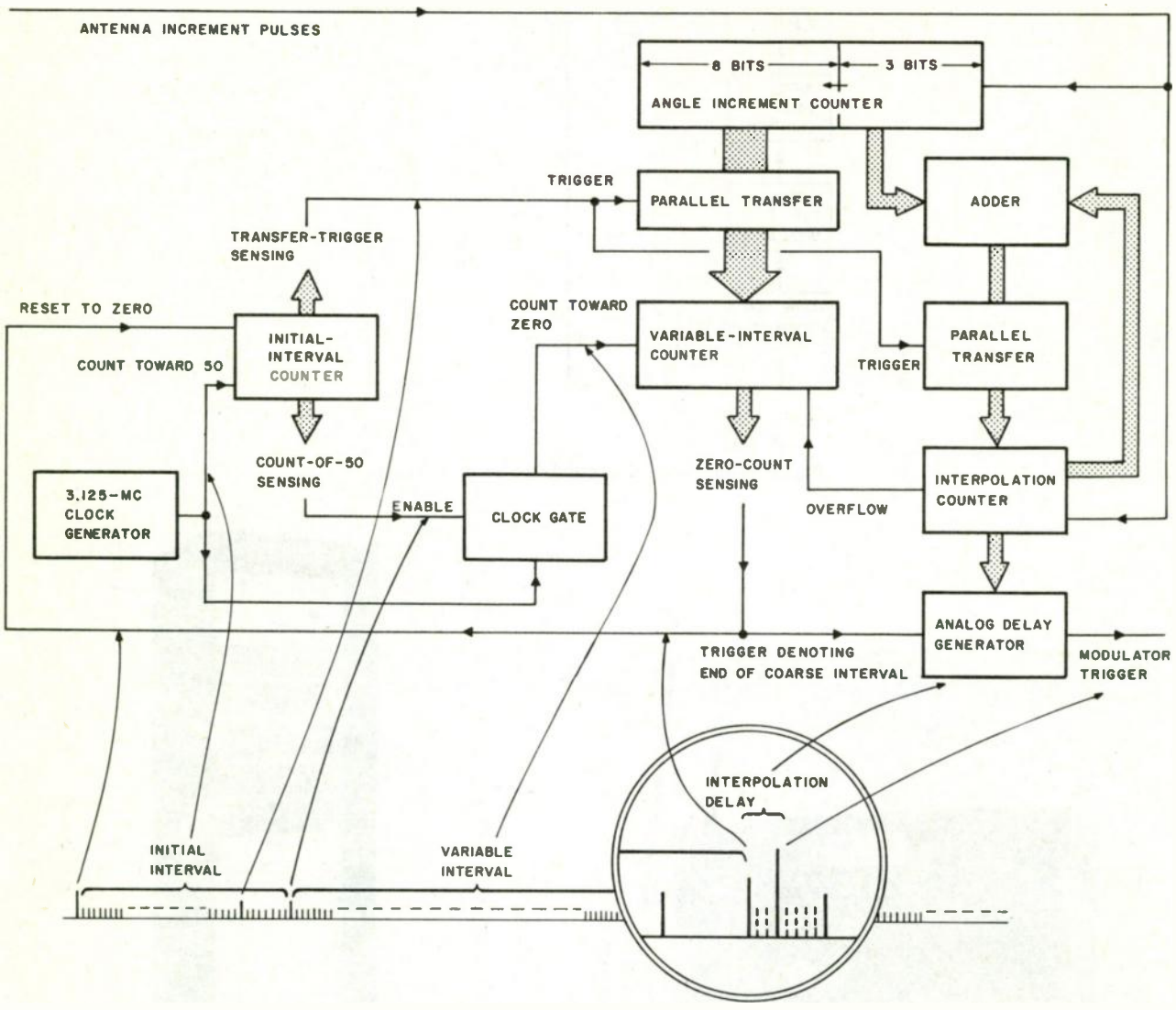


Fig. 9. Functioning of angle-data encoder (simplified).

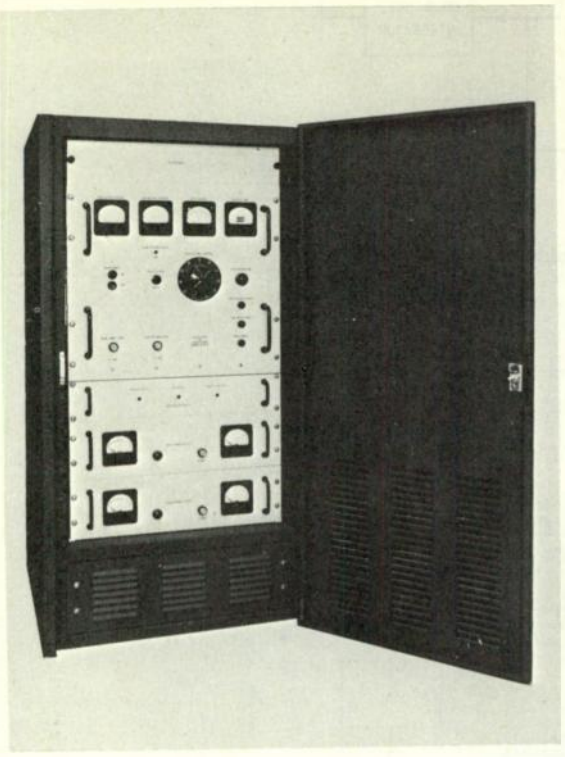


Fig. 10. Transmitter and modulator equipment.

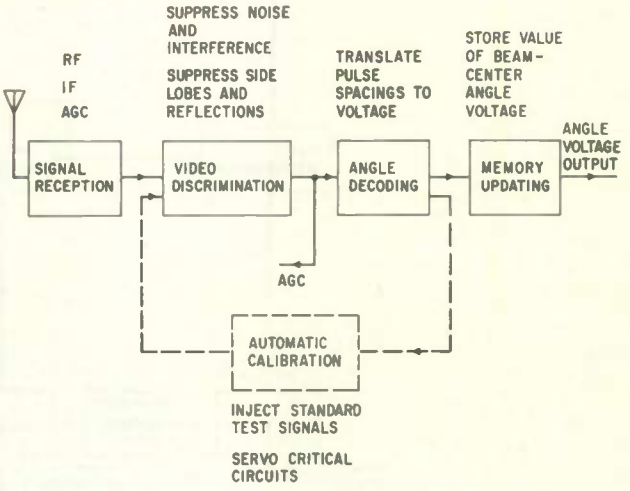


Fig. 12. Major functions of scanning-beam receiver.

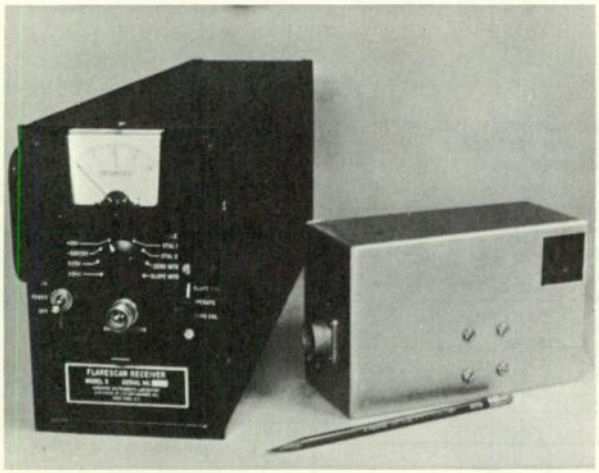


Fig. 11. Airborne receiver.

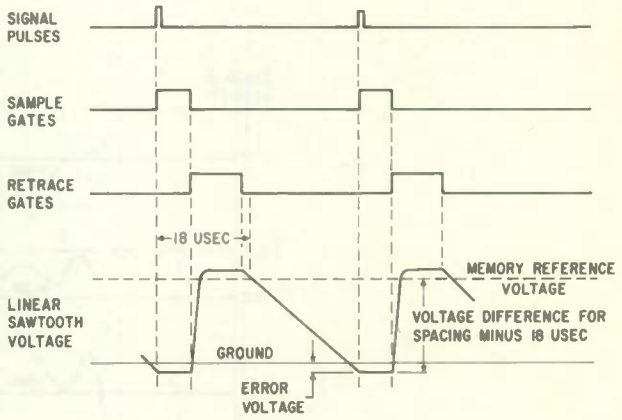


Fig. 13. Decoding waveforms.

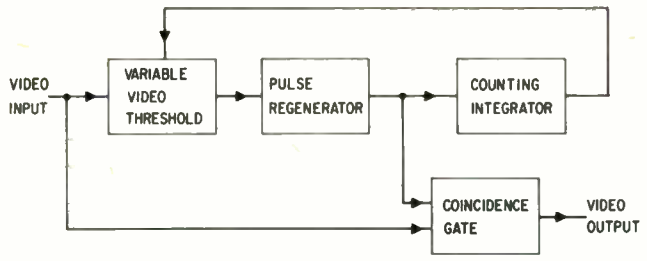


Fig. 14. Pulse-count threshold control functions.

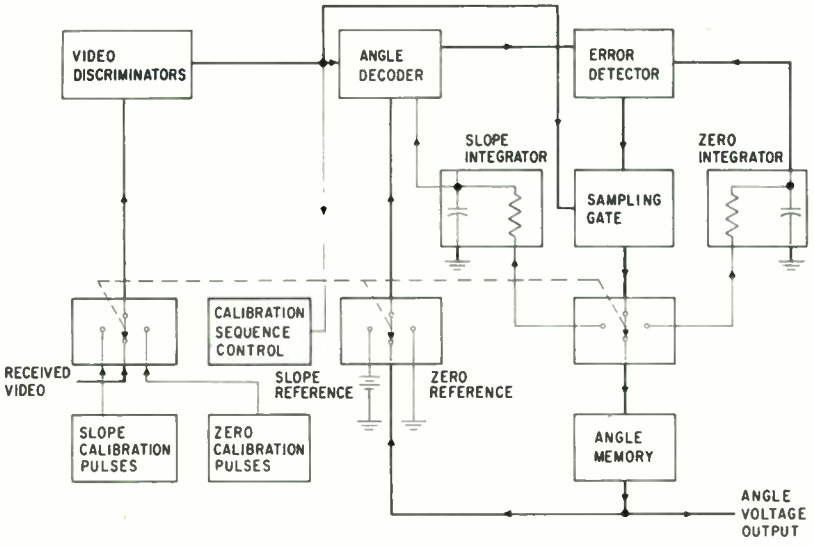


Fig. 15. Automatic calibration of basic receiver functions.

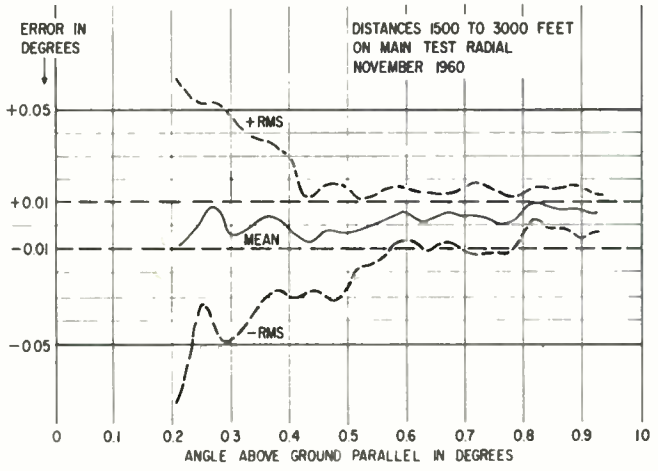


Fig. 16. Errors in receiver readings at low fixed angles.

BASIC ISSUES OF SPECTRUM CONSERVATION

Donald G. Fink
Scientific Laboratory
Philco Corporation
Blue Bell, Pa.

I have in my hand the book which has been referred to by Fred Lack, called "Radio Spectrum Conservation." This was written by members of JTAC and by consultants and experts in the field of propagation and spectrum management. When plans for this panel were announced, I went back to this book and realized that its authors had looked forward ten years when it was written in 1951. I was amazed and somewhat appalled to find that most of the problems which have been mentioned in this book, and which many of us felt something would be done about in the ten years after it was published, remain with us almost unattacked—let alone solved. It seems, therefore, appropriate in the first few minutes for me to review with you some of the points that were made in this book.

"Radio Spectrum Conservation" is now out of print, but for those of you who are interested in a detailed statement of the problem - and a lot of chapter and verse about what might be done about it - the book, which was underwritten by IRE and EIA, can be found throughout the world in technical libraries.

Background noise and interference in the spectrum have been increasing with every year. In part this is due to noise over which the FCC does not attempt to assert jurisdiction. This includes many forms of industrial noise and in particular, the type of incidental white-noise radiation produced by transmitters which do not take into account the rules and regulations of the FCC. That will be the subject, as I understand it, of Dick Gifford. I am going to talk about the interference areas over which a national policy of supervision and control does exist.

Let me say, then, that first we must realize that the electromagnetic spectrum is a bounded resource; that is, it cannot be extended much beyond its present limits. Even if we think about the new frequencies that are becoming available through the development of coherent light, we must concede that they are not all-weather frequencies. They can be used in light-conducting pipes in much the same way that lower frequency waves can be used in waveguides. The spectrum we must conserve embraces those frequencies uniquely adaptable to long-distance broadcast and mobile services. For these overcrowded regions of the "open" spectrum must be used.

In addition, whenever fixed wire and cable facilities are not available or cannot be commanded as a matter of national policy, long-distance radio communication may be the only mechanism for exercising national policy - and we are running out of the long-distance spectrum. A large part of the trouble is the highly inefficient way in which long-distance channels are used. We must find ways to use the spectrum resources that now remain idle for long periods. I need only to remind the "hams" in the room how many times they hear "di-di-di-da, di-di-di-da, di-di-di-da" repeated over and over again by the hour on channels which are used perhaps 10 or 20 per cent of the time for useful communication. That sort of thing has got to be brought under control.

Among the reasons why spectrum economy is difficult is, first, that we have to have carrier tolerances and guard bands to take account of the fact that signals cannot be kept perfectly on frequency and within fixed sideband limits. Second, we must deal with the fact that modulation methods involve concentrations of energy. If they did not, we could not transmit intelligence. Different kinds of modulation behave differently in this respect. We must increasingly make use of the more efficient methods.

Third, to be useful, signals must be stronger than the noise and interference background. Unfortunately when we propagate signals via the ionosphere or the troposphere, the interference range of a transmitter greatly exceeds its service

range. Provision must be made, therefore, to separate transmitters geographically, when they occupy the same frequency, by distances which are very large compared with the range over which they give service.

Fourth, most services broadcast over solid angles of space that are enormous compared with the angles subtended by the antennas of the intended receivers. Microwave relays are an exception, and of course, broadcast stations must perforce cover a wide horizontal angle. But the point-to-point long-distance services create interference over enormous sectors because it is technically or economically impractical to confine the emissions to the paths of actual use.

Fifth, one of the most important uses of radio is in mobile service, particularly marine and aircraft. These stations have to have command of the spectrum over the wide geographical regions they traverse, regions very large compared with the areas in which they actually render service.

There are many ways of attacking these five basic problems which are described in the book; time here permits only the briefest mention. First, we have got to use smaller guard bands. Second, we must provide better control of the constancy of carrier frequencies. Third, we must use efficient modulation methods. Fourth, the sectors and areas over which interference is caused must be restricted to make them coincide as nearly as possible with the areas over which service is rendered.

Fifth, we must realize that the use of radio, in the open spectrum, is not to be encouraged where non-radio uses, namely, transmissions by wire, cable, waveguides or light pipes will do the same job as well. This must be the policy, despite the fact that existing patterns may have to be disrupted in making the change.

Sixth, frequency assignments must be shared to the maximum possible extent.

In my last minute I would like to point out that these issues are so great that they bring into question the whole theory of democratic government. We have one such issue before us now, one which I am sure will be debated in many quarters for many months. One of the causes of spectrum crowding is the fact that a very large portion of the spectrum from 470 to 890 megacycles is reserved for UHF television and is not in fact widely used.

Things have gotten to such a pass that the Congress has finally decided to make the public a partner in spectrum conservation. Now, I am sure many in this room will share with me grave misgivings as to whether what the FCC is recommending, and what the House and the Senate appear now to be resolved to do, is in fact good democratic process. The Commission and the Congress, according to the commentators, are now resolved to make the public pay something like $20 more for every television set that is to be made after a certain date in order to ensure that the UHF spectrum channels can be received by the purchasers of these receivers.

I do not wish to bring into the debate here whether or not that is a good idea. I only wish to indicate that in order to get the spectrum fully occupied, a decision as basic as this must be made, and not by engineers—the engineers can only advise--but by the Congress and the President of the United States in a law which, as I say, brings in question the whole theory of democratic process.

SPECTRUM POLLUTION

R. P. Gifford
Engineering Communication Products Department
General Electric Company
Lynchburg, Va.

When Don Fink opened his comments he mentioned that ten years ago JTAC had compiled what I think was a magnificent piece of work regarding the radio spectrum and he showed concern for the fact that in the last ten years there are items in that list that we have not yet even attacked.

Now, my part in this Panel today will be perhaps to lay some advance groundwork for the panel that is going to be sitting up here ten years from now.

I am concerned that ten years from now someone who is asking for decisions to be made regarding the control of man-made noise will go back to his data file and find no data, for in preparing this talk this is exactly what I found. I was trying to determine the degree to which man-made noise, which is a waste product of our electrical civilization, has increased over the past decade.

Man-made noise is our own petard. It is essentially a waste product of the age of electricity, an age that has also given us radio communication. Neon lights, commutators, power transmission lines, fluorescent lights, elevators, heavy machinery controls, dielectric and inductive heating, electric arc welding, diathermy, TV sweep circuits, automobiles, planes and even all radio transmitters are sources of man-made noise. As we consume ever greater amounts of electricity, we run the risk of generating more and more noise; we run the risk of some day smothering out the utility of that part of the radio spectrum that is most useful for mobile services, with a blanket of noise that appears to be creeping up continually in frequency.

Now, my question is, does it have to do this? Do we have to set this point? And I think the answer is obviously no. Means can be found, I am sure, to reduce radio noise radiation from electrical equipment at a rate equal and opposite to the pressures of a growing population and greater use of electricity. But just as in the problems of air pollution and stream pollution, the actual existence of the danger must be recognized and the development of a concern for the effects and losses through pollution must be accomplished before we can ask for any realistic and effective controls to be applied.

I might point out that there are actually some controls available in this field today, under Part 15 of the Commission's Rules. However, the Commission is, of course, hamstrung to a considerable degree financially and manpowerwise, and even in terms of the wording of those rules, in exercising a very stringent control. For instance, it asks that no harmful interference should exist and if it does exist and if it does exist it should be corrected. Well, this requires the operator to correct an item that might have been a manufacturer fault. On top of that it requires judgment that the "harmful" interference does in fact exist. In Europe there are very stringent controls on various elements of electrical equipment and these are items that we may have to start applying here in an ever tightening range over the next ten years in order to control this growth of man-made noise.

Now, I would like to cover the man-made noise subject in some detail. Although I can't give you actual figures on what has happened over the past ten years I would like to cover enough detail to give you an idea as to why it is of concern, particularly in these low VHF and medium VHF bands.

There are, of course, literally hundreds or thousands of sources of noise at 150 megacycles but I am only going to discuss three of them:

Noise from land-mobile base station transmitters;
Noise from a Channel 7 TV transmitter; and
Noise from automobiles.

While we are inclined to think of transmitters emitting only a carrier with some discrete sidebands and an occasional spurious signal, measure-

ments will show that we are actually dealing with a broad band low level noise generator. Figure 1 shows the energy radiated from a 250 watt base station in the Land Mobile Service in terms of db above KTB (or the ideal receiver). Superimposed on that chart is another grid showing the noise level that would likely be received at a distance of .1 miles (or 2 NYC blocks) in any direction from the base station. Receivers with a noise figure of +10 db (readily obtainable in this frequency band) would notice a band of noise over a megacycle wide near the carrier at 168 Mc. In addition, it would notice smaller bands of noise around the spurious radiations. Considering that in the downtown area of large cities there may be over 30 of these transmitters operating within 1 Mc on an intermittent basis, it is quite likely that receiver performance in the downtown area will be limited by these random noise radiations - assuming of course that the receiver is trying to receive a weak signal from a transmitter located in the suburbs.

The same picture for a TV transmitter (Figure 2) may at first glance seem much more serious. But we must remember that in any one city there will be only one such "noise generator" at this particular frequency. In arriving at the overlay grid for signals received at .1 mile, an additional 10 db loss was inserted to represent a change in polarization.

As though these two sources were not enough, we must also consider the thousands of cars or "rolling noise generators." On Figure 3 are shown the peak noise level at 0.1 mile from two modern cars as measured by the proposed IRE standard. If we assume that there might be 100 cars within the 0.1 mile range then the average noise will probably come close to the peak noise of a single car. If all of the cars were like A, then the receiver will have noise interference when receiving a weak signal. Car B is considerably quieter indicating how much can be done in engine and electrical circuit design in modern cars. (Car A had a minimum of noise suppression devices.)

These are but three sources of man-made radio noise in an urban area. As stated before, there are hundreds. There is, in effect, a man-made noise smog in the vicinity 150-170 Mc over mythical Metropolis as shown in Figure 4. The first layer is made up of noise emanating from cars, neon lights, motors, generators, etc. The second layer is made up of noise from a TV transmitter atop a building off to one side. The third layer is made up of randomly spaced base station transmitters in the Land Mobile Service.

These are the three sources that I talked about and there are many others, as I have indicated before.

Now, the question I have been asking is, is this smog thickening? I am not actually positive of it and I went to get some data. I talked to Bill Crichlow, of the National Bureau of Standards, and I talked to engineers with the FCC. We dug out all the data we could get and we find that there is a pittance of such information. However, I can say from my experience in the design of urban communication systems that this noise definitely is increasing and we had better watch out.

I would like to say that the situation is analogous to the cocktail party, which has now become a famous laboratory for the communication engineer. You go into a cocktail party and find the talk at a level required by the background noise. As those present get more excited and more people enter the party, you have to talk louder and louder to be heard. Have you ever been tempted to suddenly yell, "Quiet!" and ask everyone to start over again at about 20 db down. You could continue to communicate but with considerably less effort.

I think this is significant. We need to yell "Quiet!" in the radio spectrum just for a moment and let's start over. We can get our receivers down closer to the theoretical KTB zero level and I think we ought to make use of this particular point.

We are talking about becoming more mobile in our communications in the future, we are talking about personal communications. We can't put hundreds of watts on the person to override this noise. We can only work with millewatts or watts, so I would like to yell "Quiet" for a moment in our cocktail party in the radio smog area vicinity and try and get this noise 20 db down.

Now, to do this, though, we have first got to get the data together and the purpose of this paper is to get people like the National Bureau of Standards, the FCC, Government laboratories, industry laboratories, to start collecting this data. To do this they have to collect it in a standard pattern. I am asking these people, then, to get together and establish the patterns they are going to use to collect this data so we can start to plot the growth rate of noise in our cities and be able to predict what it is going to be for the panel sitting here in 1970.

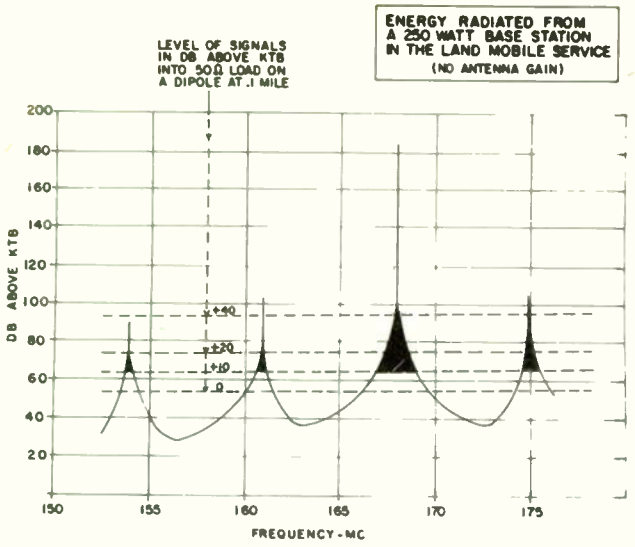


Fig. 1.

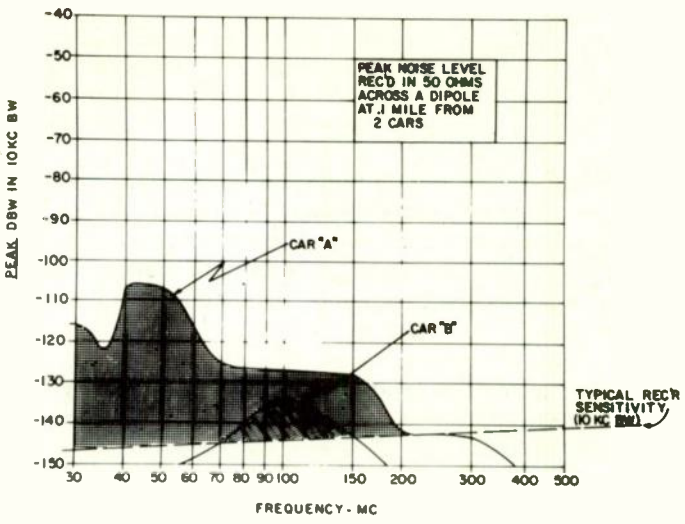


Fig. 3.

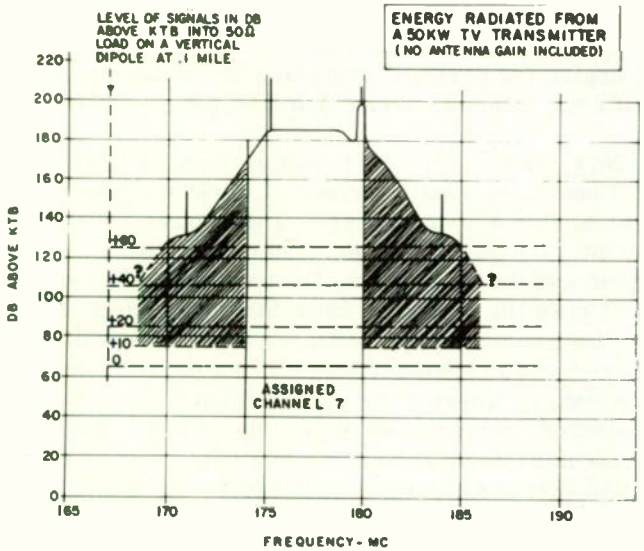


Fig. 2.

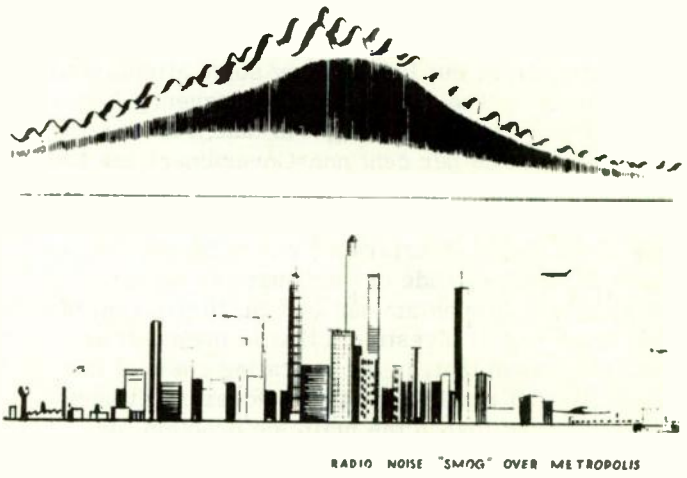


Fig. 4.

THE COST OF RADIO FREQUENCY INTERFERENCE

General F. L. Ankenbrandt
Radio Corporation of America
Camden, N. J.

I want to talk to you from the viewpoint primarily of a military man who has been involved in the management of extensive communications and electronic systems and has dealt with these problems of radio frequency interference from the user viewpoint.

I am goint to stay away from the scientific, you might say, and touch only lightly on the engineering aspects of it.

I would like to remind all of us here that the United States, whatever it is able to carve out of the world spectrum in sharing with other nations, ends up with a batch of frequencies, a band width of frequencies, which I would like to call one hundred per cent. Of that roughly 50 per cent is used by the non-Government services and is under the management of the Federal Communications Commission. The other 50 per cent is used by the Government in some form or other and the figures used to be, and I don't suppose they have changed too much, about 45 per cent--say, 45 per cent is used by the military services, 5 per cent by the other Government services, including NASA and the FAA, and so forth, and then, as you see, the other 50 per cent is in non-Government uses.

Now, I was given the title of my talk--I invited it, I will admit, but I should pay some attention to it, at least--"The Cost of Radio Frequency Interference," so I thought I ought to take a little spot check of this 50 per cent non-Government use first and in spot-checking this I came on quite a surprise. I got the clear impression that the cost of radio frequency interference on the 50 per cent non-Government side of the house is really not very great. It may amount to 5 or 10 per cent of the total capital investment that is involved, or even less than that, in the operating costs of the systems that are in use in non-Government services. I found that in the main the problem, because of reasons which I will speculate on, and these may be in contradiction with what has been said by Don Fink here and my friend on the right, they have been able to work around or through or over these interferences and to solve them to the point where you have some pretty solid circuits in the non-Government services.

Of course, the diversity of operations has been partially helpful. The international HF carrier networks are still operating quite effectively. The problems of putting your transmitters and receivers apart, your directional antennas, using filters where needed--that is, good engineering practices--have in fact solved the problems on an acceptable basis. I am not saying that everything is hotsy-totsy and that everybody loves it the way it is, but the services are working quite well, and the costs of working through these kinds of problems are not prohibitively great.

I found the same applies mostly also to the mobile and the broadcast services. Now, we all know the problems of the broadcast services but the management of those problems from an engineering viewpoint and from a frequency viewpoint has been sufficiently good so you do have good services in the main.

I hope that I have said enough on that, and also apologize for perhaps trying to give the impression that everything is rosy, which it is not.

Now, what is the real problem, then, of RFI? As I see it, as a military man, I think it is clearly over on this 45 per cent side of the house, which the military use for the defense of the United States and the free world. Over on that side you don't have the stable services that you have in the non-Government uses, you don't have the situations where you can engineer yourself out of all of these troubles, or where you can shift frequency on an organized basis, and so on, nearly as well. Tremendous systems have come into being since World War II and they have had to be fitted into a space where nothing was before, essentially. Great flexibility is needed in these systems. Many of them are hastily installed, they are moved around frequently. The power is generally higher than in comparable commercial systems, where you are

able to get a service on less power. Of course, the broadcast services are a special case in this respect.

Now, costs in this case are tremendous on the military side, the cost of the system in its degradation, whether it is an effective system or whether it is badly degraded. There is cost in the development of the system to make it an effective military adjunct, there is cost in the system operation itself. There is always built into these things system redundancy, a considerable amount of system redundancy, which, of course, adds greatly to the cost, and furthermore, there is a real cost to the proper use of the United States' part, because the military services with these additional systems coming in are exerting pressures to use more of the hundred per cent American portion of the spectrum than the 45 per cent which they are now using. They could easily use 55 per cent or 60 per cent, which means taking away frequencies from the non-Government or the other Government services.

So these are all costs which are real ones and I just wanted to remind the folks here that the military services are trying to do something, the Defense Department is trying to do something about it and this will lead directly into John McDavid's presentation here.

We are sure that Russia and her allies have got problems of this same type. They are in the main even more dependent on electronic and radio devices than are we.

As a final word, during World War II my own experience was that we were able to--time permitted, let us say, for us to find out what our problems were and correct them and to put effective radio and radar services on the air with what you might call a minimum of interference. This required, of course, among other things, frequency management and it required that these new systems--there were gaps in those days where new systems could be fitted--and they were fitted in these gaps.

Well, those days are gone, of course. After the War we had a tremendous flood of new radiation devices entering into the military services. We know that the electronic age in the military services is now an absolute must. In 1950 the projects that we had focused attention on--the radar problems--were going to come into real use ten years later, five years later, fifteen years later.

Project Monmouth focused attention on the Communications Zone radiators that are going to interact all over the place, let us say, in the battlefield area or in the war Theater Area. I was reminded, though of what--General Jimmie Gavin, at one of these presentations, said:

> "You know, we mustn't forget that this is a true problem that we have got to face, but you mustn't forget that we still have people who are able to do something about it in this particular Theater Area. You can manage your frequencies if you have the flexibility in your equipment. You can manage the frequencies. You can turn off certain things and turn on certain other ones. You can manage to do a, let's say, an effective job by careful management of these things with an intelligent point of view."

Now, a lot of us--I shouldn't say a lot of us--a few of us, perhaps, think the problem is so bad we throw up our hands and think we should start over. I must remind everyone that the management of these resources does in fact improve their effectiveness very materially.

Now, General O'Connell--I saw him over there--was very helpful to me, and as a matter of fact one of the spearheads in the earliest formal approaches of the military services to solving this incompatibility, this congestion problem on the military side. We formed, back in 1949 and 1950, the forerunner of what you are going to hear about here in a minute--and I don't want to take any thunder away from you, John.

I just wanted to mention that the Joint Spectrum Evaluation group was formed in 1949 or '50. It turned out to be a somewhat abortive effort of solving these problems. I guess it never worked very well because it disappeared and the other day, when I was talking to John McDavid, who is my successor in this same job in the Pentagon, he had never even heard of it and we couldn't find any files on the subject, so I am convinced it must have been pretty abortive but at least it was a start in this direction.

SYSTEM ELECTROMAGNETIC COMPATIBILITY - DOD's APPROACH TO THE PROBLEM

Brigadier General John A. McDavid, USAF
Communications-Electronics
Joint Chiefs of Staff
Washington, D. C.

I do want to go back a little bit to some of the things that were said this morning. I feel that in the military end of the game we are beginning to start to analyze what has been going on in these gray areas of frequency in the bands. We hope to come up with some solution.

INTRODUCTION

We could spend considerable time this morning actually documenting the problems that have occurred and now exist because of increased numbers and complexities of electromagnetic equipments attempting to operate within a relatively fixed radio-frequency spectrum. Such problems range from those that cause annoyance to commercial TV users, to some that have a severe impact on the performance of radar and communication equipment vital to national defense. I'm sure that those of you in attendance who have ever been concerned with that aspect of engineering dealing with making electronic systems compatible with each other and with the existing signal environment are well aware of the seriousness of this problem, and the need for considering the effects of unintentional interference in the assessment of over-all system performance.

The Department of Defense has long recognized that an organized program in the area of electromagnetic compatibility is essential to the successful operation of present and future military electronics systems. We spoke about this evaluation group a moment ago. It was one of the recognition areas. However, I am sure that it died because not enough people recognized the problems. We hope this one will not die. In June of 1960, DOD initiated its Radio Frequency Compatibility Program, which had three broad objectives. This program is in writing and it is approved by the Secretary of Defense. The objectives were:

a. The prediction of interference potentialities among military electronics equipments under operational environmental conditions.

b. The establishment of means whereby improved radio frequency management could be exercised within the military services, and

c. The establishment of means for predicting needs in the research and development areas so as to bring about the insurance of interference-free communications-electronics devices to military operating forces.

I don't think any of those three should be counted above the other. I think they are all equally important.

As a part of this over-all program, an Electromagnetic Compatibility Analysis Center (ECAC) has been established. This Center, which is located in Annapolis, Maryland, has been assigned responsibilities in a number of vital compatibility analysis areas, including the establishment and maintenance of equipment and environment data bases pertinent to compatibility evaluation; analysis, prediction and evaluation of present and future operational compatibility problems, and others. I would like to spend the rest of the time allotted to me discussing the Center's efforts in this regard — their approach to compatibility analysis, and their plans for the future.

ORGANIZATION OF ECAC

To clearly understand the program of the Analysis Center, it is necessary to be aware of organizational aspects of the Center itself, as well as the manner in which coordination is achieved between "in house" and various "out-of-house" efforts that are significant to the proper functioning of the Center.

The Center technical staff currently consists of approximately 50 engineers and mathematicians, about equally divided in the areas of electronic system analysis and computer techniques. Staff

size at the end of the next six months is projected at 80 technical contractor personnel plus 25 non-technical individuals such as administrative staff, secretaries, and data processors, and so forth. The facility is currently utilizing temporary quarters on the Naval Engineering Experiment Station in Annapolis, pending completion of construction of a permanent building on that base. The new building is scheduled for occupancy prior to 1 May of this year.

The Center is an amalgamation of military and contractor personnel. The Military manning is predicated on the facts that (a) the Center is assigned to a military agency for administrative support and resources, (b) data necessary for the Center base is being furnished through the Joint Chiefs of Staff, J-6, and (c) operational functions are being provided by a contractor. This complement, consisting of 15 military and civil service personnel, is concerned with the direction of the Center relative to its assigned mission, as well as serving in a coordination capacity in matters involving the three Military Services and governmental agencies.

The data base requirements for compatibility analysis prediction is known as the data basis. It can be subdivided into two major categories, spectrum signatures and environmental data. I would like to repeat that: Data basis - this is the basis of the whole thing - it consists of these two things, spectrum signatures and environmental data. Spectrum signatures are collections of specialized information indicating certain emission and susceptibility capabilities of particular communications-electronics equipment relative to unintentional interference. The most comprehensive spectrum signature data is obtained in accordance with MIL-STD-449A, through programs directed and funded by way of the Military Services' normal R&D and O&M channels.

Environmental data provides information regarding equipment deployment and operational characteristics. Files of this data will furnish such problem analysis inputs as equipment geographical location, on-off characteristics, antenna orientation information, modulation characteristics, and similar information pertinent to a particular operational or planned installation. The environmental data is supplied to the Center directly by the Military Services.

Other areas of concern to the ECAC include the results of special tests defined by the Center to validate various aspects of their interference prediction approaches, progress in military contractor R&D programs in support of the over-all Radio Frequency Compatibility Program, and similar undertakings. Representative examples of R&D coordination include liaison with Bendix Corporation on a program to develop a three-dimensional antenna pattern recording technique, and similar liaison with Airborne Instrument Laboratory on their development of waveguide power measurement techniques under multiple-mode conditions. The military offices of the three Services provide the liaison in these instances.

DEVELOPMENT OF COMPATIBILITY ANALYSIS MODELS

A major effort of ECAC is associated with the establishment of techniques for analyzing existing and potential compatibility problems, for evaluating the degree of degradation experienced by equipments in these problems, and for making recommendations leading toward the improvement of over-all system performance. The basis for such work is the development of suitable analytic models, and the processing of these models using a digital computer.

Let us assume in this case that the model of a specific type of radar system, such as a pulse-type acquisition radar with unprocessed video output is to be assembled. As will be pointed out later, other classes of models will be required in order for the Center to perform the previously specified analyses.

In order to develop an analytical model to represent a pulsed search radar, several inputs are required. Analysis of typical systems is necessary to determine the factors that must be taken into account in the model. A review of spectrum signature and other measurements will also contribute to the model development. In addition, special experiments may have to be defined in an effort to provide additional information on certain performance characteristics of the equipment being modeled.

These special experiments may be either theoretical evaluations or physical tests, but are generally a combination of both. They include such areas of investigation as the effect of emitter-receiver frequency separation on the receiver output characteristic, the signal processing characteristics of the detection process, and the effects of various modulations on receiver performance.

The resultant analytic interference prediction model consists of a series of mathematical instructions that represent the functioning of the

class of system being modeled, in this case the particular search radar class. The relationships are then programmed for processing by the computer.

To convert these expressions so that they designate a specific operational radar, the spectrum signature and environmental data on that radar are also inserted into the model. The spectrum signature information is obtained from three general sources. As previously indicated, one series of sources are equipment measurements made in accordance with MIL-STD-449A. Another source is an ECAC equipment characteristic file generated by means of an extensive literature search.

A third source involves the development of analytic techniques to synthesize spectrum signatures. This effort intends to establish empirical or mathematical relationships that describe the performance characteristics of the equipment under consideration. Examples of such programs currently being conducted at the Center include the synthesis of transmitter emission spectra, and the analytical generation of non-design frequency antenna characteristics.

Several types of prediction models are being developed on the ECAC program. These models can be classified into five categories:

1. Basic Models — These are models representing the basic emission and susceptibility characteristics of particular classes of communication-electronics equipment, for example, search radars.
2. Propagation Models — These are models describing the relationships to be employed in treating electromagnetic propagation modes.
3. Element Models — These are models of system signal processing techniques (e.g., ECCM techniques, interference fixes, etc.)
4. Equipment Models — These are models representing specific nomenclature equipment (e.g., AN/FPS-35). They are assembled using the basic models and the element models, as well as the equipment spectrum signature characteristics.
5. System Models — These are models describing comprehensive operational problems to be analyzed. They are assembled using the equipment and propagation models, as well as the environmental data file. They are designed to assess system performance, and provide a logical approach to optimizing such performance.

Within each group of models will be various levels of model sophistication, tailored to the analysis of particular types of problems.

An important task of the Center is the establishment of the validity of their compatibility analysis approaches. Since the reliability of analytical results will be no better than the reliabilities of the models and the input data, it is necessary that supporting information be supplied to determine the relative accuracy and possible limitations of the mathematical models that are employed.

Several areas of model validation testing are currently being conducted. Controlled radar tests are being performed for the Center at (a) the Navy Air Navigation Experiment Project (NANEP), Patuxent River, Maryland, (b) Rome Air Development Center (RADC), Rome, New York, and (c) the Army Electronic Proving Grounds (AEPG), Fort Huachuca, Arizona. Each of these tests is tailored to furnish specific environmental validation information, as determined by the over-all test requirements, and the facilities available at each testing laboratory. The tests are designed to provide information regarding the confidence to be expected in predicting unintentional coupling between equipments, as functions of:

a. The accuracy of the spectrum signature information and environmental data.

b. The inherent variations of equipment characteristics due to component replacement, tuning techniques, age, etc.

c. Inherent variations between the same model equipments having different serial numbers.

d. The character of the terrain in the vicinity of the test sites.

e. The assumptions made in particular analytical models employed.

Three operational problems have already been

assigned to the Center for analysis. Controlled tests to correlate predicted with actual system performance relative to the first two of these problems are underway. One of the problems involves the evaluation of L-band radar equipment performance in the San Diego-Los Angeles area. The second concerns the compatibility of radar and communications equipment in the Montgomery Air Defense Sector (MOADS). The third concerns the Cheseapeake Bay area. The associated validity tests are intended to check prediction confidence of the electromagnetic environment, specific receiver output characteristics, and the degradation to equipment performance.

Another series of tests is currently being prepared to verify a wide variety of element models. These are in general expected to be laboratory-type tests, in contrast to the more extensive field tests just described. It is expected that such tests will result in establishing the confidence of models of such signal processing techniques as MTI, sidelobe blanking, pulse integration, etc.

THE IMPACT OF THE COMPATIBILITY ANALYSIS PROGRAM

I have reviewed today some of the general approaches being employed by the Electromagnetic Compatibility Analysis Center in their analysis of unintentional interference problems. To conclude my talk, I would like to outline a few of the implications of this type of capability, particularly with regard to areas of system design and development.

It is anticipated that the analytical results of Center problem processing will not only provide an indication of over-all performance of existing systems, but will suggest techniques that might be employed to reduce or eliminate equipment degradation due to interference. Where applicable, the effects of changes in equipment deployment can be evaluated. Frequency assignment plans can be evaluated. Possible corrective measures for the reduction of undesired emissions or susceptibilities, such as equipment circuit modifications or the addition of interference "fixes," can be analyzed. In certain cases, transmitter or receiver replacements with units having different performance characteristics can be evaluated.

As far as equipment under development is concerned, compatibility analysis can be a design aid by pointing out which equipment characteristics need improvement. The predicted characteristics of equipment being proposed or in the design stage can be analyzed in its expected environment to determine potential incompatibility problems prior to actual equipment construction and installation.

It should be pointed out that the Analysis Center program is not restricted to static types of system analysis. Mobile problems, such as those encountered in the evaluation of missile, airborne, seaborne, or ground tactical system performance will be considered; these studies can lead to recommendations in both deterministic and statistical senses. In addition, the same general approaches will be employed to process space problems, as well as problems that are generated, in part, by unfriendly governments.

I have attempted to provide you with a brief picture of the objectives underlying the establishment of the Electromagnetic Compatibility Analysis Center. I have described, briefly, some of the approaches being taken toward the development of sound techniques for system analysis. I have pointed out some of the outputs we expect to achieve and some of the uses to which they will be put.

The capability that the Center is developing is one which is long overdue. Without validated analytical techniques, suitable for computerization, the task of eliminating the threat of RFI is simply overwhelming.

Establishment of the analytical capability will not of itself, of course, eliminate the interference problem. But it is certainly a keystone in the over-all compatibility effort which the DOD program represents.

The effects of this program on the communications-electronics industry can be expected to take several forms. The publication of MIL-STD 449A, which establishes standards of measurement and testing, and its implementation constitutes one of these. It and other standards will undergo revisions as the program proceeds and we begin to realize its benefits. These benefits will affect design criteria, not only for operating equipment and systems but for test equipment as well. Some of your organizations will participate directly in the data collection programs. You might be interested to note that it costs about $50,000 just to collect data on a single radar system. Some of you are already participating in these programs.

Aside from these and other specific effects, there is an over-all, broader impact on the entire communications-electronics community. Success of the DOD Compatibility Program will provide significant progress toward the removal of a

serious restriction and impediment to the realization of the full promise of the GOLDEN AGE OF ELECTRONICS.

SYSTEMS SPECIFICATIONS OF A RECOVERY CONTROL CENTER FOR HYPERSONIC VEHICLES*

Jack J. Fishman
Equipment Division, Raytheon Company
Waltham, Massachusetts

Summary

This paper summarizes a feasible configuration for a Recovery Control Center (RCC) which will be capable of effecting safe, routine recovery of manned hypersonic vehicles employing aerodynamic control. A number of critical areas are pointed out which must be considered in the design of such a center. Recovery responsibilities include the provision of advisory and/or command control of the flight from the point of atmospheric re-entry to the approach, landing, and retrieval of the spacecraft and crew on the ground. Based upon preliminary analyses, an integrated ground complex containing data processing, display, monitoring, and communications equipment is necessary to satisfy these responsibilities. This equipment will process and display incoming data so that key center flight control personnel can present advisory control data to the crew as needed. The information flow in the ground portion of the system is assumed to originate at the tracking and telemetry stations, which relay vehicle-sensed data to the RCC. Recommended major personnel responsibilities within the center are to 1) monitor the aeromedical, navigational, and aerodynamic aspects of the flight; 2) operate and maintain the data processing, communication, and display equipment; and 3) coordinate all activities with outside agencies and recovery related units.

*This paper is based, in part, on a study sponsored by the Air Research & Development Command for the Air Force Flight Test Center, Edwards AFB under Contract No. AF 04(611)-5965.

System Description

The primary mission of the Recovery Control Center (RCC) will be to effect safe, routine recovery of hypersonic vehicles throughout all aerodynamic flight regimes of the re-entry corridor. Thus, the RCC should be capable of assuming control responsibility whenever the spacecraft approaches the landing area, upon direction of the Command or Mission Control Center.

Present indications are that the spaceborne guidance and navigation equipment will, in most instances, be designed with the capability to operate as an essentially self-contained system throughout the entire flight. Thus, under normal conditions, the RCC would confine itself to providing ground-derived data to the vehicle and performing an independent check on the operation of the spacecraft and crew. In the case of a failure in the onboard system during re-entry and landing, the RCC would assume control in order to minimize the effect of such failures. The ground system would establish a completely independent source for information on vehicle position and motion in space, and provide an independent source for the computations required for generation of monitor and command information. This calls for an integrated tracking, computation, display, communications, and telemetry system.

An information flow analysis indicates that the RCC will accept telemetered data from both the onboard vehicle sensors and the ground-based tracking sites, process it, perform control decisions, and distribute or relay the results to the crew as advisory control information. The computer within the center would analyze this data for input to monitoring consoles which display:

1. Maneuver potential information (space position, energy management)

2. Vehicle status data (energy, structure, controllability, and subsystem status)

3. Aeromedical information (cabin, suit, crew functional capability and well-being)

These consoles should be monitored by aeronautical, aeromedical, and flight control officers. The operational-information flow is diagrammed in Figure 1. Based upon this flow analysis, the specific functions of the RCC have been determined to be as follows:

1. Re-entry area navigation display and monitoring
2. Recovery-related unit control and area traffic monitoring
3. Aeromedical data display and monitoring
4. Aeromedical condition advisory control
5. Vehicle status data display and monitoring
6. Vehicle status advisory control
7. Vehicle navigation data display and monitoring
8. Vehicle navigation advisory control
9. Recovery control system communications
10. Recovery Control Center data recording
11. Spacecraft interception by ground recovery units
12. Crew egress
13. Immediate post-landing medical examination
14. Emergency rescue
15. Transportation of crew to RCC
16. Post-flight medical examination
17. Debriefing and reporting
18. Transportation of spacecraft to RCC
19. Retrieval of spaceborne data

A flow diagram of these functions is given in Figure 2. In addition, it appears desirable for the RCC to maintain voice communication with the crew throughout the entire re-entry, approach, and landing phases of flight. However, the effect of the plasma sheath poses a critical problem to continuous communication. Uncertainty still remains in this area, even though a considerable effort at theoretical understanding of the plasma sheath problem has been put forth in recent years. In addition to theoretical and experimental work, predictions of plasma sheath effects must also depend heavily upon data from flights instrumented to explore these effects.

To generate realistic requirements for the RCC, the above functions have been divided into four categories: data processing complex, display complex, personnel subsystem, and voice communications complex. A discussion of each of these areas follows.

Data Processing Complex

At the RCC, the computer would accept and check vehicle sensor inputs transmitted to, and processed through, each tracking station from the ground tracking radar, and from auxiliary sites. The data-processing system would compute from these inputs and pre-determined stored data: 1) radar acquisition information (for two stations), 2) trajectory-defining parameters, 3) command signals for spacecraft control, 4) information for the center displays, 5) a simulated trajectory to predict dangerous situations, 6) abort emergency information, and 7) correction functions for pre-computed stored data derived from the error between the measured and the simulated trajectories.

Figure 3 is a general flow diagram of the proposed principle computer tasks required of the RCC data processing complex. The computer operation consists of 11 major tasks:

1. The initial sub-routine is the system's operational checkout which is executed prior to the beginning of the real-time problem. It includes such functions as setting up the control and display programs, checking out all equipment in the system, and running test problems.

2. Information is read from the tracking sites on a continuous basis. Data sampled and checked for correct format and errors includes ground tracking and telemetry data.

3. Altitude, velocity, range to go, and cross range are computed. This computation involves the conversion of the tracking site information into universal coordinates and then differentiation to derive velocity. The range to go (R_{TG}) and cross range (R_C) are computed from the difference between the vehicle's position and the position of the landing site.

4. The incoming tracking and telemetry information is cross-checked with previous readings and with itself to ensure consistency.

5. If the information is not correct, then an analysis is undertaken to determine the cause of the error and to correct it. The investigator and/or the computer should determine if the error is repetitive and if it is, what element in the system is responsible. When the fault is located, tracking site and control center personnel will be notified and action taken to eliminate the cause.

6. If the data are correct, then the vehicle's position is transformed into acquisition data (for 2 sites) and transmitted to the tracking sites in order to assist in the acquisition of the vehicle with the tracking radar.

7. The ground area attainable (GAA) is computed. This computation could be the most difficult and expensive (in time and storage) required by the computer.

8. Vehicle control signals are computed. This sub-routine generates the commands for pitch, roll and drag modulation that will bring the vehicle safely to the selected landing area.

9. The space vehicle impact point is continuously predicted, updated, and displayed.

10. Dangerous situations are to be predicted and alternate flight paths computed and displayed. This computational task is considered extremely difficult since the computation of the vehicle's flight trajectory must be accomplished faster than real-time.

11. Supplementary computer tasks to be performed include the updating of the remaining displays, and the real-time scheduling of the computational tasks.

Because of the critical real-time requirements for some of the applications, particularly in the computations of the ground area attainable and the vehicle's flight trajectory, it appears necessary to time-share the computer so that it can perform the many tasks required. A preliminary investigation indicates that a high-speed digital computer of the 7090 class, with 32,000-word core storage and having additional slow speed storage with appropriate buffering, would be able to meet the RCC computation requirements.

In addition to fulfilling the flight-path monitoring requirements, it is recommended that the RCC be mechanized to back up the normal airborne energy-management system. Thus, in situations involving airborne system failures, the RCC could provide information required to direct the spacecraft. It would consider all vehicle energy restraints and all limitations of the ground-based control system to provide all necessary control functions (i.e., to maneuver the vehicle's destination into the center of the GAA). Computations would be made to generate the necessary commands for atmospheric glide flight control and similar computations could be made to generate the retro command signals required to de-orbit the vehicle. The commands generated will be transmitted to the vehicle for display to the aircrew or be used within the spacecraft for automatic control.

In order to achieve the proposed tasks within the time restrictions, it is believed appropriate to require that the computation rate be sufficiently rapid that the total system would never respond appreciably to error or disturbance between computations. This time limitation not only includes computation time, but also data measurement, handling, and crew interpretation and decision-making time. Since the command signals are critical with respect to time, and do not require human interpretation by the RCC monitors, they could be transmitted directly to the crew and/or the spacecraft control system and be displayed in parallel to the flight control officer.

Display Complex

Since the primary mission of the Recovery Control Center would be to provide advisory control of the vehicle through the crew during the critical re-entry, approach, and landing operations, the flight control officer and other key center personnel must be presented with detailed critical information and summary-type displays so that they may be able to anticipate and recognize any potential danger to the crew, spacecraft, or flight plan. The flight control officer should have two-way voice communication with the crew so he may offer advice and instructions in a manner very similar to that of a GCA operator "talking down" an airplane pilot. To provide optimum advisory control and to maintain the confidence and morale of the crew, it is important that the officer on the ground be able to answer requests for information and guidance quickly and accurately, and be in a position to reach sound decisions rapidly. Therefore, he must be constantly aware of the status of the flight and of all important parameters. The most efficient method of communicating information to the human brain is through the sense of vision, hence the significance of a good display system.

The display system would consist essentially of two parts. The first of these, the system monitoring consoles, would show the important control parameters, such as maneuver volumes, velocity-altitude profiles, and ground area attainable, together with deviations from optimum and allowable safety limits. They would be designed for rapid accurate assessments in real or required time of the mission status. The second part would consist of various summary displays, each of which present to the center personnel information of a particular kind. These displays include a status board, status indicators, and a dynamic situation display. They would be designed for "quick-glance" summary-type monitoring by control center personnel, and be readily visible from any point in the operations area.

System Monitoring Consoles

Preliminary analysis[1,2] has determined that the system monitoring units should consist of three principle consoles - vehicle status, navigational, and aeromedical. A proposed layout of each console showing displayed parameters, scales, and location is given in Figures 4, 5 and 6. These units are to a great degree, similar to, but provide much more detailed and additional data than, the summary displays. In addition, they contain screens for displaying both analog and digital information, and possess other display devices peculiar to the RCC monitoring needs such as scopes, display meters, and continuous trace recorders to provide current status and time histories of the measured quantities. Each of these consoles would be manned by the key center operations officers. Under normal situations, the flight control officer assumes advisory control responsibilities of the flight. The aeronautical and aeromedical monitors should be able to communicate with the flight control officer and crew when necessary.

Vehicle Status Console. The proposed major flight parameters to be monitored by the aeronautical officer on the vehicle status console are illustrated in Figure 4 and are listed below.

1. Load factor - no. of g's vs velocity
2. Longitudinal mode - velocity vs angle of attack
3. Lateral mode - velocity vs angle of sideslip
4. Roll mode - velocity vs roll rate
5. Power - thrust chamber pressure
6. Skin temperature - right edge, left edge, nose, and lower surface, etc.
7. Strains - five critical areas
8. Flutter margin - wing, tail surface, and critical panels
9. Fuel quantity - propulsion, reaction control
10. Various indicators - indicate operations status

Navigation Console. The proposed navigation panel, illustrated in Figure 5 , would be monitored by the flight control officer, and displays the following information:

1. Vertical velocity - programmed and actual
2. Along course velocity - programmed and actual
3. Cross course velocity - programmed and actual
4. Flight path angle - programmed and actual
5. Angle of attack - programmed and actual
6. Bank angle - programmed and actual
7. Heading
8. Sideslip angle
9. Vector velocity
10. Altitude - programmed and actual

In addition, the navigation console would contain displays of the following parameters:

1. Maneuver volume (elevation view)- optimum and actual flight paths with reference to the altitude and range to go superimposed on the computed maneuver volume for the selected location.
2. Maneuver volume (plan view) - optimum and actual flight paths with reference to cross range and range to go superimposed on the computed maneuver volume for the selected location.
3. Velocity (altitude profile) - optimum and actual vehicle flight paths as well as limitations imposed by estimated vehicle restraints. The display will be indicated in altitude vs velocity space and is most useful to determine approach to dangerous situations.

4. Ground area attainable - a map displaying the availability of selected and alternate landing sites as a function of present energy level of the spacecraft. It will also be used to interpret astronaut spacecraft actions, since the astronaut will use this type of display within the space vehicle predominantly as his basis for decision-making.

These four navigational displays will present quick glance information of flight status within view of all key center personnel. Each display would use a large CRT (21 or 24 in) mounted above the navigation console, and would also contain its own magnetic drum unit and the electronics required to present the data described above as well as the scales of the parameters being displayed. Provision would be made to change scales automatically as a function of range to go or manually at the discretion of the flight control officer.

Aeromedical Console. The proposed aeromedical panel illustrated in Figure 6 would be monitored by the aeromedical officer. Data concerning the parameters displayed, their range, and accuracy requirements are shown in this figure.

In addition to these consoles, a fourth console would be monitored by the range communications officer who will establish and maintain communications with tracking stations, other centers and agencies, and alternate landing sites. A fifth console will be monitored by the recovery traffic officer who will handle all details relevant to chase plane vectoring, rescue and recovery units, and air-traffic control.

Dynamic Situation Display

A presentation is required which will allow the center's personnel to trace out the spacecraft's flight path on the earth's surface. This display would present general situation information on a world map display which will include the present position of the vehicle, location of the tracking sites and control centers, the flight-path history, and the predicted trajectory of the vehicle. The tracking station coverage will be indicated by lighted circles with the radii of the circles showing the approximate surface area covered by each station. One color could denote operational status; another could denote contact with the vehicle. The dynamic display need not indicate extensive detail, as this may be obtained from the other displays. It appears highly desirable to present the most probable future path of the vehicle which can be obtained from the simulated trajectory that is being computed on a faster-than-real-time basis. In addition, elapsed time and time to go could be presented in digital form as supplements to this display. Whenever a flight is aborted, a flashing red light above the center of the world map would immediately signify this situation.

Dynamic displays would be used primarily to supplement the capabilities of the status boards and the system's monitor consoles. The dynamic display will have less precision that the monitoring displays and will display curves and symbols or characters. The displays will be able to show, in either real-time or one of several multiples of real-time, the path that is being followed by the vehicle. This is accomplished by sending digital information from the computer to the dynamic displays on the wall map (and the navigation panel). This digital information will consist of the identification of the vehicle and its predicted position at specific time intervals in the future. The dynamic display will store this digital information and will perform calculations for organizing these points by means of digital equipment. In this way, the dynamic display will move a spot corresponding to the vehicle in such a manner as to give a smooth curve. The dynamic display data could be stored on a magnetic drum so that a path might be traced in accelerated time or switched to real-time. The picture displayed will show not only the present location of the space vehicle but also the optimum path it should flow. Upon requests for dynamic display, the required data concerning the space vehicle is calculated, organized, and transferred to dynamic display storage. Calculations are made from the display table of truncated elements. Starting at the time inserted by request, $\underline{x}$, $\underline{y}$, $\underline{z}$ and $\underline{t}$ coordinates of the trajectory point are calculated. These coordinates are transferred to the appropriate coordinate system for the map projection (overlay) requested.

Status Board

The status board consists of alphanumeric listings of data that convey to the viewer the general situation at the centers and tracking sites. The data presented will inform the viewer of the specific equipment available at each station and the phase (alerted, contacted) in which the station is functioning. Colored lights adjacent to each station would indicate status of equipment and completion of each specific mission.

Status Indicators

Status indicators are supplementary digital wall displays to give personnel within the system's operation room and the briefing room fundamental data of the flight situation including altitude, velocity, range to go, and cross-range of the spacecraft. This information would be derived from the center computer.

Clocks displaying the time (hours, minutes, and seconds) at the Launch Control Center, the first re-entry tracking station and the Recovery Control Center, and Greenwich Mean Time would be placed above the wall displays.

Personnel Subsystem

Although a high degree of automation is built into the system, the human operators perform many vital responsibilities of monitoring, communicating, maintaining, interpreting, and decision making. There is also a requirement to keep complete records of incoming sensor data and resulting center responses. This is necessary for scientific and mission testing analyses. The general responsibilities of the RCC personnel are summarized to be as follows:

1. Monitor the space flight - This includes monitoring the flight parameters to observe the navigational, aeromedical, and aerodynamic status of the spacecraft and aircrew.

2. Communicate with the crew - This involves two-way voice communications with the crew to advise them of impending dangers and to recommend improved flight paths.

3. Monitor system equipment operation - This includes checking the status of both control center equipment and sensor site equipment.

4. Systems analysis and programming - This entails the development of diagnostic, simulation and systems operation computer routines to facilitate in the operation, analysis and checkout of the components and total system.

5. Maintain and repair - This involves preventive maintenance and emergency repairs.

6. Operate computer and communications equipment - This requires the starting or stopping of the center's equipment or the checking of a new operation and communicating with adjacent tracking stations and the Command or Mission Control Center.

7. Collect data - This involves collecting data from heterogeneous sensors, intelligence, and other sources and inserting it into computer files.

8. Publish data - This entails publishing and disseminating bulletins and reports.

9. Training and briefing - This involves training of new personnel and briefing of visitors.

Detailed task analyses for many of the center's operational personnel including the aeromedical, aeronautical, flight control, range communication, and recovery traffic officers have been developed.

Voice Communication Complex

Based on the functional requirements for communications and the recovery system personnel responsibilities, the voice communication complex requirements have been determined in terms of which recovery control system personnel must be in communication with other recovery control system personnel, and what types of communications they must have. The information presented herein specifies the minimum voice communication requirements for recovery of a spacecraft and does not preclude other communications if desired. The proposed RCC voice communication complex is presented graphically on Figure 7. Four types of voice communication requirements have been identified and specified as follows:

1. Private - These refer to communication channels between two persons providing communications between these personnel exclusively, and not interfering with other communication channels.

2. Party - These refer to communication channels between many personnel, all of whom can talk or listen at the same time. In a party communication link only one channel is required, but only one person may talk at any one time, although all personnel on the linkage may listen.

3. Broadcast - These communication refer to loudspeaker or intercom type communications where one person may direct communications to all other persons in the linkage at the same time.

4. Direct voice - These communications refer to direct person-to-person verbal communications with no equipment interceding.

In addition to the four types of communications described above, sending and receiving requirements are specified. The accompanying chart (Figure 7) indicates by means of arrowheads which personnel can send or receive to other personnel through the various communication channels.

Private Communications

There are five private communication channels required. These channels are as follows:

Channel 1 -Communication between the recovery traffic officer of the RCC and the local area air traffic control center

Channel 2 -Communications between the range communications officer and possibly the aeronautical or flight control officer at the RCC and other tracking stations or alternate recovery sites

Channel 3 -Communications between the recovery traffic officer at the RCC and the pilot and crew of the rescue helicopter

Channel 4 -Communications between the recovery traffic officer and the pilot of the chase aircraft

Channel 5 -Communications between the recovery traffic officer and the ground recovery unit controller in the control van.

Channel 2 will be used principally to maintain voice communications with the tracking stations to obtain reports on flight performance and telemetered data not transmitted directly to the RCC (i.e. comparison of actual and average vehicle attitudes) and to provide information to alternate recovery sites.

Party Communications

There are two party communication channels required. Each of these channels allows several personnel to communicate among themselves as required. The two necessary channels are:

Channel P-1 -Communications among the recovery traffic officer, the aeromedical officer, the aeronautical officer and the flight control officer.

Channel P-2 -Communications among the spacecraft, the chase aircraft, the rescue helicopter, the ground recovery unit controller, the flight control officer, the aeromedical officer, the aeronautical officer, and the recovery traffic officer. Tracking stations and alternate recovery sites would also be included in this channel under emergency conditions.

Channel P-1 would be used principally to coordinate activities of the four primary recovery personnel at the RCC. Channel P-2 would be used principally as the primary recovery operations channel. This channel will permit all recovery system personnel to monitor advisory control communications to the astronaut. This channel will also permit recovery control system personnel to communicate directly with the astronaut as required. Under normal conditions, the flight control officer will use this channel to communicate with the spacecraft from acquisition to the point where the chase aircraft intercepts the space vehicle. The chase aircraft will use this channel to provide advisory control communications to the astronaut from the interception point until touchdown of the spacecraft. The ground recovery unit controller will use this channel to communicate with the astronaut after touchdown. No other personnel will communicate directly with the spacecraft unless specifically directed to do so by the flight control officer.

Broadcast Communications

Broadcast communications are necessary for the ground recovery unit controller to direct and control all ground recovery related groups such as the crash units, the vehicle handling units, and all personnel assisting with handling the pilot and the vehicle.

Direct Voice Communications

Direct voice communications should be possible among the flight control officer, the aeromedical officer, the aeronautical officer, the range communications officer, and the recovery traffic officer. It appears that by optimum arrangement of key personnel and their monitoring consoles in the RCC layout, this direct voice communication could readily be accomplished.

Conclusions

This paper has presented the systems' requirements, and specified the personnel and equipment functions and requirements needed to achieve an operational Recovery Control Center.

To provide real-time advisory or command control of a maneuverable hypersonic vehicle involves computations and control problems far

more complex than those required for ballistic re-entry capsules. Voice communications through the plasma sheath and continuous computations in real-time of the spacecraft's ground area attainable and flight trajectories are some critical areas which require further investigation. The major tasks of the center are to:

1. Monitor displays of the spacecraft trajectory throughout re-entry, approach, and landing.

2. Monitor spacecraft status and perform computations which will provide data to advise the crew of present status and the corrective actions required to achieve a safe recovery.

3. Monitor vehicle crew status and be capable of assuming control of the spacecraft if the crew is incapacitated or if such control is requested by the pilot of the vehicle.

4. Maintain voice communications with the crew throughout the entire re-entry and landing phases.

5. Direct the deployment of recovery forces.

To accomplish these tasks, an integrated complex of data processing, display, monitoring, and communications is necessary. Implementation of such a recovery complex described here appears to be feasible and necessary to ensure a high probability of success in the recovery of future spacecraft.

Acknowledgements

The author wishes to acknowledge and express appreciation to the following personnel who, as a result of their specialties, significantly contributed to the various aspects of this paper:
Mr. Robert L. Schroeder, Dr. Julius King, Jr., Mr. Larry Monroe, and Mr. John Zvara of Raytheon Company (systems requirements); Dr. Thomas A. Hussman, Jr. of the Matrix Corporation (human factors and displays); and Dr. Paul Webb of Webb Associates (biomedical requirements).

References

1. Inba, K. et al, Human Factors Research Support in the Development of a Flight Safety Monitoring Console, Psychological Research Associates, September 1959.

2. Investigation of Media and Design of a Console for Real-Time Presentation of In-Flight Telemetered Data, Final Report, ITTL Avionics Laboratory, November 1959.

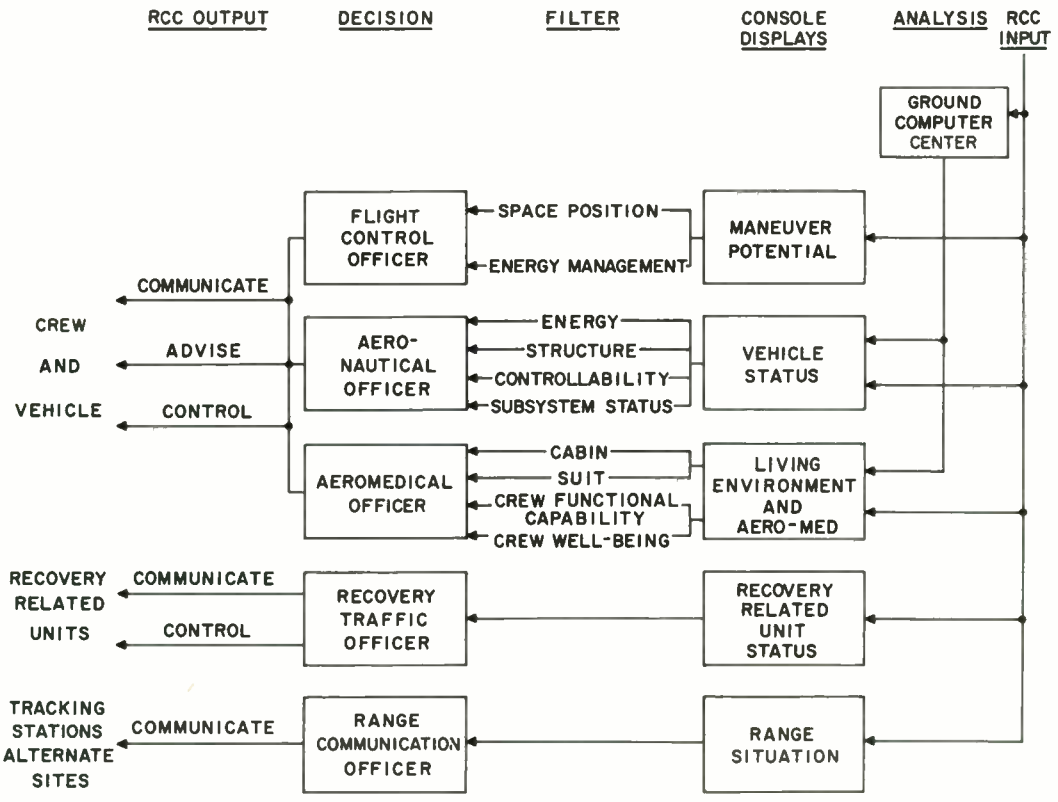


Fig. 1. RCC operational information flow.

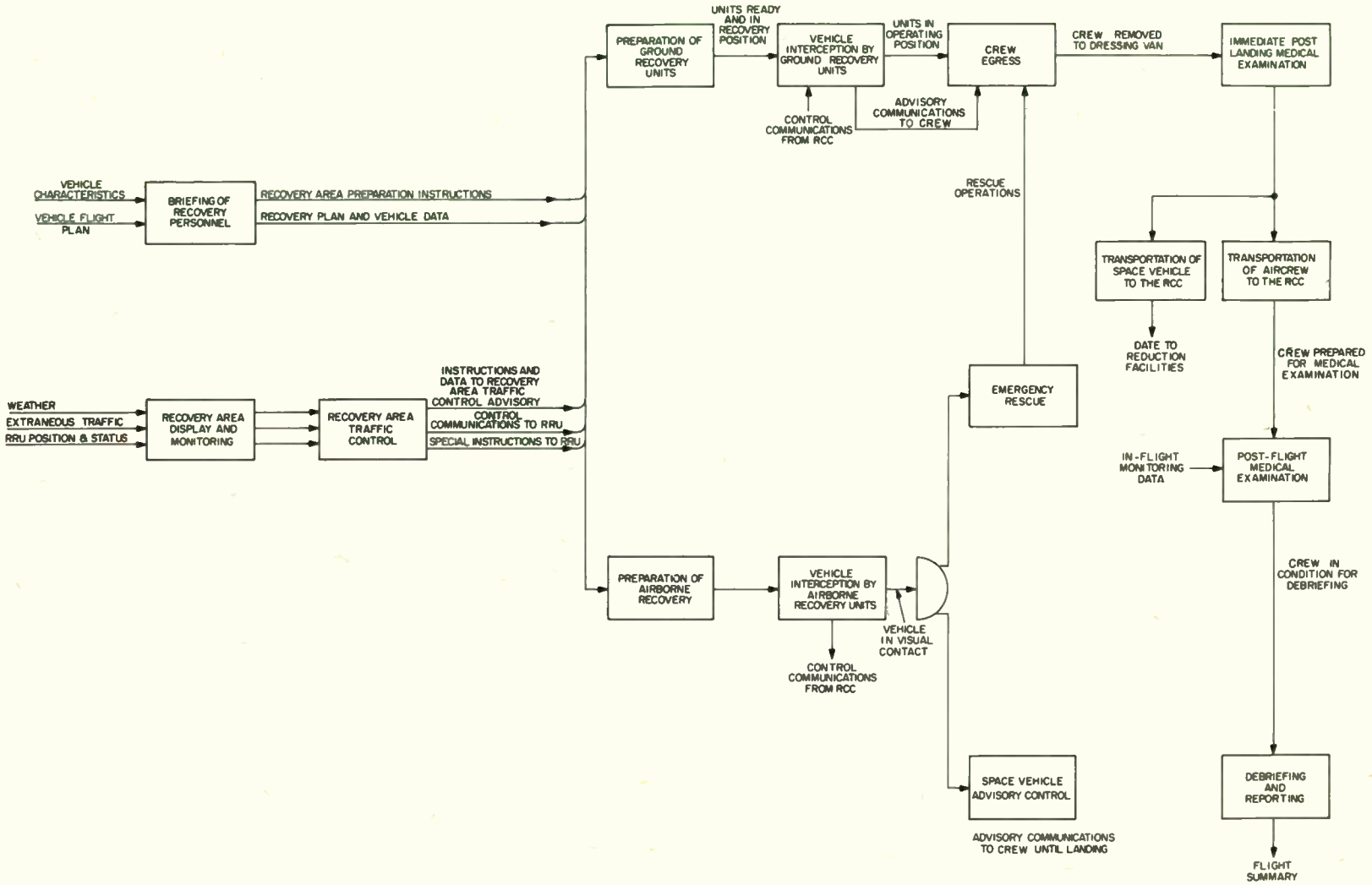


Fig. 2. RCC functional requirements.

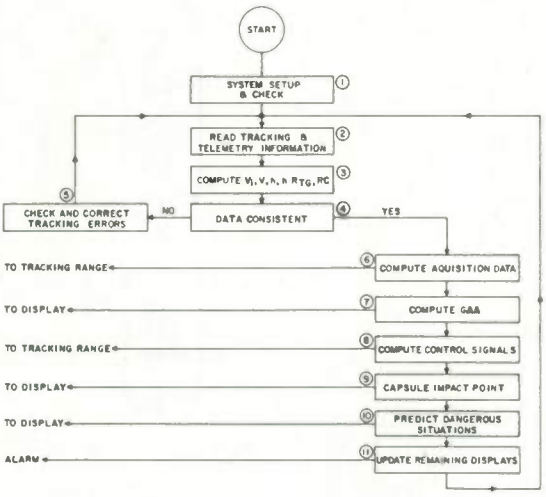


Fig. 3. RCC computational requirements.

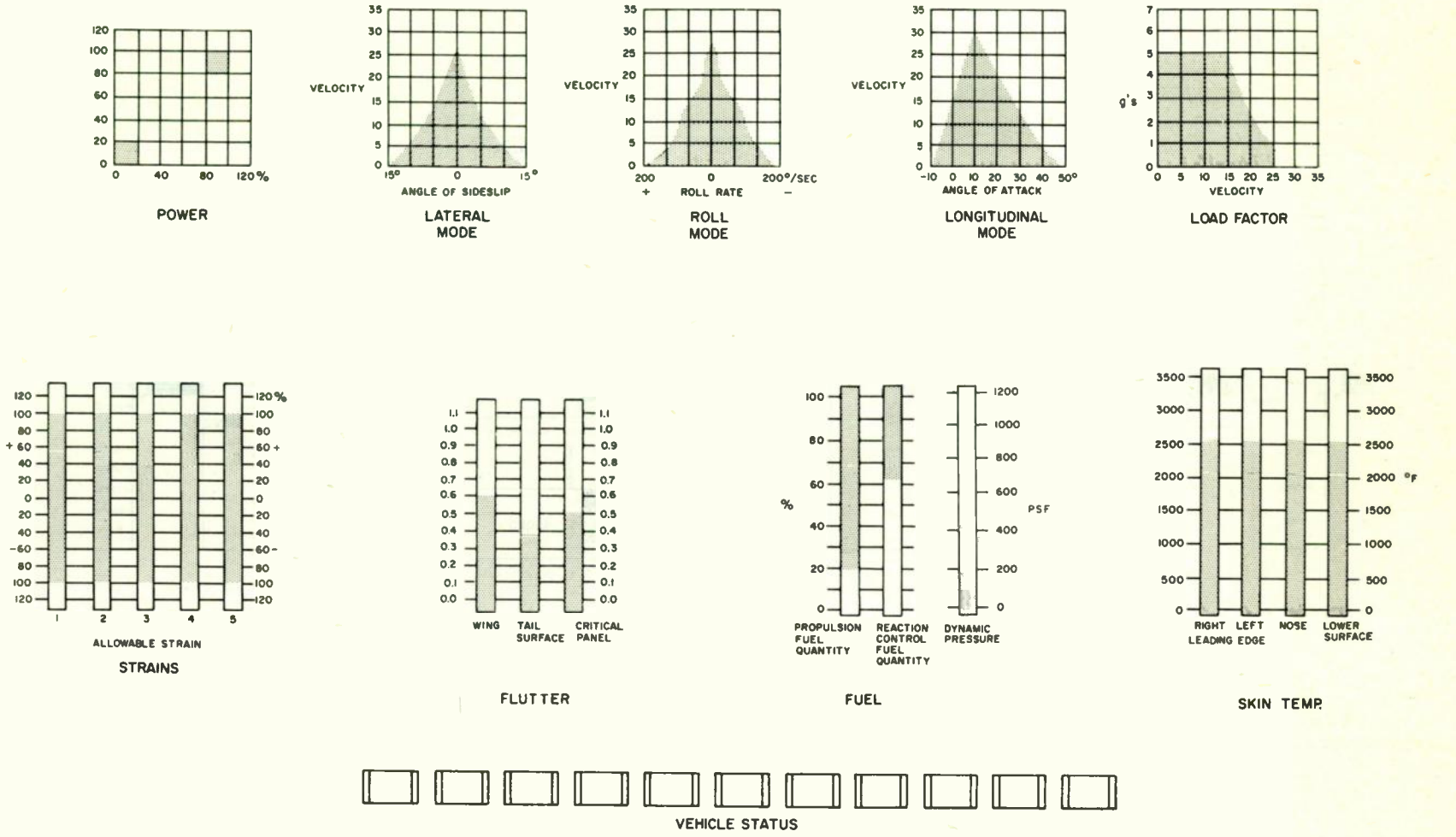


Fig. 4. Vehicle status console layout.

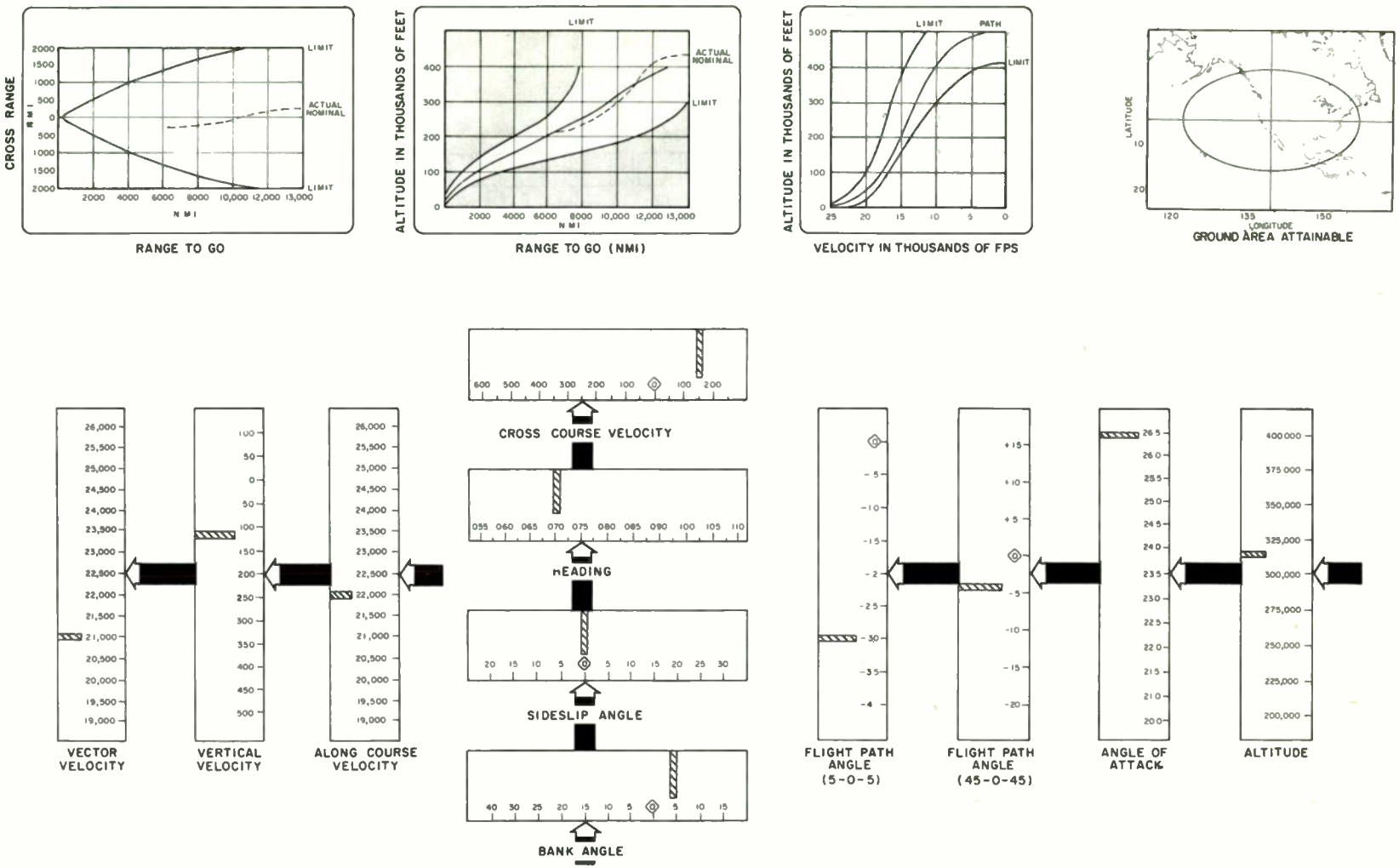


Fig. 5. Navigational console layout.

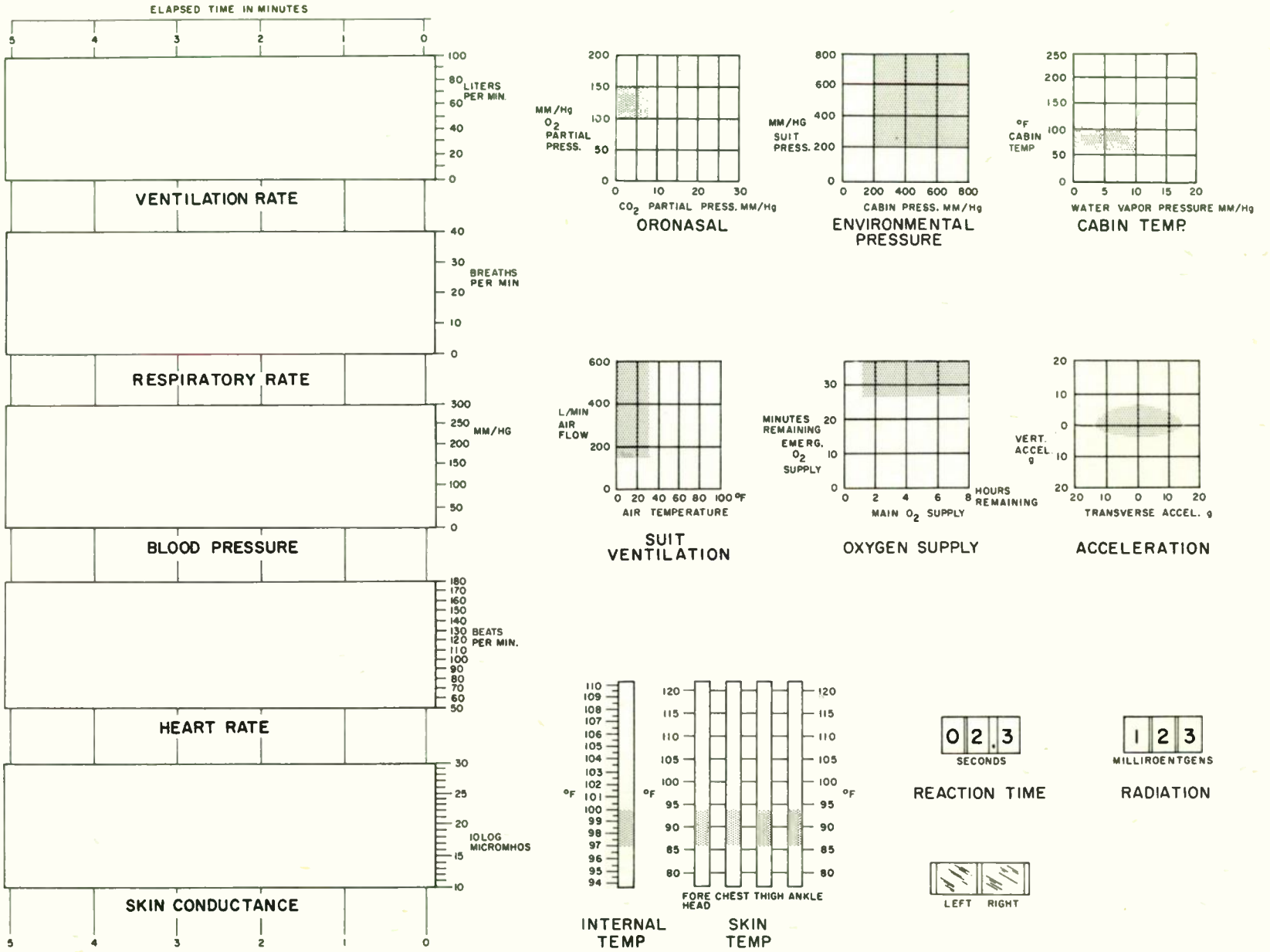


Fig. 6. Aeromedical console layout.

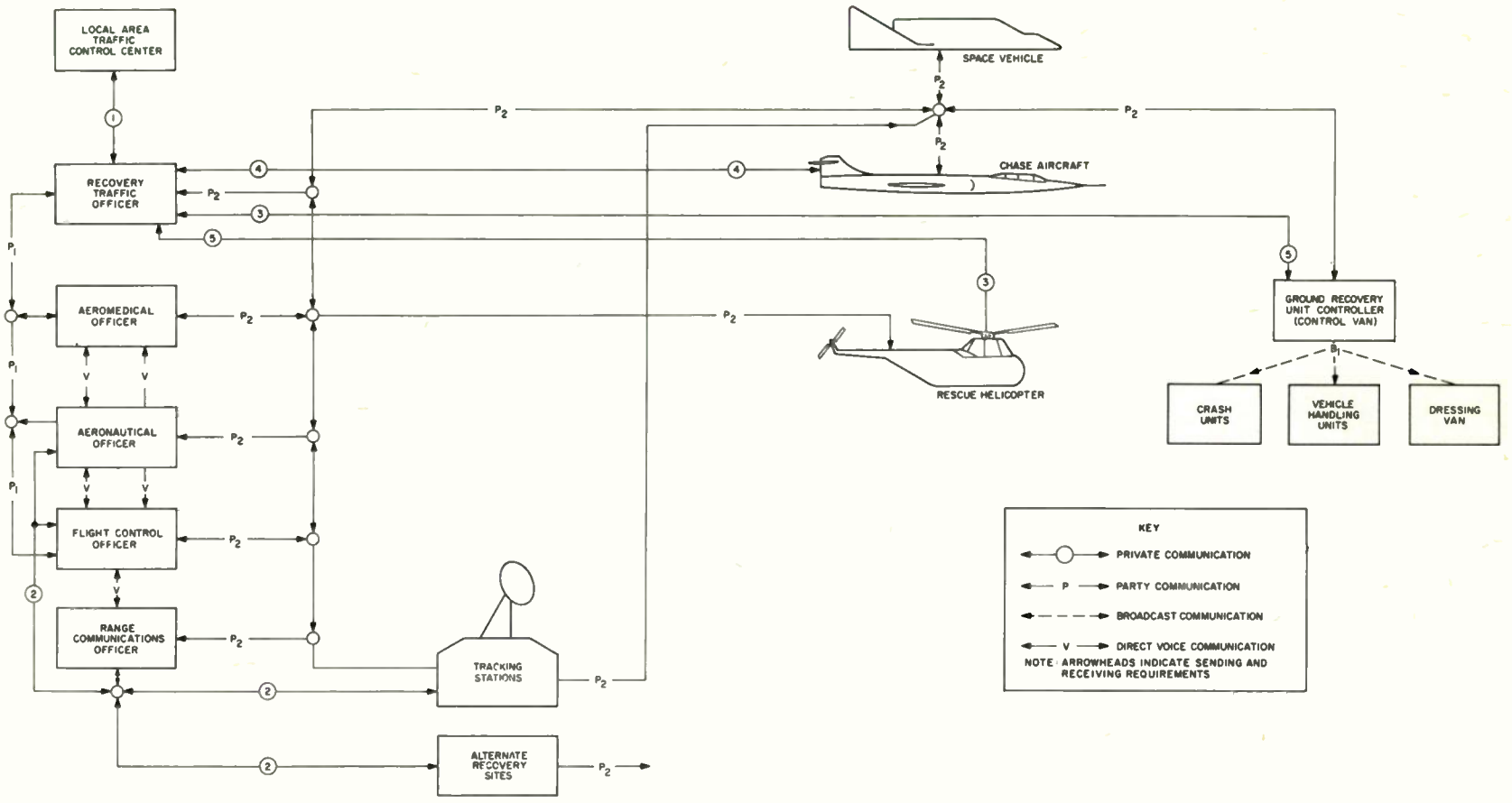


Fig. 7. Proposed communication system for the recovery complex.

MANNED-SATELLITE RE-ENTRY and LANDING GUIDANCE SYSTEM*

D.K. Breseke, W.G. Green**, R.J. Reis, B.K. Robrock, & J.R. Steinhilber
ITT Federal Laboratories
Astrionics Laboratory
Fort Wayne, Indiana

Summary

A system is described and discussed for achieving satellite departure from an earth orbit, atmospheric re-entry and a soft landing at a predetermined geographic location. Its self-contained guidance allows considerable flexibility of initial orbit conditions. This provides the pilot-passenger with enhanced freedom of choice and non-dependence upon ground tracking and communication. This could be important for abort capabilities or during severe ground weather conditions. The vehicle's non-lifting characteristics and reliance upon drag modulation only eliminates the hazards of skipping re-entry and its consequential landing point range control problems. The ability of the system to land within a time interval of less than an orbit period is also an important feature under abort conditions. The combination of shorter time-of-flight and improved landing point control are major improvements over other drag modulation systems presented in the literature.

Introduction

A system will be described and discussed for achieving satellite departure from an earth orbit, atmospheric re-entry and a soft-landing at a predetermined geographic location.[1] The specifications for system design are that the landing occur within a few nautical miles of a specified location, with a touchdown velocity less than 25 fps. Maximum deceleration is 8 g and the guidance system is self-contained in that all intelligence gathering ability is in the satellite. This does not rule out the possibility of using ground transmitters simply as signal sources.

The system must perform the following basic functions:

a. Determine its own orbit and its position in that orbit relative to the landing site (orbit determination phase).

b. Determine a suitable transfer orbit, which will ballistically convey the satellite from the initial orbit to acceptable re-entry conditions, and then eject the vehicle upon that transfer orbit. (orbit departure phase).

c. Predict the vehicle's position relative to the landing site at re-entry such that re-entry guidance can steer vehicle properly. (ballistic flight and re-entry prediction phase).

d. Control the vehicle's path during re-entry and atmospheric flight to bring vehicle to zero horizontal velocity over landing site. (re-entry guidance and control phase).

e. Provide a final descent through the lower atmosphere, and decelerate vehicle to provide soft, vertical landing at desired location. (final descent phase).

Table 1 shows these five phases of operation as a function of altitude. Figure 1 provides a two-dimensional view of the application of the different phases. The section on the re-entry guidance and control phase includes the analysis of the application of one guidance law to the drag modulation guidance technique.

Orbit Determination Phase

In this phase the system must determine the satellite's position in orbit as a function of time relative to the desired landing site. The orbit must be determined very precisely since it represents an initial condition for the other phases of the system.

Of primary concern, in the development of a system which is to terminate in a particular target area, is the determination of a reference from which measurements and computations can proceed. In the case of recovering an orbiting satellite, a logical reference is the orbit itself which includes all possible initial conditions from which the maneuver may initiate. The current method of determining a satellite orbit with ground tracking is not considered, since a primary requirement of the system under study is that it be as self-contained as possible. The lack of communication between satellite and ground tracking centers, except possibly as a backup, implies that onboard sensor equipment and computers are required to determine the reference.

*The research, reported in this paper, was partially sponsored by the Navigation and Guidance Laboratory, Aeronautical Systems Division, Air Force Systems Command, under Contract No. AF33(616)-7049 with ITT Federal Laboratories.

** Bendix Corporation, Bendix Systems Division, Ann Arbor, Michigan

Several possible orbital determination techniques employ physical characteristics of the earth: the occultation of diametrically opposed stars by the earth's shadow (abberascope or occultascope),[2] the gravitational force field about the earth (accelerometer mapping), surface characteristics (optical mapping) and surface reflectivity of radar signals (Doppler velocity). The list may be extended to include man-made information sources of a passive nature such as a single beacon at the target (Doppler frequency matching) and arrays of strategically placed beacons. Of these various techniques, only the Doppler frequency mapping, which is the technique used in the Transit program,[3] appears to have the required accuracy, simplicity, and availability within the near future.

The Reverse Transit process,[1] as is proposed for this particular application, reverses the roles of the satellite and ground stations in that the transmitter is based at an accurately known ground site; the signal reception, Doppler frequency shift determination and data reduction to orbital parameters occur on board the satellite. This portion of the system reduces the uncertainties of orbit parameters from almost complete uncertainty to knowledge of orbit and vehicle position in orbit to better than 0.5 nautical mile.[3] This orbit determination procedure requires about 1/3 of an orbit period prior to the initiation of the orbit departure phase.

Orbit Departure Phase

During this phase the capability will be provided to compute an acceptable ballistic transfer orbit, the time to initiate departure from the initial orbit, and the thrusting requirements to accomplish the ejection maneuver. Computations may begin as soon as the function of resolving the position of the satellite in its orbit is completed. This is accomplished during the orbit determination phase. The computation time may require as much as 1/3 of an orbit period predicated on the number of cyclic computations necessary to find an acceptable transfer condition.

The execution of the ejection maneuver is required during this phase. This maneuver requires the erection of the inertial platform and control of the timing, direction and magnitude of the thrusting operation. A local vertical tracker[4, 5,6,7] is required to initially erect the inertial system (platform) which provides the inertial space referenced coordinate system. The plane of the satellite orbit is resolved with a null-seeking gyro compass located on the platform. A computer is required to provide the coordinate transformation function in addition to the computations indicated above. Orientation of the vehicle during the platform erection phase and the orbit ejection thrusting phase is maintained by the attitude control system. During these critical stages the attitude of the vehicle should be maintained very accurately. The accelerometers of the inertial system will be used to monitor the magnitude of the thrusting required to eject the vehicle from the initial orbit. The thrusting must be controlled through the integrated outputs of the platform accelerometers to provide velocity impulses to an accuracy commensurate with the Reverse Transit data. This may require restraint on the thrusting levels through vernier control.

The satellite vehicle is removed, during this phase, from the initial orbit onto a ballistic transfer orbit that will achieve re-entry with specified re-entry conditions.

Ballistic Flight and Re-Entry Prediction Phase

A system requirement exists for the prediction of the re-entry conditions based upon the data available at the end of thrusting of the orbit departure phase. Execution of this function will reduce the effects of the system errors that exist at the re-entry window. These errors result from the inaccuracies of both the instrumentation and orbit determination data. Uncompensated system errors at this point could introduce an error of some 15 percent (of the total ground range at re-entry) between the actual re-entry position of the vehicle and the predicted re-entry position.

Output data from the orbit departure system is used to predict the re-entry conditions that should exist at some predetermined altitude after an elapsed time interval. The local altitude, the elapsed time interval, the transfer orbit ephemeris and the conditions at the end of the orbit departure phase provide sufficient data to compute the more accurate prediction of the re-entry conditions. The clock used in the Reverse Transit system would be applicable to initiate the altitude measurement. Attitude control is required since the altimeter operation requires a specific vehicle orientation because of the directivity of the altimeter propagation pattern. Execution of this phase reduces the unknown position error in the transfer orbit plane from 15 percent to less than 1 percent.

Re-Entry Guidance and Control Phase

In this phase the system is required to accept the variations in re-entry conditions, constrain the deceleration and heating to tolerable values and, at the same time, maintain range control of the resulting trajectory.

The inertial platform provides the computer with measured data of the vehicle's accelerations which are integrated to obtain velocities and positions relative to the predicted landing site. Thus, a measure of the ground range from the vehicle subpoint to the landing area and the corresponding range rate are available for calculating the desired acceleration utilizing a guidance law. The computed acceleration is compared with the actual vehicle acceleration obtained from the inertial system and the difference, if any, is used as an error signal to control the variable drag device[8] to null out the difference.

The inability to accurately predetermine at-

mospheric density and wind effects dictates the use of a closed loop guidance technique which can also remove known initial condition errors effectively. Since precision range control through drag variation is the desired objective, it is obvious that the guidance law should be functionally dependent upon range and/or range rate. From a limited investigation of all pertinent guidance laws, the hyperbolic cosecant (CSCH) modulation law was chosen as being particularly suitable for range control of the nominal trajectory conditions. This law involves controlling range acceleration as a function of range and range rate. The law is logarithmic in nature in that simultaneous zero values for range acceleration, range rate, and range occur at termination and it is insensitive to initial conditions. The formulation of the law includes two independent shaping parameters which allow a close fit of the nominal ballistic trajectory. Thus, the modulation capability is reserved for initial condition errors and atmospheric uncertainties.

Although no explicit criteria was chosen to determine acceptability of a particular law, it is clear that a "good" law will require close agreement with the nominal ballistic trajectory in peak acceleration, peak heating, and terminal velocity and, also, that required drag variation will be low. Minimizing drag variation required to follow the nominal ballistic trajectory will reserve the drag variation capability for atmospheric uncertainty and initial condition errors. This minimization was investigated using the angular range to the target, β, as the controlled variable and selecting the parameters of the guidance law to produce a match between the $\ddot{\beta}$ vs β curve of the guidance law and the corresponding curve for the nominal ballistic trajectory. The criterion of the match was taken as minimization of the maximum departure rather than an average departure.

An indication of the match obtainable with the CSCH modulation law, equation 1

$$\ddot{\beta} = k\ \dot{\beta}^2\ \ \mathrm{CSCH}\ \ a\beta \qquad (1)$$

is shown in Figure 2. The match obtained with the CSCH law was superior to the match obtained from other laws studied for the selected nominal ballistic trajectory.

A variation of the parameter values by 10 per cent in the range of $k/a = 0.808$, and $a\beta_0/2 = 2.75$, changes the fit very slightly. The curve fitting is but an indication of proper parameter values, and this fit is considered adequate pending further analysis.

The effects of various deviations from the standard conditions were evaluated by analog and digital computer simulations to show the drag variations required to compensate for uncertainties in initial conditions and instrumentation errors. An indication of the capabilities of the CSCH law to correct for atmospheric variations (seasonal, diurnal, and other uncertainties in knowledge of density) is given in Table 2.

Re-entry velocity magnitude deviations of ±1 percent, ±5 percent, and ±10 percent were injected and the resulting variations in drag as a function of range for the initial velocity variations is shown in Figure 3. More than a 20 to 1 variation in drag is required for the ±10 percent initial velocity change. The ±5 percent change requires about a 20 to 1 drag variation and the ±1 percent change requires only a 1.5 to 1 drag variation.

Sensitivity to re-entry flight path angle deviations was determined by changing the angle ±25 percent and ±50 percent. It was evident early in the ±50 percent runs that this large change would require far more than a 20 to 1 variation in drag and hence the run was terminated. The ±25 percent change was marginal and required very nearly a 20 to 1 drag variation. The drag factor versus range curves for the re-entry flight path angle deviations are shown in Figure 4.

Effects of changing the re-entry range to target were investigated. The required drag variation to compensate for a ±15 percent variation of re-entry range to the target from the nominal value would be about 5.5 to 1, while the drag variation to correct for a ±25 percent range change would be slightly greater than 20 to 1.

In order to determine the effects of the dynamic response of the drag brake and associated control elements, a second-order lag network was used to simulate the dynamic error between the calculated desired deceleration and the actual deceleration. A network with a natural frequency of 0.1 cycle per second and a damping factor of 0.7 was used. The total drag variation required with the lag network in the control loop was 1.6 to 1.

Final Descent Phase

This phase of the system covers the vehicle flight during the vertical descent from 100,000 feet to the landing area; and the deceleration of the vehicle at the terminal point to effect a soft-vertical landing.

Investigations indicate that horizontal disturbances have little or no effect upon the vertical flight in this low altitude region, thus allowing the vertical and horizontal problems to be analyzed separately. The horizontal problem is affected by the vertical profile history, since a long time of fall increases the time interval that the wind forces operate upon the vehicle. Minimizing the drag at 100,000 feet (versus falling with maximum drag) reduces the wind produced range error at impact by a factor of four; however, the impact velocity is increased from 36 fps to 180 fps. A compromise between impact velocity and range errors may be obtained by minimizing drag from about 100,000 feet altitude to 5,000 feet altitude and maximizing the drag again to decelerate the vehicle. This compromise allows the impact velocity to be about 36 fps while the effects of winds produce a range error of only a few nautical miles. Reduction of the impact

velocity to a value below 25 fps requires the addition of retro-thrusting or other type of decelerating force. Analysis of the problem indicates that a constant thrust engine ignited at about 200 feet altitude, with thrusting force less than gravitational forces, would materially augment the decelerating forces on the vehicle and produce touchdown velocities of less than 25 fps with rather loose tolerances upon ignition delays, thrust levels, etc.

The operation in the final descent phase in the region of the landing area requires a knowledge of the altitude around 5000 feet and 200 feet. State of the art, operational, radar altimeters[9] using triangular wave modulation would more than meet the requirements for the final descent phase.

System Instrumentation

Figure 5 is an instrumentation block diagram of the complete system showing the basic ties between the vehicle functions and the ground based equipments. The computer computation functions are enclosed in a separate block within the vehicle system block to illustrate the importance of the computer as well as the relative simplicity of the remaining functions. Table 3 is a listing of the instrumentation requirements of the system in accordance with the five basic phases of the re-entry mission. Detailed equipments specifications are presented in the classified Final Report, Reference 1.

Conclusion

A system has been described and discussed for achieving satellite departure from an earth orbit, atmospheric re-entry and a soft-landing at a predetermined geographic location. Its self-contained guidance system allows considerable flexibility of initial orbit conditions. This provides the pilot-passenger with enhanced freedom of choice and non-dependence upon ground tracking and communication. This could be important for abort capabilities or during severe ground weather conditions. The vehicle's non-lifting characteristics and sole reliance upon drag modulation eliminates the hazards of skipping re-entry and its consequential landing point range control problems. The ability of the system to land within a time interval of less than an orbit period is also an important feature under abort conditions. The combination of shorter time-of-flight and improved landing point control are major improvements over other drag modulation systems presented in the literature.

References

1 Astrionics Laboratory, "Final Report, Study and Development of Terminal Guidance for Interplanetary Vehicle, Vol. II Terminal Guidance for Re-Entry and Landing of Earth Satellites", ASD Technical Report 61-418 Contract No. AF33(616)-7049, ITT Federal Laboratories, Fort Wayne, Indiana, October 1961 (Classified).

2 General Mills, Inc., "Stellar Aberrascope Study Final Report", Contract AF33(616)-7400, December 1960.

3 Guier, W. H., Weiffenbach, G. C., "A Satellite Doppler Navigation System", Proc IRE, Vol. 48, No. 4, April 1960.

4 Astrionics Laboratory, "Technical Proposal for a Vertical Reference System for Space Vehicles," ITT Federal Laboratories, Fort Wayne, Indiana, February 1959.

5 Astrionics Laboratory, "Technical Proposal for a Horizon Sensor System", ITT Laboratories, Fort Wayne, Indiana, December, 1959.

6 Roberson, Robert E., "Methods for the Control of Satellites and Space Vehicles", Systems Corporation of America, 31 July 1960.

7 Research Division, Radiation Incorporated, "Final Technical Report Study of Satellite Guidance Sensors", (AF33(616)-5988) 30 January 1961.

8 Avco-Everett Research Laboratory, "Study of a Drag Brake Satellite Recovery System", for Dyna-Soar Weapons Systems Project Office, WPAFB, Ohio, July 1960.

9 Astrionics Laboratory, "Final Report, Interplanetary Terminal Guidance Techniques Study", (AF33(616)-6083) ITT Laboratories, Fort Wayne, Indiana, October 1959 (Classified).

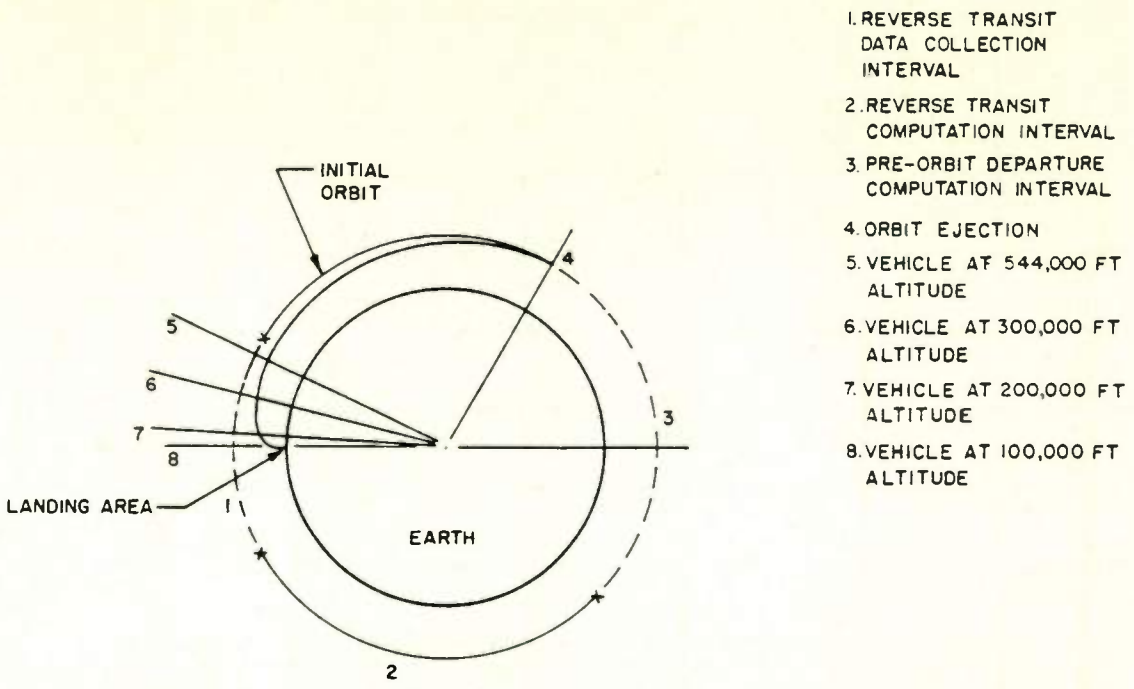


Fig. 1. Two-dimensional display system concept.

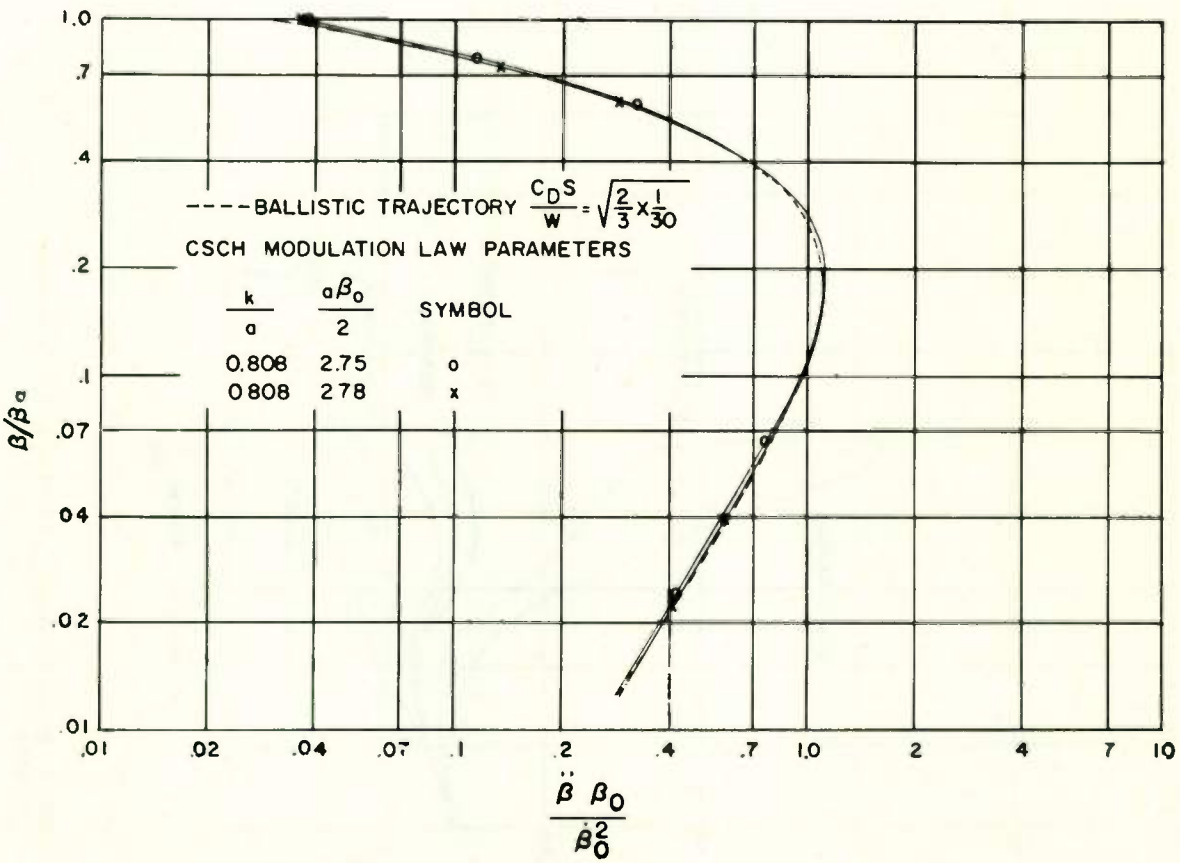


Fig. 2. Comparison of $\frac{\ddot{\beta}\beta_0}{\dot{\beta}_0^2}$ of ballistic trajectory and CSCH modulation law trajectory.

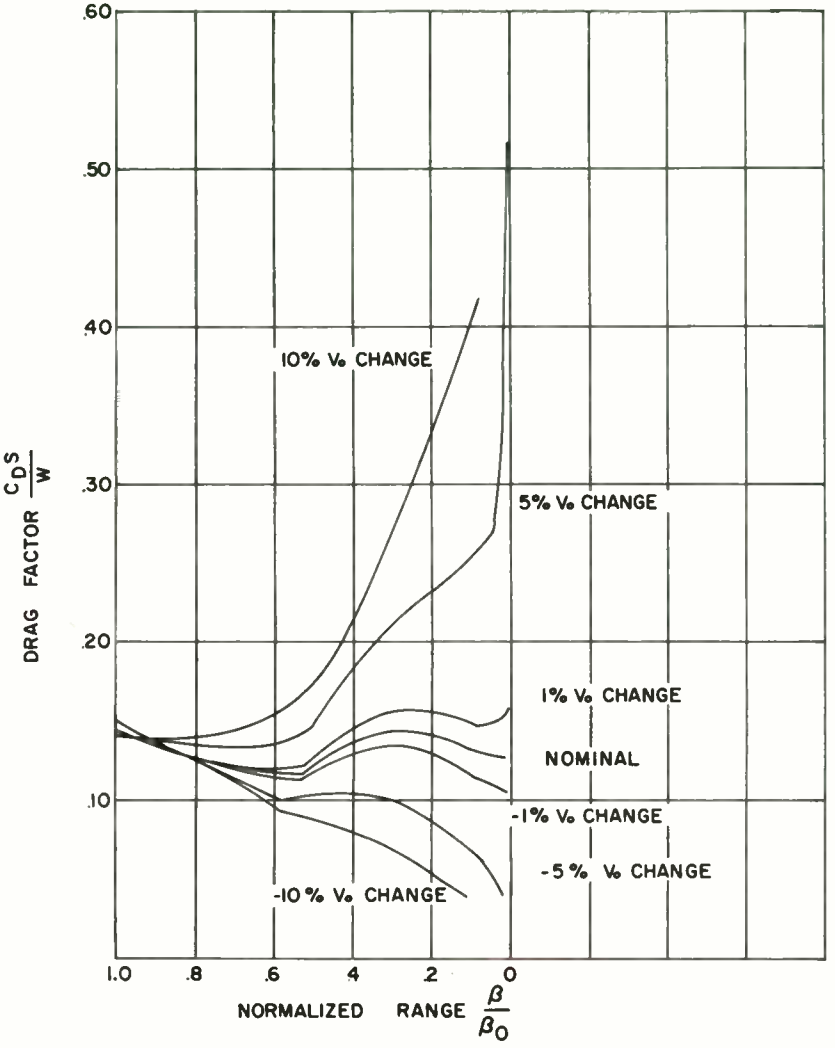


Fig. 3. Initial velocity (V_0) variation effects on drag profile.

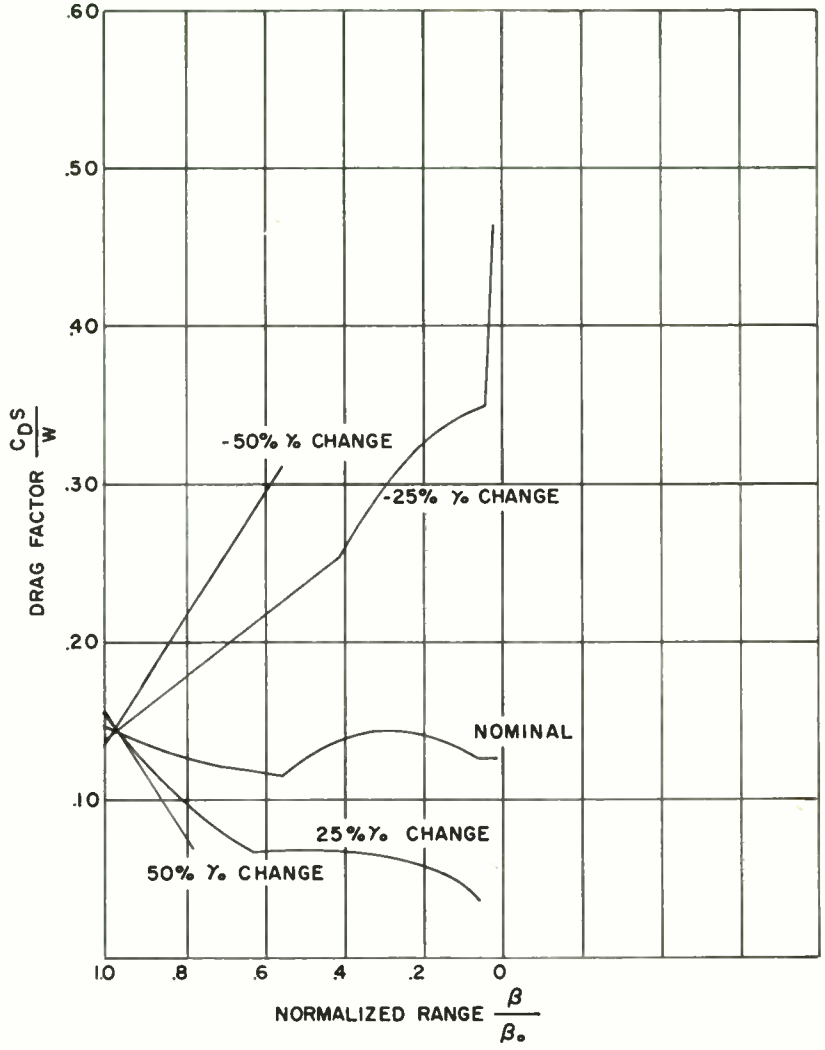


Fig. 4. Initial flight path angle (γ_0) variation effects on drag profile.

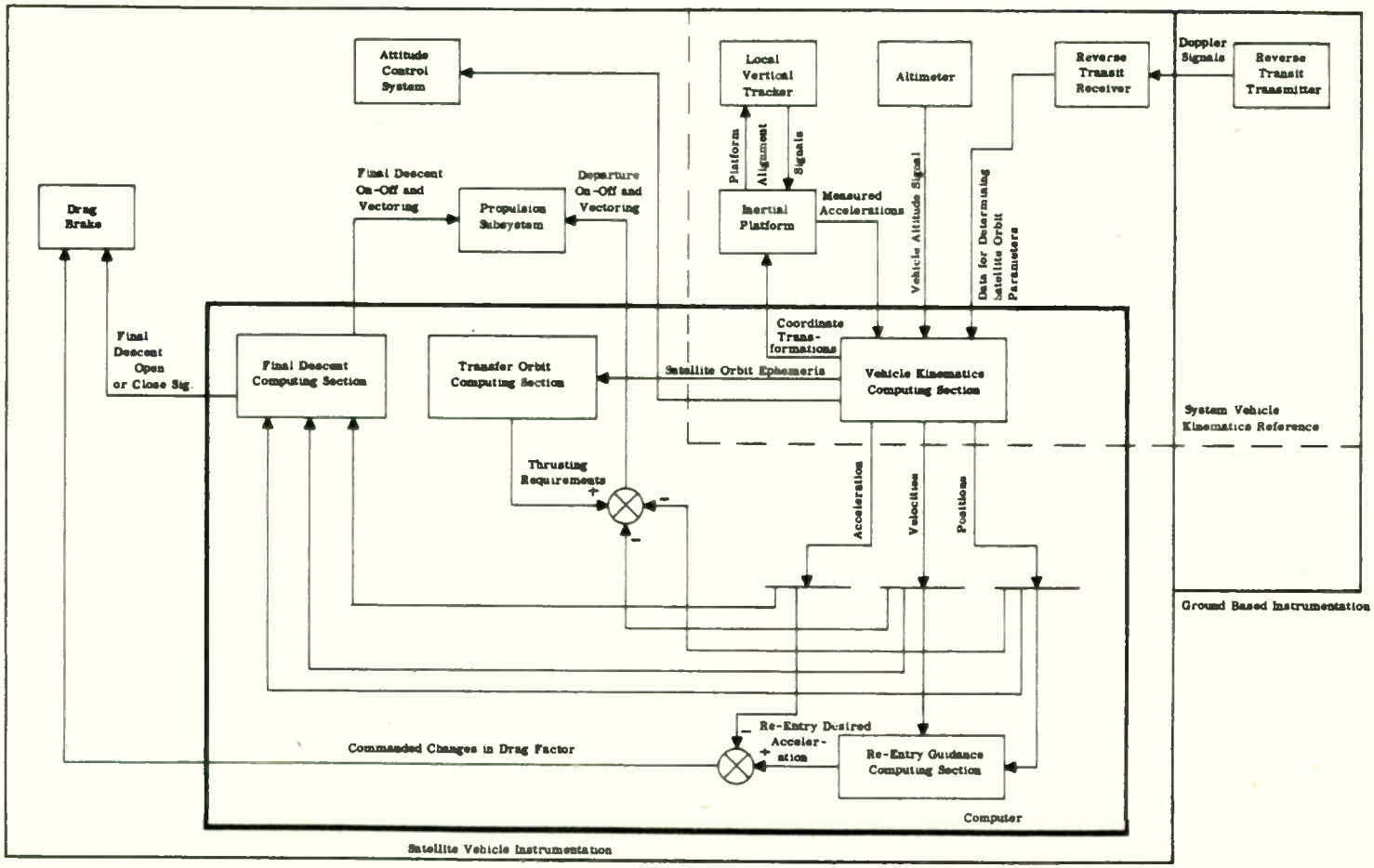


Fig. 5. System instrumentation block diagram.

Table 1

Altitude Profile of Events

Subsystem Phase	Altitude (ft)	Event
Orbit Determination	orbit altitude	Orbit determination (utilizes 1/3 of orbit period prior to initiation of orbit departure phase)
Orbit Departure	orbit altitude	Orbit departure (utilizes 1/3 of orbit period prior to orbit departure maneuver)(See notes below)
Ballistic Flight and Re-entry Prediction	554,000	Predicted re-entry initial conditions correction
	300,000	Drag modulation guidance system initiated
Re-Entry Guidance and Control		
	100,000	Drag brake closed
Final Descent	5,000	Drag brake opened
	200	Landing retro-thrusting initiated
	0	Touchdown

Note:

1. Geocentric angular range from orbit departure to touchdown is approximately 1/3 of a revolution.

2. Total time from orbit departure to touchdown is approximately 1/2 of orbit period.

Table 2

CSCH Modulation Law Capabilities in Presence of Atmospheric Variation

		Ballistic	CSCH	CSCH	CSCH
Atmosphere		Nominal	Nominal	Dense	Thin
Initial Altitude(h_o)	ft	298,566	298,566	298,566	298,566
Initial Velocity(V_o)	fps	25,809	25,809	25,809	25,809
Initial Flight Angle(γ_o)	Degrees	-2.146	-2.146	-2.146	-2.146
Initial Drag Factor $(C_D S/W)_o$	ft^2/lbs	0.1491	0.1491	0.1491	0.1491
Initial Temperature	^oR	671	671	671	671
Peak Temperature	^oR	1040	1038	2072	1782
Alt of Peak Temperature	ft	153,000	157,800	179,500	168,700
Peak Acceleration	ft/sec^2	270	263	275	276
Alt of Peak Acceleration	ft	200,300	200,700	200,700	202,800
Max Drag Factor (Above 100,000 ft)	ft^2/lbs	--	.1723	.1841	.6892
Alt of Max Drag Factor	ft	--	Initial	100,000	Initial
Min. Drag Factor (Above 100,000 ft)	ft^2/lbs	--	.1369	.09378	.1858
Alt of Min Drag Factor	ft	--	100,000	Initial	248,700
Velocity at 100,000 ft	fps	754	803	635	248,700
Flight Angle at 100,000ft	Degrees	-65.2	-62.2	-68.2	-67.4
Drag Factor at 100,000 ft	ft^2/lbs	--	0.1369	.1841	0.2811
Temperature at 100,000 ft	^oR	1031	1030	1902	1690
Range Angle change to 100,000 ft	% of initial reentry range	99.8	99.8	99.8	99.9
Required Drag Variation		--	1.26:1	1.96:1	3.71:1

The CSCH modulation law used for all of these trajectories has the parameter values of k/a = .8075 and $a\beta_o$ = 5.5

Table 3

System Instrumentation Diagram

	Instrumentation On Board Satellite								Instrumentation on Ground
	Computer	Inertial System	Reverse Transit	Altimeter	Local Vertical Tracker	Propulsion	Drag Brake	Attitude Control	Reverse Transit Transmitter
Orbit Determination	X		X						X
Orbit Departure	X	X			X	X		X	
Ballistic Flight and Re-entry Prediction	X			X				X	
Re-entry Guidance and Control	X	X		X			X	X	
Final Descent Phase	X	X		X		X	X	X	

TRENDS IN SPACE NAVIGATION

K. N. Satyendra
Director of Research
Nortronics, A Division of
Northrop Corporation
Research Park, Palos Verdes Estates, California

Summary

The theme of the paper is based on the growing use of self-contained navigational systems in space navigation. A figure of merit, called the Self-Containment Index (SCI), is defined in order to define quantitatively the degree of autonomy for which a system design is directed.

The navigational problem of spacecraft is divided into problems of point-dynamics and rigid-body dynamics and the major factors are defined which deserve consideration in the understanding of the space navigational problem. Consistent with the requirements for low size, weight and power allocations and for very high accuracy and reliability tolerances, optical techniques are found ideally suited for space navigation. Open-loop and closed-loop methods of earth satellite navigation are presented and extended to the lunar navigation problem.

Some advanced navigational concepts are given using nonrotating and track-while-scan instruments. The importance is emphasized of the increase in data processing requirements and their part in accuracy enhancement through statistical techniques. A review of optical Doppler and lasers for measurement of distance and velocity is given. The concept of guidance management is proposed for consolidation of all the launch, orbital, mid-course, return and other guidance phases of a future multistage spacecraft into a single, integrated, self-contained navigation system within the final upper stage.

Introduction

The purpose of this paper is to point out some trends in space navigation and some advanced navigational concepts which appear to be feasible and practicable, especially in the light of recent improvements in signal-sensing and signal-processing techniques. The paper is therefore not an original and detailed discussion of any specific navigational method; however, heavy emphasis is placed on optical celestial sensing techniques since these techniques appear to be an essential aspect of nearly all future space systems.

With increased skills in astronautics, the space traffic consisting of scientific and military payloads is rapidly increasing. An educated guess of five to six thousand objects in orbit during the next decade is not unreasonable. Some of these objects may be useful payloads which may be manned or unmanned scientific or military payloads while the others may be inactive payloads and space debris - booster tankages - and the like. The tracking and cataloging by ephemeris prediction of such a vast number of space objects creates a data processing problem at a ground station resulting in possible denial of navigational data to useful orbital payloads which depend on tracking and computations made on the ground. Unlike the air traffic problem over modern airports where aircraft can be made to fly a pattern while awaiting landing instructions, orbital vehicles have a limited time for communication over the ground station before they get lost over the horizon.

The manned spacecraft, very much like the manned aircraft, demand on-board guidance and control capability. The term "self-contained navigation" has become a familiar term in the aerospace industry and an ever-increasing demand is made on the electronics industry to provide all navigational measuring and computational instrumentation on-board the spacecraft itself with very little dependence on the ground complex. The extent of "self-containment" or the degree to which the spacecraft can navigate without receiving inputs from outside depends on the mission requirements, payload and equipment limitations, cost and reliability considerations. No clear definition of the term "self-contained navigation" appears to be present and therefore the following definition, which is especially illustrative for earth satellites, is offered:

Definition

A self-contained navigation system is defined as one which, for a given portion of the mission flight path, functions independently of concurrent man-made friendly signals originating from points external to the spacecraft.

For a long time to come, the earth will still be the main source of friendly signals which will be generated in ground-based communication and control stations. With the above definition of self-containment, three classes of self-contained or autonomous navigation systems may be studied:

Class I: Autonomous outside the horizon of ground communication and Control Center. This class covers those space vehicles which are independent of communication with the earth when outside the line-of-sight of a set of predefined ground tracking and communication control centers.

During this time any navigation performed is accomplished by the on-board systems.

Class II: Autonomous for more than one pass over the ground communication and Control Center. This degree of autonomy may be necessary to accommodate those instances when the ground-based communication stations are saturated because of too many space vehicles overhead and cannot transmit updated navigational data to an individual spacecraft.

Class III: Indefinitely autonomous or continuous systems. This class covers those systems which are completely autonomous and do not depend upon earth-based tracking and communication networks to navigate. The degree of self-containment for this class actually depends on the reliability of the system.

A self-contained navigation system for Class III, requiring more advanced and sophisticated components, will chronologically follow the development of simple systems to cover Classes I and II. A fully self-contained navigation system is therefore the ultimate one can hope to accomplish and therefore suffers from all the headaches such aspirations usually involve.

Quantitatively, one can define a figure of merit to distinguish between the classes of self-containment in terms of the ratio T_Q/T_o, where T_Q is the quiescent period during which there is no communication between the space vehicle and the earth, and T_o is the orbital period or time for a round trip. This ratio, rounded off to the nearest integer, may be defined as the self-containment index, and abbreviated as SCI. For example, if an earth satellite were autonomous for 260 minutes and its orbital time period were 90 minutes, then its SCI = 3. With this definition, the SCI of Class I, II, and III systems would respectively be less than unity, greater than unity and approach infinity.

The methods involved in solving the three classes of problems differ, not in principle, but in technique and detail because of the differences in accuracies, component life times, duty cycles and degree of sophistication permissible. Present day technology is adequate to achieve values of SCI between 1 and 3.

Higher degrees of self-containment await improved techniques for developing components with reduced size, weight and power allocations and increased reliability in a hostile space environment. Almost all aerospace systems of the future will demand a fully self-contained navigation (SCI = 3) as the primary system and a ground based or earth-dependent navigation system as a backup. Some illustrative examples of systems with different requirements for the SCI will be presented in this paper.

The Navigation Problem and Its Solution

In its simplest form, the navigational problem of a spacecraft consists of two basic parts:

1. Point Dynamics, defining the motion around the geocenter of the spacecraft treated as a point-mass.

2. Rigid-body Dynamics, defining the interrelation between the body axes of the spacecraft and certain directional space references. This aspect of navigation includes the definition of a system of coordinate axes within the spacecraft, and their relationship to one or more systems of directional reference in space.

Before proceeding with the detailed exposition of the two problems, it is desirable to distinguish between the terms, navigation, guidance, and control, at least to classify the instrumentation associated with each of these terms.

The origin of the term navigation lies in the French words "navis" meaning ship, and "agere," meaning to move or direct. Space navigation may be defined by analogy with sea navigation, by a modification of Webster as "the art or science of conducting a spacecraft from one place to another, including the method of determining position, course, distance passed over, etc., by the principles of geometry and astronomy. . ." Revising Webster's order, one may single out as navigation functions:

1. The determination of position, present course, distance covered, etc.

2. The action of "conducting."

Based on the line of reasoning above, Hoelker[1] defines Guidance as the technological field concerned with the following sequence of action:

1. Measuring the data for the purpose of the determination of the current state of flight

2. Determining the current state of flight from the measured data

3. Determining the path ahead that joins the current state of flight with the desired terminal state of flight, and

4. Providing to the control mechanism an immediate reference that serves for implementation of the determined path ahead.

The action of conducting may be interpreted as the enforcement of the course of action through the use of the "control" system. The above definition separates the functions of guidance and control, and guidance terminates at providing references, computations of maneuver ballistics and inputs to the control loop or loops.

With the concepts previously stated, it may be desirable to relate the two broad problems of navigation with guidance and control as follows:

Guidance implies preparation of the inputs for solving the point-dynamics and rigid-body dynamics of a spacecraft. Control implies enforcement, through the use of control loops and additional force, of the point-dynamics and rigid-body dynamics of the spacecraft to a prescribed course of action.

The solution of the guidance problems and the preparation of the necessary inputs requires the use of instrumentation involving two basic functions, namely sensing and signal processing. The sensing function involves establishment of directional and positional references through gyroscopic or celestial (planetary or stellar) means, or man made optical and/or radio frequency means. The signal processing function involves the computation or extraction, usually by means of a computer, of the navigational parameters of the spacecraft from the sensed signals. The most demanding problem is the establishment of the orbital elements of a spacecraft to a high degree of accuracy in the shortest possible time. The accuracy in sensing direction references or velocity or position of a spacecraft has a physical limit set by the equipment and laws of physics. Space missions, especially those requiring self-contained navigation techniques, demand refinements in accuracy which, therefore, must be obtained by statistical techniques, yielding what is usually referred to as "smoothed" data. Such refinements are achieved mainly through high-speed computers and the trend in space navigation is therefore an optimization of the signal-sensing and signal-processing techniques, usually resulting in a higher burden being placed on the latter.

In summary, the following major factors deserve consideration in the understanding of the space navigation problems and their solution:

I. DIRECTIONAL REFERENCES
   - A. Inertial
     1. Celestial (stellar)
     2. Gyroscopic
     3. Combination or astronertial
   - B. Local Vertical

II. POSITION DETERMINATION
   - A. Static
     1. Triangulation
     2. Stadiametry
   - B. Dynamic
     1. Active ranging
     2. Explicit solution of point-dynamics

III. TYPES OF MEASUREMENT
   - A. Angular - angle, angle-rate, angular acceleration, etc
   - B. Range - range, range-rate, range acceleration, etc.

IV. ACCURACY REFINEMENTS
   - A. Systematic Error Isolation - intrinsic and external to instruments
   - B. Statistical Techniques - time smoothing and space smoothing

V. FORMS OF SOLUTION
   - A. Time
     1. Real time solution
     2. Delayed time solutions
   - B. Location
     1. Earth-based
     2. On-board or self-contained

VI. SIGNAL SOURCES
   1. Natural - stars, planets, fields of force
   2. Man-made - radio or optical both passive and active

VII. PHYSICAL DESCRIPTION
   - A. Size, Weight, Area, Power, Volume
   - B. Reliability, Cost, etc.

Self-Contained Navigation of Earth Satellites and Lunar Spacecraft

Earth Satellites

The most difficult problem of navigation is the solution of the point-mass dynamics. In general, the motion of a point-mass is governed by Newton's and Kepler's laws which state, in effect, that the rate of change of momentum is equal to the total force acting on the point-mass. The total force F_T on a satellite, treated as a point-mass, is given by the following equation:

$$F_T = F_E + F_D + F_M + F_P \qquad (1)$$

where

F_E = Most complete force field of the earth, such as the Jeffreys model of the oblate earth

F_D = Atmospheric drag

F_M = Force due to maneuvering rocket thrusts, such as those due to retro-rockets, ion-engines, etc, used for rapid position changes of the satellite

F_P = Perturbative forces due to solar photon pressure, lunar gravitational effects, rotation of earth, etc.

It is a fact that if F_T is completely known, and the initial values of position and velocity of the point-mass are known, the solution of the Newtonian force equation (1) is theoretically possible. However, uncertainties in several factors make it impossible to provide an ideal and true theoretical solution of the force equation and thus predict the position of the satellite point-mass. It is therefore necessary to make compromises in practice and be satisfied with the most reasonable, if not the optimum solution. The position and velocity accuracies of a satellite, based on such solutions, will be better for short flight times, as in Class I and Class II type systems than for long flight times, as in Class III self-contained navigation systems. The mechanization of the computer for solving these equations depends also on the degree of sophistication employed. Two such methods are described in the

following paragraphs -- the open-loop method for Class I and Class II types of satellite navigation and the closed-loop method for Class III type of navigation.

Open-Loop Method. For Class I and II types of self-contained navigation schemes for which the SCI is between 1 and 3, the explicit solution of the force equation (1) may be simplified by the help of some assumptions, such as the following:

1. The earth is perfectly spherical
2. The earth's force field is perfectly central
3. The orbital elements of the satellite orbit may be accurately computed on the ground and transmitted to the satellite
4. That there are no external perturbative, thrust maneuvering or atmospheric drag forces, and
5. That the satellite is truly a point-mass.

The above simplifying assumptions yield a circular or elliptic orbit for the satellite. With known initial conditions, i.e., with known position and velocity at any time of such an orbit, the subsequent orbital parameters after a time interval, t, may be easily computed algebraically rather than through a differential equation solution. The time difference, t, can be measured in the satellite by a simple clock or oscillator. The effects of earth oblateness, atmospheric drag, solar winds, etc, could be added as refinements in the form of empirical corrections. Specifically, the vehicle's position in spherical coordinates may be programmed into a small on-board function generator as trigonometric or power series. Thus, the latitude and/or longitude might be represented in the form

$$\theta = \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 t^3 \qquad (2)$$

$$\theta = \theta_0 + \theta_m \sin(\omega t + \alpha) \qquad (3)$$

The form of the function will be determined by eccentricity, inclination of the orbit, altitude, etc. The block diagram for such a simplified navigation system is shown in Figure 1. The output from the clock to the function generator serves to define the position and velocity. Position and velocity are channeled into the guidance computer along with vehicle attitude and thrust acceleration to be included in the guidance equations. Attitude control signals and thrust commands are generated by the guidance computer and are fed to the attitude control device and thrust actuating device, respectively. A complex, high-data rate, general purpose computer is not necessary.

Closed Loop Solution - Class III. In the previous method, position and velocity predictions are made by open-loop methods, with periodic updating from the ground. To obtain the accuracy and independence required for Class III type navigation systems, a closed-loop system will be necessary to improve system accuracy by up-dating refinements made by means of on-board measurements.

The basic refinement necessary for Class III type navigation schemes is the explicit solution of the force equation (1), taking into account as many perturbative factors as possible.

One such method has been presented by Satyendra and Bradford.[3] This method is built around the concept of sensing the satellite attitude by means of a prime reference established by a pair of star trackers; direct measurements, over a period of time, of the local geocentric vertical by means of a horizon tracker; and determining the orbital elements by a precision time reference and a data-processing computer.

The basic conceptual feature of this approach is that suitable stars are preselected and tracked. The initial star lock-on can be achieved by orienting the satellite from the injection guidance subsystem's gyroscopes. Once this lock-on is achieved, it is continually maintained to obviate a complex search and identification process implied by loss of orientation resulting from extended shutdowns of operation.

A typical design of a lightweight star tracker is described by Stevens.[4] Two star trackers are locked on to two known stars, S_1 and S_2, as shown in Figure 2. The selection of the stars S_1 and S_2 may be based on the techniques described by Stevens.[5] A coordinate reference frame consisting of three mutually perpendicular vectors, $\hat{C}_1$, $\hat{C}_2$, $\hat{C}_3$, is rigidly connected to the two star-tracker telescopes, which are represented by the vectors $\hat{P}_1$ and $\hat{P}_2$ in Figure 2. With the star trackers locked on to stars S_1 and S_2, the orientation of the frame ($\hat{C}_1$, $\hat{C}_2$, $\hat{C}_3$) in inertial space is known. For convenience in analysis and mechanization, it is assumed that this orientation is such that the vector $\hat{C}_3$ is approximately normal to the plane of the orbit. The azimuth and elevation angles θ and ϕ between the inertial frame ($\hat{C}_1$, $\hat{C}_2$, $\hat{C}_3$) and the local vertical are measured by means of a horizon tracker.

In the ideal case of perfectly accurate measurements and an exact knowledge of the force field about the earth, one would be able to determine an earth satellite's trajectory by measuring the direction of the local vertical at four distinct times. In such an ideal situation, no new information could be obtained by collecting a greater amount of data. In reality, however, all measurements will be corrupted by noise; hence the above reasoning is not applicable to any practical situation. It can be clearly demonstrated that, if one has a large set of measurements of the local vertical, statistical decision techniques can be used to obtain a much more accurate estimate of the trajectory than could be obtained by using only four measurements. For this reason the use of approximately 1000 measurements of the local vertical is envisioned to obtain a good estimate of a satellite's orbit.

To estimate a trajectory it is necessary first to choose a model for the force field about the earth. Under such a choice, every possible trajectory will be completely characterized by six initial conditions -- three of position and three of velocity. These six initial conditions will completely characterize a trajectory, for the period of time for which the model is applicable.

For this reason the problem of trajectory determination reduces to the problem of determining the six initial conditions of position and velocity.

The force model chosen for this method was the Jeffreys model of the oblate earth. This model neglects atmospheric drag and perturbations due to the influence of the moon; but these effects appear to be negligible for typical earth satellite trajectories for periods of time corresponding to at least several cycles of the satellite. If the only requirement is that the model give a satisfactory representation of a satellite's trajectory for a few periods, the assumed model is probably adequate for the present purpose.

The second step is to measure the various angles θ and ϕ, using the basic geometry shown in Figure 2. Suppose that n measurements are taken of these angles. The data will then consist of the triples

$$(\theta_1, \phi_1, t_1), (\theta_2, \phi_2, t_2), (\theta_3, \phi_3, t_3),$$

$$\ldots \quad , \quad \ldots , (\theta_n, \phi_n, t_n),$$

where t_i is the time at which the i^{th} measurement was taken.

In the absence of any errors in the horizon tracker and inertial reference, the angles θ and ϕ would truly represent the direction of the geocentric vector. The graphs of θ or ϕ versus time would be straight lines or periodic curves for circular or elliptic orbits. The presence of instrument errors discussed in Reference 6 results in data which are shown as points distributed along these curves for θ and ϕ as shown in Figure 3.

The basis of the method involves computing the difference between the calculated values of θ and ϕ from the force field model and the horizon tracker gimbal-pickoff measured values. These difference values are smoothed over as many full or fractional cycles to yield averaging values of θ and ϕ, from which the orbital elements can be calculated.

Error analysis based on a 400 statute mile circular orbit and assuming a 1 σ error in angle determination of 1 arc minute, indicate the following 1 σ errors in position and velocity over a 1 revolution smoothing period:

TABLE 1 TYPICAL ACCURACIES OF 2-STAR HORIZON TRACKER METHOD

	1σ Position Error	1σ Velocity Error
X Component	335 ft	0.67 fps
Y Component	674 ft	0.30 fps
Z Component	291 ft	0.31 fps

Of course, the accuracy figures depend somewhat upon initial conditions defining the orbital elements. These initial conditions can be established by ground-based measurements or by simple onboard techniques. Even with crude initial conditions, the smoothing techniques can be used to effect major improvements in updating the values.

The functional block diagram for a closed-loop Class III navigation system is used in Figure 4. The basic equipment consists of an astronertial platform, a horizon tracker, a clock and a computer.

The astronertial platform is used to maintain an inertial reference frame in terms of which the horizon tracker measures the angle describing the satellite-geocenter vector. The computer combines the measured angles with angles calculated from preliminary position predictions to obtain updated initial conditions.

These updated initial conditions are used by the computer to make more accurate position and velocity predictions than can be made by an open-loop system. This method of periodic updating of initial conditions reduces inaccuracies due to errors in initial conditions and lack of completely accurate force field information.

The computer shown in Figure 4 replaces the function generator shown in the block diagram of the open-loop system. In addition to this change from open-loop to closed-loop position and velocity determination, the system for the Class III satellites will require an attitude control system operable for a much longer time period than that required for Classes I and II.

The system described above is presented as an example of the methods of closed-loop position and velocity determination to be studied for Class III satellites. No attempt has been made here to describe a specific system, but rather the general method has been indicated. It is recognized that various instruments might be used, alternatively, in each of the instrument blocks.

Lunar Navigation. The navigational problems of lunar probes and manned lunar vehicles are receiving such detailed attention today that one cannot do justice to any given method or technique without a very detailed discussion. The purpose of this paper is not to go into such discussion but to point out the trend in lunar navigation, which like earth-satellite navigation, once again shows a great emphasis on the concept of self-containment. Several types of lunar missions are

in various stages of planning and progress. The guidance systems which have been used for space probes and those which are presently planned for other early programs have taken advantage of the current state of the art in systems and use off-the-shelf hardware. This generally results in a heterogeneous system consisting of independent launch and injection systems, in some cases separate for each boost vehicle stage; radio-controlled midcourse guidance with ground-based computation; and a variety of combination radio-optical-inertial systems for operation in the vicinity of the moon. The future trend will definitely be toward a nonradiating, fully integrated, self-contained system. An excellent example is given by R. C. Hakes[7] in which the integrated guidance subsystem consists of an astronertial platform, a two-star tracker, wide-angle and narrow-angle optical earth and moon trackers, an integrated computer-clock, and a radio altimeter. The utilization of these instruments by mission phase is shown in Table 2. Perhaps the most demanding of problems is the enhancement of navigational accuracies by self-contained methods. It is becoming increasingly clear that such accuracy enhancements can come more from employing statistical techniques than from over-design of sensors to achieve extreme accuracies, possibly jeopardizing cost and lifetime figures.

The statistical methods in the determination of the trajectory of a spacecraft are concerned with the problem of finding the best (in the sense of the statistically most significant) values of the position and velocity of a spacecraft, when the available data are redundant and, on account of the errors of measurement, contradictory.

As an example (Figure 5), consider the radar R, at some point on the earth, which measures the range and direction (the azimuth and elevation angles) of a vehicle at V. From this data, a determination of the position of the vehicle (its x, y, and z coordinates, for example, in some reference frame) can be found. Here there is no problem, since one has a minimum of data. Three measurements, range, azimuth and elevation, are used to compute 3 position coordinates, and there is no redundancy.

Now, however, assume that another radar R', at another spot on the earth measures the corresponding quantities at the same time. From these measurements one can compute the x, y, and z coordinates of the vehicle in the same reference frame as before. Call these x', y', and z'.

Since each of these six measurements will inevitably be subject to error, no matter how carefully the measurements are made or how accurate the system, the position as calculated from R will not coincide exactly with the position calculated from R', and the question arises as to the choice of the position of the vehicle.

One simple answer would be to average the two positions:

$$\bar{x} = \frac{1}{2}(x + x')$$

$$\bar{y} = \frac{1}{2}(y + y') \qquad (4)$$

$$\bar{z} = \frac{1}{2}(z + z')$$

Although this position would be better than either position separately, it does not generally represent the statistically best determination, i.e., the most probable position in view of all available information.

Statisticians have worked with this problem, and have developed a simple criterion, referred to as the least squares criterion, by which the position of the spacecraft is so chosen that the sum of the squares of the errors is a minimum.

One complication immediately arises. The data include both distance and angle measurements. The question arises as to how to combine errors measured in feet with errors measured in degrees. To account for this difference, as well as possible differences in precision of various measurements of the same type, one minimizes the sum of the weighted residuals, where the weights are given by

$$w_i = \frac{1}{\sigma_i^2} \qquad (5)$$

where w_i and σ_i are the weight and the standard derivation of the i^{th} observation.

Suppose now that one makes a number of such observations at specified times. Since the vehicle during this interval must move in accordance with Newton's laws, the sum total of the known information consists of all the measured data and the laws of motion governing the trajectory, including all the forces acting on the vehicle.

A criterion for selection can thus be stated as follows:

Select the trajectory which satisfies the laws of motion and for which the sum of the squares of the weighted residuals is a minimum.

In the case of a linear function, a least squares solution is a simple and straightforward computation to make. Unfortunately, in the case of a space vehicle, the differential equations of motion cannot be solved explicitly. This means that the trajectory must be calculated numerically and approximately, and the direct least squares solution cannot be made. Instead, one is forced to expand the various functions in Taylor Series, of which one uses only the first term. For a given first approximation, the equations give an improved value for the trajectory. A second correction can give a still better estimate, and so on.

In many cases, the knowledge of the accuracy of the position and velocity determination is as important as the quantities themselves. Two methods are commonly used for finding these errors. The first, a numerical method, uses the Monte Carlo principle, in which a considerable number of trajectory cases are run with random errors added to the measurements. In the second, the usual formula for the propagation of errors is used. If the x-coordinate of position at a given time, for example, is thought of as a function of the measurements,

$$x = x(\psi_1, \psi_2, \ldots, \psi_n)$$

then the standard deviation of x, σ_x, is given by

$$\sigma_x^2 = \sum \left(\frac{\partial x}{\partial \psi_i}\right)^2 \sigma\psi_i^2 \qquad (6)$$

These calculations have been performed for a typical case of the reentry vehicle from the moon. The Propagation of Errors Method gave the standard deviations in the six variables as

$\sigma_x = \pm 155$ ft $\quad \sigma_{v_x} = \pm 0.50$ ft/sec

$\sigma_y = 575$ ft $\quad \sigma_{v_y} = 0.20$ ft/sec

$\sigma_z = 385$ ft $\quad \sigma_{v_z} = 0.52$ ft/sec

The Monte Carlo method yielded

$\sigma_x = \pm 172$ ft $\quad \sigma_{v_x} = \pm 0.41$ ft/sec

$\sigma_y = \pm 438$ ft $\quad \sigma_{v_y} = \pm 0.22$ ft/sec

$\sigma_z = \pm 411$ ft $\quad \sigma_{v_z} = \pm 0.48$ ft/sec

This agreement is considered to be very good in view of the small number of samples.

Some Advanced Navigational Concepts

Advanced Navigation Requirements

The navigational requirements for aerospace vehicles differ from mission to mission and from phase to phase within a particular mission. In each mission, the navigation instruments, including the computer, must be capable of providing current position and velocity information and, in many cases, they must be capable of predicting the future position and velocity of the spacecraft. In addition, the navigation system in many cases is called upon to calculate maneuver ballistics and specify time, magnitude, and direction of corrective thrusts.

The major variations in navigation requirements arise when one considers the accuracy and speed with which navigational data must be acquired and processed to provide the necessary guidance functions. In certain satellite missions, the navigational system must be capable of positional accuracies measured in feet, velocity accuracies of less than one foot per second with a data acquiring and processing time measured in seconds. On the other hand, the navigational system of a deep-space mission could tolerate midcourse positional errors measured in miles and require minutes, if not hours, to perform the necessary computations. However, in the interest of minimizing fuel requirements, maximizing payload, etc, it is desirable that the maximum accuracy and minimum time be obtained with due consideration for size, weight, power, cost and reliability. The consideration of these parameters and their trade-offs often lead to interesting and new navigational concepts. The trend is definitely pointed toward novel optical techniques which offer increased accuracy and reliability with reduced size, weight, and power allocations.

Some Optical Navigational Techniques

An important distinction has to be made between an optical system in a vehicle within the atmosphere and a space vehicle. For example, a star tracker in a typical vehicle, which has most of its flight profile within the earth's atmosphere, usually has a narrow field of view intended to minimize sky brightness and background noise and consists of a telescope system rigidly pointed toward a given star and held in track mode by a servo motor. The sky brightness and background noise are not as serious when the vehicle is above the atmosphere and hence the optical field of view can be increased, giving the opportunity to observe many celestial bodies. Mixed with celestial bodies are man-made objects which, in many instances, reflect solar energy, appearing like stars. Wide-angle optical navigation systems must, therefore, have the ability to discriminate between natural celestial bodies and man-made moving targets. Unless compensated for, a typical servo control system which would enforce the pointing to a specified star of a tracker in an atmospheric vehicle would create angular momentum unbalance in an exo-atmospheric spacecraft. The need for preserving angular momentum equilibrium or balanced torques is therefore essential in spacecraft.

Some of these conflicting requirements and properties of space navigation systems suggest two possible configurations, a wide-angle nonrotating system and a rotating track-while-scan system.

Nonrotating Sensor Concept. In principle, this method considers the acquisition and monitoring of an adequate image of the entire spherical field of view available to a space vehicle. A conceptual sketch describing this concept is shown in Figure 6. Such a device, which may be called a

Fly's Eye Camera because of its ability to look through a wide angle and track different objects independently may conceivably be developed by a mosaic type or sensor array in which multiple fields of view may be overlapped to give a broad picture of the celestial sphere. From a knowledge of the solar system and the celestial sphere, which may be carried in its memory, the fly's eye camera could provide many valuable navigational data without serious control system requirements. There is no doubt that the on-board computer will have to be sophisticated because of the many functions and transformations it must perform.

Track-While-Scan (TWS) Concept. This concept is an extension of the method of self-contained navigation presented by Lillestrand and Carroll.[2] In principle, the TWS instrument observes a portion of the celestial sphere by a rotating telescope. The instantaneous portion of the celestial sphere under observation is defined by a slit or slits in the image plane of the telescope. Figure 7 illustrates a belt-like projection of the celestial sphere observed by a single scanning slit. The corresponding object space image of the slit is shown superimposed on the celestial sphere.

The azimuth angle between any two stars in the field of view may be measured in two ways:

1. If the platform supporting the rotating telescope remains fixed in space, an angle encoder may be employed and the angle of rotation of the telescope measured at the instant of transit of each star.

2. Alternatively, if the rotation rate ω is a constant, the periodic nature of the observed star pattern will permit an accurate calibration of this rate. The difference in azimuth angle $(\theta_2 - \theta_1)$ between any two stars may then be determined by measuring the time interval $(t_2 - t_1)$ between each star crossing the slit, since $(\theta_2 - \theta_1) = \omega(T_2 - t_1)$.

The technique for precise measurement of azimuth angular separation of celestial bodies by means of a telescope rotating with a constant angular rate is similar to the well-known transit telescope. The aligned transit telescope is constrained so that its line of sight is always in the plane of the meridian. The constant angular rate in the azimuth direction is supplied by the angular rotation of the earth. During the manual operation of a transit telescope, the observer records the time at which each star crosses the instrument crosshair. In the TWS automatic instrument, the star transit across a narrow slit may be recorded as a photoelectric pulse from a photodetector placed behind the slit.

Advanced Concepts in Data Processing Techniques

The successful deployment of self-contained navigation systems depends on successful development of reliable components of small size, weight, and power allocations. The increased emphasis of accuracy and redundant logic, as well as explicit trajectory solution in completely self-contained systems, requires the use of a sophisticated, large-scale general-purpose data processing machine. Some trends in data processing computers may be pointed out.

Some factors which may be considered in optimizing the data processing are:

1. Equipment time sharing: The possibility of time sharing of an arithmetic module between two computational processes, where each process has a different time when peak load occurs.
2. Computer control: The problem may lend itself to unique addressing and instruction coding.
3. Input-Output: The realm of data conversion where perhaps the input-output unit can be greatly reduced. This appears particularly feasible if the system is self-contained. The problem then becomes one of designing the subsystems to operate using machine language inputs.

In addition to the factors listed above, the design and development necessary to build a digital computer that will perform within high reliability standards, perform all desired functions with a speed and accuracy dictated by system specifications, and within critical weight and size limits, seems to center around the application of state-of-the-art considerations of packaging philosophy, electronic component microminiaturization construction, advanced circuit and logical design techniques, and a thorough testing and quality control program.

To obtain maximum packaging density, a required amount of development must be expended in the area of efficient packaging of miniature and microminiature components. The state of the art in electronic packaging is changing at a rapid rate as shown in Figure 8, which is a graphical representation of the variation of component packing density as a function of time. As indicated in Figure 8, there are three methods of achieving microminiaturization of electronic equipment. All may be considered for application to space flight, and may be grouped as follows:

1. Single Dimension: This method uses thin films for resistors and capacitors and separate microcomponents for inductances and semiconductors. A single type of element (resistor or capacitor) is deposited on each wafer. This method is currently in use and promises a density of up to 600,000 components per cubic foot.
2. Two Dimension: This method uses thin films for resistors and capacitors, but combines them on a single wafer. This combining reduces external interconnections leading to an overall increase in packing density. The disadvantage is that the yield is reduced due to the multiple operations involved. When these problems are solved, packing densities of up to 2.5 million components per cubic foot can be realized.

3. Molecular: In this method, resistors, capacitors, inductors, diodes, transistors, and other electronic devices are mounted together on one surface of a semiconductor wafer. Packing densities of from 20 to 100 million components per cubic foot are expected; however, major improvements in connectors, heat transfer methods, and intercommunication between circuits are required to fully utilize the possible high densities.

Advanced Concepts Using Optical Doppler Techniques and Lasers for Space Navigation

A comparison between radar and optical techniques for making velocity and distance measurements will soon convince the reader that the optical techniques offer higher accuracies, narrower beamwidths and all-round better performance than radar, especially in terms of size, weight, and power requirements. A brief review is now presented of the use of optical Doppler techniques and lasers for space navigation.

Two papers by Franklin and Birx [8,9] describe studies carried out for the USAF investigating the use of natural electromagnetic radiation for navigational purposes. The emphasis is on the measurement of velocity utilizing the Doppler phenomenon.

Franklin and Birx use the well-known fact that when there is relative motion between a source of radiation and an observer, the radiation observed is shifted an amount $\Delta\lambda$ which is proportional to the relative velocity according to the equation

$$\Delta\lambda = \lambda_1 - \lambda_2 = \frac{v\lambda_1}{c} \qquad (7)$$

where

v = relative velocity of source with respect to observer

c = velocity of light

λ_1 = wavelength of radiation observed with no relative motion

λ_2 = wavelength observed with relative motion

from which one obtains the velocity equation

$$v = \frac{c(\lambda_1 - \lambda_2)}{\lambda_1} \qquad (8)$$

Franklin and Birx base their error analysis upon the velocity error equation

$$\delta v = \frac{2c\,\delta\lambda}{\lambda_1} \qquad (9)$$

where $\delta\lambda$ = error in determination of wavelength. In the first paper by Franklin and Birx[8] a method is described using a slit system constructed to allow the radiation from a spectral line to be split so that each half of the flux falls on a separate detector. The outputs of the two detectors would be balanced at zero velocity. Utilizing such a system, a limitation on accuracy of ± 60 ft/sec was the theoretical figure obtained. The second paper by Franklin and Birx[9] describes a method of computing the shift by comparing the spectral lines of the received radiation with a template consisting of a photographic negative of the spectral lines and measuring the displacement of the template necessary to return to a matched condition. The authors find that with such a system it should be possible to obtain sensitivities of from 1 to 10 meters per second but at the time of the writing of the paper, work had not progressed far enough to provide evidence of this order of sensitivity.

It should be noted that all such methods of utilizing the Doppler phenomenon can give only the component of velocity in a radial direction from the source. Also, all measurements made thus far have been made with respect to the sun and great advancements in the state of the art are required if radiation from other stellar bodies is to be used.

A paper by Norton and Wildey[10] is concerned with fundamental limitations to optical Doppler measurements for space navigation. The authors consider the result of various physical phenomenon which produce asymmetric broadening of the spectral lines in the radiation from stellar bodies. The phenomenon considered are: second order Zeeman effect, quadratic Stark effect, pressure shift, sunspot effect, and variability of radius. Calculations by the authors, using the sun as an example, result in their concluding that the physical characteristics of stellar electromagnetic sources such as the sun vary in such a manner that a variability of approximately ±200 ft/sec may be expected in the observer's measured optical Doppler velocity. Norton and Wildey express the opinion that it is this, rather than the instrumentation, which imposes a fundamental limitation on the accuracy with which Doppler measurements may be made optically.

One of the most interesting discussions on the determination of distance and velocity using optical maser principles is found in a paper by Ellis and Greenwood.[11] The particular features contributing to the advantageous application of optical masers to space navigation are:

1. The systems could be very small in size, and self-contained
2. Power requirements could be low even for long distances
3. Velocity measurements will be very accurate.

The availability of the extraordinary degree of parallelism and the narrowness of frequency bandwidth achievable in the laser makes this instrument extremely noteworthy for space navigation applications. Another attractive feature of the laser is the extraordinary capacity of the laser receiver for discriminating against extraneous background light and to its extremely low internal noise level.

The determination of distance from a laser transmitter to some scattering object can be accomplished by standard methods similar to radar and sonar. Ellis and Greenwood present an example of a two-foot diameter transmitter and receiver optical maser consuming 66 watts of average power yielding an accuracy of one part in 100,000 for the measurement of distance between two spacecraft 160,000 kilometers apart. The most serious disadvantage for many space applications is the requirement for transmitter and receiver beams to be trained in the same direction, and its poor performance in the search mode.

Another interesting application of the laser is the high accuracy with which one can measure the altitude of a spacecraft above a very small spot on the surface of a planet. This is in contrast to the microwave situation where, because of unavoidable beamwidth, the measurement always constitutes an average of distances to a broad range of points below.

Ellis and Greenwood[11] extol the virtues of the laser as a Doppler velocity measuring device. They present an interesting example of a vehicle for soft lunar landing and point out that an accuracy of one-tenth of one percent in velocity is reasonable for lunar approach velocity of about 2300 knots at an altitude of about 1600 kilometers above the moon's surface. Using a similar example for a planetary approach, they predict a self-contained system of about one cubic foot volume and about 10 pounds weight.

The laser transmitter is now available; a space-worthy laser receiver is still not available. The future, however, holds promise for more advanced laser developments and hence their use in space navigation.

Advanced Concepts of Guidance Management

The majority of space boosters which are deployed for launching orbital payloads were developed originally for military missile applications. It is therefore quite common to see each stage of a multistage vehicle equipped with its own guidance and control system for thrust vectoring, booster control and engine cutoff. The payload seldom has a major part in the navigational functions of the individual stages of the booster which places the payload along its path. It seems possible that savings in weight, instrumentation costs, and ground-based equipment could be obtained by "bringing the payload into the picture," especially in advanced and future space missions for which both the payload and the boosters can be specially designed to fit each other's requirements. If one accepts the premise that future loads will include the requirement of self-contained contained navigation, the final upper stage (Figure 9) will have all the usual equipment, such as gyros, accelerometers, star and planet trackers, clocks, radio frequency equipment and a fairly large general-purpose computer, necessary to guide the spacecraft during all phases of flight. Such an integrated, self-contained system would seem very desirable for providing all the necessary navigation and guidance facilities for the boost phase, the orbital phase, midcourse, landing or return phase, etc. Thoughts such as these have been expressed by Cashmore.[12]

A basic problem is created when the inertial sensors of the multistage system are located at a station or stations which are an appreciable distance from the center of gravity of the vehicle. During launch operations, the payload at the top of the booster does not experience the bending and c.g. separations and accelerations. This compensation or correction must be allowed for in the thrust-vector control and engine cutoff of the multistages. Secondly, the optical windows of the optical instrumentation contribute errors during launch and the tracking of celestial bodies is subject to refraction and shimmering until the vehicle approaches very high altitudes. Compensations for these errors will be necessary. Third, aborts on the launch pad are still better handled by ground crew than by the payload, and provisions for individual engine controls are necessary.

The future trend will, however, be an evolution of the guidance management concept by which the self-contained payload guidance and the ground-based command and control center jointly "manage" the guidance functions to achieve thrust-vector control and engine cutoff of the individual stages, as well as the navigation of the orbital spacecraft, all functions to be performed mostly by the use of and within the available instrumentation of the final upper stage itself.

Acknowledgment

A review paper such as the preceding necessarily reflects the concepts and creative efforts of many individuals, frequently informally transmitted and therefore difficult to acknowledge. Specifically cited, however, are the contributions of Mr. W. J. Reilly, Dr. J. Titus, Mr. G. D. Barnes, Mr. R. B. Freund, Dr. J. K. Robe, and Mr. R. C. Hakes.

References

1. R. F. Hoelker, "The Evolution of Guidance Theory and Trajectory Analysis Into a Single Scientific Discipline", Institute of Navigation Meeting, Williamsburg, Virginia, June 1961

2. R. L. Lillestrand and J. E. Carroll, "Self-Contained System for Interplanetary Navigation", American Astronautical Society Meeting, August 1961

3. K. N. Satyendra and R. E. Bradford, "A Self-Contained Navigational System for Determination of Orbital Elements of a Satellite", ARS Paper, Number 1495-60, December 5-8, 1960

4. Frederick Stevens, "Space Navigation by Star and Planet Tracking Techniques", Proceedings of the 5th Annual National Space Electronics and Telemetry Symposium, IRE, Washington, D. C., September 1960; Section 10.3

5. Frederick Stevens, "Application of Optical Techniques to Interplanetary Navigation", Third AFOSR Symposium, Los Angeles, October 1960

6. T. J. Jaramillo, "The Combined Non-Linear Effects of Earth Oblateness and Atmospheric Drag on a Near Earth Satellite", O.T.S. 161494, 250, U.S. Department of Commerce, Washington 25, D.C.

7. R. C. Hakes, "Guidance and Navigation Systems for Lunar Missions", ARS Space Flight Report to the Nation, October 1961

8. R. G. Franklin and D. L. Birx, "A Study of Natural Electromagnetic Phenomena for Space Navigation", Proceedings of the IRE, Volume 48, Number 4, Pages 532-541, April 1960

9. R. G. Franklin and D. L. Birx, "Optical Doppler for Space Navigation", ARS Preprint, Number 1940-61, Presented to ARS Guidance, Control and Navigation Conference, Stanford University, August 1961

10. R. H. Norton and R. L. Wildey, "Fundamental Limitations to Optical Doppler Measurements for Space Navigation", ARS Preprint, Number 1942-61, Presented at ARS Guidance, Control and Navigation Conference, Stanford University, August 1961

11. Cecil B. Ellis and Ivan A. Greenwood, "Optical Masers in Space Navigation", Institute of Navigation Journal, Navigation, 17th Annual Meeting, 30 June 1961

12. D. J. Cashmore, "The Role of Inertial Equipment", Journal of the British Interplanetary Society, July-August 1961, Volume 18, Number 4, Page 146

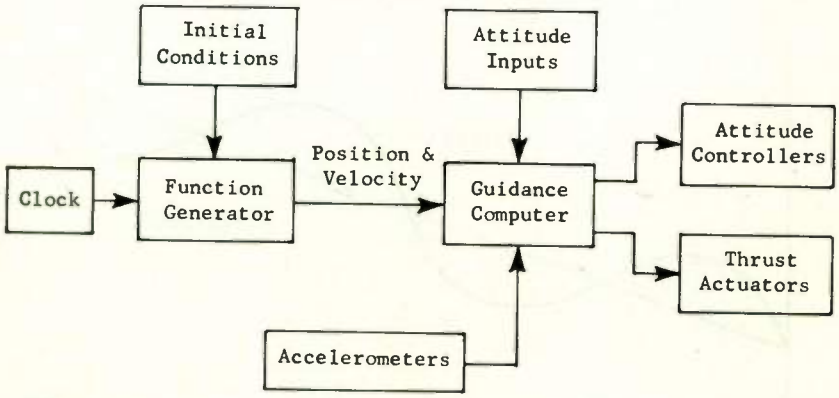


Fig. 1—Open-loop method—functional block diagram of self-contained navigation (Class I and Class II).

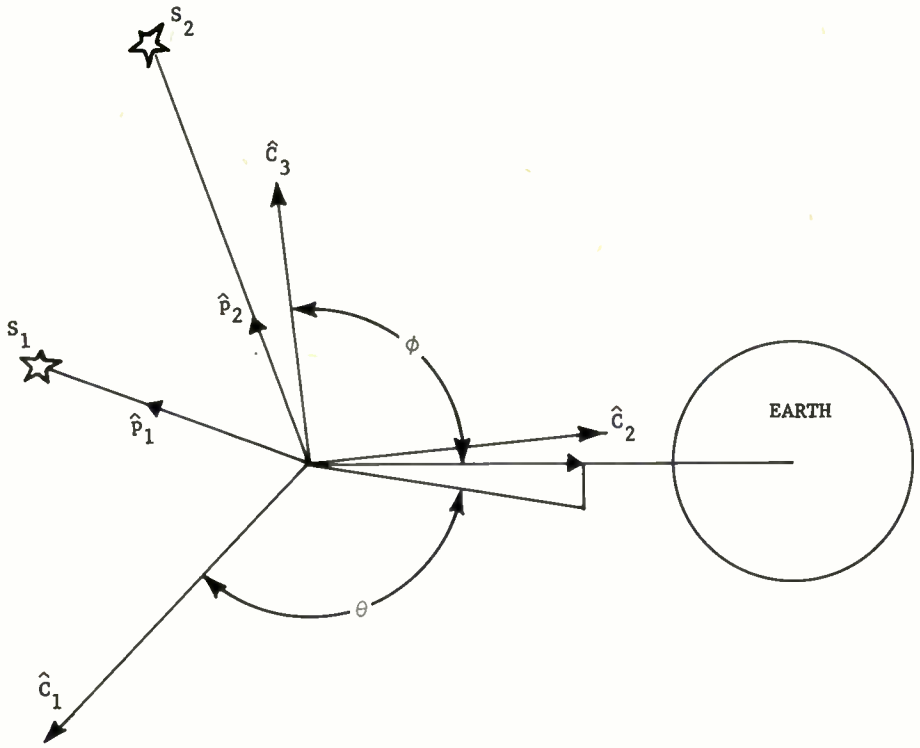


Fig. 2— Basic geometry of the Satyendra and Bradford method.

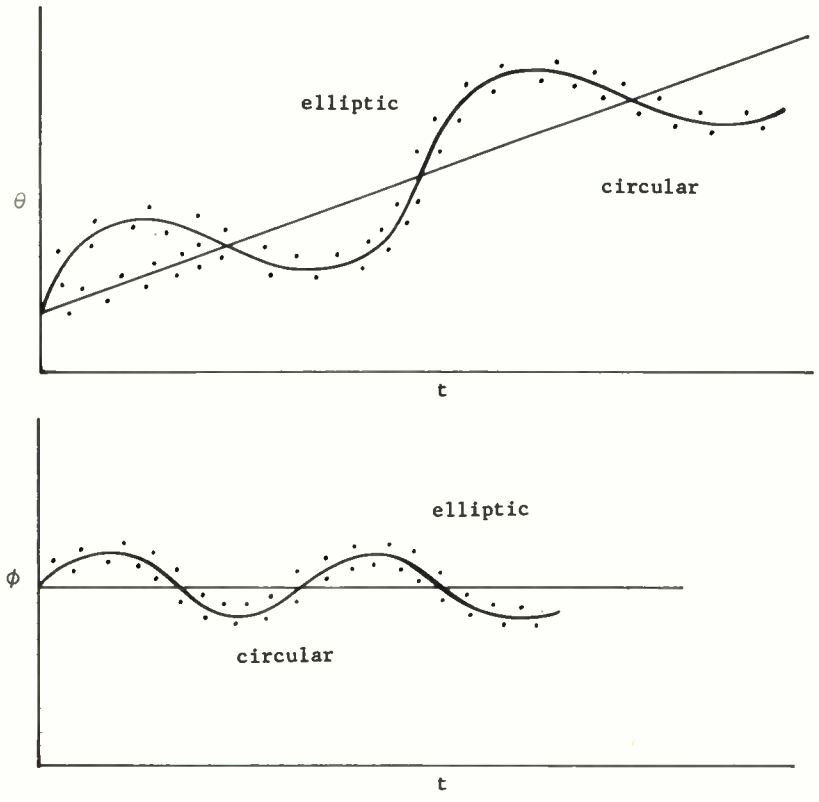


Fig. 3— Variation of θ and ϕ with time.

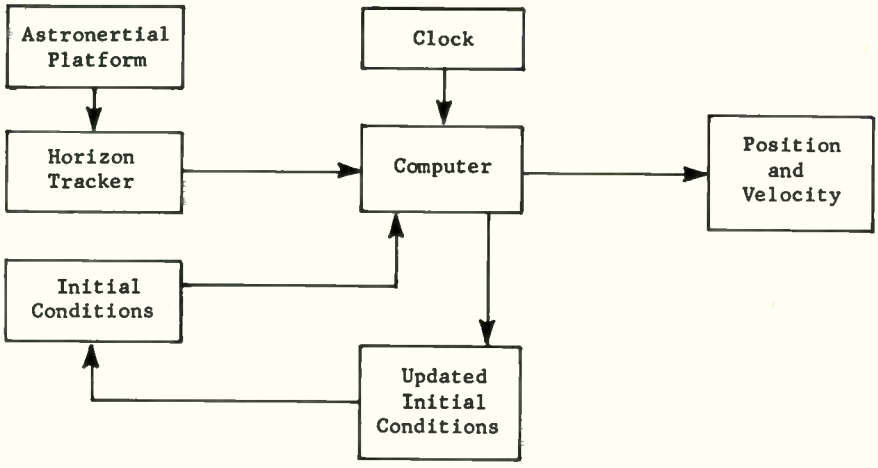


Fig. 4—Functional block diagram of a self-contained earth satellite navigation system (closed-loop method)—Class III.

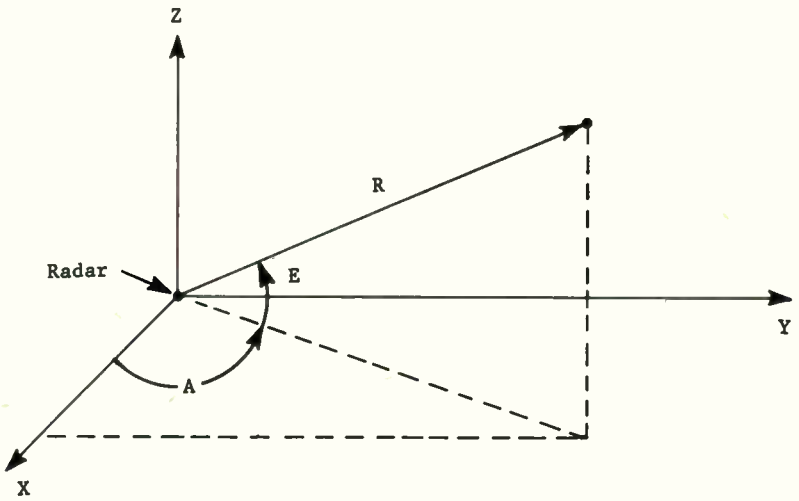


Fig. 5—Coordinates of spacecraft relative to ground-based radar.

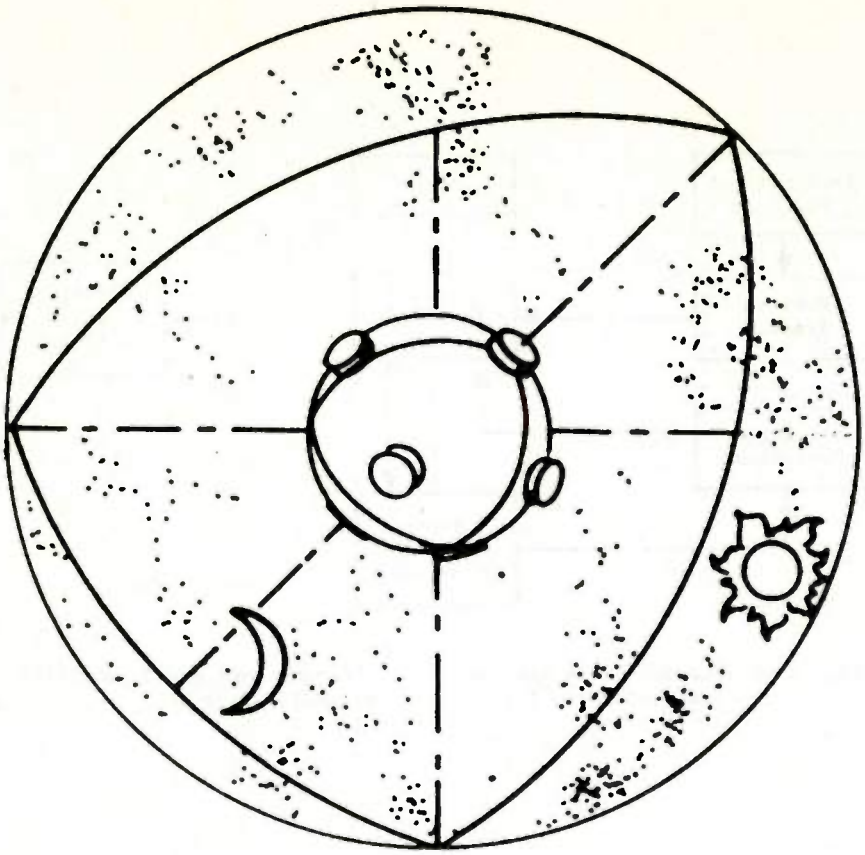


Fig. 6 - Nonrotating tracking instrument

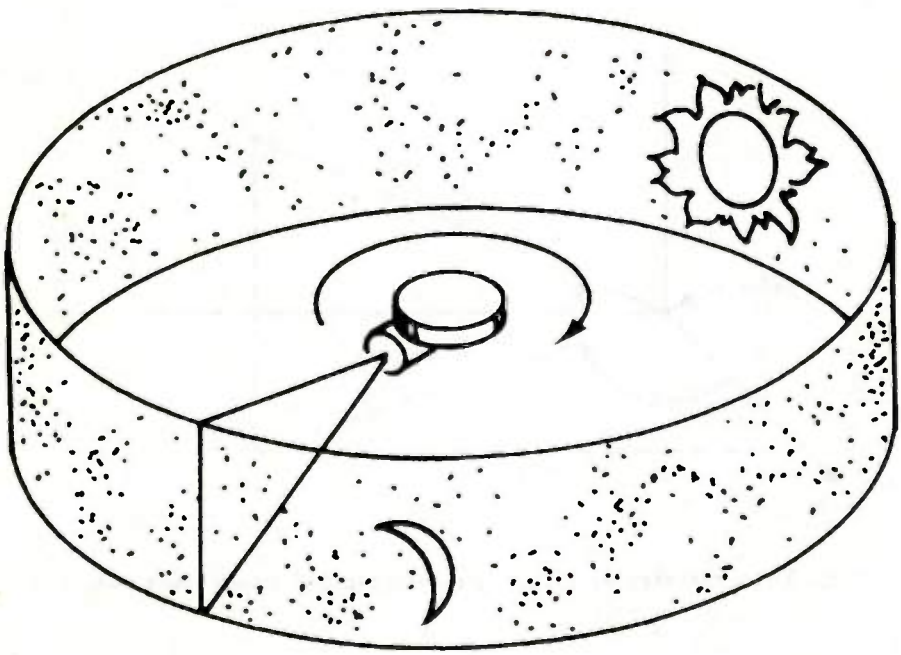


Fig. 7—Track-while-scan instrument.

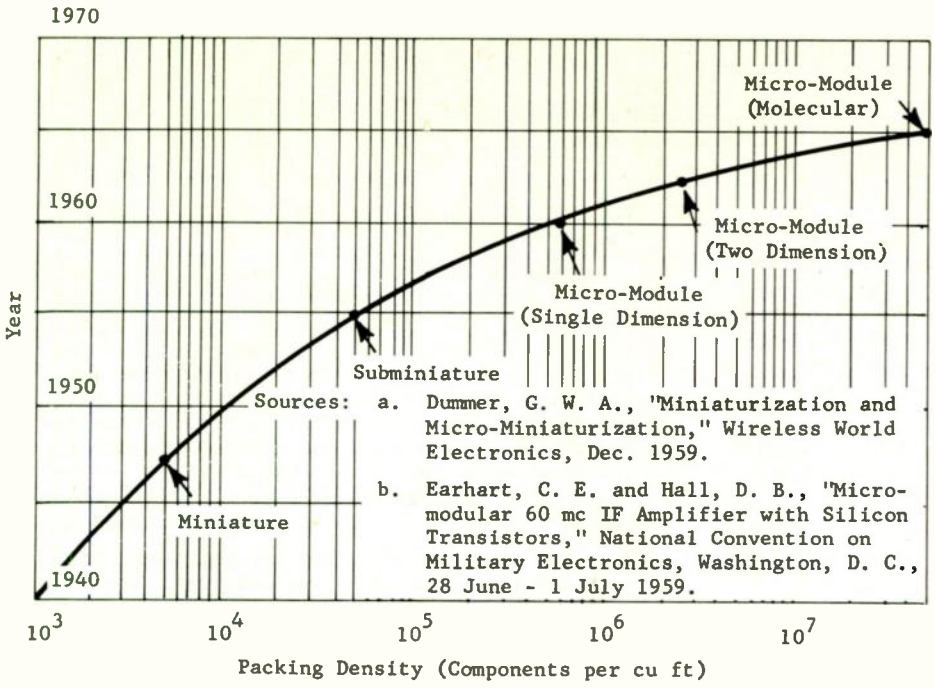


Fig. 8—Trend of packing density.

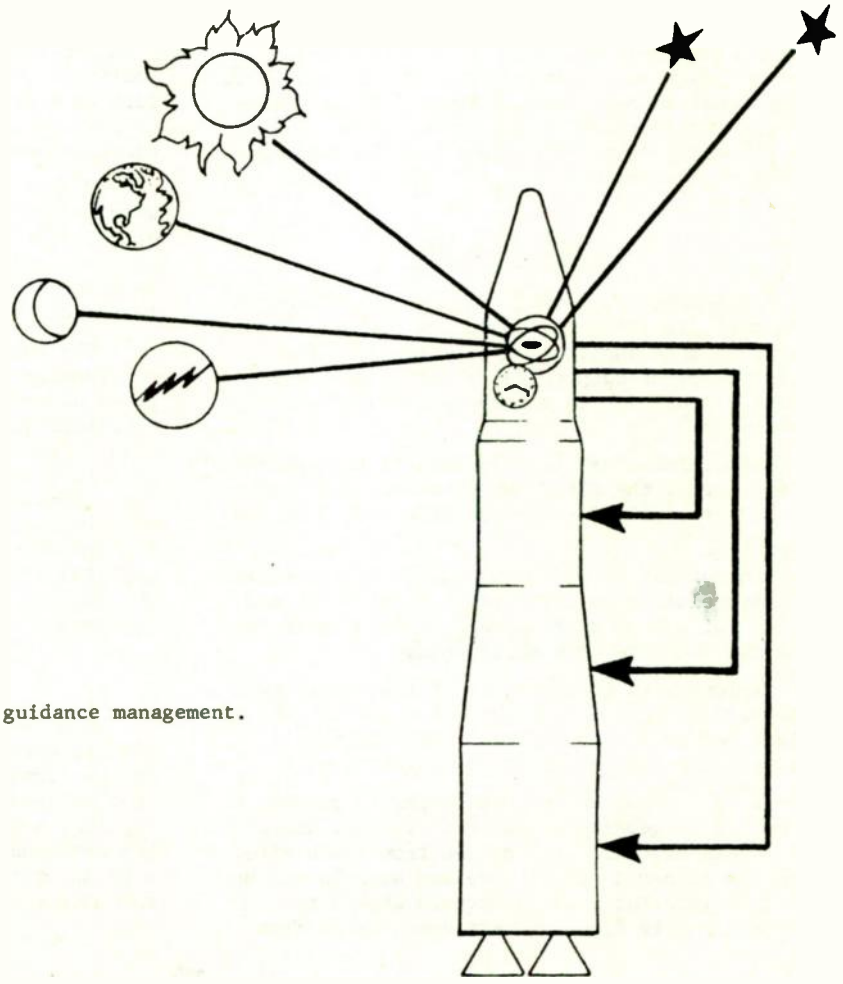


Fig. 9— Payload guidance management.

SPACESCAN--A RADIO BROADCAST SYSTEM OF SPACE NAVIGATION

A. Tatz
Airborne Instruments Laboratory
Deer Park, New York

Abstract

SPACESCAN is a radio-broadcast system of navigation guidance with total-volume coverage for on-board space-position measurement of geographical position, velocity and path angle for orbit, re-entry, and recovery of manned vehicles.

SPACESCAN provides a system of radio navigation aids that broadcast position signals useful in forming a space-position, three-dimensional grid with total-volume radio coverage. The vehicle equipment determines its position in the grid by receiving and processing signals that are radiated from earth-based sites over volume-scanning plane beams.

The beams are essentially thin, flat fans of K_u-band energy, modulated with elevation-angle data, and scanning in elevation angle from horizon to horizon over the site of an earth-based transmitter.

The vehicle-borne receiver measures elevation angle by decoding the beam contents as the beam scans through the vehicle. The elevation angle defines a plane of position; three intersecting planes of position are thus measured from two sites as far apart as permitted by line-of-sight coverage; the intersection of the three planes of position determines the vehicle position in rectangular coordinates relative to the center of the earth.

I. Introduction

For manned space vehicles that are either in earth-orbit or in earth-atmosphere re-entry, it is desired to make continuously available in the vehicle a set of outputs representing the vehicle's geographic position in a recognizable set of coordinates.

A coordinate set that is readily recognizable is, of course, the altitude, latitude, and longitude of the vertical reference line down from the vehicle.

Another set of useful outputs consists of the velocity relative to the center of the earth and the rate of change of altitude, from which is computed the relative path angle (gamma).

Various ways to accomplish the on-board navigation functions outlined above include the use of optical sensors, inertial platform, and radio communication with earth-based tracking nets.

It is the purpose of this paper to present a new method of geographic position fixing, using radio beams that are transmitted from earth sites in such a manner that the received signals can be processed into the desired outputs aboard the vehicle directly from the data demodulated from the radio beams without using tracking radars at the earth sites.

The system under discussion can be characterized as a radio-navigational aid, with the same connotation that is evoked when speaking of the network of VOR stations as a radio-navigational aid for aircraft. The correct implication is that navigation service is provided to any user who is equipped with compatible radio receiving equipment, provided that the vehicle is within radio coverage from two ground stations, each of which broadcasts modulated energy in a hemisphere centered over the zenith of each station site.

The usual way to visualize radio-navigation position fixing is to regard a datum of measurement as the basis for defining a surface of position. It is then axiomatic that three surfaces of position, intersecting in a point, are sufficient to define the vehicle's geographic position.

For convenience in visualizing the point of intersection, and for mathematical simplicity in solving for position coordinates, it is advantageous to use flat planes for the surfaces of position. Then a triplet of equations of the form $Ax + By + Cz = D$ expresses the three planes of position, and the solution for x, y, and z is straightforward.

The set of rectangular coordinates is taken with respect to the origin at the center of the earth, with the x-axis passing through the point of zero latitude and zero longitude. Then the geographic coordinates are obtained by a transformation to spherical coordinates.

The radio-navigational aid to be described provides geographic position fixing directly computed aboard the vehicle without dependence on computing look information, such as is needed for radar tracking techniques, and without dependence on initial-condition values, such as are needed for inertial devices. For these reasons, it is worth-while to consider this new system as a useful complement to the other prevailing systems, especially where emphasis is placed on maximum reliability and safety through redundancy of independent techniques.

II. Techniques

A predesign analysis of a radio-navigational aid for geographic position fixing has resulted in the formulation of a set of techniques embodied in a project called SPACESCAN. As the name implies, a beam-scanning principle is involved in the broadcasting of surface-of-position data from a ground site into the hemisphere of radio coverage over the site.

The beams are essentially thin, flat fans of K_u-band energy, modulated with elevation-angle data, and scanning in elevation angle from horizon to horizon over the site of an earth-based transmitter (see Figure 1).

The vehicle-borne receiver measures elevation angle by decoding the beam contents as the beam scans through the vehicle. The elevation angle defines a plane of position; three intersecting planes of position are thus measured from two sites as far apart as permitted by line-of-sight coverage; the intersection of the three planes of position determines the vehicle position in rectangular coordinates relative to the center of the earth (see Figure 2).

A SPACESCAN radiation pattern is very narrow (0.5 degree) in elevation angle and very wide (170 degrees) in azimuth angle. The beam can be approximately visualized as a flat sheet, like the page of a book. This analogy is useful in describing the scanning action. Imagine an open book lying horizontally on the table. As each page is turned, it represents the scanning of a beam, the center of the book binding representing the site of the station. In this context, the binding represents the axis of the scanning motion, and the line of scan is taken as the horizontal perpendicular bisector of the axis of scan. As each page is turned, the beam position is defined by the elevation angle between the plane of the beam and the horizontal plane of the book covers.

The peak-amplitude point on the radiation pattern is a point on the scanning plane of the page, directly over the horizontal line of scan. The peak-amplitude point is made to scan in a semicircle through 180 degrees of elevation angle, from one horizontal direction of the line of scan, through the station zenith, and over to the opposite direction of the line of scan (see Figure 1).

The scan semicircle lies in the vertical plane (perpendicular to the earth-tangent plane of the station site). Thus, the vertical semicircle is directly over the line of scan, and an elevation angle is formed between the line of scan and the semicircle's rotating radius vector. This elevation angle defines the plane of position of the vehicle when the center of the scanning beam is detected in the vehicle receiver.

The microwave energy in the scanning beam is pulse-modulated at the transmitter site, so that the time between pulses accurately represents the elevation angle of the peak of the beam. This beam is only 0.5 degree wide in the direction in which it scans 180 degrees in 1 second; therefore, the beam scans through the vehicle in a period of about 3000 μsec, the exact duration depending on the level of receiver detection below the peak amplitude of the beam.

The beam modulation consists of an interpulse spacing of between 30 and 150 μsec, where the interpulse spacing is linearly related to the elevation angle of scan at a scale factor of 4 μsec for degree of scan* (see Figure 3). Taking into account the period of scan through the vehicle (3000 μsec) and the interpulse spacing, it is seen that between 20 and 100 samples of elevation angle are measured by the vehicle receiver during each passage of a beam.

As each pulse is detected in the receiver, the interpulse spacing is measured and compared with the previously tracked and stored memory analog of the elevation angle. Throughout the entire train of received pulses, a tracking error voltage is generated, so that after the beam contents have been sampled in each scan, the corrected memory represents the plane of position with an accuracy of 0.25 milliradian in terms of elevation angle.

An elevation angle defining a plane of position is measured at the rate of three times per second, since three beams of the same type sweep across the vehicle each second. The tracking circuit in the receiver contains angle-velocity memory as well as angle memory; the angle-velocity data are used to update the angle data, so that the angle readout is current at the time it is sampled by the coordinate computer.

III. Geometry

In this discussion, a plane of position has thus far been defined by an elevation-angle value relative to a known line of scan in the earth-tangent plane containing the transmitter site.

Ultimately, the position of the vehicle must be computed from the intersection of three planes of position, and it is desired to express the computed position in a universal geographic coordinate system.

The universal geographic coordinate system is a rectangular coordinate system with the origin at the center of the earth. The positive X-axis passes through the zero meridian at the equator; the positive Y-axis is oriented 90 degrees east in the earth's equatorial plane; the positive Z-axis is oriented northward, perpendicular to the equatorial plane.

A given scan line is specified in terms of its three direction cosines relative to the XYZ axes; the station zenith is specified by three other direction cosines relative to the XYZ axes; the seventh constant required for coordinate transformation is the radius of the earth at the station site.

The following steps are used in deriving the mathematical expressions for a plane of position in geographic coordinates (XYZ).

---

* The modulation range over 30 degrees is repeated six times to cover 180 degrees of scan.

1. State the plane of position in terms of the sine and cosine of the elevation angle relative to the earth-tangent-plane origin.
2. Translate the origin along the earth's radius to the center of the earth.
3. Rotate the axes of the scan line and the station zenith to the XYZ axes, using the six direction-cosine constants for the rotational transformation.

After these steps are performed, the plane of position is expressed in the following form:

$$Ax + By + Cz = D \qquad (1)$$

where the coefficients A, B, C, and D are arithmetic combinations of the seven constants and the sine and cosine of the elevation angle.

Three equations of the form of equation 1 make up a system of linear equations from which the XYZ coordinates are computed, provided that the three equations represent three planes of position that intersect in a point fix.

In the foregoing discussion, consideration was given to one line of scan and one axis of scan, both lying in the earth-tangent plane, and one scanning beam whose peak-amplitude point moved in a semicircle over the line of scan.

If the roles of the line of scan and the axis of scan were interchanged, the result would be a scanning beam whose semicircle lay in a vertical plane rotated 90 degrees around the station-zenith axis.

A complete SPACESCAN station uses two mutually perpendicular lines of scan, and thus two simultaneous systems of scanning beams. The line of scan for one scan serves as the axis of scan for the other scan. For example, if one scan line runs north and south from the transmitter site, the second line of scan runs east and west from the same transmitter site (see Figure 4).

A plane of position is measured from each of the two scans, and the result is that two intersecting planes of position are measured from a single transmitter site.

In a typical situation, one line of scan is oriented so that it lies nearly in the plane of the expected trajectory of the vehicle. This arrangement signifies an along-trajectory scanning system. Its orthogonal counterpart, the second scan, signifies the across-trajectory scanning system (see Figure 2, planes A and C).

The intersection of the along-trajectory plane of position and the across-trajectory plane of position is a line of position from the transmitter site to the vehicle (see Figure 2, line $\overline{AC}$).

The third plane of position is measured advantageously from an along-trajectory scan that originates from another transmitter site. Preferably, one of the sites would be behind the vehicle, and its second site ahead of the vehicle in its trajectory. The intersection of the two along-trajectory planes of position is a line of position approximately perpendicular to the trajectory (see Figure 2, line $\overline{BC}$).

For best results in terms of minimum geometric dilution of position fix and maximum utilization of site configurations, the sites should be as far apart as line-of-sight radio coverage permits, taking into account the curvature of the earth.

In low-altitude orbits and for a re-entry trajectory, the expected altitude of the trajectory over the approaching site will be the principal limiting factor in siting for optimum coverage.

A second limiting factor is the change in the index of refraction for radio propagation at low elevation angles. The region of most useful coverage is limited to elevation angles above 6 degrees in order to minimize the angular errors introduced by errors in predicting the change in the index of refraction.

Taking into account the earth curvature and the 6-degree lower limit, the maximum site spacing is approximately 500 miles for 100-mile altitudes, 300 miles for 50-mile altitudes, and 200 miles for 30-mile altitudes.

A chain of about 12 stations is required to completely cover a typical re-entry trajectory of about 3000 miles and to provide continuous re-entry guidance data by SPACESCAN, from re-entry to the recovery area. (Trajectory data for this estimate were supplied by the Navigation and Control Section, Space Task Group, NASA, Langley Field, Virginia, during the study program.)

Figure 2 shows three planes of position intersecting at the circled point. Planes A and C are the planes of position for the across-trajectory and along-trajectory scans from site 1. Plane B is the plane of position for the along-trajectory scan from site 2. The circled position fix is visualized as the intersection of two lines of position, one formed by planes A and C (line $\overline{AC}$) and the other formed by planes B and C (line $\overline{BC}$).

The elevation angles that define the planes of position are each measured three times per second. This rate is estimated to be sufficiently rapid so that rates of change of X, Y, and Z are computable with about 1-second smoothing. Thus, the directly computed outputs are the rectangular coordinates of geographic position (XYZ) and their rates of change ($\dot{X}$, $\dot{Y}$, and $\dot{Z}$).

Guidance Outputs

The following positional data outputs are computed from the rectangular coordinate data:

1. Altitude (H) and its rate ($\dot{H}$), where H is the root sum square of X, Y, and Z (range R_o) minus the earth radius.
2. Relative velocity (V), which is the root sum square of $\dot{X}$, $\dot{Y}$, and $\dot{Z}$.
3. Relative flight path angle (gamma), which is the angle whose sine is $\dot{H}/V$.
4. Geocentric longitude, which is the angle whose tangent is Y/X.
5. Geocentric latitude, which is the angle whose sine is Z/R_o.

As an incidental feature, if it is desired to align the vehicle trajectory with a particular vertical plane containing a landing site, it is feasible to obtain lateral-alignment guidance directly from the across-trajectory scanning beam. The station site for this localizer-type operation would be in the desired vertical alignment plane, located on the earth's surface at a point to be passed over before reaching the landing area. The 90-degree value on the across-trajectory scan is taken as the reference plane, and left-right guidance is obtained as the deviation from 90 degrees of the elevation-angle output of the across-trajectory plane of position.

IV. Block Diagrams

Figure 5 is a block diagram of the major functional features of a SPACESCAN station. At each station, there are two systems--one for an along-trajectory scan and one for an across-trajectory scan. As the antenna rotates in a vertical circle, the scan of the peak of the beam is picked off the rotating mechanism at a rate of one pulse per 0.01 degree of movement. The accumulated pulse train is encoded to a time-delay scale (4 μsec per degree) by a digital-counting encoder. The time delay establishes the interpulse spacing between successive pulses of the modulator output and thus between successive pulses of the transmitter output. The K_u-band (16-kMc 0.35-μsec) pulses are radiated over the scanning beam. The result is that the pulse contents of the scanning beam contain the data representing the elevation angle of the beam throughout 180 degrees of scan.

The modulation code is repeated over six 30-degree sectors. The sector number (one to six) is encoded on a beam that precedes the angle-data beam by 1 degree in the scan (see Figure 6). The advance beam also contains pulse-pair coding that identifies the station site and the scan direction (along- or across-trajectory). The identity code is used in setting up the vehicle receiver circuits for unambiguous angle tracking, and in the vehicle coordinate computer for selecting the station constants involved in computing geographical coordinates of position.

Figure 7 is a block diagram of the major functional features of the vehicle equipment. The RF and IF sections of the receiver supply detected pulses to the identity decoder and the angle-data decoder. The decoder supplies a voltage analog of the interpulse spacing to one of three angle-tracking memory circuits, correcting the angle memory to define the plane of position of the vehicle. The automatic calibration operates just before the angle beam is received, establishing the decoder levels for highly accurate tracking outputs.

The computer contains stored constants relating to the identity of both the station and the scan; it operates on the three angle-data inputs to provide a position fix in earth-centered rectangular coordinates. These coordinates are then processed into the outputs shown in Figure 7. The outputs are used as back-up for the inertial guidance system, and are displayed for monitoring and for manual-control back-up.

V. Antenna

The 0.5-degree beamwidth is formed by an 8-foot antenna aperture in the vertical plane of the semicircular scan. This aperture is energized as soon as the beam rises above the plus axis, and the energy is cut off when the beam drops below the minus axis. Another antenna aperture, the back-to-back mate of its forerunner, will now rise above the plus axis, and the transmitted energy will be switched to the newly rising beam, as shown in Figure 8.

The antenna system for each line of scan consists of six linear apertures, arranged in the shape of two overlapping equilateral triangles (Figure 9). While their beams are above the tangent plane, three of the six apertures are energized. Since their beams would be directed below the tangent plane, the remaining three apertures are de-energized.

As the beam of an active aperture scans the semicircle over the tangent plane, the elevation angle of the pointing line of the beam is picked off the antenna structure at the rate of one pulse per 0.01 degree of antenna motion. The pulse count is run up on a register during a 30-degree sector of scan, run down on the next 30-degree sector of scan, run up again, and so forth. The pulse count, therefore, represents the angle of the pointing line in a 30-degree sector.

In a parallel register, the angle pulse count is run down at a standard rate (25 counts per microsecond). The time required to run down the count is taken as the measure of the interpulse spacing for modulating the transmitter.

The interpulse spacing is arranged to vary from 30 to 150 μsec, corresponding to a scanned sector of 30 degrees.

It is convenient to use a code sector that is one-half the angular spacing between beams in each of the semi-circular scans. As shown in Figure 9, the angle-data beams for the three active antenna apertures are separated by 60 degrees of scan angle. The range of interpulse spacing (30 to 150 μsec) encodes a 30-degree sector of scan. Under these conditions, the three beams that are on the air carry the same angle-data code (except for sector number). This permits some simplification in the techniques of picking off the rotational angle from the antenna system and for encoding the modulators of the RF transmitters.

The scanning speed is 180 degrees per second, so that one degree of scan takes 5555 μsec. With an interpulse spacing of 150 μsec, a receiver will detect a train of about 40 pulses within 1 degree of scan. The half-power beamwidth is 0.5 degree, and the system can be designed so that only pulses more than 16 db below the strongest ones in a beam fail to contribute to the angle-data measurement. Consequently, the effective measurement beamwidth is practically

1 degree, and at least 35 to 40 pulses are contained in the effective beam envelope. Based on experience with angle measurements in AIL Landing System equipment, it is expected that the angular accuracy will not be significantly affected by variations in interpulse spacing for the system of SPACESCAN techniques described here and that an angular accuracy of 0.25 milliradian is technically feasible.

VI. Identity Coding

To interpret the angle readout operationally, the vehicle equipment must have information relative to the identity of the station (one of about twelve stations), the identity of the scan (one of two possible directions for each station), and the identity of the 30-degree encoded sector (one of six possible sectors for each scan).

The information relating to station identity, scan identity, and sector identity is contained on a microwave beam that scans over the vehicle just before the reception of the angle-data beam. An advance beam is associated, on a one-for-one basis, with each of the six angle-data beams from a station. Using a separate transmitter and a microwave antenna feed slightly displaced from the focal point of the angle-data feed, the advance beam is formed by the same antenna aperture that forms the associated angle-data beam. The small displacement from the aperture focus gives the advance beam pattern a 1-degree spatial displacement toward the direction of the final end of the semicircle of scan. Consequently, the vehicle equipment will first receive the advance beam and then the angle-data beam, all within a period of about 1/60 second (Figure 6).

The principal function of the advance beam is to establish the identity of the station, the scan, and the sector. Secondary functions of the advance beam are to set the automatic gain control of the receiver for the associated angle-data beam and to gate on the receiver circuits to anticipate the advent of a train of angle-data pulses.

Taking advantage of the fact that the angle-data pulses are single pulses spaced not less than 30 μsec apart, the identity messages in the advance beam are distinguished by pulse pairs whose interpulse pair spacing is 20 μsec for the station and scan identity and 24 μsec for the sector identity.

Each scan will have a constant address message for its site number (one of about ten) and its scan direction (one of two). Accordingly, the 20-μsec address pairs of pulses are spaced 41 to 65 μsec apart (leading pulse to leading pulse), thus allowing 24 possible addresses at increments of 1 μsec. A typical address message thus consists of two 20-μsec pulse pairs within an interval of 85 μsec maximum.

The address messages are interlaced, on a one-for-one basis, with sector identity messages. These are 24-μsec pulse pairs whose interpair spacing (leading pulse to leading pulse) is between 40 and 46 μsec. This allows six possible sector identities at increments of 1 μsec per sector. A typical sector identity message thus consists of two 24-μsec pulse pairs within an interval of 70 μsec.

A total set of messages (station, scan, and sector) is thus completed within less than 200 μsec. Since a 1/2-degree beam takes about 2800 μsec to scan through the vehicle equipment, the advance beam will deliver at least 14 complete scan address and sector identity messages before the pulses of the angle-data beam are received by the vehicle equipment.

Since the three message-data beams for one scanner carry the same information relating to site identity and scan identity, the only difference among the transmitted data for a given scanner is the sector identity number; the difference between sector identity numbers is precisely two at any instant. As previously discussed, the angle-data beams for one scanner carry the same angle-data code. These universal features among the data are useful in simplifying the monitoring equipment required at the station sites.

To obtain more accurate updating of the angle data during the condition of angular acceleration, it is advantageous to obtain several beam measurements per second. Accordingly, the system is designed to provide three beam measurements per second per scan, using the dual-equilateral-shape antenna structure that radiates three beams 60 degrees apart (Figure 9).

VII. Performance Analysis

Figure 10 shows the predicted acquisition ranges for the achievable transmission and reception characteristics and for a relatively severe weather model. It is seen that the K_u-band system is good for maximum ranges of about 1000 to 3000 miles, depending on weather and angle of elevation. This range capability is more than adequate for re-entry applications, where the curvature of the earth limits radio coverage to ranges less than 1000 miles.

Accuracy of position fix is determined by the angular accuracy of the system and by the geometric dilution inherent in the intersections of the three planes of position. The angular system accuracy is 0.25 milliradian, so that very roughly the position accuracy is 0.025 percent of the distance to the more remote station--for example, 0.1 mile for a site spacing of 400 miles.

A more precise error analysis requires that the partial errors due to angular error in each plane of position be combined by taking into account the geometric relations between the trajectory and the site locations. A precisely computed error analysis is shown in Figure 11. A possible trajectory is shown for a 2000-mile re-entry toward a landing area.

The station sites are shown on the bottom scale of the figure. Stations 1 through 7 are spaced 240 miles apart, stations 7 through 9 are spaced 210 miles apart, and station 10 is located

90 miles from station 9. For the re-entry trajectory shown at the top of the figure, the chain of stations provides solid coverage for three planes of position for re-entry trajectories that are within about 100 miles to either side of the line of stations.

Altitude error, along-course error, and across-course error were computed for a trajectory directly over the station chain. The errors were computed as the square root of the total variance resulting from a standard deviation of 0.25 milliradian for each of the three angle values that define the three planes of position used in determining a position fix.

From the figure it is seen that the maximum altitude errors (about 350 feet) occur directly above the stations; the minimum altitude errors (about 120 to 150 feet) occur midway between stations.

The maximum along-trajectory errors (about 450 feet) occur midway between stations and the minimum along-course errors (about 50 to 100 feet) occur directly over a station.

The across-trajectory errors (about 50 to 100 feet) are the smallest of the three errors, mainly because the trajectory is over the chain of stations.

The differences among the distributions for the three types of errors are due to the differences in geometric dilution at different positions with respect to the pairs of sites from which the planes of position originate. The transients in the error plots are due to changes in the particular pairs of sites used along the trajectory.

Throughout the entire trajectory over the chain of stations, the position error in any direction is uniformly less than 500 feet.

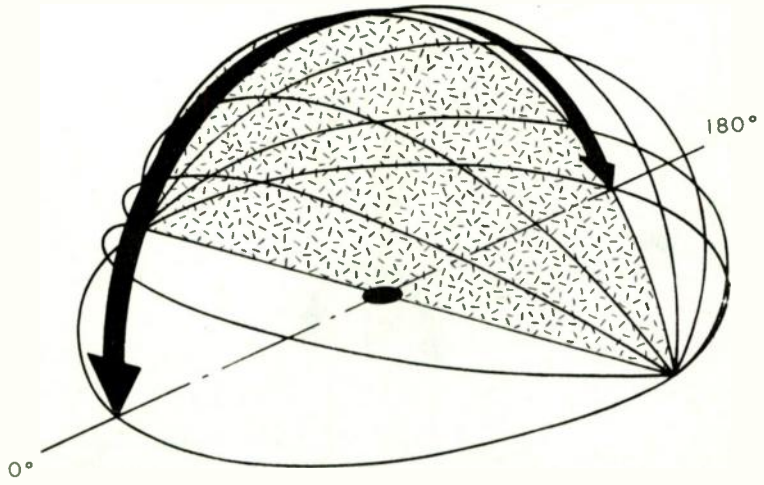


Fig. 1. One-plane scanning.

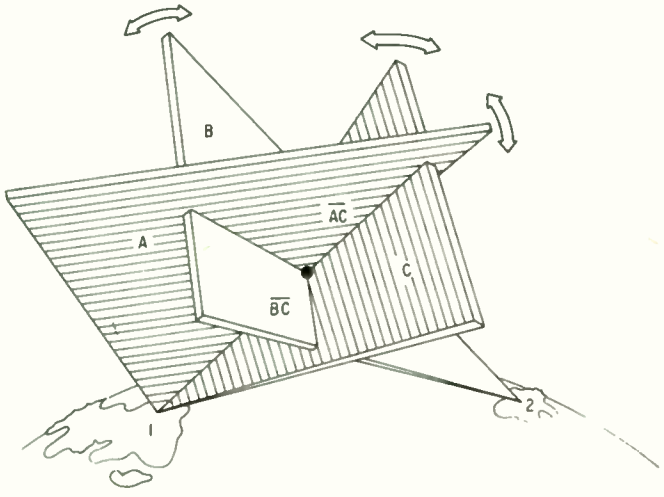


Fig. 2. Intersecting planes of position.

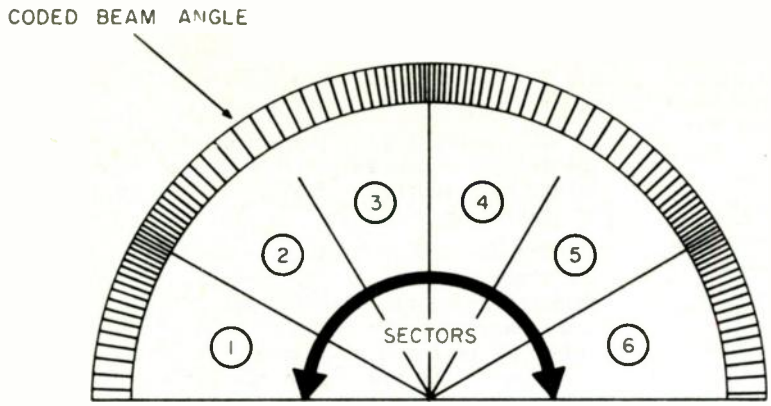


Fig. 3. Angle data transmission and reception.

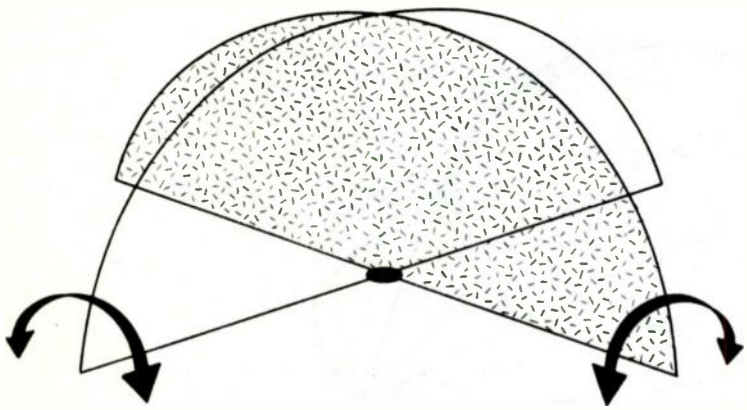


Fig. 4. Scanning beams from each ground site.

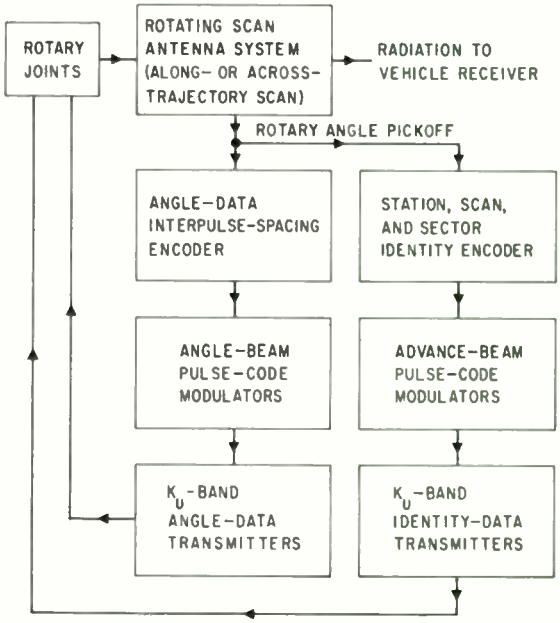


Fig. 5. Functional block diagram of SPACESCAN ground station.

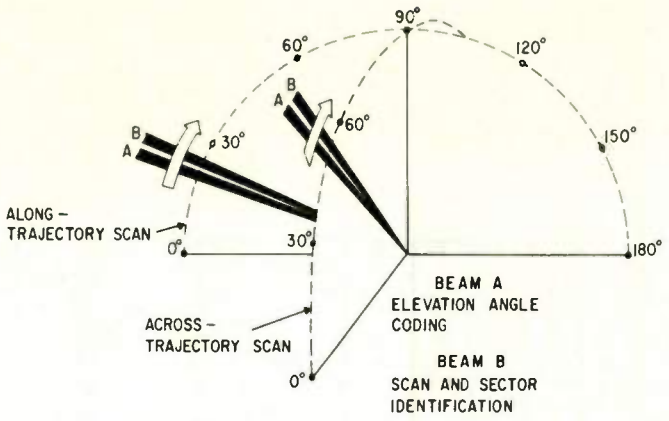


Fig. 6. Relationship between advance (identity) beam and angle-data beam.

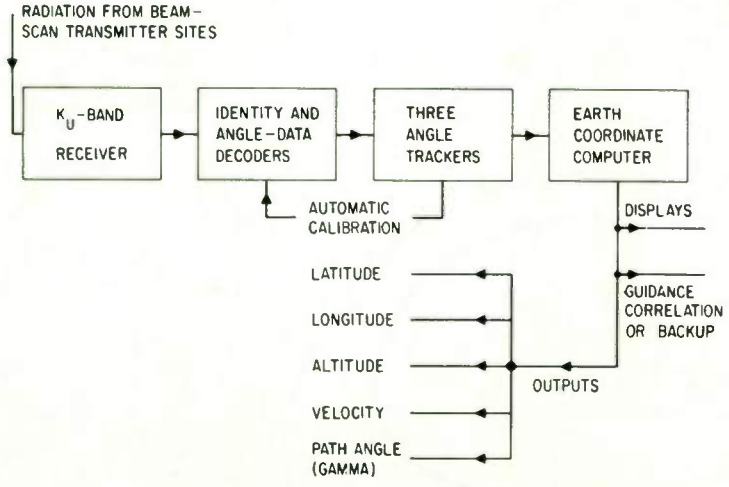


Fig. 7. Functional block diagram of SPACESCAN vehicle receiver.

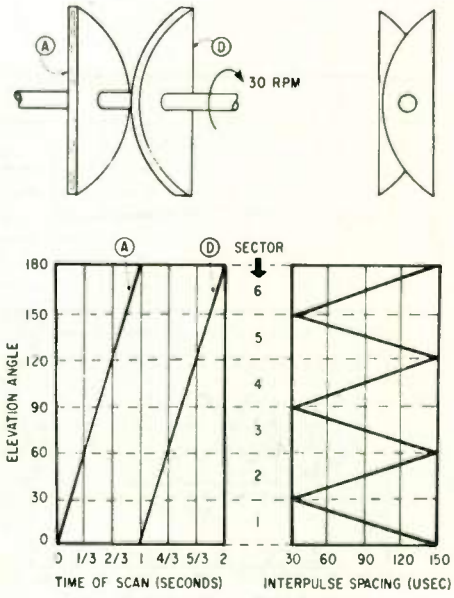


Fig. 8. Back-to back scanner.

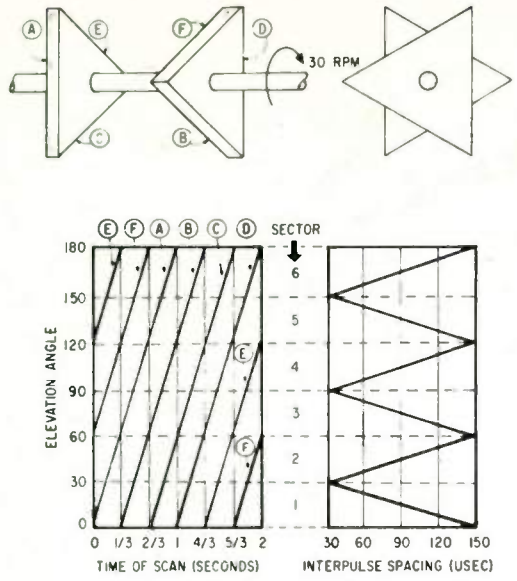


Fig. 9. Dual-equilateral scanner.

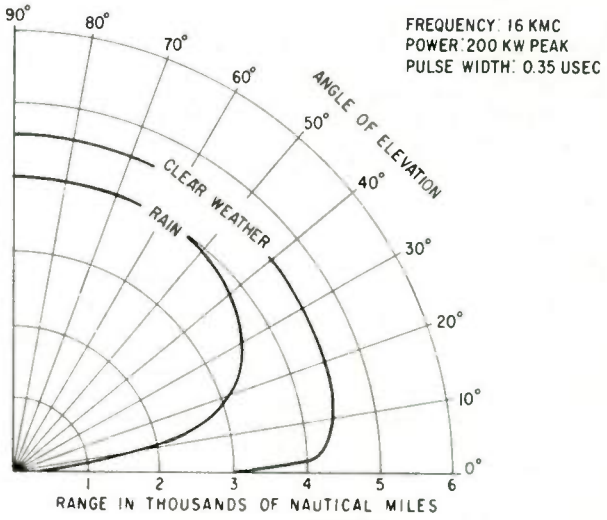


Fig. 10. Predicted acquisition ranges for SPACE-SCAN.

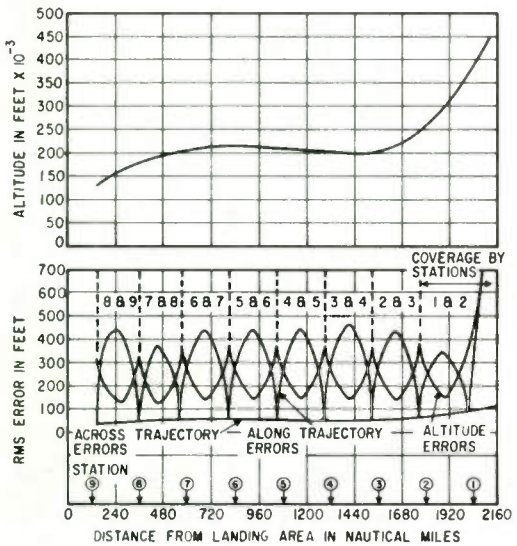


Fig. 11. Re-entry trajectory and error plot.

THE DESIGN AND SUPPORT OF A TOPSIDE SOUNDER

W. B. Offutt and S. Russell, Jr.
Airborne Instruments Lab.
Deer Park, N. Y.

Abstract

The principal factors determining the operational parameters and important aspects of the payload, checkout and data processing equipment for an ionospheric topside sounding experiment are described. This represents a cooperative program with the Central Radio Propagation Laboratory of the National Bureau of Standards, under the management and technical direction of the Goddard Space Flight Center of NASA. Equipment for both the rocket probe phase and the satellite phase of the program is described, with emphasis on novel circuit details and reliability aspects.

DYNAMIC ANALYSIS OF OAO SPACECRAFT MOTION BY ANALOG-DIGITAL SIMULATION *

George Zetkov and Ross Fleisig
Grumman Aircraft Engineering Corporation
Bethpage, New York

Summary

The Orbiting Astronomical Observatory is a 3300-pound satellite capable of accommodating a variety of astronomical experiments. It is to orbit the earth at an altitude of 500 statute miles in an approximately circular orbit which is inclined 32 degrees to the equatorial plane. The first observatory will be launched from the Atlantic Missile Range in 1964, and will be boosted to orbital velocity by an Atlas-Agena B vehicle.

Two main component systems comprise the observatory: the spacecraft - being developed by Grumman for the Goddard Space Flight Center of NASA - and the experiment packages. The spacecraft is to have a 70% probability of operating in orbit for a lifetime of one year. The experiment packages, different for each OAO, will be supplied to NASA by leading astronomers.

One of the major subsystems of the OAO spacecraft is the Stabilization and Control Subsystem. While an experiment is being conducted, the OAO Stabilization and Control Subsystem will point the satellite, thus sighting the experimenter's telescope at a point in space. Coarse Pointing, an operational mode of the Stabilization and Control Subsystem, involves control of spacecraft attitude about all three of its axes and is utilized when a particular class of experiments is being performed and also during the transition to the fine pointing mode, in which other experiments are performed. The coarse pointing mode utilizes gimballed star trackers to sense spacecraft angular motion, stability compensation networks, amplifiers, and drive motors coupled to three orthogonally-mounted inertia wheels which torque the spaceframe. Separate channels of control are available for the spacecraft yaw, pitch, and roll axes. An attitude command is inserted into the system in the form of star tracker command gimbal angles.

Selection of the combined analog-digital simulation technique for the dynamic analysis of coarse pointing control was based mainly on the considerations: (1) ease of mechanizing logic, (2) machine time, and (3) fidelity of simulation. Review of these factors led to simulating the major portion of the control loop in analog form and the balance in a digital program. A data link consisting of control equipment and analog-to-digital and digital-to-analog converters tied together general purpose analog and digital computers. The analog-digital simulation technique was found to be an effective method of analysis.

Results of the dynamic analysis demonstrate that the spacecraft motion for proper selections of guide stars is satisfactory with respect to precision pointing and attitude holding requirements. The torques due to inertia coupling, and precession torques due to fine wheel rotation relative to the spacecraft are negligible during normal attitude hold operation. Further, the gimbal angle sampling rate, for trackers in the tracking mode, is adequate. The spacecraft error about each control axis due to gimbal angle quantization is approximately 5 seconds of arc or less. The maximum value of this type of error corresponds to the resolution of the gimbal angle transducers.

* This work was done under Contract NAS 5-814 with the Goddard Space Flight Center of NASA.

Introduction

The specific control configuration which is analyzed in this paper is that of the coarse pointing mode - one mode of the OAO Stabilization and Control Subsystem. The major objective of this paper is to demonstrate by analysis the capability of the coarse pointing control system to meet extremely exacting satellite pointing and attitude holding requirements. These requirements are: the control system shall be capable of holding the spacecraft attitude with an accuracy of 1' of arc (rms circular error), and shall maintain this position within 15 seconds of arc for no less than 50 minutes.

A secondary objective of this paper is to comment on the analog-digital simulation technique as it was applied to the analysis of the coarse pointing mode.

Prior to the discussion of the coarse pointing mode analysis, general background information on the spacecraft with emphasis on the Stabilization and Control Subsystem is provided in order that the analysis may be viewed with the proper perspective.

OAO Configuration

Figure I is a photograph of a full-scale mockup of the OAO spacecraft. The main body is 116 inches long and has an octagonal cross section 80 inches across the flats. Rectangularly-shaped solar cell paddles, approximately 52 by 138 inches, are folded flat against the side during launch and are erected following burnout of the booster. The satellite weighs 3300 pounds including an allowance of 1000 pounds for the experimenter's payload. Housed in a rigidly-supported 40-inch diameter central tube, each payload package contains a specially-designed telescope or imaging sensor together with supporting optical equipment.

Major subsystems of the OAO spacecraft include: Structure, Data Processing and Instrumentation, Communications, Power Supply, Thermal Control, and Stabilization and Control. The Stabilization and Control Subsystem is now described briefly.

Stabilization and Control Subsystem

The functional operation of the OAO Stabilization and Control Subsystem is described briefly in terms of sensor and actuator components. Depending on the operational mode that is programmed, one or more of the sensors (rate gyros, sun sensors, star trackers, etc.) will activate a particular actuator (a set of jets or

an inertia wheel). The modes are described in sequence.

Initial Stabilization and Orientation. Initial tumbling rates are reduced and, at the same time, the spacecraft roll axis is locked on the sun. This is accomplished by initially combining coarse sun sensor and rate gyro signals to control the corresponding high level jets in the pitch and yaw axes and by feeding the roll rate gyro signal to the roll high level jet. After rate null and fine sun sensor operative signals are experienced, pitch and yaw motions are controlled by the fine inertia wheels in response to the corresponding fine sun sensor outputs. When null signals occur in the pitch and yaw loops, acquisition of the guide stars by the star trackers is initiated. This results in rolling the spacecraft at a constant rate about the sunline by activating the roll high level jets in response to a biased output signal from the roll rate gyro. By gimballing, each of the star trackers possesses two degrees of rotational freedom. The gimbals for each tracker are fixed at pre-computed angles with respect to spacecraft axes. Since these commanded angles correspond to known guide star positions when the spacecraft roll axis is directed at the sun and the spacecraft is at some particular roll angle, acquisition will result when the trackers' detector outputs give a simultaneous indication that stars are present in their respective optical fields of view. Following star acquisition, the commanded roll rate signal to the roll jet is removed and roll rate is reduced to the null level of the rate gyro. At the same time, the gimbals are freed, allowing the star trackers to track their acquired guide stars. At this stage, control of the fine inertia wheels in pitch and yaw is switched from the fine solar sensors to the star trackers. In addition the roll fine inertia wheel is controlled by the star trackers.

Slewing (Reorientation). The roll axis of the spacecraft is slewed from the line of sight to the sun (or any other point in space) to the line of sight of another point in space. This is achieved by commanding a specified number of rotations for each of the three coarse inertia wheels. Since the ratio of spacecraft angular displacement to inertia wheel rotation depends on the ratio of wheel to spacecraft moment of inertia, specific spacecraft attitude changes may be obtained by commanding pre-computed numbers of wheel rotations. This mode of operation is performed one axis at a time on an open loop control basis. During slewing, the star trackers are tracking but the tracker output signals are disconnected from the fine inertia wheels.

Coarse Pointing. In the coarse pointing mode, star trackers are utilized as sensors, and the fine inertia wheels are utilized as actuators. Six gimballed star trackers are mounted on the spacecraft. Under favorable conditions, the minimum number required for three axis attitude control is two. The extra trackers are provided: (1) to substitute for those which occasionally are occulted by the earth, (2) to increase reliability by redundancy, and (3) to improve accuracy by averaging the outputs of trackers which are effective. Detailed description of this mode is given in the next section.

Fine Pointing. During this mode, pitching and yawing motions are sensed by special optical equipment provided by the experimenter. Signals from these sensors are fed to the pitch and yaw fine inertia wheels. The roll fine inertia wheel is controlled by the star trackers. This type of control is designed to yield much higher pointing accuracy in the pitch and yaw channels than that resulting from Coarse Pointing Mode operation. The increased accuracy is specified for certain astronomical experiments.

Description of Coarse Pointing Control

In this section, reference coordinate systems which are utilized in the mathematical model of the coarse pointing control configuration are described. A functional description of coarse pointing control is then provided to lead into the presentation of the mathematical model. The mathematical model includes analytical representations of the formation of control error signals as developed by the star trackers, the action of the inertia wheel actuators in response to control error signals, and the spacecraft motion resulting from actuator and other torques.

Reference Systems.

The basic reference system is a geocentric, quasi-inertial, equatorial coordinate system with axis E_1 directed toward the first point of Aries, axis E_3 directed toward Polaris, and axis E_2 directed so that it completes an orthogonal right-handed coordinate system.

The spacecraft control axes form an orthogonal right-handed coordinate system fixed in the OAO as shown in Figure 2.

The definition of the star tracker reference axes with respect to the spacecraft control axes is presented in Figure 3. Each set of tracker reference axes x, y, z is a right-handed set of orthogonal axes. The x axis of each set points along the positive or negative direction of one of the three control axes, i.e., x_c, $-x_c$, y_c, $-y_c$, z_c, $-z_c$.

The outer gimbal axis corresponds to the z axis in each tracker reference set and remains fixed regardless of tracker position. The inner gimbal axis corresponds to the y axis in each tracker reference set at the tracker null position only. The inner gimbal axis rotates in the tracker xy reference plane from the y reference axis by an amount equal to the outer gimbal angle.

With respect to any tracker reference set of axes, the outer and inner gimbal angles are measured about the outer and inner gimbal axes, respectively, and therefore correspond to Euler yaw and pitch angles. The outer and inner gimbal angles of the y tracker are illustrated in Figure 3.

Functional Description

A guide star is one of a set of stars acceptable with respect to tracker detection capability and is utilized by a tracker so that its optical axis remains fixed in inertial space. The motion of the spacecraft, relative to the stationary trackers, is detected by changes in the measurements of tracker positions with respect to the spacecraft.

A tracker is driven by pointing error signals developed in the tracker optical detector. In order to simulate the generation of pointing error signals, the actual tracker position must be compared with the line of sight to that guide star being utilized by the tracker.

Reference is now made to the functional block diagram in Figure 4. Utilizing knowledge of the spacecraft attitude and line of sight to the j^{th} guide star, both with respect to an inertial reference frame, the line of sight to the j^{th} guide star with respect to the k^{th} tracker reference axes may be determined. That is, the inner and outer gimbal angles that are required to point the k^{th} tracker precisely at the j^{th} guide star are determined. Comparison of the inner gimbal required to point the k^{th} tracker at the j^{th} guide star with the actual inner gimbal angle yields a measurement of one of the tracker pointing error signals. This pointing error is utilized to drive the tracker about the inner gimbal axis towards the line of sight to the j^{th} guide star. Similarly, a pointing error is generated simultaneously to rotate the tracker about the outer gimbal axis. The inner and outer gimbal angle commands are the tracker gimbal angles that correspond to the desired spacecraft attitude. Hence, for a specified spacecraft attitude, the fixed gimbal angle commands provide a reference for the actual inner and outer gimbal angles. In order to compute the gimbal angle commands, a desired spacecraft attitude is inserted at point A in Figure 4, the guide star position is set in at point B, and the command inner and outer gimbal angles are obtained at points C and D.

By comparison between command gimbal angles and actual gimbal angles, gimbal angle errors are computed. These errors are then transformed into error signals about the roll, pitch, and yaw spacecraft control axes as seen by the k^{th} tracker. The error signals about the roll control axis as seen by each of the operating trackers are averaged to form the error signal used for control about the roll axis. Similarly, pitch and yaw error signals are formed. Due to the acceleration or deceleration of the fine inertia wheels in responding to the roll, pitch and yaw errors, reaction torques are applied to the spacecraft. The spacecraft attitude, determined as a function of all torques applied to the spacecraft, feeds back into the computation of inner and outer gimbal angles required by the k^{th} tracker in order to point at the j^{th} guide star.

Mathematical Model

Details left out of the preceding functional description are now included. In the mathematical model of Figure 5, the arrangement of functional elements is based upon the flow of signals just as it was in the functional block diagram.

The gimbal angles of the k^{th} tracker, when pointing at star j, are computed by means of the matrix equation at the top of Figure 5[1]. A unit vector pointing toward star j is projected upon the inertial reference frame. The resulting vectors along the inertial coordinates are in turn projected upon the spacecraft control axes. The process is continued, moving through appropriate rotations to the k^{th} tracker reference axes and then through particular gimbal angles ψ_{jk} and θ_{jk} to a right-handed orthogonal reference system with x axis pointed in the same direction as the direction of the original vector. The projection along the x axis of the last reference system must be 1, and the projections upon the y and z axes must be zero.

The gimbal angles of the k^{th} tracker when pointing at star j, θ_{jk} and ψ_{jk}, are compared with the actual tracker gimbal angles, θ_{Gk} and ψ_{Gk}, in the k^{th} tracker optical detector. Error signals are generated which drive the k^{th} tracker to the j^{th} star through the use of gimbal torquers.

The k^{th} tracker gimbal angles θ_{Gk} and ψ_{Gk} are measured by gimbal angle transducers and each gimbal angle is sampled every T seconds. Each of the twelve gimbal angles is sampled in turn, a time period of τ_s seconds separating the successive samples of different gimbal angles. The k^{th} tracker gimbal angles are compared with the k^{th} tracker command gimbal angles to form the k^{th} tracker gimbal angle errors, $e\,\theta_{Gk}$ and $e\,\psi_{Gk}$. The k^{th} tracker gimbal angle errors are transformed into spacecraft control error signals as seen by the k^{th} tracker. After averaging spacecraft control error signal contributions from each operating tracker, the final spacecraft control error signals, through reaction torques applied by the fine inertia wheels, drive the vehicle to some new attitude, expressed in terms of the Euler angles ψ, θ, ϕ.

Simulation of Coarse Pointing Control

As indicated in the section above, the mathematical model for coarse pointing control involves a large number of equations which define the loop dynamics, non-linearities, reference axes transformations, and other kinematic relations. In view of the magnitude of the overall computational requirements for solution of these equations, the most effective approach to the determination of control dynamic performance appeared to be the use of simulation techniques.

Selection of Computation Facility

Both analog and digital computer simulation methods were considered. In order to simulate the characteristics of gimbal angle quantization, sampling, zero-order hold, and time delay, the use of a general purpose digital computer was considered highly desirable. Although these characteristics may also be simulated on an analog computer, the fidelity of simulation, operating at real time or faster, was considered to be less than that obtainable from a digital computer. The convenience of utilizing the built-in logic capability of the digital computer was also desirable. For accuracy reasons, gimbal angle data are processed digitally in actual system operation so it is natural to resort to digital techniques for this part of the problem simulation.

However, it was also desired to take advantage of the general purpose analog computer as a "development tool". That is, continuous line recordings of the important control variables are immediately available as solutions to the equations. This permits rapid interpretation of the results. New investigation goals may be established and system parameters may be adjusted to fulfill the new goals; hence, a run may be taken shortly after a previous run. Also, almost continuous

operation through the day each day is a normal procedure. Further, in general, analog operation requires significantly less machine time than digital operation.

In view of the above, the desirability of a combined analog-digital simulation was apparent. Moreover, an initial effort toward development of this simulation technique was underway.[2]

Hence, the decision was made to simulate the problem with analog-digital equipment during short periods of each day, the remaining part of the day to be used for all-analog simulation. All-analog simulation, which excludes the effects of gimbal angle quantization, sampling, time delays, and zero order holds, was also included for two reasons. It was desirable to establish satisfactory operation of the analog portion of the simulation prior to incorporating the digital program. This procedure minimizes the utilization of the digital computer. The second reason is that a comparison would be made between analog-digital and all-analog simulation results. If all-analog response data yielded satisfactory results in certain areas, then the all-analog simulation would be employed to obtain part of the required results.

The units of equipment in the hybrid simulation were an IBM 704 general purpose digital computer, a Reeves REAC 400 general purpose analog computer, and an Adage data link. The data link, through analog-to-digital and digital-to-analog converters, provided communication between the digital and analog computers. Five variables could be transmitted from the analog to the digital computer, processed in the digital computer, and then returned to the analog computer. The data link converted an analog signal into a binary signal of 11 bits plus a sign bit.

Application of Analog-Digital Simulation Technique

The digital program was coded in machine language because other coding systems could not handle the necessary data link control instructions.

After the digital computer sent an "operate" instruction to the analog computer, gimbal angle errors were not fed into the analog portion of the simulation until a short time interval, 0.1 second, had elapsed. The wait period was provided in order to put the analog computer into full operation and allow the recorder to establish proper speed. The transients during the wait period due to analog computer initial conditions were negligible.

A "sample" problem was run prior to the simulation of coarse pointing control in order to test the operation of the data link and digital computer program. The analog computer portion of the sample problem simulation was a simplified version of the analog computer portion of the coarse pointing simulation. In other respects the sample problem and coarse pointing simulations were the same. Sample problem results were satisfactorily compared with predicted results obtained by use of z transform theory.

The computation rate utilizing the analog-digital facility could not be increased significantly beyond the rate of real time operation. This limitation was a function of the time required for gimbal angle processing as it was performed in the digital computer program. (The machine time required when utilizing the analog-digital facility, however, is less than that which would be required by an all-digital simulation.)

Control of the analog-digital facility could be exercised essentially in the same manner as an analog computer is controlled. However, more planning and coordination were required prior to performing a series of runs in order to prevent the use of the digital computer over extended intervals of time.

The simulation accuracy of gimbal angle quantization, sampling, zero-order hold, and time delay as provided by the data link and digital program was satisfactory.

To exploit the analog-digital simulation technique fully, a capacity to transmit a large number of variables back and forth between the analog and digital computers is desirable.

Results of Analysis

Various approximations to the mathematical model of Figure 5 were utilized in early simulations. In these initial efforts, certain parts of the model were found to have negligible effect upon system operation within restricted regions of particular variables. As a result, in a second series of analog-digital runs, those portions of the model to which system performance was not sensitive were not included.

Some of the characteristics included in the early simulations were the following:

a) the determination of two guide star positions in terms of tracker gimbal angles, where the gimbal angles for the two trackers were restricted to small magnitudes;
b) inclusion of the field of view limitation in the tracker optical detector characteristic;
c) adoption of an assumed form for the gimbal dynamics which included a gimbal torquer with torque and speed saturation and tachometer feedback;
d) use of command gimbal angles corresponding to a desired spacecraft attitude;
e) inclusion of coupling torques due to differences in spacecraft moment of inertia and crossproducts of inertia;
f) inclusion of coupling torques due to precession of the fine inertia wheels;
g) simulation of reaction torques arising from star tracker gimbal motion, and
h) nonlinear transformation of spacecraft angular rates about control axes to obtain attitude relative to inertial reference coordinates.

For spacecraft motion restricted in its departure from a nominal attitude, no noticeable effects upon spacecraft motion were obtained due to inclusion of tracker servos (characteristics (b) and (c) above) and inertia coupling, precession, and tracker gimballing reaction torques (characteristics (e), (f) and (g)). Further, if attention was focused on attitude hold as contrasted to transient response to relatively large initial attitude errors, the excursion of the vehicle

attitude was generally extremely small - much less than 1 minute of arc.

Revised Simulation Diagram

In view of the above, the mathematical model was simplified for additional simulation. Tracker servos were assumed to be perfect. Inertia coupling, precession, and tracker gimballing reaction torques were neglected. To obtain a major reduction in model detail, a "perturbation" concept was applied to the mathematical model within a restricted region of spacecraft motion.* The perturbation concept was applied to the description of satellite motion and to the description of the line of sight to the j^{th} star in terms of tracker gimbal angles. Thus characteristics (a), (d) and (h) above were modified. Perturbation relations were used in (a); command gimbal angles of (d) were matched with nominal gimbal angles and set equal to zero; and the transformation of (h) was linearized.

The resulting simulation diagram for two tracker operation is shown in Figures 6a and 6b. In Figure 6a, the spacecraft rates are integrated directly to give angular perturbations $d\phi$, $d\theta$, $d\psi$ about the control axes from some nominal attitude. If the Euler angles ψ, θ, ϕ and gimbal angles θ_{jk}, ψ_{jk} in the matrix equation which appears at the top part of the mathematical model of Figure 5 are each set equal to a nominal value plus a corresponding perturbation, relationships between the spacecraft angular perturbations $d\phi$, $d\theta$, $d\psi$ and gimbal angle perturbations $d\theta_{jk}$, $d\psi_{jk}$ are obtained. By neglecting secondary effects, these relationships are linearized. The same relationships are obtained systematically by the following procedure:** (1) postulate gimbal angle perturbations (including roll gimbal angle perturbation which is not instrumented in the control loop), (2) determine from the geometry the spacecraft rotations that would cause the postulated gimbal angle perturbations of (1), (3) compute the inverse of the matrix obtained in step (2). The result of step (3) is given in Table 1.

In Figure 6b, the time sequence of gimbal angle sampling and error release is given for a particular pair of trackers.

Results from Use of Revised Simulation Diagram

Utilizing the simulation diagram of Figures 6a and 6b, the basic objectives of the study were to (1) examine the effect of gimbal angle sampling rate upon loop performance, (2) observe the effect of quantization upon control behavior, (3) examine the sensitivity of loop performance to variation in the gimbal angle sampling rate, and (4) investigate the accuracy and drift of the spacecraft attitude control.

To perform the first two objectives, responses utilizing the analog-digital simulation diagram of Figure 6 were compared with responses from an all-analog simulation. In the all-analog runs, the control loop configuration was the same as for the analog-digital simulation, except that quantization, sampling, time delays, and zero order holds were omitted. A typical run is given in Figure 7. Values of parameters and initial conditions for this run are given in Table 2. In this example, the three spacecraft angular perturbations were initially set at a value of 20 seconds of arc. The spacecraft angular perturbations and the envelopes of the control error signals for the analog-digital simulation differ from the spacecraft angular perturbations and control error signals of the all-analog simulation by less than 5 seconds arc. After transformation and averaging, the gimbal angle error signals contribute to the total control error signals, $e\phi$, $e\theta$, $e\psi$. The control error signals cycle between two values which are approximately one quantum or 5 seconds of arc apart. As may be noted from Figure 7, the minimum half period of control error limit cycling is approximately 1 second, an interval far greater than the gimbal angle sampling period, which is 0.048 seconds. The above results indicate that the sampling rate is adequate with respect to attitude control, and that differences between spacecraft motions obtained by analog-digital simulation and spacecraft motions determined from all-analog simulation are due to gimbal angle quantization. Spacecraft attitude perturbations in response to control error limit cycling are greatly attenuated at the cycling frequency of about 1/2 cycle per second.

In order to test the sensitivity of control loop behavior to a variation in the gimbal angle sampling rate, the simulated sampling period was increased from 48 milliseconds to a level at which dynamic performance was significantly changed. At a sampling period of as much as 1.5 seconds, roughly 20% overshoot was obtained in the initial transient response in contrast to no overshoot in the initial transient for a sampling period of 48 milliseconds. Hence, loop performance is not sensitive to a variation in gimbal angle sampling rate.

In order to examine spacecraft pointing and attitude holding capability over a real time interval of one hour, it would have been necessary to operate the analog-digital simulation for at least the same period. To conserve machine time, an all-analog simulation was performed with the computer running 50 times faster than real time. However, while the all-analog simulation is satisfactory for the determination of spacecraft motion, this simulation is not, of course, applicable for the detailed examination of gimbal angle measurements, gimbal angle errors, and the resultant control error signals.

Typical results are shown in Figure 8. The pointing error between the line of sight to a target in space and the optical axis of the experiment telescope is approximately equal to $\sqrt{(d\psi)^2 + (d\theta)^2}$. The maximum value of this error in Figure 8 is 28.3 seconds arc. The drift rate in pointing error is then $\sqrt{(d\psi^I - d\psi^F)^2 + (d\theta^I - d\theta^F)^2}$/time where I and F are initial and final states. In Figure 8, the drift rate is 7.8 seconds of arc in 50 minutes of time. The

*The application of this concept to the description of coarse pointing control was introduced by W. Seawell, Missile and Space Vehicle Department, General Electric Company, Philadelphia, Pennsylvania.

**This procedure essentially follows that utilized by W. Seawell.

total rms error, considering perturbations about all 3 control axes, is $\sqrt{(d\psi)^2 + (d\theta)^2 + (d\phi)^2}$. The maximum value of this error in Figure 8 is 29.7 seconds. Since spacecraft motions obtained from analog-digital simulation may differ from spacecraft motions obtained from all-analog simulation by as much as 5 seconds of arc, corrections of 5 seconds arc must be made to the angular motions about each axis. With these corrections, the maximum error is still less than 1 minute of arc, and the drift rate is then just under 15 seconds of arc in 50 minutes of time. Hence, for the typical configuration simulated, the pointing and attitude holding requirements were satisfied.

Conclusions

The capability of the OAO spacecraft to meet pointing and attitude holding requirements while in the Coarse Pointing Mode has been substantially verified for the proper choice of star tracker guide stars.

Torques due to inertia coupling, and precession due to fine wheel rotation relative to the spacecraft are negligible during normal attitude hold operation.

The gimbal angle sampling rate, for trackers in the tracking mode, is adequate. The spacecraft error about each control axis due to gimbal angle quantization is approximately 5 seconds of arc or less. The maximum value of this type of error corresponds to the resolution of the gimbal angle transducers.

Application of the analog-digital simulation technique provided a relatively rapid, flexible, and accurate method of determining the performance of the coarse pointing control system.

Acknowledgements

The authors wish to express their gratitude to the following Grumman engineering personnel: Erick Stern, for many helpful suggestions given in directing the OAO Stabilization and Control Subsystem effort; George Mayer, for extensive aid in the area of systems simulation; Hugh Winton, for digital programming and detailed explanation of the data link and digital computer characteristics; Arthur J. Burns, for consultations and direct aid on analog-digital simulation; and Robert Alleva, for his considerable effort in analog programming.

References

1. A. S. Lange, "Automatic Control of Vector Quantities," IRE Transactions on Automatic Control, Vol. AC-4, May 1959.

2. A. J. Burns and R. E. Kopp, "Combined Analog-Digital Simulation," presented at the 1961 Eastern Joint Computer Conference, Washington, D. C., December 12-14, 1961.

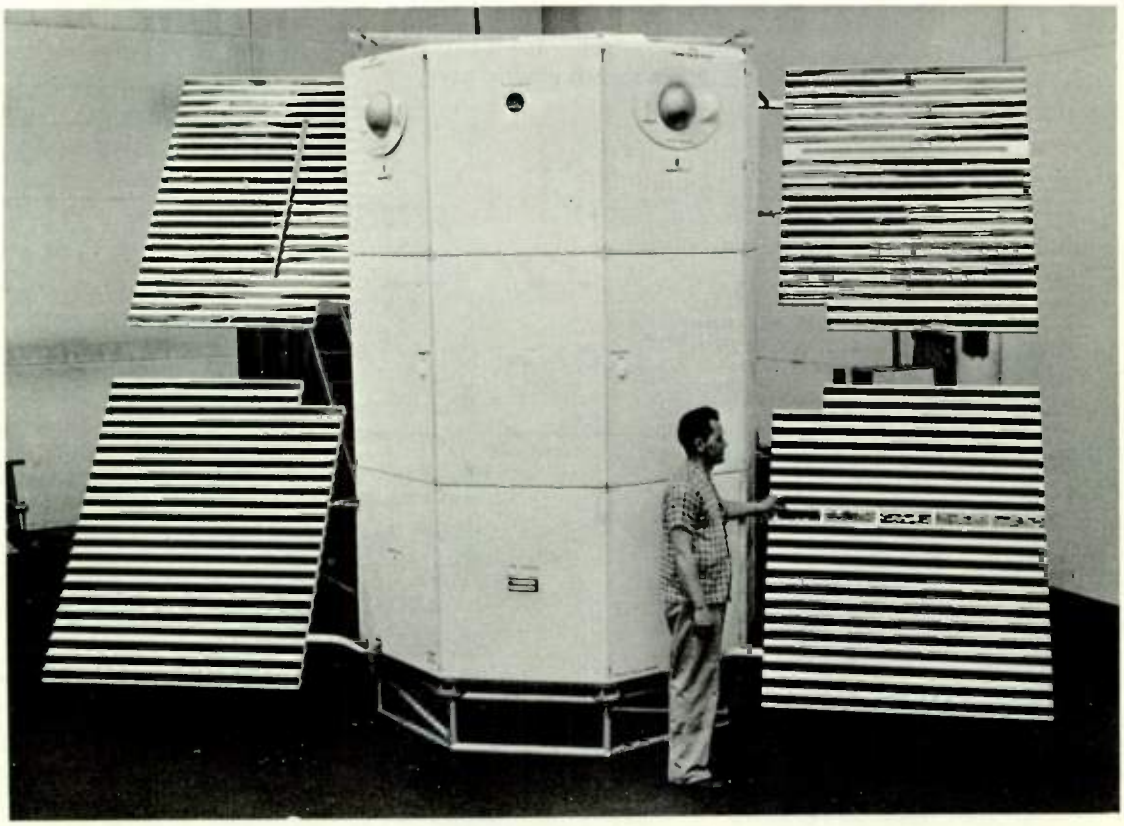


Fig. 1. Full-scale mock up of OAO spacecraft.

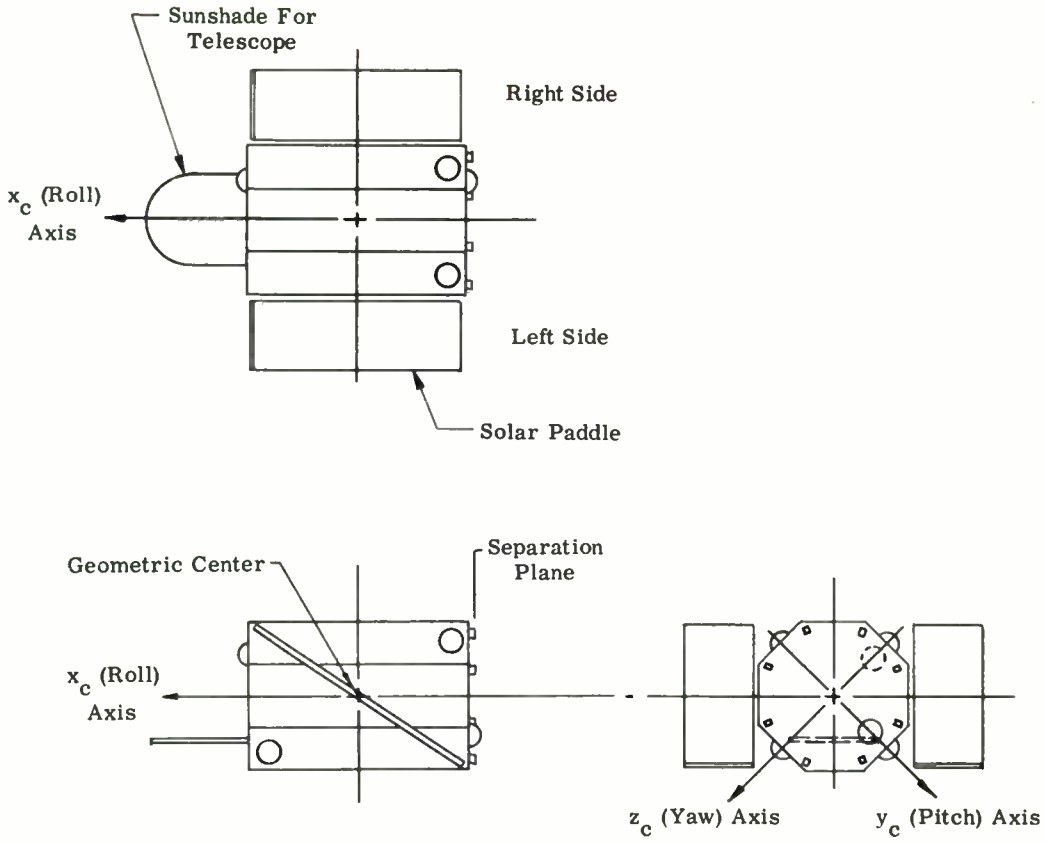
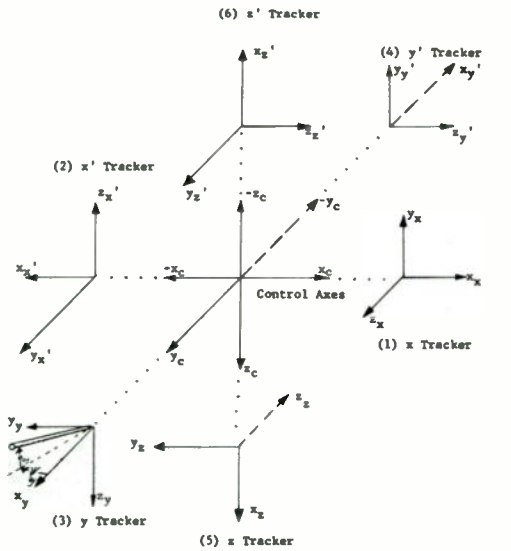


Fig. 2. Spacecraft control axes.



NOTES:
(1) On all Tracker Reference Axes:
All x axes are optical axes at tracker null position.
All y axes are inner gimbal axes at tracker null position.(Plus y points toward single lug).
All z axes are outer gimbal axes regardless of tracker position (Plus z points toward phasolver.)
(2) All solid lines at an angle are directed up from paper.

Fig. 3. Tracker reference axes.

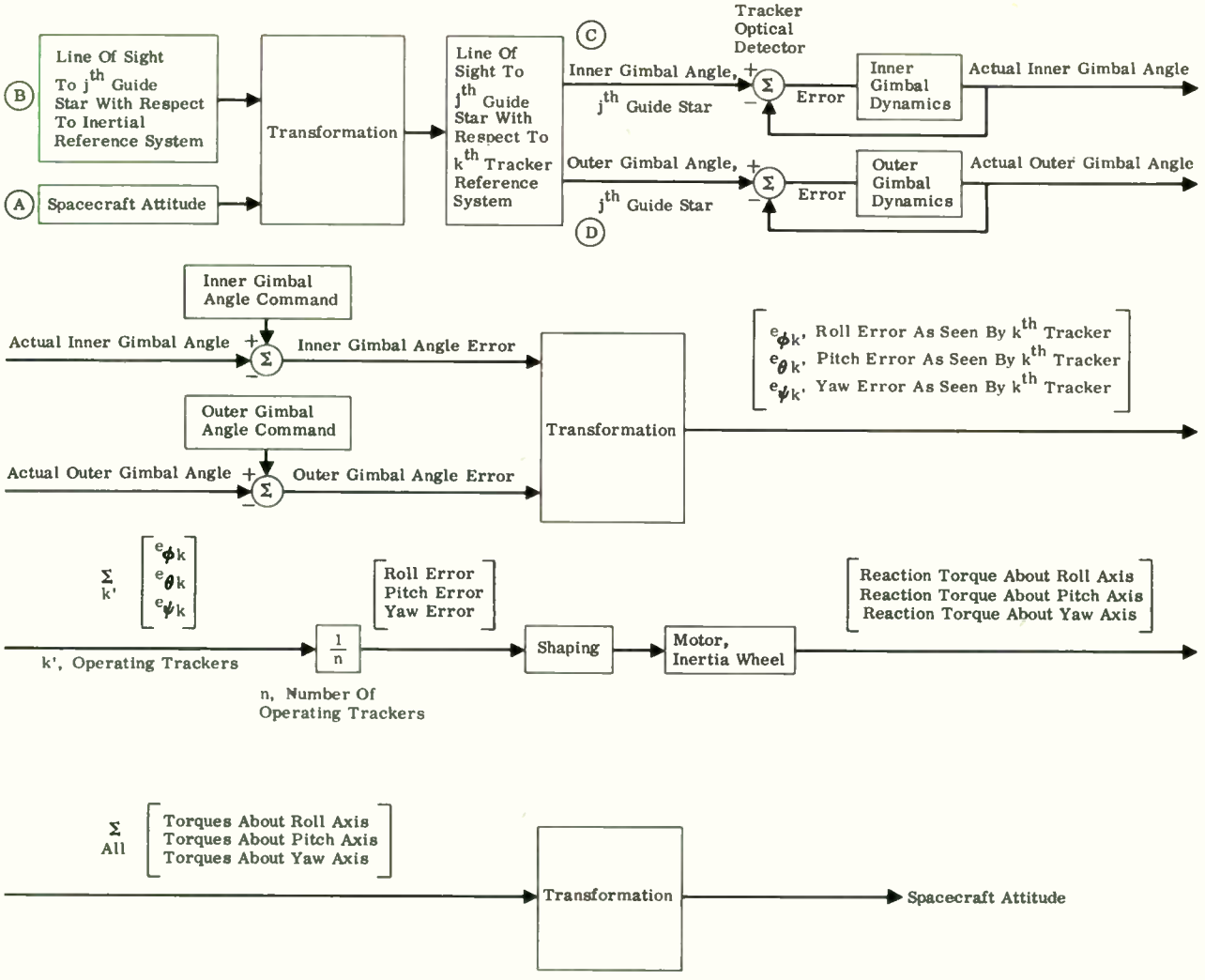


Fig. 4. Functional block diagram, coarse pointing.

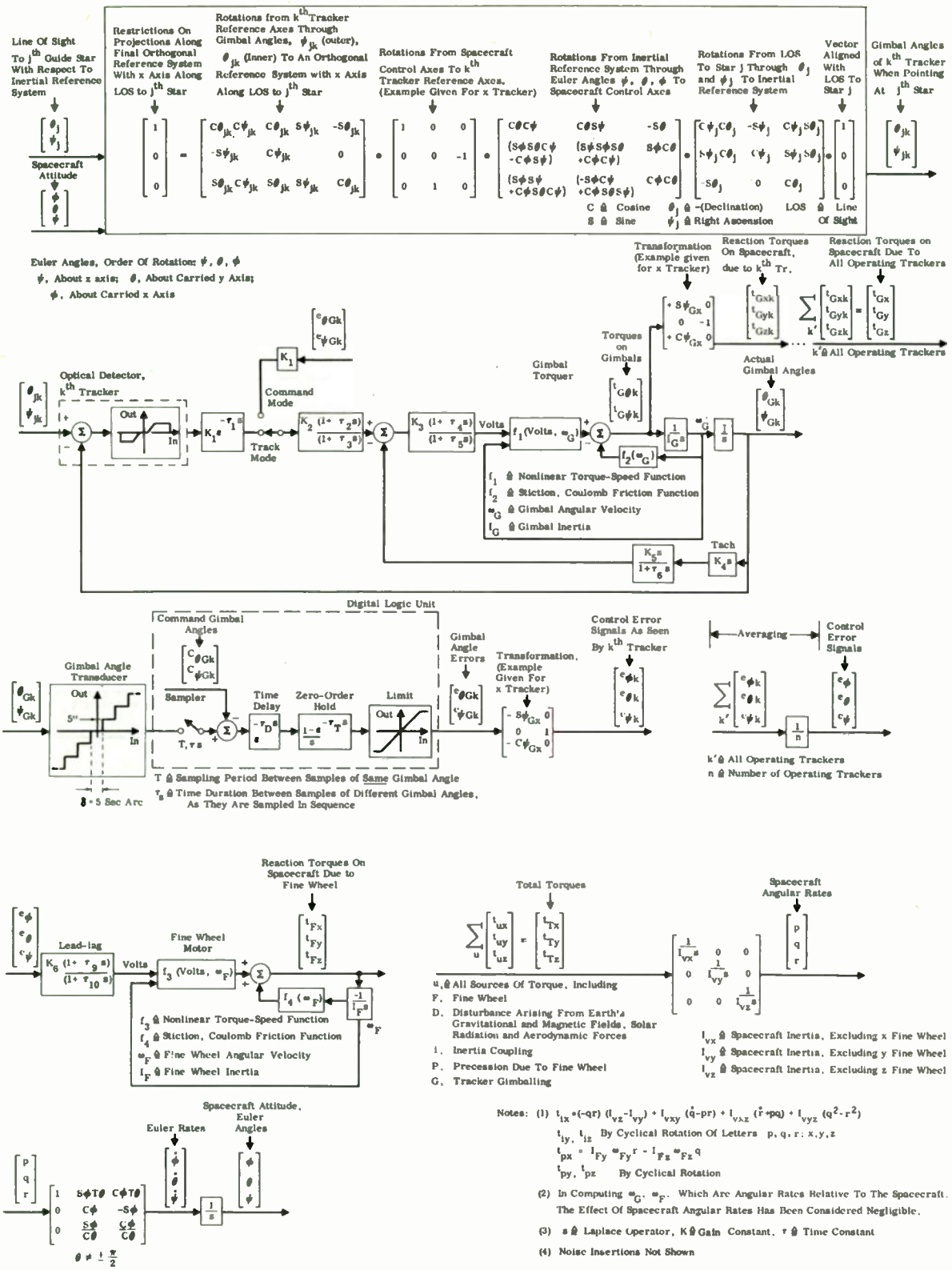


Fig. 5. Mathematical model of Coarse pointing system.

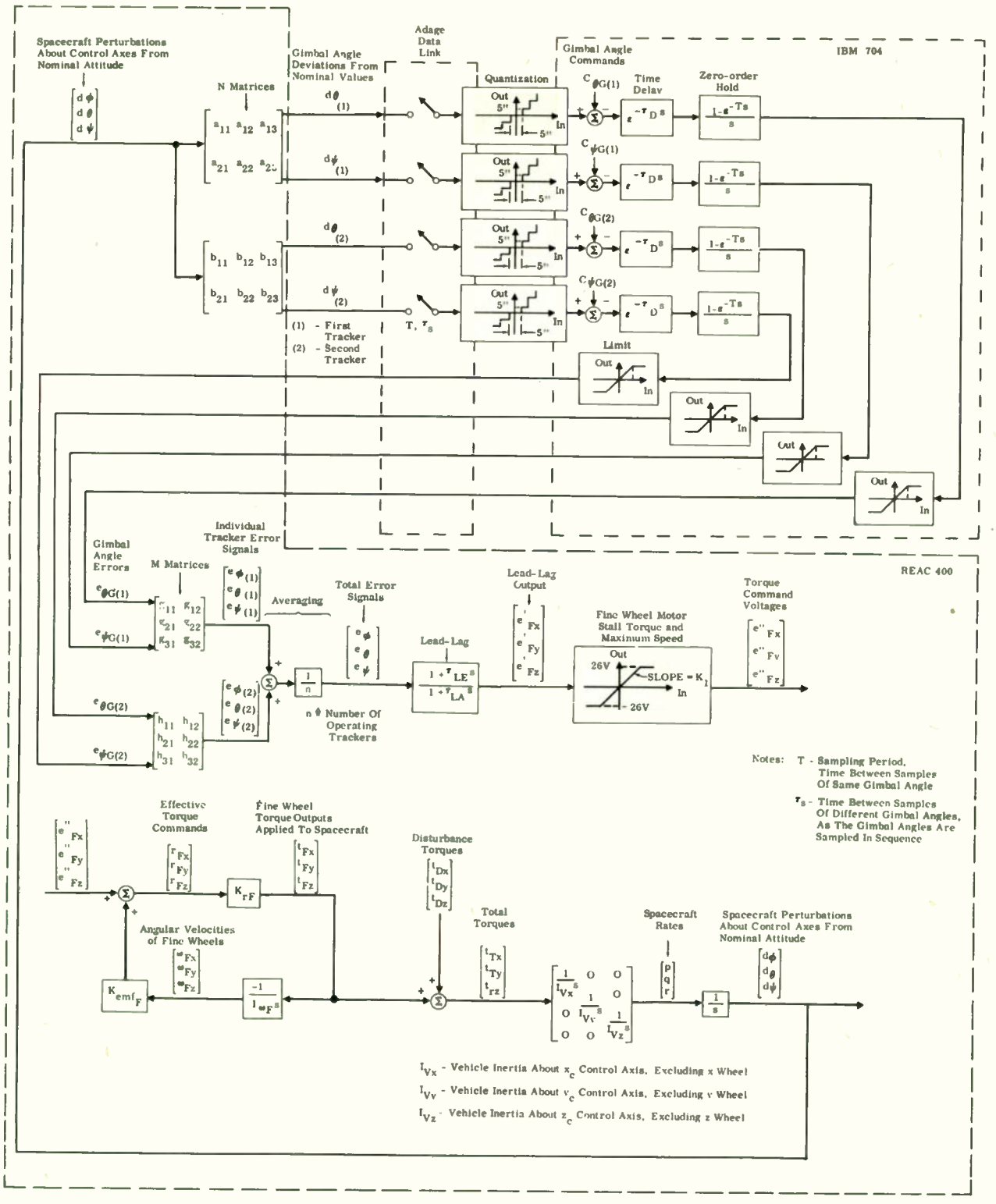
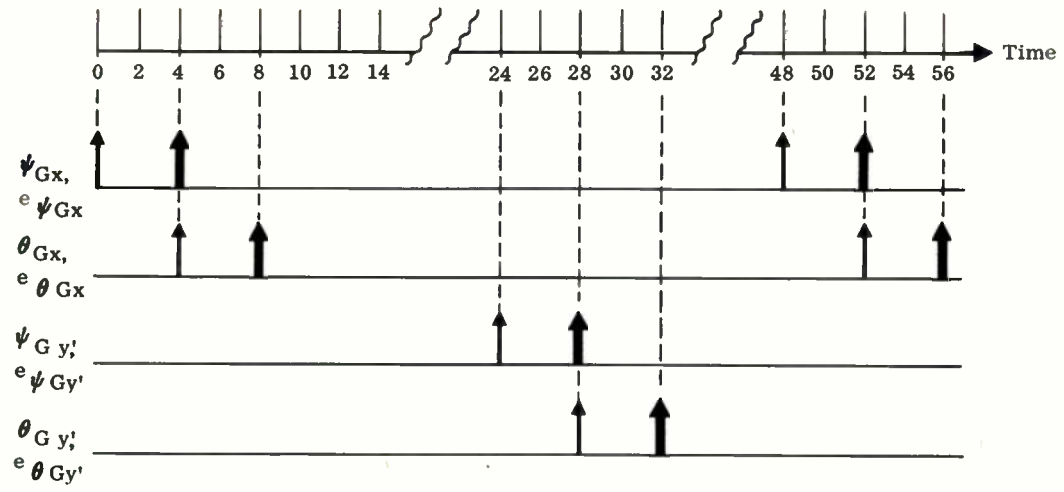


Fig. 6(a). Simulation diagram.



Note: The time between the release of the inner gimbal angle of the first tracker and the sampling of the outer gimbal angle of the second tracker depends upon which trackers are used. The timing diagram above is an example of the use of the x and y' pair of trackers.

↑ - Gimbal Angle Sampled

↑ - Gimbal Angle Error Released

Fig. 6(b). Gimbal angle sampling and error release.

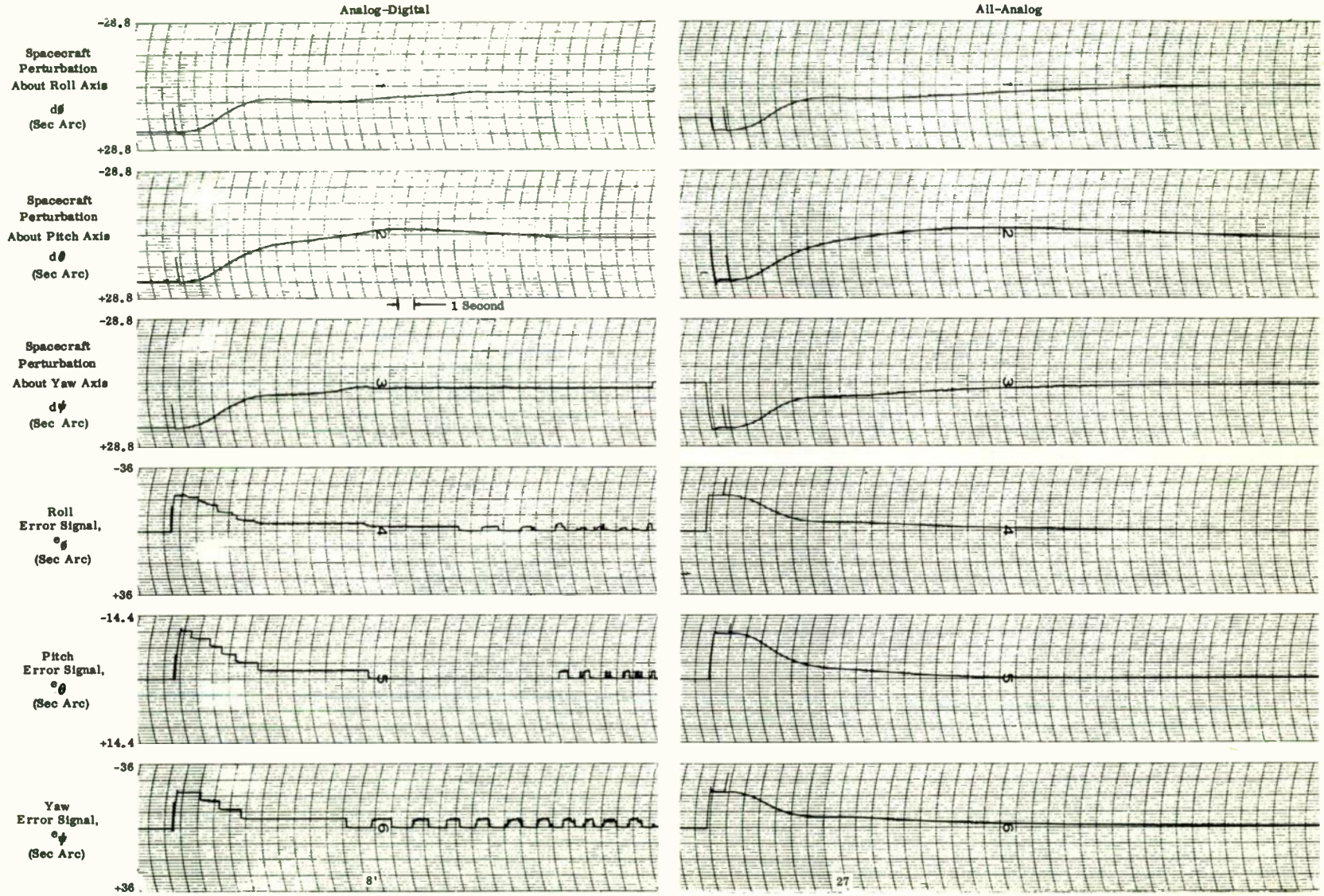


Fig. 7. Comparison between analog-digital and all-analog simulation.

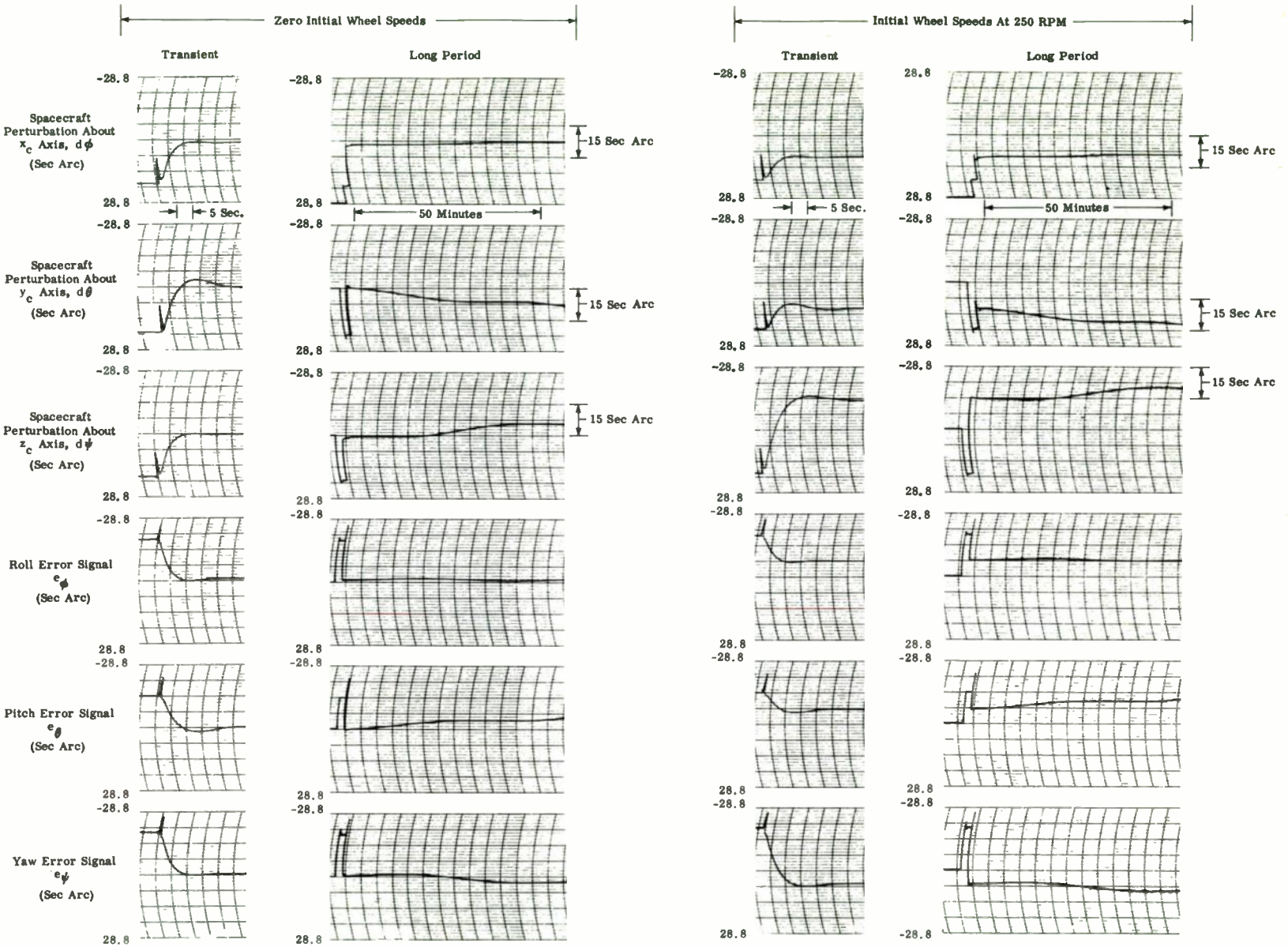


Fig. 8. Typical transient responses and attitude hold control.

Table 1

MATRICES RELATING SPACECRAFT PERTURBATIONS, GIMBAL ANGLE ERRORS, AND ERROR SIGNALS

No.	Tracker	Error Signals	M_i Matrices	Gimbal Angle Errors	N_i Matrices	Spacecraft Perturbations

$$\begin{bmatrix} e_{\phi_x} \\ e_{\theta_x} \\ e_{\psi_x} \end{bmatrix} = \begin{bmatrix} -S\psi_x & 0 \\ 0 & 1 \\ -C\psi_x & 0 \end{bmatrix} \cdot \left\{ \begin{bmatrix} d\theta_x \\ d\psi_x \end{bmatrix} = \begin{bmatrix} S\psi_x & 0 & C\psi_x \\ -T\theta_x C\psi_x & -1 & T\theta_x S\psi_x \end{bmatrix} \cdot \begin{bmatrix} d\phi \\ d\theta \\ d\psi \end{bmatrix} \right\}$$

1 x

$$\begin{bmatrix} e_{\phi_{x'}} \\ e_{\theta_{x'}} \\ e_{\psi_{x'}} \end{bmatrix} = \begin{bmatrix} S\psi_{x'} & 0 \\ C\psi_{x'} & 0 \\ 0 & -1 \end{bmatrix} \cdot \left\{ \begin{bmatrix} d\theta_{x'} \\ d\psi_{x'} \end{bmatrix} = \begin{bmatrix} -S\psi_{x'} & -C\psi_{x'} & 0 \\ T\theta_{x'} C\psi_{x'} & -T\theta_{x'} S\psi_{x'} & 1 \end{bmatrix} \cdot \begin{bmatrix} d\phi \\ d\theta \\ d\psi \end{bmatrix} \right\}$$

2 x'

$$\begin{bmatrix} e_{\phi_y} \\ e_{\theta_y} \\ e_{\psi_y} \end{bmatrix} = \begin{bmatrix} -C\psi_y & 0 \\ -S\psi_y & 0 \\ 0 & 1 \end{bmatrix} \cdot \left\{ \begin{bmatrix} d\theta_y \\ d\psi_y \end{bmatrix} = \begin{bmatrix} C\psi_y & S\psi_y & 0 \\ T\theta_y S\psi_y & -T\theta_y C\psi_y & -1 \end{bmatrix} \cdot \begin{bmatrix} d\phi \\ d\theta \\ d\psi \end{bmatrix} \right\}$$

3 y

$$\begin{bmatrix} e_{\phi_{y'}} \\ e_{\theta_{y'}} \\ e_{\psi_{y'}} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ S\psi_{y'} & 0 \\ -C\psi_{y'} & 0 \end{bmatrix} \cdot \left\{ \begin{bmatrix} d\theta_{y'} \\ d\psi_{y'} \end{bmatrix} = \begin{bmatrix} 0 & -S\psi_{y'} & C\psi_{y'} \\ -1 & T\theta_{y'} C\psi_{y'} & T\theta_{y'} S\psi_{y'} \end{bmatrix} \cdot \begin{bmatrix} d\phi \\ d\theta \\ d\psi \end{bmatrix} \right\}$$

4 y'

$$\begin{bmatrix} e_{\phi_z} \\ e_{\theta_z} \\ e_{\psi_z} \end{bmatrix} = \begin{bmatrix} -C\psi_z & 0 \\ 0 & -1 \\ -S\psi_z & 0 \end{bmatrix} \cdot \left\{ \begin{bmatrix} d\theta_z \\ d\psi_z \end{bmatrix} = \begin{bmatrix} C\psi_z & 0 & S\psi_z \\ T\theta_z S\psi_z & 1 & -T\theta_z C\psi_z \end{bmatrix} \cdot \begin{bmatrix} d\phi \\ d\theta \\ d\psi \end{bmatrix} \right\}$$

5 z

$$\begin{bmatrix} e_{\phi_{z'}} \\ e_{\theta_{z'}} \\ e_{\psi_{z'}} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ C\psi_{z'} & 0 \\ S\psi_{z'} & 0 \end{bmatrix} \cdot \left\{ \begin{bmatrix} d\theta_{z'} \\ d\psi_{z'} \end{bmatrix} = \begin{bmatrix} 0 & -C\psi_{z'} & -S\psi_{z'} \\ -1 & -T\theta_{z'} S\psi_{z'} & T\theta_{z'} C\psi_{z'} \end{bmatrix} \cdot \begin{bmatrix} d\phi \\ d\theta \\ d\psi \end{bmatrix} \right\}$$

6 z'

NOTE: The subscript G has been intentionally omitted in the M and N matrix elements above to simplify the expressions. Normally the Symbols, Θ_{Gk} and Ψ_{Gk}, are used to represent the outer and inner gimbal angles of the k^{th} tracker.

Table 2. Values of Parameters and Initial Conditions

Parameters in Analog Portion of Simulation	Value, Figure 7	Value, Figure 8
Time Scale	$t = \overline{T}$ (Real) / Machine	$t = 50\,\overline{T}$
n	2	3
τ_{LE} (sec)	10	5
τ_{LA} (sec)	1	.5
K_1 (volts/sec arc)	26/22.5 →	
K_{rF} (dy-cm/volt)	138,312/26 →	
I_{WF} (Sl-ft^2)	.0197 →	
K_{emf_F} (volts/RPM)	26/1000 →	
t_{Dx} (dy-cm)	950 sin (.00218 t) →	
t_{Dy}	[1436 sin (.00218 t) -1042 cos (.00218 t) +1525] →	
t_{Dz}	[1979 sin (.00218 t) +532 cos (.00218 t) -1200] →	
I_{vx} (Sl-ft^2)	1000 →	
I_{vy}	995	950
I_{vz}	990	900
a_{11}	.70711	.17365
a_{12}	-.70711	0
a_{13}	0	.98481
a_{21}	0	.33910
a_{22}	0	-1
a_{23}	-1	-.059793
b_{11}	0	0
b_{12}	0	.34202
b_{13}	1	.93969
b_{21}	-1	-1
b_{22}	0	-.56462
b_{23}	0	.20551

Parameters in Analog Portion of Simulation	Value, Figure 7	Value, Figure 8
C_{11}		.93969
C_{12}		0
C_{13}		.34202
C_{21}		.05417
C_{22}		1
C_{23}		-.14883
g_{11}	-.70711	-.17365
g_{12}	0	0
g_{21}	.70711	0
g_{22}	0	1
g_{31}	0	-.98481
g_{32}	1	0
h_{11}	0	0
h_{12}	1	1
h_{21}	0	-.34202
h_{22}	0	0
h_{31}	-1	-.93969
h_{32}	0	0
i_{11}		-.93969
i_{12}		0
i_{21}		0
i_{22}		-1
i_{31}		-.34202
i_{32}		0
Nominal ψ_{Gx} (deg)		10
θ_{Gx}		-19
ψ_{Gy}	-45	
θ_{Gy}	0	
$\psi_{Gy'}$	0	-20
$\theta_{Gy'}$	0	-31
ψ_{Gz}		20
θ_{Gz}		9

Parameters in Digital Portion of Simulation	Value, Figure 7	Value, Figure 8
T (millisec)	48	NA
τ_D "	4	
τ_s "	4	
δ (sec arc)	5	
$C_{\psi Gy}$	0	
$C_{\theta Gy}$	0	
$C_{\psi Gy'}$	0	
$C_{\theta Gy'}$	0	

Initial Conditions	Value, Figure 7	Value, Figure 8
p (sec arc/sec)	0 →	
q	0 →	
r	0 →	
$d\phi$ (sec arc)	20 →	
$d\theta$	20 →	
$d\psi$	20 →	
ω_{Fx} (RPM)	0	0,250
ω_{Fy}	0	0,250
ω_{Fz}	0	0,-250

Notes:

(1) The "first" tracker is the simulated tracker first appearing in the tracker set: x, x′, y, y′, z, z′. The N type matrix with "a" elements is that N matrix corresponding to the "first" tracker. The N matrix with "b" elements is that N matrix corresponding to the "second" tracker, etc.

(2) The M type matrix with "g" elements is that M matrix corresponding to the "first" tracker, etc.

NEW PROBLEMS IN SPACE TELEMETRY

Conrad Hoeppner
Electronetics Corporation
Melbourne, Fla.

Abstract

The problems of telemetering from a number of satellites and space probes all of which have different missions are discussed together with the ground nets necessary to receive the data and the instrumentation in the space vehicle.

A LASER DESIGN FOR SPACE COMMUNICATIONS

Lawrence Goldmuntz
TRG, Incorporated
Syosset, New York

A LASER communication system for use either from vehicle-to-vehicle or from vehicle-to-earth has many advantages and disadvantages. If the full spatial coherence of the LASER is to be utilized the optical beam must be space stabilized to seconds of arc. This is expensive but not impossible. To the extent that the LASER communication system is to be utilized from the vehicle-to-earth the earth station must be spaced in such a climate or at such an altitude that the probability of clear weather is large. The attenuation of an optical signal in a heavy fog is 300db/kilometer. Of course there is no all-weather problem from vehicle-to-vehicle in space.

The advantages of a LASER system derive from the narrow beamwidth that is obtainable with a small aperture radiator which is compatible with a space vehicle, and also from the extraordinary modula - tion bandwidths that an optical communication make possible. Coherent optical communication is in its infancy, and much analysis as well as many experimental techniques have to be developed before the full implications of optical systems for space communications will be understood.

In this paper we are going to sketch some rough performance characteristics in various missions for simple LASER systems utilizing both incoherent and coherent detection. Coherent detection or optical heterodyning implies coherence in both time and space on the detecting element. This is not simple to achieve but we will describe some techniques under investigation at TRG towards this objective. The purpose of an optical coherent detector is to permit discrimination against background current whether it arises from the detecting element or from the sun or daylight sky. We will show that with a perfect optical heterodyne, it is possible to achieve a radiation limited system, almost independent of the practical background that one might encounter. By radiation limited we mean that the dominant noise is the shot or fluctuation noise in the signal itself.

Let us first compare a radio telemetry system with an optical telemetry system. Since the minimum detectable signal varies approximately inversely with wavelength and since the gain of a transmitter also varies inversely but as the square of the wavelength, the power required for a given channel capacity in an ideal system varies directly as the wavelength. It is then desirable in a fair-weather system to operate at the shortest wavelength at which you can

generate power efficiently. Since the ratio of radio wavelengths, to optical wavelengths is of the order of 10^4, this would tend to indicate that in any fair-weather system it would always be desirable to utilize an optical communication link.

However, since it is possible to build radio frequency reflectors of much larger aperture than an optical collector, maintaining the tolerances and seeing ability that are required, one can diminish the advantage of the optical system in terms of power transmitted by constructing larger radio apertures. On the ground one might consider using one of the larger radio frequency antennas with an aperture of 10^4 meter^2. At optical frequencies because of seeing conditions through the atmosphere as well as difficulties of fabrication one would probably consider an aperture no larger than 1 square meter. Thus the factor of 10^4 in the radio of wavelenths is overcome by brute force. Such a technique is not possible when both receiver and transmitter are on a vehicle. In this situation it is probably unreasonable to consider radio antennas with apertures larger than 1 meter^2 or optical apertures greater than 0.1 of a meter^2. In this case the optical system would have an advantage of 1,000 in the power requirements for a communication link.

The power advantage of an optical telemetry system as compared to radio frequency systems assumes the availability of ideal detectors and the absence of background noise due either to the detectors or solar scattering.

Such ideal detectors are unfortunately unavailable. Figure 1 summarizes the properties of currently available detectors.

In Figure 1 we list appropriate detectors for operation at various well-known LASER frequencies, the S-20 surface for ruby which oscillates at .7 microns, a silicon photocell appropriate for the neodymium glass LASER or helium-neon LASER oscillating at approximately 1 micron and the gold-doped germanium photoconductor for such LASERs as the helium pumped cesium cell operating at 7 microns. The quantum efficiency of the photo-emitters is at best 3% whereas for the photoconductive detectors it becomes of the order of unity. However, the "dark" current for these detectors is many orders of magnitude larger than the dark current for the photoemissive surface. This dark current is expressed as an equivalent photon flux per square centimeter of detector surface in Figure 1.

There is an equivalent dark current due to the optical backgrounds in which the communication system may have to operate, see Figure 2. This is due either to having the sun directly on the

axis of the communication system or perhaps the earth or the moon bathed in reflected sunlight is in the field of view. The scattering from the daylight sky is another source of background for an earth based system.

One can calculate the flux on a detector at a given wavelength assuming diffraction limited optics and assuming that the detector area is 10 times the diffraction limited spot to permit scanning for acquisition of the signal. This calculation can be accomplished in terms of the known brightness of the background and the optical filter width. We have assumed a 10Å filter because the temperature stability of narrower filters is questionable. It can be seen that the background current in a communication system due to the daylight sky is of the order of magnitude of the background current due to dark current from the photo-emissive surface and the background current due to the sun is of the order of the background current that would obtain in a silicon cell.

As a point of departure we postulate three characteristic missions for a space vehicle system, one operating at orbital ranges of the order of 500 kilometers, one at Moon to Earth ranges approximately 4×10^5 kilometers and one operating at Mars to Earth ranges approximately 5×10^8 kilometers. These missions are outlined in Figure 3.

We assume 0.1 meter^2 transmitting and receiving mirrors and a system operating at 5000Å. The signal flux in photons per second per watt transmitted, for each mission are calculated in the right hand column.

Now let us make some simple minded assumptions about the reliability of the telemetry system. Let us assume that in order to have a high probability of detection of every transmitted bit and a low false alarm rate, 100 signal photons must be received for each background or dark current photon in the system.

In the Figure 4, we list the channel capacity against the sun background where the receiver utilizes the S-20 surface and an 0.1 meter^2 transmitting aperture and a 1 milliwatt transmitter. It can be seen from the first row that for a 10Å filter discriminating against the sun background, reliable communication is possible only for Earth-Orbital missions. For these system parameters it is not possible to communicate with the reliability that 100 signal photons per background photon indicates, on an Earth-Moon mission or an Earth-Mars mission.

This is where the utility of an optical heterodyne becomes apparent. Up to this point we have not described an optical heterodyne but in theory if not in practice its operation is similar to an

r.f. heterodyne. We will describe possible configurations later. At this point it is sufficient to say that while an optical filter with an equivalent bandwidth at 7 microns of approximately 10^{12} cycles is about the limit that one can obtain with standard temperature stable optics, we feel that it is possible to obtain a 10^6 cycle system bandwidth with an optical heterodyne. One heterodynes the optical signal to a lower frequency, so as to obtain the narrow system bandwidth at the lower frequency with electrical filters. At optical frequencies this filter is not practically realizable.

If the system is band limited by use of the optical heterodyne to a one megacycle width, then this optical telemetry system can achieve a 10^6 bit/sec channel capacity against the sun background on Earth-Moon missions. When one uses the 1 megacycle optical heterodyne in the situation that we have outlined the channel capacity at Earth-Orbital ranges is naturally reduced from 10^{12} bits/sec to 10^6 bits/sec. It is still not possible with the 1 milliwatt transmitter and a 0.1 meter^2 apertures to communicate reliably on an Earth-Mars mission against the sun background.

In Figure 5, we compare the channel capacity for the same system with the sky as background rather than the sun. The same trend that was indicated in the previous figure is demonstrated again here. Namely, by using the optical heterodyne it is possible to filter the background sufficiently so that, while with the 10Å filter in the receiver it would not be possible to communicate reliably on a Mars mission, it would be possible with a 1 megacycle optical heterodyne receiver. Of course the information capacity of the 1 milliwatt transmitter is quite small in this case, of the order of 1 bit/sec.

There are two components to the optical heterodyne, the optical portion that controls the spatial coherence and the electrical proportion that controls the temporal coherence (automatic frequency control) and detects the beat frequency. Two types of electrical systems are possible. The one indicated in Figure 6 does not tune the heterodyne beam to the signal beam but operates on whatever the optical beat frequency might be. The voltage controlled oscillator signal is mixed against the optical beat note. The difference frequency is applied to a frequency discriminator whose output controls the voltage controlled oscillator. This is the simplest type of system if the optical sources are sufficiently stable. However, it may be that the optical beat frequency will wander over such a large range as to make the electronics impractical. Figure 7

indicates the more desirable type of electrical circuitry for use with the optical heterodyne where the voltage controlled optical local oscillator is tuned in frequency by an adjustment of its reflectors so that the radio frequency beat note between the signal beam and the local controlled optical oscillator is maintained constant. Such a device is under construction at TRG and it is hoped to be able to maintain the frequency of the local optical oscillator sufficiently tightly so that the discriminator filter need be no wider than 1 megacycle.

A most important difference between radio frequency or microwave crystal detectors and photodetectors is their size relative to the radiation wavelength. Even the smallest practical photodetector is many times larger than a wavelength of light. Thus the phase difference between the electric fields due to two different spatially incoherent light sources varies over the detector surface. The resulting cancellation of temporally coherent interference reduces the signal-to-noise ratio and heterodyne detection by a factor of 10^{-5} from the theoretical optimum. The theoretical optimum can be derived as suggested in Figure 8. The local oscillator signal is assumed to be much larger than the incoming signal or any of the background induced signals. The photosensitive surface is assumed to be square law. The ratio of the signal current to the fluctuation in the steady current, i.e., noise, is computed. As long as the local oscillator signal is the dominant signal in producing photocurrent, the signal-to-noise per unit cycle is the theoretical limit, due to the noise-in-signal, i.e., the quantum efficiency of the detector, η divided by 2 times the number of signal photons arriving at the detector per second.

This is a most important result, because it indicates, at least in theory, that as long as a local oscillator signal produces more photocurrent than either the dark current in the detector or the photocurrent induced by the background, it is possible to achieve a signal-to-noise ratio determined soley by the noise-in-signal or radiation noise.

The equation in Figure 8 can be generalized to take into account the polarization and spatial phase over the photodetector surface. The result, when this is done, reduces to that of Figure 8 as long as the spatial phase difference between the local oscillator and the incoming signal are constant at each point on the photosensitive surface. If two waves are spatially coherent across any surface in space they will be coherent at all subsequent surfaces in an optical train.

We now consider the optical

component of the heterodyne that controls the spatial coherence. A potential system is shown in Figure 9, that achieves spatial coherence by mode selection. The photodetector, pin hole and lens assembly is aligned to focus the plane wave communication signal through the pin hole. A semi-transparent mirror reflects the plane wave from the local oscillator into the pin hole. The plane wave generated by the local oscillator can be assumed to be made up of a set of spherical waves. The pin holes selects precisely that spherical wave which will be spatially coherent with the incoming signal wave over the surface of the photodetector. This system is stabilized against angular misalignments of various components such as the local oscillator or semi-transparent mirror. Such misalignments are compensated by the pin-hole mode selector.

We can consider briefly the effects of various types of imperfections in the spatial coherence between the local oscillator and the incoming signal waves. These examples should make clear the requirements on a useful optical heterodyne detector.

Figure 10 demonstrates the situation when a local oscillator uniformly illuminates the photodetector but the plane wave is brought to a focus on the photodetector. Analysis indicates that the contribution to the beat frequency signal comes only from the coherence area on the photodetector within the Fraunhofer spot. Local oscillator light falling on the remainder of the photodetector only adds noise. It turns out that the optimum signal-to-noise ratio is decreased by the ratio of the total illuminated detected area to the effective coherence area.

In Figure 11 there is shown the effect of a broad background source such as scattering particles in the immediate vicinity of the receiver. If the pinhole selector is not utilized, the background will add to the noise and cannot be coherently detected and suppressed.

Figure 12 indicates the difficulties if the plane wave local oscillator is not aligned at precisely right angles with the incoming signal and the local oscillator eliminates these alignment problems.

To sum up, we have attempted to show in optical communication or telemetry systems, certain situations in which an optical heterodyne receiver has importance. We have sketched an electrical circuit that is achievable to control the temporal stability of current gaseous c.w. LASERs. An optical train involving the pin hole mode selector has further been suggested to make the problem of optical alignment of local oscillator with the incoming wave as automatic as possible. Analysis that has only been alluded to in this

paper indicates that as long as the local oscillator signal is intense enough to provide a photocurrent larger than the background derived current due to detector noise or background radiation, the communication system can be radiation limited, that is the major source of noise can be due to the fluctuations in the signal itself.

CHARACTERISTICS OF DETECTORS

TYPE	LASER FREQUENCY	QUANTUM EFFICIENCY	DARK CURRENT (EQUIVALENT PHOTONS PER SECOND/cm^2)
PHOTO-EMITTER S-20 SURFACE	RUBY 6943Å	0.03	10^6
PHOTO-VOLTAIC SOLAR CELL	NEODYMIUM $\simeq 1u$ He-Ne	0.50	10^{14}
PHOTO-CONDUCTOR Ge(Au)	HELIUM PUMPED CESIUM $7u$	1.00	10^{18}

Fig. 1.

OPTICAL BACKGROUNDS

SOURCE	BRIGHTNESS (I_b WATTS/cm^2)	FILTERED BRIGHTNESS $\Delta\lambda = 10$Å $\lambda_o = 5000$ Å WATTS/cm^2	DETECTOR FLUX $\zeta^2 = 10$ $\lambda_o = 5000$ Å P_b PHOTONS/sec
SUN	6200	8	2.5×10^{12}
EARTH, MOON	0.14	10^{-4}	3×10^7
DAYLIGHT SKY	0.01	10^{-5}	3×10^6

Fig. 2.

SIGNAL FLUX

MISSION	DISTANCE (KILOMETERS)	SIGNAL PER WATT TRANSMITTED (PHOTONS/sec)
ORBITAL FLIGHT	5×10^2	3×10^{17}
MOON TO EARTH	4×10^5	5×10^{11}
MARS TO EARTH	4×10^8	5×10^5

TRANSMITTING AND RECEIVING MIRRORS = 30 cm
$\lambda = 5000$ Å

Fig. 3.

CHANNEL CAPACITY AGAINST SUN BACKGROUND ONE MILLIWATT TRANSMITTER (S-20 SURFACE) (BITS/SEC)

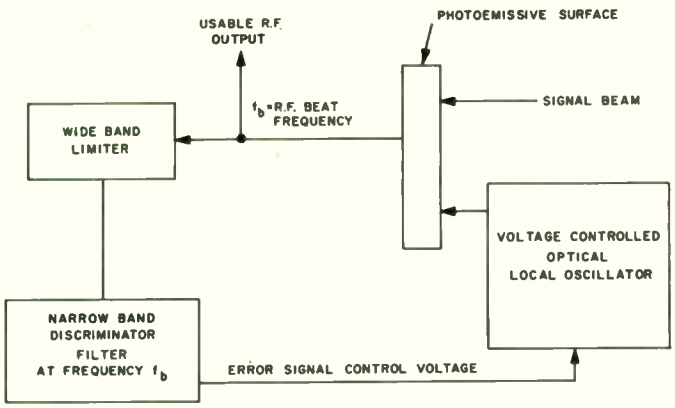
	10 Å FILTER	1 Mc OPTICAL HET.
ORBITAL MISSION	10^{12}	10^6
MOON MISSION	X	10^6
MARS MISSION	X	X

Fig. 4.

CHANNEL CAPACITY AGAINST CLEAR SKY BACKGROUND ONE MILLIWATT TRANSMITTER (S-20 SURFACE) (BITS/SEC)

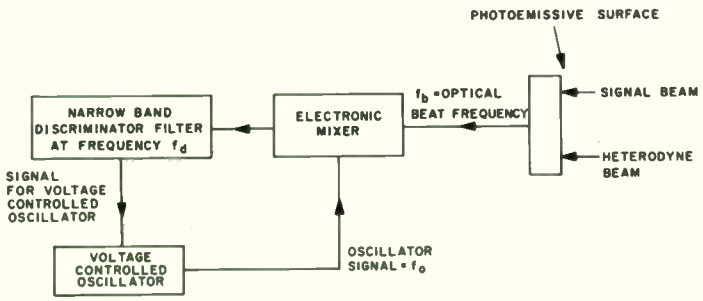
	10 Å FILTER	1 Mc OPTICAL HET.
ORBITAL MISSION	10^{12}	10^6
MOON MISSION	10^6	10^6
MARS MISSION	X	1

Fig. 5.



NARROW BAND OPTICAL RECEIVER TYPE II

Fig. 6.



NARROW BAND OPTICAL RECEIVER TYPE I

Fig. 7.

OPTICAL HETERODYNING

$$E = E_{Lo} \cos w_{Lo} t + E_s \cos w_s t$$

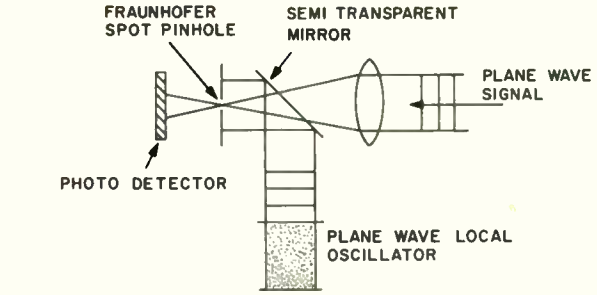
$$\overline{E^2} = \frac{1}{2} E_{Lo}^2 + E_{Lo} E_s \cos(w_{Lo} - w_s) t + \frac{1}{2} E_s^2$$

$$\overline{i_{AC}^2} = 2 \frac{P_s}{P_{Lo}} i_{DC}^2 \qquad i_{DC} = \frac{\eta e P_{Lo}}{h\nu}$$

$$\overline{i_N^2} = 2 e i_{DC} \Delta\nu$$

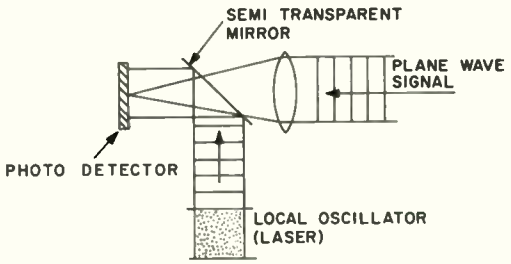
$$S/N_o = \frac{\eta}{2} \frac{P_s}{h\nu} = \frac{\eta}{2} \text{ (NOS. OF SIGNAL PHOTONS/SEC.)}$$

Fig. 8.



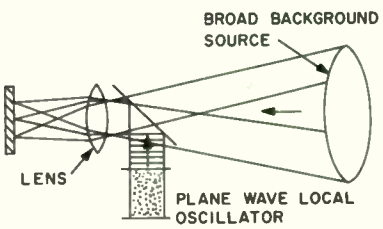
SPACIAL COHERENCE ACHIEVED BY MODE SELECTION

Fig. 9.



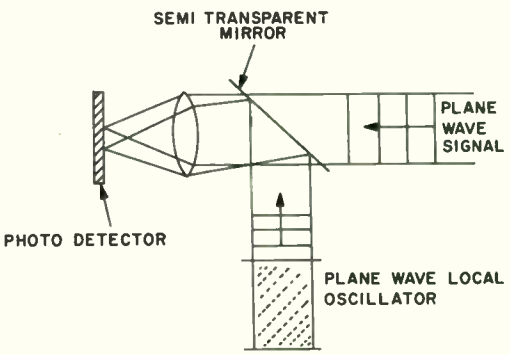
IMPERFECT SPATIAL COHERENCE TYPE III

Fig. 10.



IMPERFECT SPATIAL COHERENCE TYPE II

Fig. 11.



IMPERFECT SPATIAL COHERENCE TYPE I

Fig. 12.

A DIGITAL DATA TELEMETRY SYSTEM FOR AN ATOMIC-POWERED AUTOMATIC WEATHER STATION

Albert M. Sullivan, Jr.
Electronic Systems & Products Division
Martin Company
Baltimore 3, Maryland

Introduction

The worldwide network of weather-observing stations is incomplete. Many serious discontinuities in the network prevail in areas which, because of their weather conditions, are nearly inaccessible. One such area is the Canadian Arctic. At present, the Meteorological Branch of Canada's Department of Transport and the United States Weather Bureau jointly maintain a system of five manned weather stations in the Canadian Arctic. These stations are situated about 200 to 300 miles from one another, giving data which would be comparable to that gathered along the Atlantic Coast if there were stations only at Boston, Philadelphia and Richmond.

Additional--gap-filling--weather stations would obviously be expedient, but manned stations are costly. To establish and maintain living quarters and to supply food, heat and power for additional remote Arctic installations would be extremely difficult. The Arctic area is therefore ideally suited to small, automatic weather stations. In the past, automatic weather stations tried in the Arctic were found to be of limited value. Either the primary power source or the electronic system failed, generally within a few months, and in every case, in less than six months.

Two developments, then, were necessary to make automatic Arctic weather stations a practical reality: (1) a reliable, long-lived source of primary power, and (2) a reliable telemetry system compatible with the power source. As the primary power source, some form of atomic energy would be ideal.

Primary Power Source

Strontium-90

The radioisotope Strontium-90, a reactor fission product whose half life is 28 years, was chosen by the U. S. Atomic Energy Commission* to be developed into a safe, useful source of energy. The strontium ion behaves much like the calcium ion and tends to deposit in the bones if ingested. It is quite soluble in water and therefore highly mobile. A high priority was assigned to finding a compound of strontium which was insoluble in water. Further, the compound had to exhibit a high melting point, stability over its solid temperature range, and ease of compounding and fabricating.

Titanate Compound

Ten strontium compounds were investigated; strontium metatitanate ($SrTiO_3$) was selected. After 100 days in tap water at 150° F, a metatitanate sample showed no detectable dissolution in the water. A second sample, after 102 days in natural sea water at 150° F, had dissolved at the rate of only six parts per million. Strontium metatitanate is easily compounded and fabricated. Its melting point, 1670° C, is well above the maximum expected fuel temperature in the thermoelectric generator, 600° C.

Generator

Fuel Source. A fuel inventory of 17,500 curies of strontium in the form $SrTiO_3$, producing 108 watts of thermal energy, was fabricated into a cylinder about 2 inches in diameter and 5 inches in length. This operation was performed at Oak Ridge National Laboratory. Through a series of tests, the nickel alloy Hastelloy C was determined to be the most suitable encapsulating material for the fuel compound. (Hastelloy C corrodes in sea water at the rate of 0.0001 inch per year.) The fuel was sealed inside a Hastelloy C capsule with walls 1/4 inch in thickness.

Energy Conversion System. The energy conversion system of the generator consists of a hot junction, 60 pairs of thermoelectric elements, thermal insulation, and a cold junction. The hot junction is a second Hastelloy C can, also 1/4 inch thick, machined to an extremely close fit with the fuel capsule. The outer surface of the hot junction is machined to provide 12 flat surfaces to accommodate the hot shoes of the 60 thermoelectric pairs. Each thermoelectric pair is made up of a pure iron hot shoe, N- and P-type thermoelectric elements and copper terminals at the cold ends. Both the N- and the P-type elements are bonded to the iron hot shoe and to the copper cold end terminals. The terminals provide both spherical seats to allow for misalignment, and good conduction paths to the cold junction. The heat sink, or cold junction, is a cylinder of aluminum alloy, 1/2 inch thick, bored to accommodate the 60 radial pairs of thermoelectric elements. The cylinder is Hardcoated* to provide good electrical insulation to the metallic surface without impairing its thermal conductivity. A cylinder of thermal insulating material is used to completely fill the void area, between the hot and cold junctions, not occupied by thermoelectric elements. The biological radiation shield, a cylinder of lead 4 inches thick, surrounds the generator. The shield is encased in a stainless steel jacket to prevent its loss in the event of fire. The

*U. S. Atomic Energy Commission, Division of Isotopes Development, Contract AT(30-1)-2519, with the Martin Company.

*Martin Hardcoat Process.

dose rate from gamma radiation one yard from the generator is only 4 mr/hr. The generator produces a continuous electrical output of 5 watts. Thus, the first requirement for reliable Arctic automatic weather stations, a long-lived primary power source, is met.

Figure 1 is a cutaway view of the generator; Fig. 2 shows the completed generator. It weighs 1675 pounds, of which the lead shielding accounts for 1620 pounds.

Telemetry System

Concurrent with the development of the generator, the telemetry system was designed.

Weather Bureau Specifications

The U.S. Weather Bureau listed the following specifications for the telemetry system:

(1) Two-year unattended operation in the Arctic.
(2) Transmission range of 200 miles, with a high probability of successful reception.
(3) Measurement of:
  (a) Instantaneous temperature from -80° to +120° F in one-degree increments.
  (b) Instantaneous barometric pressure from 28.00 to 31.00 in. Hg in 0.02-inch increments.
  (c) Instantaneous wind direction to the nearest 10 degrees.
  (d) Wind speed from 0 to 150 knots in one-knot increments, averaged over eight-minute and one-minute periods.
(4) Data sampled and transmitted every three hours.
(5) Provisions for adding up to three weather instruments in the future.
(6) Receiving station operable by meteorological technicians, and recorded data reducible without calibration curves or charts.
(7) Station transportable by icebreaker, and easily installed.

General Description

The station designed to meet these specifications is shown in Fig. 3. A steel cylinder 8 feet in height and 26 inches in diameter forms the station housing. It is buried in the ground to its full depth. In the permafrost regions of the Arctic, the temperature several feet below the surface remains relatively constant throughout the year. The thermoelectric generator, which radiates about 100 watts of normally wasted thermal energy, is placed in the bottom of the housing. This thermal energy is used to warm the electronic equipment; and, because of the relatively constant heat sink provided by the permafrost, the temperature is maintained between +40° and +80° F.

The electronic deck containing the station power, programming, data processing, and transmitting subsystems is located immediately above the generator. The electronic components are not damaged by radiation from the generator; tests of representative components in Martin's gamma pool confirmed this. A three-foot styrofoam insulating plug attached to the underside of the lid minimizes heat loss to the atmosphere. The barometer is also located on the electronic deck. It senses outside pressure through the open inverted "J" conduit. This conduit carries the antenna feeds and the cable from the other instruments.

The wind speed and direction instruments are located atop a 37-foot telescoping tower. The thermometer is located on this tower, 15 feet above the ground. It is provided with a self-aspirating solar radiation shield. A junction box ties the three individual instruments to a single multiconductor cable.

System Considerations

One of the primary telemetry system design factors was the choice between the analog and digital systems of data handling. The weather instruments, with the exception of the wind speed instrument (a contacting anemometer), were inherently analog, having outputs in the form of shaft rotations. But analog systems are far from ideal for long-term, unattended operation. They place more stringent demands on circuit performance, and they introduce error at each data handling point--for instance, in the multiplexing. And at the receiving station, analog recordings are somewhat more difficult to read precisely and more subject to human error. On the other hand, digital systems require simple on-off circuits, introduce no error once the data is in digital form, can easily be multiplexed, and can readily be interpreted with less chance of human error.

Accordingly, a digital system was chosen for the automatic weather station. Figure 4 is a block diagram of that system. It comprises the four weather instruments, a power subsystem, a programming subsystem, a data processing subsystem, two transmitting subsystems, and a simple receiving station.

Instrumentation

Because stability and accuracy in electronic analog-to-digital conversion would be difficult to maintain for two service-free years, it was desirable that the instruments present data directly in digital form. The anemometer output was already in digital form--an output of pulses proportional to wind speed. Shaft position encoders would directly convert the shaft rotations of the thermometer, barometer, and wind direction indicator to digital data. By matching the code disc patterns to the instruments, it was possible to linearize the thermometer and barometer, thus answering the requirement for direct data reduction without calibration curves or charts.

The temperature range of 200 counts (-80° to +120° F in one-degree increments) required an eight-bit binary system to encode since a seven-bit system yields only 128 discrete steps. Eight bits were also required for the barometer range of 150 increments of 0.02 inch (28.00 to 31.00 inches Hg) and for the wind speed range of 150 knots in one-knot increments. To avoid ambiguity, the encoders were developed with the reflected binary "Gray" code.

Thermometer Encoder. The thermometer code disc starts at a count of 80, descends to zero, and then ascends to 120. Seven tracks are required; the eighth track is used as a sign bit to denote whether the reading is above or below 0° F.

Thermometer. The thermometer supplied by the Weather Bureau was of the small, bimetallic, helix type. It produced insufficient torque to overcome the friction of a brush encoder. Optical encoders use a frictionless system with glass code disc, light source, and photocell detectors. Since they require very little torque, the thermometer could have driven one; but optical encoders are not suited to the severe environment outside the station housing. Therefore, a special high torque thermometer with an oversized bimetallic helix with sufficient torque to drive a brush encoder was developed. It has proved highly satisfactory.

Barometer. The barometer is also a low-torque instrument. However, since it is located within the mild environment of the station housing, an optical encoder is suitable. A lightweight glass code disc, a redundant lamp arrangement, and eight photodiodes, one for each track, are used in the encoder. Each photodiode feeds a Schmitt trigger; the eight Schmitt trigger outputs are sent to the data processing subsystem.

Wind Direction Indicator. This instrument was fitted with a six-track brush encoder with 10-degree angular spacing.

Anemometer. The anemometer supplied by the Weather Bureau was of the contacting type. However, the contact arrangement was not sufficiently reliable for two-year operation; a magnetic dry-reed switch was therefore installed. This snap-action switch is actuated by an external permanent magnet rotated by the anemometer shaft. Gear ratios were chosen to give one switch closure per knot per minute. By counting closures for one minute with a binary counter, the average wind speed during that minute is stored. As an example of the switch reliability required, an average wind speed of 10 knots would result in about 10^7 switch closures in two years. The switch is rated for billions of operations.

Station Power Subsystem

This subsystem comprises, in addition to the thermoelectric generator, a transistorized dc-to-dc converter, charging limiters, and a 20-cell nickel-cadmium battery. The dc-to-dc converter steps up the 4-volt generator output to 28 volts, maintaining a continuous charge on the battery. Four charging limiters are used, each across five cells. (Charging limiters are needed because the data telemetry package uses only slightly more than half the power available from the generator. This means that, although the generator output is decreasing exponentially, it will remain sufficient to operate this station for many years.) The charging current is shunted through the limiter when the voltage across its five cells reaches 6.9 volts.

Station Programming Subsystem

A master timer, a sequencer, a solid-state timer and several power relays make up this subsystem.

Master Timer. The master timer, which is the basic time reference for the station, is a marine chronometer with a printed circuit face and dual gold plated "cat's whisker" contacts on the hour and minute hands. Every three hours, the hands close the circuit which initiates station operation. The chronometer is rated at ±2 seconds per day and has a 56-hour movement. A winding indicator operates an electric rewind motor at the 24-hour rundown position. The spring is thus kept in the upper 24-hour position, where best accuracy is attained.

Sequencer. When the master timer hand circuit closes, the sequencer is started. The sequencer is a motor-driven cam switch device, with a chronometric regulator to hold the motor speed constant within ±0.3% when the input voltage variation is ±20%. The sequencer provides:

(1) Eight-minute and one-minute time bases for wind speed averaging.
(2) Power to the averaging circuits.
(3) At the end of the averaging period, power to the transmitter filaments.
(4) Two seconds after filament turn-on, high voltage to the transmitter through the power relays.

Solid-State Timer. This unit is turned on with the filaments, and controls the actual broadcast interval, which lasts for 10 seconds.

Data Processing Subsystem

After the transmitters have broadcasted an unmodulated carrier for three seconds, power is applied to the data processing subsystem. This unit generates the station call sign (Morse letters VA); counts and stores anemometer switch closures; reads and stores Gray-encoded data from the thermometer, barometer, and wind direction indicator, as well as from three spare channels; converts the Gray-encoded data to pure binary data; multiplexes these data, along

with the station call letters and the two wind speed averages, to the transmitter modulators. All of these multiplexed data are sent sequentially to the transmitters twice, so that the entire message is repeated. This process requires six seconds.

The five data processing functions are accomplished electronically by transistorized circuits contained on six printed cards.

Wind Speed Counters. Two of the six cards are the two wind speed counters. These are identical circuits; however, since the anemometer output is in closures per knot per minute, the input to the eight-minute counter is first divided by 8. These counters are also shift registers. At the end of the counting period, carry-pulse gates are inhibited. The circuits can then be operated as conventional serial output shift registers.

Gray Data Storage. Data from the remaining, Gray encoded, instruments are all stored in a single, eight-stage, parallel input, serial output shift register through the use of time multiplexing. The station call sign letters are generated in this register by interrogating two resistor matrices. To the shift register, these appear to be two more brush encoders.

The three remaining printed cards contain the clock and reset pulse generators, the multiplexer, and the Gray-to-binary converter.

Data in the wind speed counter-shift registers are already in pure binary form, and, as such, are sent directly to the transmitter modulators. All remaining data, in Gray binary form, pass through the Gray-to-binary converter before reaching the modulators.

Data Format

Each data "word" is allotted eight information bits and a ninth, or space, bit. The space bit is always a zero and serves to separate the data words. Figure 5 shows the data format. For the call sign (Morse letters VA), a single one is defined as a dot; two successive ones, as a dash. Single zeros separate the dots and dashes in a letter, and three zeros separate the letters. For the temperature reading, the first, or most significant bit has a value of 64; the eighth bit is used to denote whether the reading is above or below 0° F. For all other readings, the first bit has a value of 128; the second bit, a value of 64, and so on in normal binary fashion. The first two bits of the wind direction reading are always zeros since there are only six tracks on the encoder.

Data Reduction. The oscillograph chart on which the final weather data are recorded is marked off for the eight information bits and one space bit of each data word. Since the Morse letters VA are always the first two words and are of a known shape, they can be used to establish bit width. The bits in the three spare channel words have also been arranged to give a characteristic shape. They therefore serve as a further aid in finding the data words.

Wind speed answere are obtained directly in knots. The wind direction head is obtained by multiplying the reading by 10 degrees. For temperature, the value is determined as shown previously, the first bit representing 64, etc. This gives the value in degrees Fahrenheit; examining the space bit determines whether the reading is above or below 0° F. Barometric pressure is determined by multiplying the barometer word number by 0.02 inch and adding the product to 28.00 inches Hg.

Meteorological technicians having no previous experience with the binary system can, after an hour's training, reduce all the data in less than two minutes. With practice, reduction time can be decreased to 30 seconds. The data, as received, can be directly entered, without reduction, in a digital computer.

Transmitting Subsystem

Propagation Considerations. An analysis revealed that, in the Arctic, no single frequency in the allotted band (2 to 8 megacycles) would propagate reliably for 200 miles throughout the year. Since the station's primary power was limited, there was a tradeoff between frequency diversity and transmitter power. The best compromise was found to be two transmitters, one at 3.36 mc, and one at 4.97 mc, each with 250-watt outputs. In an area notorious for difficult communication, about 90% of the broadcasts have been received with usable signal strength.

Transmitters. The two transmitters are identical, except for frequency. Each has an oscillator-driver-modulator, a final amplifier, and a high voltage power supply of the dc-to-dc converter type. The oscillators are crystal-controlled; by changing the crystal, each transmitter can be tuned over the 2- to 8-mc band. The only tubes in the entire weather station are in the final amplifiers of the transmitters. Each employs a pair of 4-65A tetrodes in parallel. Should one tube fail, the other can still supply one-half the rated power.

Modulation. The modulation technique is frequency-shift keying (FSK), a technique which allows an extremely simple modulator. The unshifted carrier represents binary zero; shifted 600 cps, it represents a binary one.

Antennas. The transmitting antennas were furnished by the Weather Bureau. These are half-wave dipoles suspended between two 71-foot telescoping towers, one dipole at the top and one ten feet below the top. While it is recognized that this is not the deal configuration, it is quite adequate and was relatively simple to install.

Receiving Station

Equipment for the receiving station comprises two half-wave dipoles, two commercial receivers, a dual-channel, frequency-shift-keyed converter, and a dual-pen oscillograph. Each receiver employs a crystal-controlled local oscillator and a beat-frequency oscillator (BFO). Each BFO is tuned to produce a nominal 2100-cps audio output for the unshifted carrier, representing a binary zero. A binary one results in a 2700-cps tone. The three seconds of unmodulated carrier at the beginning of transmission allow the operator time to start the oscillograph chart drive. The station technician warms up the receivers and FSK converter prior to the data-gathering period. The FSK converter translates the 2100- and 2700-cps tones to a digital pulse train identical to that which modulates the transmitters. These pulse trains are recorded on the oscillograph. If both frequencies are received, four sets of data are recorded since the entire message is sent twice.

Installation

The station was sent by track to Montreal in July 1961. Here, it was placed aboard a Canadian icebreaker, the John A. MacDonald. Canadian and United States weather personnel flew by commercial and RCAF aircraft to the main Arctic weather station at Resolute, on Cornwallis Island, about 600 miles above the Arctic Circle. The receiving station was installed here. The party then traveled north on the icebreaker, through seas frozen to a depth of eight feet, to Sherwood Head on Axel Heiberg Island. This island, almost 200 miles long and 100 miles wide, is completely uninhabited. Sherwood Head, at the southern end of the island, lies 710 miles from the North Pole and about midway between the manned stations at Resolute and at Eureka.

Because of the perpetual daylight at that time of the year, it was possible to work around the clock. The installation was completed in less than 40 hours. Automatic operation began at 1200, Greenwich Mean Time, on August 17, 1961. Data from the station are being used in the world synoptic weather network. On the last day of 1961, the temperature at Sherwood Head was -36° F; the barometric pressure was 30.02 inches Hg; and the wind was from 60 degrees at three knots, steady.

Conclusions

This station has now operated, unattended, in the Arctic longer than any previous automatic weather station anywhere in the world. The chief of the Polar Operations Project of the U.S. Weather Bureau, J. Glenn Dyer, recently remarked that the station has proved very reliable and troublefree, having received no attention since its installation. He noted further that this is the most successful automatic weather station ever developed.

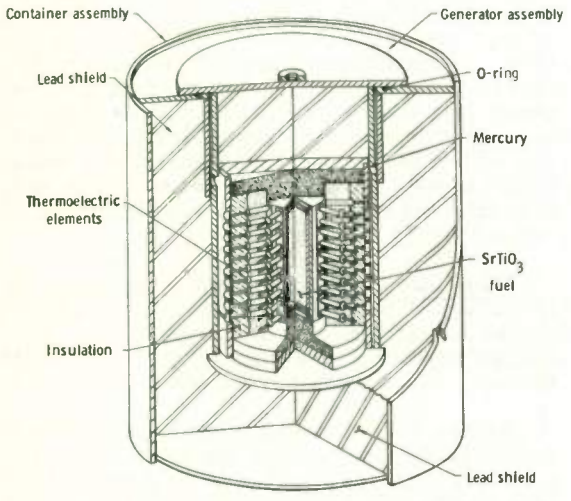


Fig. 1. Cutaway View of Thermoelectric Generator and Shield Assembly

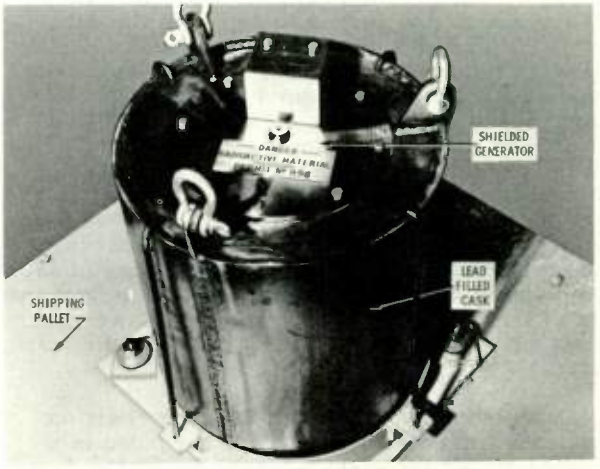


Fig. 2. Lead-shielded thermoelectric generator fueled with strontium-90.

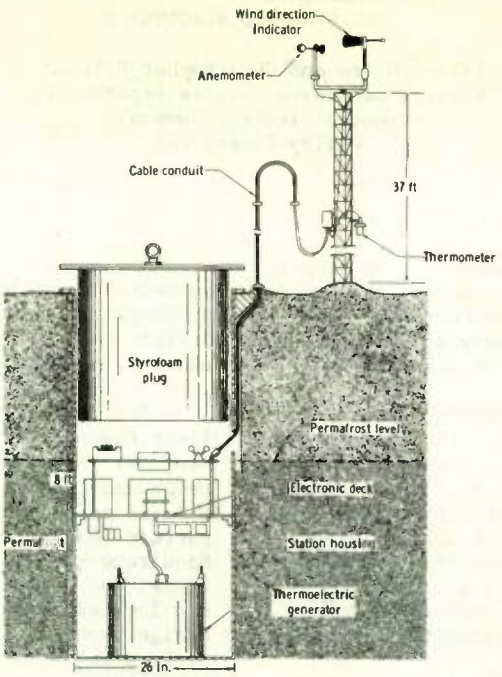


Fig. 3. Weather Station

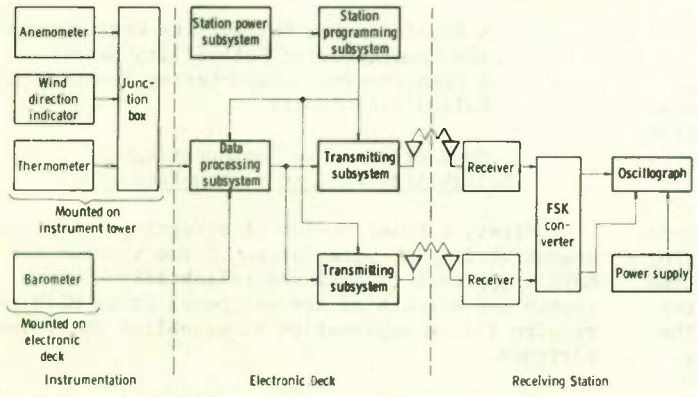


Fig. 4. Block Diagram of Data Telemetry System

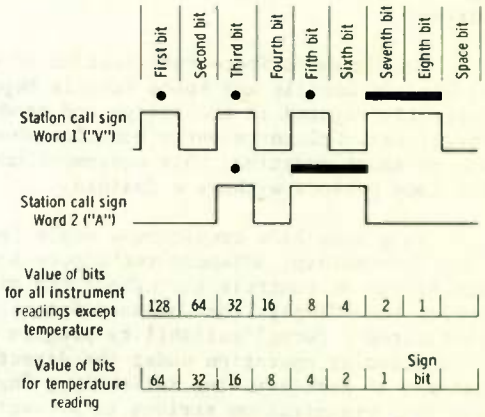


Fig. 5. Data Format

A MANUFACTURING APPROACH TO LONG LIFE REQUIREMENTS FOR UNMANNED SPACECRAFT

Robert Waite and Christopher Balzano
Missile and Space Vehicle Department
General Electric Company
Valley Forge, Pa.

Summary

This paper covers MSVD's experience in preparing processes, personnel, and procedures in advance for the production of unmanned spacecraft. It describes a program for getting "ready to do by doing" in a preprototype manufacturing demonstration.

The philosophy underlying MSVD's approach to achieved reliability is founded on the premise that reliability is built into, not inspected into, a product. Thus, the prime objective of the program is to protect the inherent reliability based on design from degradation during the manufacturing cycle. In addition to the aforementioned preprototype manufacturing demonstration, the means used for accomplishing this objective are the manufacturing reliability matrix and degradation controls.

Each of these elements of the manufacturing reliability program is described and evaluated.

Introduction

In a recent interview published by Aerospace Management, Hilliard W. Paige, General Manager of General Electric's Missile and Space Vehicle Department declared that "the electronics industry has not covered itself with glory in the field of reliability." He warned that there must be a hundred-fold increase in the level of reliability and stated that "our products must have nothing less than zero failure reliability."

This raises the question - what is being done at MSVD toward this end, particularly by manufacturing?

The Unmanned Spacecraft Section of General Electric's Missile and Space Vehicle Department is currently engaged in the design and production of satellites. Unlike re-entry vehicles whose missions are of short duration, this equipment must function for long periods without a failure.

This long life requirement calls for meticulous workmanship, advanced techniques and processes, and stringent controls throughout the manufacturing cycle. Accordingly, the Unmanned Spacecraft Section instituted a formal reliability program within the manufacturing operation under the direction of the Manager of Manufacturing Reliability Engineering. How this organization strives to protect the inherent design reliability throughout the manufacturing cycle is the subject matter of this paper.

Philosophy

Customarily, the literature on missiles and spacecraft distinguishes between inherent and achieved reliability. The former implies good design; the latter depends on manufacturing workmanship, processes, and controls.

Ordinarily, formal reliability activities limit themselves to inherent reliability. Most programs emphasize design, dwelling on feasibility, redundancy, merit analysis, etc. Up to a point the Missile and Space Vehicle Department follows this pattern. Groups such as Reliability and Technical Requirements, Reliability and Analysis of Operations, and Technical Requirements and Applications, are located in engineering and deal primarily with design problems. Beyond this, however, MSVD supports a full-scale manufacturing reliability program based on the premise that reliability is built into - never inspected into - a product.

The Manager of Manufacturing Reliability Engineering summed up MSVD's philosophy as follows: "Our ultimate objective or goal is zero defects rather than a high output at an acceptable quality level."

Scope

Practically, what is involved in making a philosophy or concept of zero defects an operating reality?

This presentation describes the major elements of MSVD's manufacturing reliability program under the following headings -

A Manufacturing Reliability Organization
The Manufacturing Reliability Matrix
A Preprototype Manufacturing Demonstration
Reliability Audits

Some Problems in Implementing a Reliability Program in Manufacturing

First, a brief review of several problems is essential for an appreciation of the reasons for MSVD's approach to achieved reliability. Those chosen for discussion are not peculiar to MSVD and require little explanation to establish their importance.

For convenience, these problems are grouped as follows -

Misunderstanding and Misguided Enthusiasm
Design Delays and Instability
Manufacturing Schedules

Misunderstanding and Misguided Enthusiasm

The notion that reliability is everybody's business often leads to the situation when everybody's business becomes nobody's business. For this reason, MSVD has institutionalized responsibility for manufacturing reliability. The resulting organization, MRE, functions as a reminder to the other manufacturing operations of their involvement in the difficult tasks of meeting long life requirements.

Nevertheless, one must guard against misunderstanding which finds expression in the tendency to regard the manufacturing reliability organization as the sole custodian, warrantor, and fountainhead of reliability. The idea that an identifiable group has specific responsibility for reliability can evolve into the attitude that these specialists have exclusive responsibility, and encourages others to disassociate themselves from the problem. Unless this misconception is checked, the manufacturing function becomes a dualism, with manufacturing operations dedicated to output and the manufacturing reliability engineers handicapped in their efforts to protect long life requirements.

Related to misunderstanding is misguided enthusiasm. An example is the reliability engineer whose yen to "do" produces an ill-considered and impulsive approach to reliability. Ideally, a reliability organization's mission lies in the field of "fire prevention". By responding haphazardly to every crisis, the reliability engineer can transform himself unwittingly into a "fire fighter." In his zeal, he can lose his image as a guardian of reliability and become identified with some particular process or procedure. By "doing" rather than guiding and monitoring, the engineer can work himself right out of the reliability business into the field of process development.

Channeling reliability efforts so that they augment, without usurping, the function of other groups in the manufacturing organization is a major problem. Without proper direction a reliability program dissipates its energies, becomes enmeshed and entangled by irrelevant commitments, and bogs down in unrelated fields of activity. The question --what does a manufacturing reliability engineer do - is readily answered in theory. His role as a planner, coordinator, persuader, monitor, and critic is easy to portray in a job description. But, in the real world of action, things sometimes get out of hand.

Design Delays and Instability

Rigid inherent reliability requirements sometimes cause delays in the release of engineering designs. As a result, manufacturing is compelled at times to undertake its planning without knowing exactly what to plan for, and frequently finds itself in the predicament of the French General Staff, whose Maginot line was built for the wrong war to meet contingencies that never materialized.

Manufacturing Schedules

Meeting schedules is an important manufacturing goal, despite the difficulties, uncertainties, delays, and changes inherent in a research and development operation. Regardless of the novelties involved and the limitations imposed by the "state of the art", the end goal of the manufacturing effort is usable equipment. So, unless top management stresses the importance of long life requirements, the manufacturing reliability organization may receive plaudits and the support of production personnel for advocating perfection only as long as the exhortation doesn't interfere with the business of delivering hardware.

These are the problems - misunderstanding, misguided enthusiasm, design changes, and manufacturing schedules - that Manufacturing Reliability Engineering must overcome. The following sections describe and evaluate MSVD's manufacturing reliability program.

A Manufacturing Reliability Organization

The keystone of the manufacturing reliability program is the Manufacturing Reliability Engineering unit. At MSVD it is under the direction of a manager who reports to the Manager of Manufacturing Engineering (Figure 1). Other organizational arrangements are a possibility.

In addition to the manager, the reliability unit consists of six engineers, two specialists, and one analyst.

Manufacturing Reliability Engineering's responsibilities include -

1. Preparation and implementation of the manufacturing reliability matrix for each project
2. Representation on project reliability board
3. Analysis of engineering drawings to identify and define problem areas
4. Initiation of process developments
5. Investigation of the need for reliability tooling and fixturing
6. Determination of needed levels of controlled environment
7. Establishment of critical storage and handling requirements
8. Consultation on "make" or "buy" decisions
9. Delineation of manufacturing procedures needed to assure reliability and repeatability
10. Review of methods and planning
11. Auditing of manufacturing activities
12. Initiation of human factors studies
13. Coordinate the talents and "know-how" of manufacturing personnel for the solution of unique and complex problems
14. Establishment and operation of a preprototype manufacturing demonstration

15. Personnel training, indoctrination, and motivation
16. Participation in vendor survey and surveillance

Ideally, Manufacturing Reliability Engineering plans, guides, coordinates, advises, directs, suggests, monitors, audits, and persuades. Other groups, such as Advanced Planning, Producibility Engineering, Applications and Processes, and Production Control, have the primary responsibility for conducting the manufacturing sequence so as to enhance reliability.

In addition, it makes plans, develops strategies, and gives training, and then monitors and guides. Except in the areas of training and preprototype demonstrations, MRE accomplishes its goals by advising manufacturing's conventional organizations. At times this ideal is threatened, particularly when delivery dates loom ahead. How to remain entirely free from the entanglements created by expediencies is one of the more difficult problems that remain to be solved in the pioneering field of manufacturing reliability. MSVD accomplishes this end through the use of the manufacturing reliability matrix, the preprototype manufacturing demonstration and the degradation controls.

The Manufacturing Reliability Matrix

The matrix (Figure 2) is one of three major reliability tools used by Manufacturing Reliability Engineering. Essentially, it is a master control chart that records the key manufacturing events and product components in an orderly, integrated manner. As an organized work plan, it enables MRE to carry out and monitor the progress of its tasks. Though a mere tabulation in appearance, it initiates and ties together innumerable programs. It is not a self-enforcing document, but requires continual selling; least of all is it a cure-all. Much work remains to be done before the matrix becomes fully effective.

Meanwhile, the matrix provides a starting point for the creation of viable programs. The manufacturing reliability matrix gives stability and direction to the engineer's steps in setting up his program and forestalls the kind of endeavor where the reliability engineer ends up doing jobs other than his own. By particularizing reliability and making of it a complex of specific namable and scheduled events, the matrix gives shape and substance to the reliability concept. Above all, it applies a harness to the manufacturing effort and it minimizes the tendency for production to separate into distinct endeavours with the majority intent on finished hardware, and a small minority devoted to reliability.

When properly applied, the matrix provides the following benefits:

1. It translates a nebulous concept of reliability into a definite, understandable plan of action.
2. It replaces conflicting pressures, interpretations, and personal inclinations that might otherwise shape the course of a reliability program with documents that can be read, endorsed, and adhered to.
3. The matrix defines the job of a manufacturing reliability engineer through the assignment of specific tasks to be performed on a schedule.
4. It defines, coordinates, and schedules the reliability activities of all of manufacturing.

The several steps involved in preparing the matrix are now described in brief outline:

1. The manufacturing reliability engineer first secures a list of the vital in-line parts, components, subsystems, and systems from design engineering. The designation, vital and in-line, means that these items are interconnected in series so that the faulty operation or failure of any element can abort the mission.

   A typical, but partial, list of vital in-line parts, components, and assemblies would read as follows:

   Power Supply Subsystem

   Solar array
   Regulator
   Power control unit

   Attitude Control Subsystem

   Horizon attitude computer
   Programmer
   Power supply
   Rate gyroscope

   Telemetry

   Antenna
   Transmitter
   Power supply
   Digital telemetry

   The purpose in listing all the vital in-line components is to make sure that potential problems are recognized early and that solutions are prepared in advance of production schedules. Unless this step is carried out conscientiously, Manufacturing Reliability Engineering will be forced to resort sooner or later to improvisations to meet emergencies. Failure to include an item is bound to degrade reliability because of a gap in the planning.

2. Once the vital in-line components are identified and catalogued, Producibility Engineering and MRE agree on a schedule for the "Make" or "Buy" decisions. These commitments are recorded on the matrix,

and later when the decisions are made, the outcomes of the deliberations are, likewise, posted.

The important factors that influence a "Make or Buy" decision are:

Reliability considerations
MSVD facilities and resources
Manufacturing flexibility required
Firmness of design
Shop and test facility load
Company proprietary nature of the items
Security classifications
Cost

3. Next, a search for sources of unreliability is made through the review and analysis of engineering designs. This is a major task designed to spot potential problem areas in manufacturing and to develop ways and means to reduce this hazard through recommended design changes and provisions for protecting the design against degradation during the manufacturing cycle. Considerable time is expended on these activities since they entail meetings, planning, developments, and related undertakings. The resulting key events are incorporated in the matrix. In content and impact they vary widely. Typical examples are -- redesign floating terminals, develop automatic equipment for solar cell sub-module fabrication, provide packaging for honeycomb panels, design fixturing for electronic module assembly, and construct storage facilities for electronic parts. On the basis of the needs uncovered by the design analysis, programs, schedules, procedures, instructions, developments, etc. are prepared under the direction of Manufacturing Reliability Engineering.

4. The scheduled dates and the responsibilities for all the aforementioned action items or key events are established at the conclusion of the analysis, and this information is also entered in the matrix.

5. The first release of the matrix to key personnel in Manufacturing and Quality Control requests concurrence. This puts everybody on notice, telling them what is expected and what is going to happen.

6. Next, MRE draws up a manufacturing reliability plan for each vital in-line part, component, subsystem and system. The basic elements of a plan to cover in-house or "Make" items include:

Process requirements
Facilities requirements
Material handling and flow requirements
Instructions for manufacturing planning reviews

The plan for "Buy" items covers:

Selection and qualification of vendors
Interface requirements and/or special instructions
Shipping and handling requirements
Instructions for material request reviews
Instructions for vendor surveillance

These plans stimulate the formulation of procedures, instructions, etc.

7. As the program proceeds the matrix is periodically amended, enlarged, and reissued. Continual checking assures that each manufacturing reliability key event occurs in accord with the schedule.

The second document, the degradation report, that supplements the matrix and provides it with "teeth" will be covered under the section on Manufacturing Reliability Audits.

A Preprototype Manufacturing Demonstration

General Electric's Missile and Space Vehicle Department did pioneering work on a concept of a preprototype manufacturing demonstration. Since maintenance is presently impossible on unmanned spacecraft while in flight, long life requirements demand more advanced techniques and meticulous manufacturing methods and processes. Thus, in October 1960, a program was instituted to protect the design inherent reliability throughout the manufacturing and assembly process by preparing people, processes, and procedures in advance for the manufacture of prime hardware. Recognizing that failure to control the human element is a major cause of reliability degradation in manufacturing, the program preplanned and rehearsed the entire manufacturing cycle in order to achieve long life reliability.

The philosophy underlying the preprototype manufacturing demonstration is summed up in the expression, "getting ready to do by doing". This approach to long life requirements recognizes the urgent necessity for identifying and eliminating the causes of reliability degradation before the prime hardware production stage is reached. It is based on the premise that a solution of the reliability problem should place more emphasis on devising tools and methods and doing a job correctly, than on doing it repeatedly. Thus, PMD sets up, evaluates, and practices every detail of the manufacturing cycle from the procurement of materials to the shipment of equipment in advance of prime hardware production.

The planning and execution of PMD is under the direction of MRE. Help in this effort is provided by the other components of MSVD's Manufacturing Engineering organization. This includes

Advanced Planning, Producibility Engineering, Applications and Processes, and Manufacturing Programming. Additional participation is forthcoming from Production Control, Shop Operations, Quality Control, Reliability Programs, Purchasing, Sub-Contracts, and Engineering.

Typical PMD activities include - investigation of criteria for personnel selection, operator training and indoctrination, development of processes, tools and fixtures, identification of problem areas, and round table work sessions.

Admittedly, the several elements of PMD as enumerated above resemble activities familiar to every manufacturing enterprise. Personnel selection, training, process development and other such items are not new in themselves, but are, on the contrary, well-established practices in most industries. The novelty of PMD lies in its overall structure and purpose, not in its discrete elements. The PMD program encompasses all the manufacturing functions and runs the gamut of the production cycle in a purposeful, integrated rehearsal of the manufacturing cycle.

The important question is - how successfully can such a complex enterprise be carried out in the circumstances that characterize a research and development program? To what degree can Manufacturing Engineering, Purchasing, Production Control, Production Support, Shop Operations, and other manufacturing functions rehearse in advance and in unison their respective responsibilities? The answer is that PMD's accomplishments were multi-valued, with more creditable contributions in some areas than in others.

As conducted at MSVD the program took as its focal point the production of electronic modules, solar cell sub-modules, and pneumatic assemblies. As a result, PMD operated almost exclusively in and around the lower echelons of the organization. And, as might be expected, it was in these areas that the most was accomplished. As an exercise in making modules, PMD established a record in minimizing defects. Specifically, PMD

1. provided basic experience in the making of electronic modules.
2. developed criteria for the design and construction of controlled environment facilities.
3. initiated standards for the design of modules from a manufacturing standpoint.
4. issued a tentative welding standing instruction.
5. trained a cadre of operators.

Where the initial PMD fell short of expectations was in its impact on the supporting functions that aid shop operations. Some who participated were engaged in two jobs - their own and the PMD assignment - and they carried them out with diligence as if they were completely separate endeavors. As a consequence, only a limited number of procedures and other documents was developed to serve as controls for subsequent making of prime hardware, and so the impact of a basically excellent program was restricted.

MSVD's manufacturing operation learned the following important lessons as a result of its experience to date in conducting a preprototype manufacturing demonstration:

1. The leadership from within Manufacturing Reliability Engineering must be expanded.
2. Management must take a more active part in the program.
3. Greater emphasis should be placed on the development of documents rather than hardware.

Recently, by shifting the emphasis from output of hardware to implementation of the matrix, MSVD broadened PMD's effectiveness and accomplished a greater number of its objectives.

Manufacturing Degradation Reports and Audits

The sequel to the matrix is the manufacturing degradation analysis (Figure 3), the degradation status report (Figure 4) and the full-scale audit. Whereas the matrix establishes the ground rules, these documents measure the performance. In general, the analysis and status reports are used to take care of the lesser mishaps and infractions that degrade reliability, whereas the audit is carried out whenever the threat to reliability becomes critical.

This part of the program bears some resemblance to the discrepancy and defect reports used in other programs. Greater emphasis, however, is placed on anticipating trouble and making potential degradation the subject of the report, in contrast to the practice of recording a defect after it has happened.

Conclusion

The MSVD approach to achieved reliability has approximately a year and a half experience behind it. As might be expected in a pioneering effort, performance was uneven. But deficiencies are being eliminated rapidly with continued application of the program.

The MSVD concept emphasizes prevention, rather than documentation and analysis, of defects. The importance attached to planning and rehearsal, to preparing processes, procedures, and personnel in advance of prime hardware production through an organization located within manufacturing operations is one of its characteristics that distinguishes the MSVD manufacturing reliability effort from Dr. A. V. Feigenbaums' Total Quality Control, PERT, PRISM, SCANS, and those reliability plans that motivate superior workmanship by charging the costs of human errors back to the offending function.

To accomplish its goals, Manufacturing Reliability Engineering must exercise considerable influence throughout the manufacturing organization.

Accordingly, the location of the reliability group within the manufacturing structure is an issue of prime importance. This and other problems are under continual study in order to make MSVD's program fully effective for long-life unmanned spacecraft.

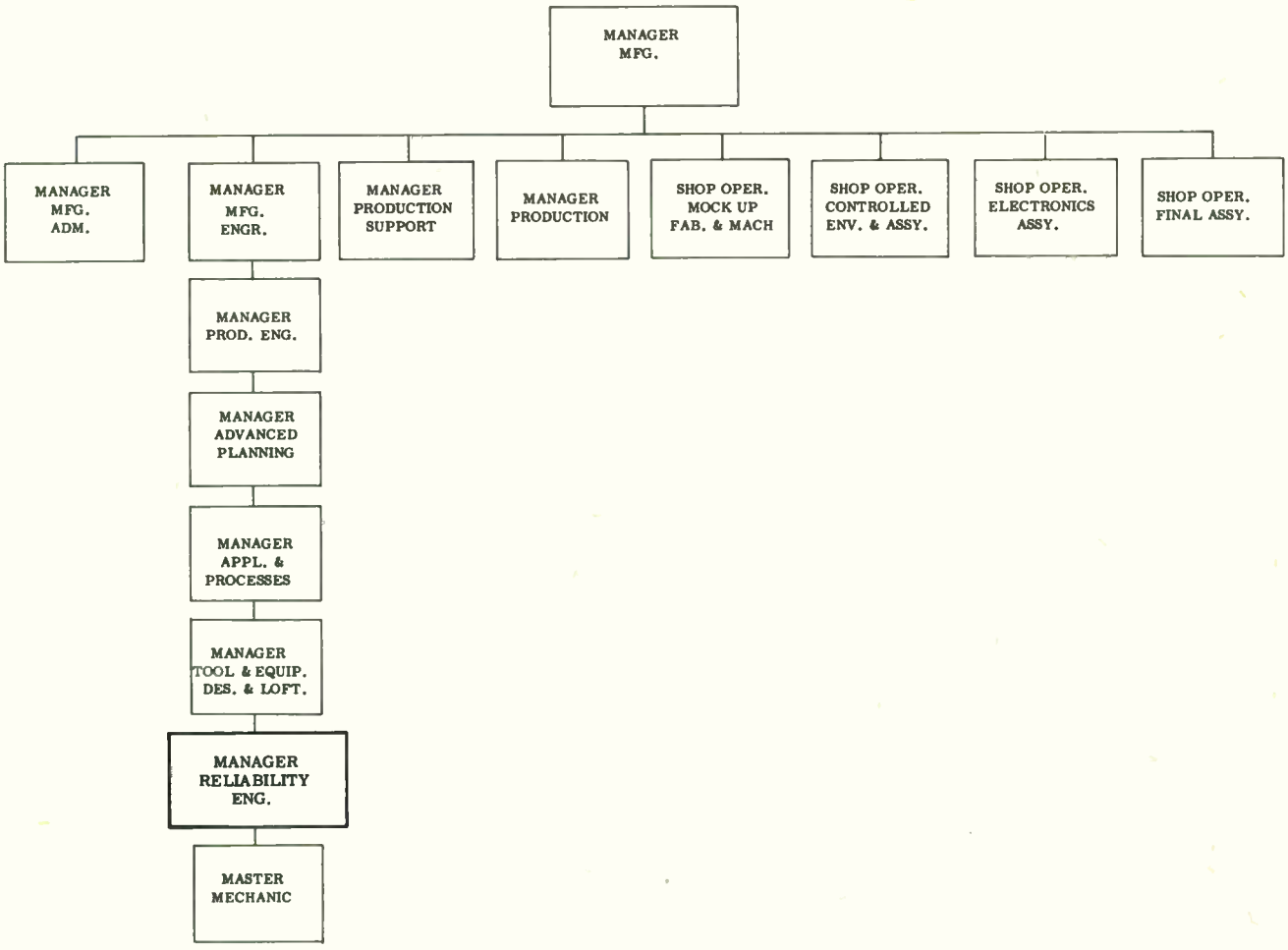


Figure 1
Manufacturing Organization Chart

MANUFACTURING RELIABILITY MATRIX

$^{3}{}_{5}$ = START WK 3
COMP WK 5
M – MAKE
B – BUY

PROGRAM			PER																					SHEET OF
ITEM	VITAL	DRAWING NO.	REL.	DESIGN	Q.C.	Q.A.P.	MFG REL. PLAN	P.C. STATUS		MFG	MFG STAND'G INSTR.	MFG PLAN REVIEW	SUB CONTRACT		VENDOR		IND VEND SURVEY		SHOP OPERATION FACILITY STATUS				REMARKS
	PARTS, COMPONENTS OR ASSY'S		INDEX	ENGR.	ENGR.	NUMBER		M.R. REVIEW	M.R. RELEASE	PLN'S			NAME	BUYER	NAME	BUYER	REQ'D	COMP	MACHINE SHOP	SHT MAT'L & FAB	ASSEMBLY	CONTROLLED ENVIR. ASSEMBLY	

(9–[illegible]–61) [illegible]

Figure 2

MANUFACTURING DEGRADATION ANALYSIS

FORM 1-9707 (7-61)

		SHEET OF
MANUFACTURING RELIABILITY ENGINEERING		DATE
PROGRAM	DESCRIPTION	DWG. NO.
TYPE OF DEGRADATION		
CAUSES: PLANNING		SHOP OPERATION
CAUSES: PROCESS		
CAUSES: TOOLING		
CAUSES: EQUIPMENT		
CAUSES: OPERATOR		
CAUSES: SUPERVISION		
CORRECTIVE ACTION REQUIRED		
CORRECTIVE ACTION COMPLETED		

Figure 3

9-13-61 500V

SHEET OF

REPORT NO	SUBMITTED BY AND DATE	RESPONSIBILITY	CORRECTIVE ACTION SCHEDULE	COMMENTS

Figure 4
Degradation Status Form

C W REDEMPTION

A NEW TECHNIQUE FOR HIGH VELOCITY MISSILE SCORING

Frank C. Lanza

Dalmo Victor Company
Belmont, California

and

George M. Hahn

U.S. Naval Post Graduate School
Monterey, California

Introduction

The measurement of missile miss distance has persisted as one of the most annoying and difficult instrumentation problems in the testing of air-to-air and ground-to-air missiles. The reasons for the difficulty are inherent in the physical characteristics of the missile, in the relative velocity of missile with respect to the drone, and in the desire of the test engineer not to change the characteristics of the missile under test (prohibiting the use of cooperative devices). Furthermore, elaborate airborne electronics and other airborne equipment is ruled out because in many cases the drone is expendable.

The two techniques used most frequently in instrumentation of MDI's have been optical and radar. Camera pods, C W - F M radar, C W doppler and extremely short pulse non-coherent radar are among the approaches instrumented to date. It is not the purpose of this paper to assess the merits and demerits of the various MDI's. Suffice it to say that none have met successfully all the operational requirements.

In the subsequent paragraphs, a novel combination of non-coherent pulse and narrow band C W radar is described which appears to circumvent many of the problems associated with the traditional approaches.

An MDI system incorporating the C W redemption technique is being developed under Air Force Contract AF 08(635)-2116 for Target Development Laboratory, Detachment 4, ASD, Eglin AFB, Florida.

C W Redemption

The transmitted signal due to a pulsed radar operating at frequency f_c is given by the usual Fourier representation:

$$f(t) = \sum_{n=-\infty}^{\infty} \frac{A\tau}{T} \frac{\sin \frac{n\pi\tau}{T}}{\frac{n\pi\tau}{T}} e^{j\left(\frac{f_c}{2\pi} + \frac{n\pi}{T}\right)t}$$

$f(t)$ = pulse train
A = pulse amplitude
T = interpulse period
τ = pulse width

Here it is of course assumed that the pulse shape is rectangular. Different pulse shapes vary the behavior of the envelope of the amplitudes of the spectral lines. However, the position of the spectral lines depends only on the transmitted frequency and the interpulse period, τ, or on the Pulse Repetition Frequency (PRF), which is, of course, the inverse of the interpulse period. Figure 1a is an illustration of the R.F. spectrum for a rectangular pulse.

The received signal is of the same form as the transmitted signal. However, the amplitude of each spectral component is reduced in accordance with the range equation. Furthermore, the phase of each component undergoes a change. The magnitude of the phase change is proportional to range.

If the signal is detected synchronously, each spectral component is translated linearly to center about zero frequency. For a moving target, causing a doppler shift, f_D , with respect to the transmitted frequency, the translated spectrum is centered about f_D . This is illustrated in Figure 1b. Figure 1 c shows the effect of a non-linear detector. The output of the detector is centered about zero frequency; however, each spectral component is shifted by an amount proportional to f_D .

The basic idea of C W redemption is to look at the first of the spectral lines only and make use of the phase information it contains to accomplish ranging. This amounts to making the system coherent by redeeming the C W information at the PRF frequency from the pulsed signal; phase comparison between a reference signal and the redemption signal then provides the range information.

Application to MDI and Scoring Problem

The application of the redemption technique to miss distance measurement is straight forward.

The MDI is a device which measures the distance of closest approach of the missile to the target drone. A Scorer simply determines whether or not a missile came within a predetermined distance of the drone. Both devices require an airborne unit to effect the skin tracking, telemetry equipment, and a ground station to perform the phase comparison. Figure 2 illustrates the non-cooperative system. Since it is clear that the Scorer is a simplified version of an MDI, discussion from now on will be restricted to the MDI.

The airborne unit in Figure 2 generates a train of R.F. pulses at a carefully controlled PRF. The signal generating the PRF is transmitted to the ground station and constitutes the reference signal for ranging.

The reflected pulses are amplified and detected in the air and that portion of the amplified video that contains the PRF information is also transmitted to the ground. After proper filtering, it is then phase compared to the reference signal, and the phase difference, translated to range difference, is displayed on the ground station.

There are several factors unique to the MDI problem which allow C W redemption to be used and, in fact, make it appear as an extremely attractive approach.

Limited Range Requirement

Since the phase measurement is repetitive in multiples of 2π , the wavelength of the "ranging tone," i.e. $\frac{c}{PRF}$, must be kept appreciably greater than the desired detection range. For ranges in excess of a few miles, this would require a PRF so low as to make the technique as described unacceptable. MDI applications typically require a detection range of only a few hundred feet.

Single Traget Detection

Multiple targets require considerable data reduction before phase measurements can be related unambiguously to ranges of the individual targets. The MDI problem concerns detection and ranging at any one time of one target only. Hence, data reduction is straight forward.

Omni-directional Coverage

Most MDI applications require full spherical coverage, leading to difficult antenna problems. In a C W system the isolation between the receiver and transmitter becomes critical and the antenna problems are magnified. Since the C W redemption transmitter is pulsed, gating can be used to accomplish much of the required isolation.

Limited Airborne Equipment

It has been stated earlier that for economic reasons the airborne equipment must be kept to a minimum. The C W redemption system mechanization lends itself very well to meeting such a requirement. The airborne transmitter and receiver is quite simple. The telemetry unit of the drone can be utilized to transmit the reference and target signals, without taking up a large portion of the telemetry bandwidth. Sensitive and relatively expensive circuitry is limited to the ground station.

System Mechanization

An MDI system incorporating the C W redemption ideas is being developed for the Air Force under Contract AF 08(635)-2116.

Parameter Selection

Some of the design considerations of this system will now be discussed since they illustrate the features of C W redemption.

Ranging Frequency (PRF). Three factors influence selection of the PRF. These are:

Maximum range to which unambiguous ranging is to be accomplished, systems resolution as determined by the ability of the ground station to perform the phase comparison, and data rate required by missile-target dynamics.

A range of 100' would allow a maximum ranging frequency of 5 mcs. The ground station is capable of resolving phase shifts of $\pm$ .1°; hence, a resolution of 1.5 feet requires a minimum PRF of about 100 kc. This latter figure is also consistent with the data rate required by the dynamics of the missile.

Pulse Shape. The C W redemption system utilizes only the power in the fundamental of the Fourier expansion of the pulse train. Hence, the ratio of power in the fundamental to total power in one pulse represents a figure of merit of that pulse shape for redemption applications. A few figures of merit for various pulse shapes are listed below. The figure of merit for the rectangular pulse is normalized to unity to eliminate dependence on pulse parameters. All pulses are assumed to have the same total energy. The pulsewidth of the triangular, the cosine pulse, and cosine squared pulse is measured between zero crossings, while that of the Gaussian between 3 db points.

Pulse Shape	Figure of Merit
Rectangular	1
Isosceles triangle	1
Cosine	$\pi^2/4$
Cosine Squared	4
Gaussian	2π

As can be seen, the Gaussian pulse is optimum. The pulse shape of a narrow pulse (15 nanoseconds) is inherently somewhat bell shaped. Therefore, no additional pulse shaping is attempted.

Pulse Width. In order to minimize the effect of ground reflection and vehicle surface reflection, the pulse width should be as small as possible. In the present system, 15 nanoseconds is used.

Ground Station Bandwidth. The filter bandwidth of the bandpass filter about the 100 kc fundamental (See Figure 2) is determined by practical filter considerations. The doppler variations about 100 kc due to missile-target motion are only of the order of a few cycles. A commercially available narrowband crystal in a temperature controlled environment with a bandwidth of 1000 cps is used. This 1000 cps bandwidth is the noise bandwidth of the overall system as far as accuracy considerations are concerned.

Airborne Receiver Bandwidth. The primary function of the airborne radar is to achieve detection. Hence its bandwidth, prior to the threshold detector must be consistent with the matched filter requirement for a 15 nanosecond pulse, i.e., approximately 50 mc. However, since the information transmitted to ground requires but a sideband of approximately 100 kc about the telemetry carrier, the system can be narrowbanded behind the second detector. (Figure 3).

Transmitter Power Requirements. Since the detection criterion of the airborne receiver is the same as that for any pulse radar, the minimum power required for the redemption system would appear to be the same as that for an equivalent pulsed system. However, the improvement in signal-to-noise ratio achieved by narrowbanding in the ground station permits the signal-to-noise ratio at the output of the second detector to be quite poor. For the system under discussion, an average power of 1 watt for detection and ranging of a .1 m^2 target at 100 feet is adequate.

Ground Computing Equipment

The ground computing console contains all the assemblies necessary for the reception, processing, readout and storage of the ranging information. Figure 4 illustrates a typical range computing console. The active circuitry is completely solid state and is packaged on computer type plug in assemblies.

The heart of the ground computing equipment is a specialized digital processor which computes the range word by operating on the pre-selected ranging harmonic generated in the airborne sensor. The peripheral hardware consists of the readout, control, and recording equipments. The input to the ground console is multiplexed in the airborne sensor and relayed through a

narrowband telemetry link to the ground station. Conventional demodulation incorporating stable phase lock discriminators extract the ranging tones which in turn are filtered and sent to the computing logic.

The range interval can be expressed by the following expression:

$$\Theta = nRf_r \times 360^\circ$$

where:

R = Range interval
f_r = Ranging oscillator fundamental frequency
n = Harmonic selected to process

The maximum range angle to be computed for a specific range gate is then

$$\Theta_{MAX} = \frac{2L_m f_r}{c} \times 360^\circ$$

where:

L_m = Maximum range interval
c = Velocity of light

The error contributed by the measuring equipment is made a negligible part of the total allowable system error.

A maximum range of 200 feet utilizing a range tone of 100 kcps with an overall system accuracy of 6 feet requires a maximum phase uncertainty of less than 0.4°. The resolution of the system, of course, contributes along with other anomolies to be described under "Error Analysis" to this phase uncertainty. Selecting 0.1° is the range resolution and solving for R in the following expression yield the time delay associated with a system resolution of 1 foot.

$$R = \frac{\Theta}{f_r} \times 360^\circ$$

From the above equations it can be seen that any phase shift can be correlated to a time delay when referenced to a specific periodic wave. Thus, to maintain a range resolution of .002 μ sec. or 1 foot, the equipment must resolve the period of the ranging tone into .002 μ sec. increments.

The instrumentation of such a device is not practical in this form. The frequency, therefore, is translated to a more workable region. The method selected for this translation before the actual computation is initiated is to expand the time base containing the range measurement by frequency conversion and compute the expanded range increments at a more convenient frequency compatible with the information or sampling rates required by the system dynamics.

Design Criteria

A few pertinent design parameters are now considered.

The time interval which represents the radar range measurement after redemption and translation, is taken at the output of a high speed gate. This interval is digitized by utilizing a 5 mc clock pulse and sent to a storage register. The register stores for one sampling period the 12 bit range word. The register is also used to drive a decimal and binary display and paper recorder. The information displayed on the recorder is expressed in decimal form with three 10 line columns reading units, tens and hundreds of feet. Computation occurs on a real time axis and is read out instantaneously.

The total ground station is kept in synchronism by the built in timing generator also located in the console. The sync pulse for the generator is derived from the reference redeemed harmonic wave. The start of each range measurement is then in phase lock with the airborne sensor.

Calibration of the ground equipment and pre-flight and/or in-flight calibration of the complete system is easily accomplished. Standby oscillators and an electromechanical phase shifter are used to simulate range measurements to pre check the equipment and also to periodically pre align the system. Accurate operation of the ground station over a large dynamic range is accomplished by incorporating a simple IAGC loop. Manual switches are incorporated in the console to program into the device accurate range words in order to check out the accuracies of the digital and analogue circuits.

Airborne Equipment

The airborne equipment is a low noise, light weight, self contained package enclosed in an 8" x 8" cylinder and weighing 10 lbs.

The major assemblies which comprise the airborne sensor are:

1. Short pulse transmitter
2. Crystal controlled range oscillator
3. Receiver
4. Post redemption circuits
5. Telemetry transmitter
6. Power supply

Figure 3 is a simplified diagram of the airborne unit.

A crystal controlled ranging oscillator operating at 100 kcps is used to drive the short pulse transmitter. The transmitter is connected to a circulator which in turn drives the antenna. The transmitted energy upon traversing the round trip transit time to the vehicle is passed through the transmit antenna, through the circulator and to the receiver. Range gating and receive transmitter isolation may be accomplished by incorporating the following:

1. I F Gating
2. R.F. gating utilizing crystal microwave switches
3. Ferrite circulator

These methods of isolation are common in radar systems, but are not possible in C W radar. Thus by using a hybrid system as the C W redemption system, the good features of a conventional pulsed radar are utilized. Perturbations caused by gating the receiver system as the same PRF in the signal are circumvented by gating the system at twice the transmitted PRF.

Post filtering of the redeemed continuous wave is accomplished by dividing the I F amplifier into a wide band preamplifier and a narrow band post amplifier. The wide band preamplifier is necessitated by the video pulse spectrum required to gate the receiver. Second time around echoes are eliminated by selecting a PRF which is a compromise between the system accuracy and the sensitivity of the receiver. By operating with a PRF, which places the second time around echoes below the receiver threshold, ambiguities are eliminated. Pulse stretching caused by target clutter, ground clutter or receiver saturation which would contribute to a range error, are eliminated by passing the video pulse through a high pass filter.

The output of the I F amplifier is detected and sent to the pre-redemption filters and on to the telemetry transmitter. The two tracking filters are 100 kc low pass filters and used to pre-filter the telemetered tones. The relative phase stability of the filters is better than 0.1° throughout the operating temperature range.

The short pulse transmitter has been designed in several frequency bands. The X Band Unit utilizes a magnetron originally developed under Diamond Ordnance Fuze Laboratories for Eglin Air Force Base. This device is capable of generating 15 nanosecond pulses. The C Band and L Band units utilize special coaxial cavities designed around a ceramic 6442 tube. The short pulse modulator is used to initiate the gating pulses for the I F preamplifier and the r. f. switches.

System Accuracy

This paper will close with a discussion of the factors affecting system accuracy, and indicate their quantative contributions to overall systems error. A listing of the major contributors to ranging errors follows:

1. Effect of ground (ocean) reflections
2. Effect of multipath reflections caused by the vehicle itself.
3. Phase uncertainty due to noise within the passband of the 100 kcs bandpass filter
4. Uncertainty in zero crossing of the 100 kc wave due to finite pulsewidth

In addition to these, consideration may also be given to the effect of doppler, transmission anomalies and errors due to the finite data rate of the ground station. However, the magnitudes of these errors are negligible, compared to those listed earlier.

Effect of Ground Interference

The phase error introduced by ground interference is due to the change in position of the zero-crossing of the signal plus ground return as compared to signal alone. To obtain a quantitative idea, assume a reasonably calm ocean (worst case), with the drone at an altitude of h meters. It is well known that the amount

of ground reflection is proportional to the component of instantaneously illuminated area at right angles to the direction of propagation. For salt water, the proportionality constant is close to one. The area instantaneously illuminated is limited by the finite pulse width. The absolute error in the range measurement is due to the combination of ground clutter with signals; the worst situation occurs when these components are 90° out of phase.

$$P_{IN} = \frac{P_{TR}\, G_T\, G_r\, \lambda^2\, \sigma_o}{(4\pi)^3\, R4\ clutter}$$

$$P_T = \frac{P_{TR}\, G_T\, G_r\, \lambda^2}{(4\pi)^3} \times \frac{\sigma_T}{R^4}$$

$$\phi = \frac{2\pi c \tau}{\sigma target} \times \frac{R_T^4}{h^3 \sec 4\theta}$$

$$\tau = 10^{-8}$$

$$\sigma_T = .05 m^2$$

$$R_T = 33 m$$

Figure 5 is obtained by solving for the above parameters. Figure 6 is obtained by solving for a range of 150 meters. These graphs illustrate the errors which can be expected as the physical parameter altitude and look angle are varied.

Error Due to Drone Reflections

The erroneous range errors caused by multipath reflections resulting from the physical characteristics of the drone vehicle can be attributed to the same phenomena as discerned in the case of ground reflections. Unfortunately, drone reflections do not readily lend themselves to analysis.

Noncoherent scattering has only small effects and is negligible compared to the errors which can be caused by optical (coherent) scattering from specific surfaces into the receiving antenna. Furthermore, these errors are strongly a function of antenna design, drone geometry, etc. Careful control of available factors will permit minimization of drone reflections.

Phase Error Due to Noise

The presence of narrow band noise causes an error to be introduced into the range measurement due to the random displacement of resulting phase angle representative of the desired range measurement.

Since noise is a random process, the error can be described only in statistical terms. The output of a square law device consists of spectral densities due to the interaction of noise times noise, noise times signal and signal with itself.

The output of the second detector will be the desired signal amplitude corrected for the pulse shape plus the noise/signal products. The output of the narrow band filter following the detector will be that part of the signal with (S x N) and (N x N) passed by the narrow band filter.

Using Rice's equation for the phase distribution at the output of a square law device with a signal to noise ratio of 2000 results in a 95% probability that the phase error will not exceed 0.2°.

Summary

A ranging technique has been presented which is being developed for use in the scoring of high velocity noncooperative missiles. The system described is in essence a hybrid device which combines the good features inherent in both CW and pulse type radars.

References

RICE: "Mathematical Analysis of Random Noise;" BSTJ 23, 282.

DAVENPORT and ROOT: "Random Signals and Noise."

The authors would like to acknowledge the efforts of Diamond Ordnance Fuze Laboratories, around whose Short Pulse Transmitter the Redemption System was developed; also, the guidance of Target Development Laboratory, Detachment 4, ASD, under whose direction the Scoring System was developed; specifically: Captain Scholtz, William Robinette and James Bloomquist.

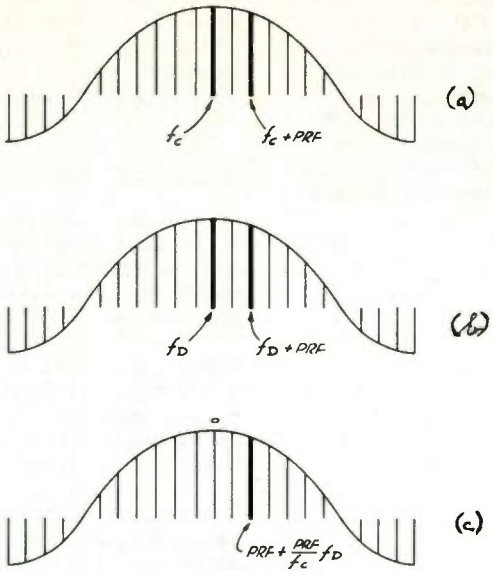


Fig. 1. Pulse spectra. (a) RF spectrum--stationary target. (b) Synchronous detection--spectrum due to moving target. (c) Nonlinear detection--spectrum due to moving target.

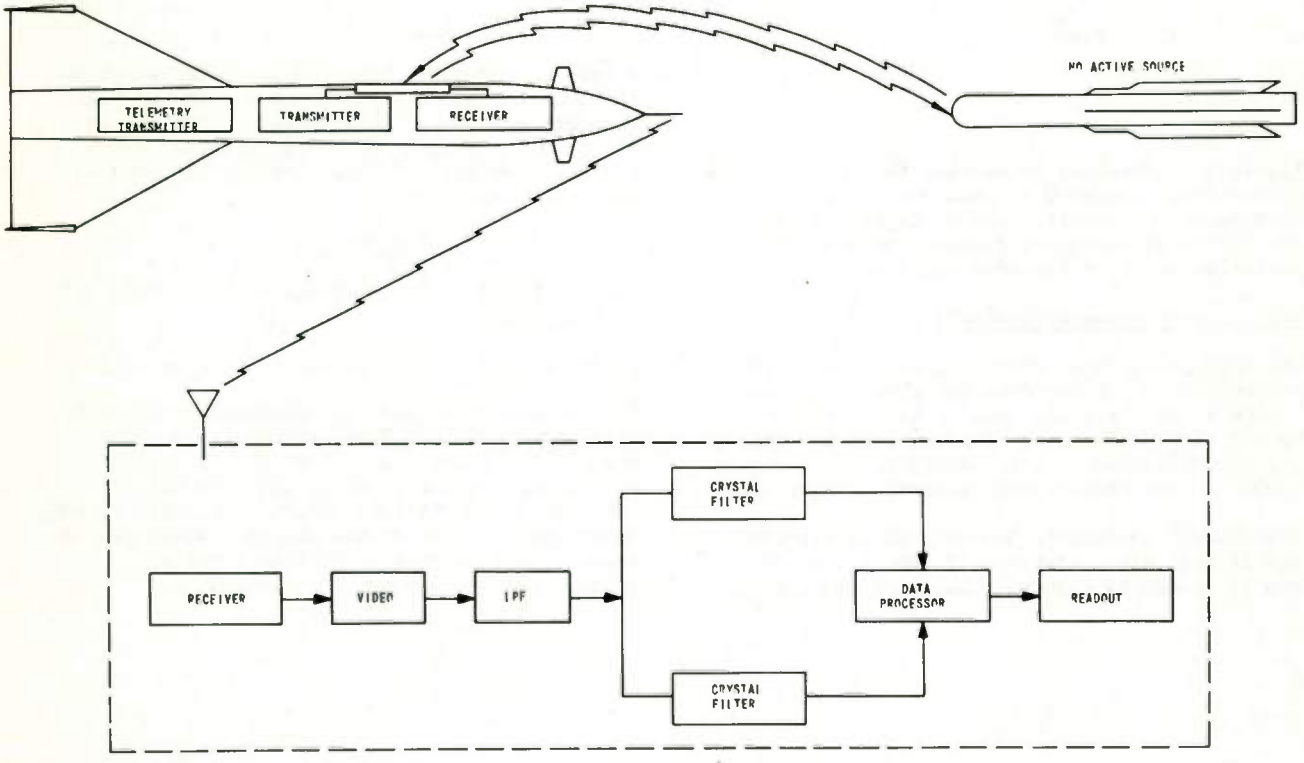


Fig. 2. Non-cooperative MDI.

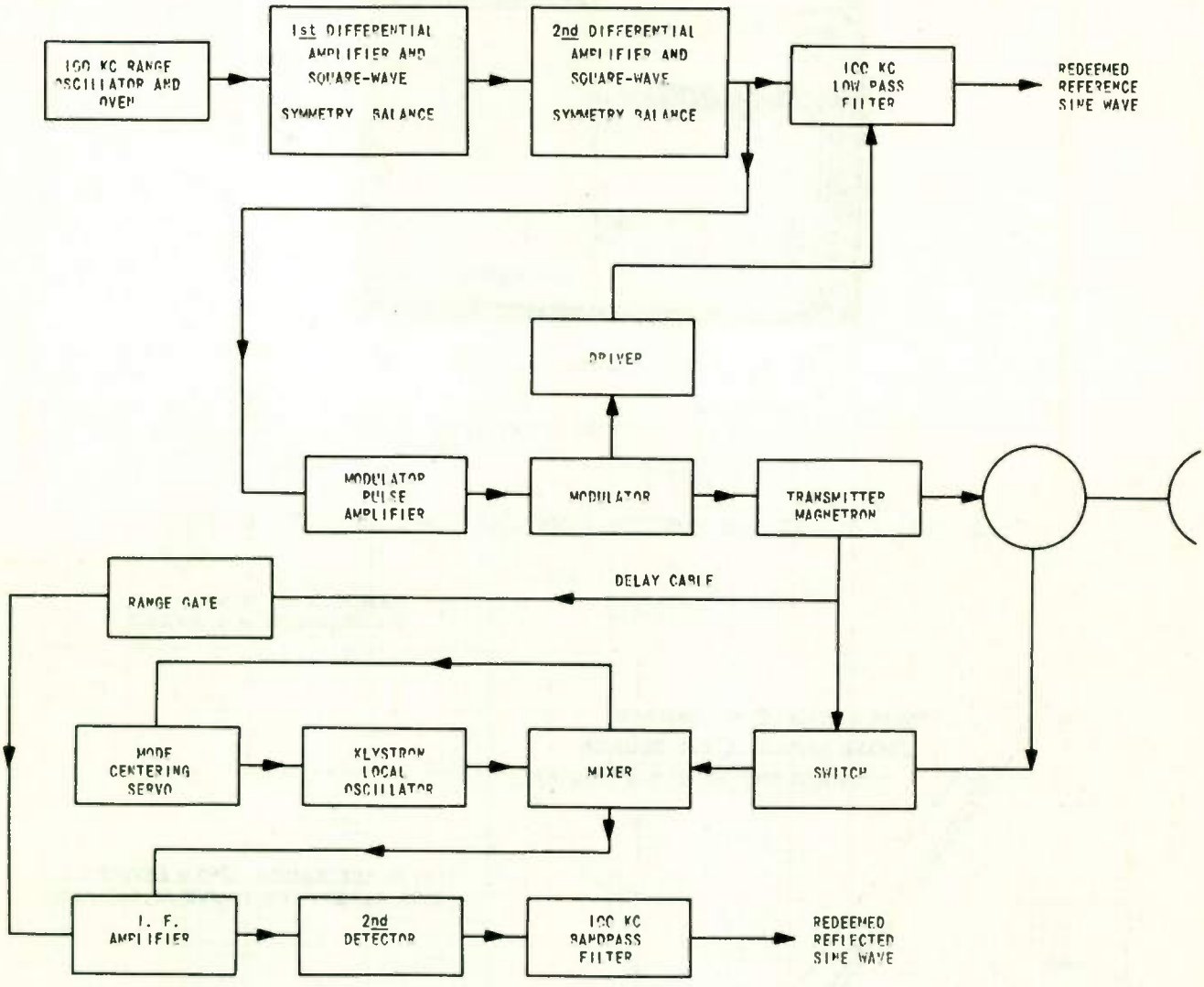


Fig. 3. Airborne system block diagram.

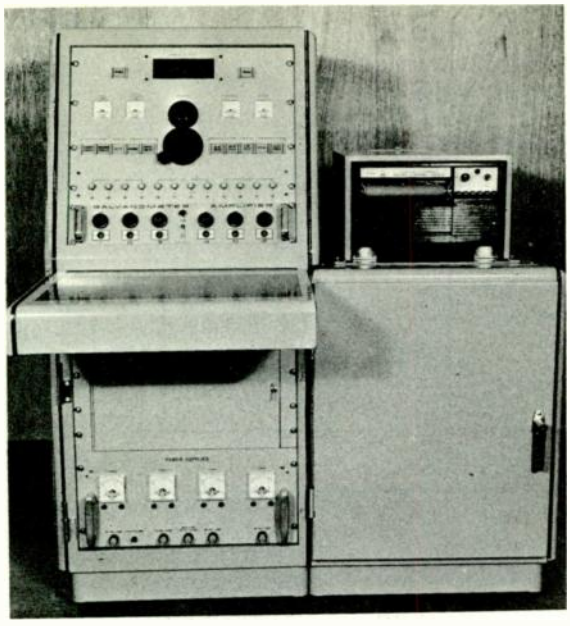


Fig. 4. Ground station equipment.

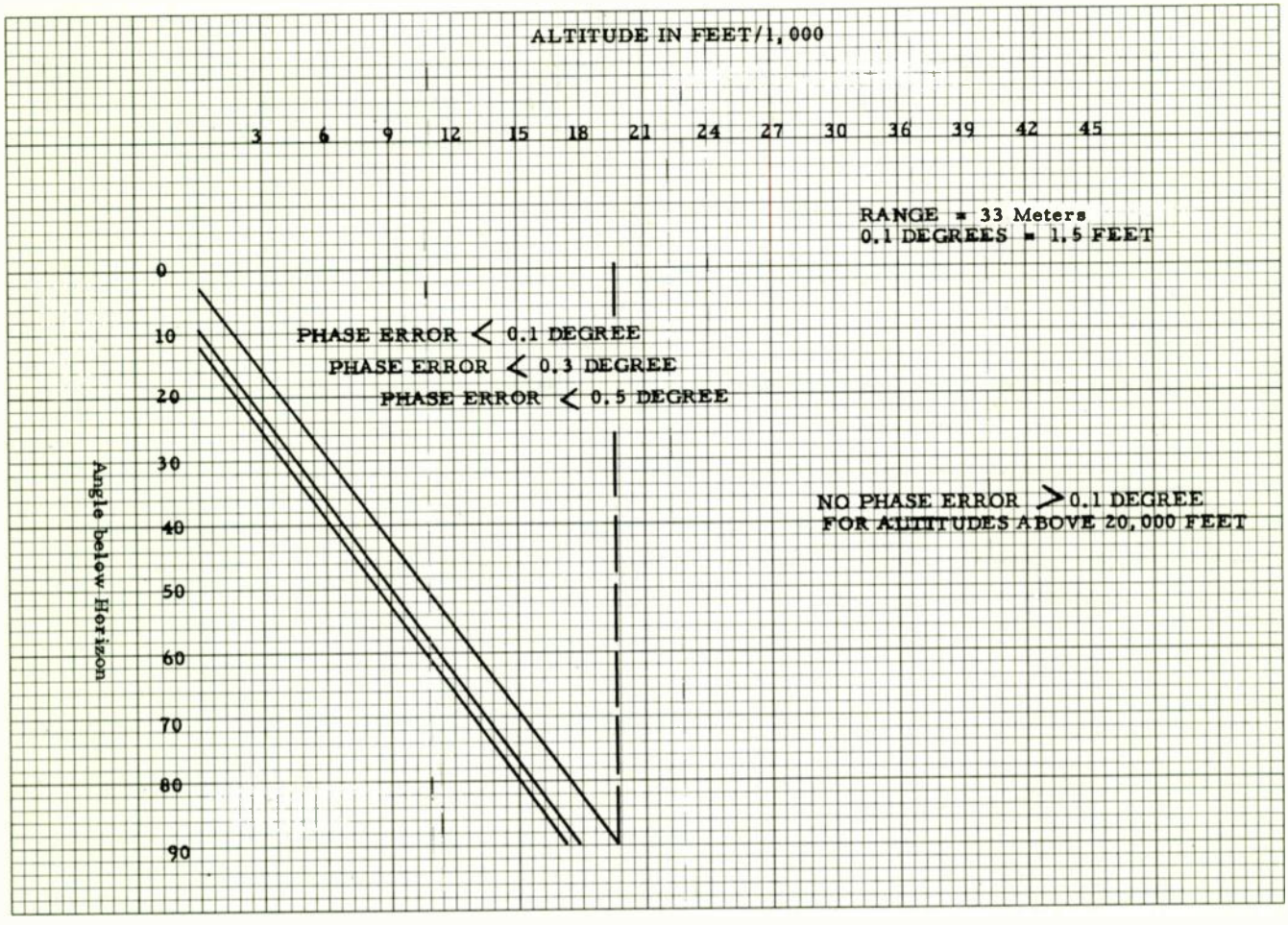


Fig. 5. Phase error due to ground reflection 33 meter target range.

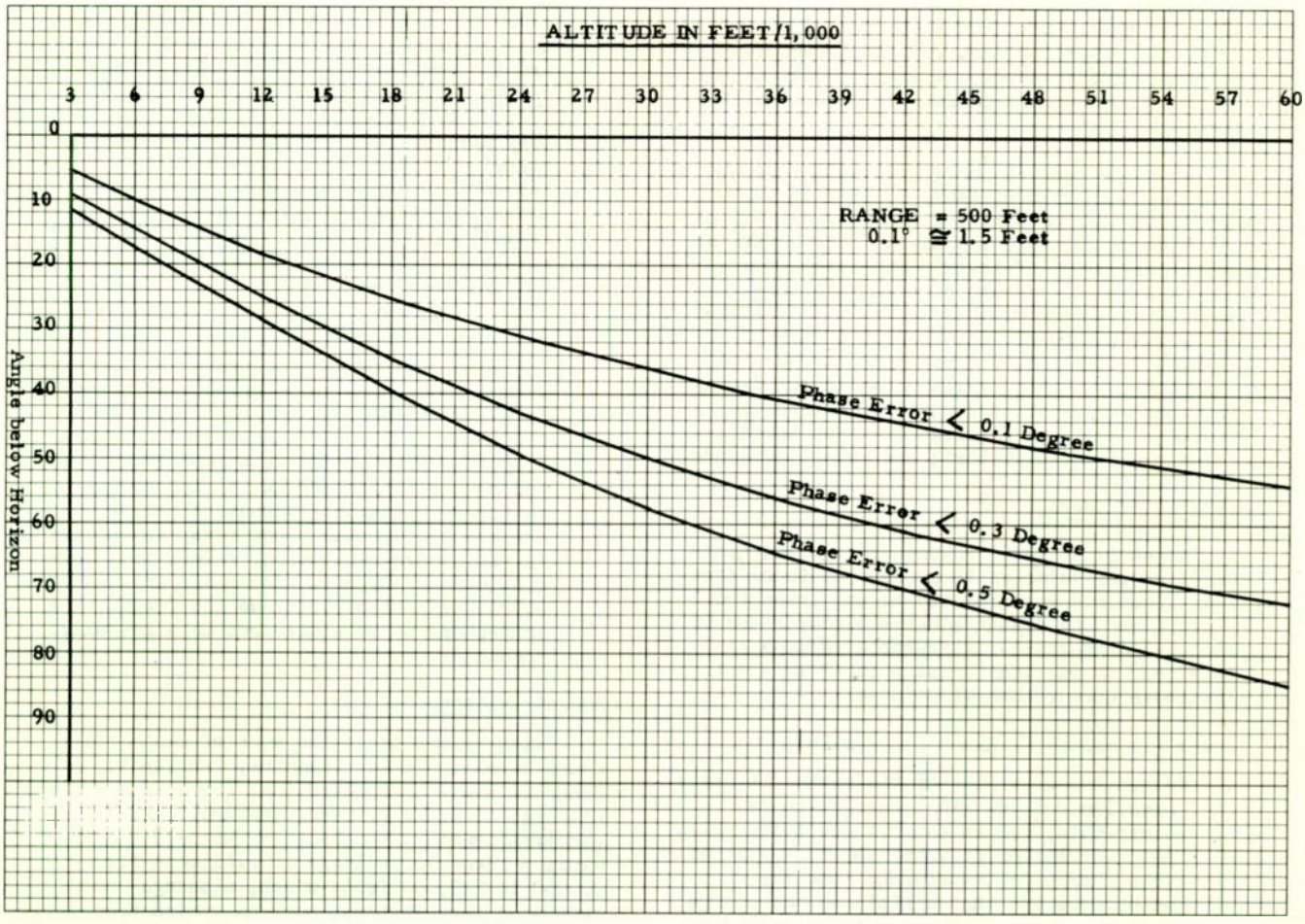


Fig. 6. Phase error due to ground reflection 150 meter target range.

A. C. FLASHOVER AT INTERMEDIATE ALTITUDES

By Dettmer H. Otto
Lockheed Missiles and Space Company
Sunnyvale, California

Is Ohm's Law Valid in Space

The question of is Ohm's Law valid in space can best be answered by defining the purpose of this paper, in which an attempt is made to describe the effects of ever-changing environmental conditions upon the supplies of electrical power for instrumentation and controls of a space vehicle. These conditions may determine the success or failure of a satellite mission.

Much the same as "more information can be generated than transmitted", it is also true that electrical sensors and controls are worthless without reliable supplies and transfers of electrical power. Today's earth satellites require several thousand electrical interconnections. Many of these carry a.c. voltages from supplying-to-using equipment through terminal boards and multipin connectors. The compatibility of a.c. voltages at intermediate altitudes has been the subject of extensive investigation. The findings of these investigation concur with historic and recent evidence.

Historic Evidence

As early as 1889, Paschen recognized minimum flashover voltages in gases. He found that their constants were related to the product of pressure and the distance between electrodes of potential difference. At certain reduced pressures, it was immaterial whether adjacent electrodes were smooth or pointed.

Israel (1) quotes F. Exner (1849-1926):

> "Since the beginning of the earth, it has a negative charge. This charge can redistribute, but it can not get lost. With the circulation of water vapor, a part of this charge is transferred into the atmosphere and thus diminishes the charge near the ground. Because of diurnal and seasonal variations, the electric field in the atmosphere must behave, or change, similarily. It follows (C.A. Coulomb, 1785) that meterorological and atmospheric-electrical phenomena are closely connected."

The frequency and extent of these variations are described by H.H. Koelle, e.a., (2) as resulting from extra-terrestrial causes:

> "There are periodic variations, with a period of 27 days, which are most certainly tied to the rotational period of the sun, and indicate the effects of the sun in changing the structure of the interplanetary magnetic field. The diurnal variation of cosmic rays is also partly due to extra-terrestrial causes, but the detailed cause has not yet been established. Among the nonperiodic changes, the most important are the decreases of cosmic rays, the so-called Forbush decreases, which last for periods of the order of hours or days, and they are very often correlated with geomagnetic storms, i.e. large disturbances of the earth's magnetic field. They are certainly due to solar emission of corpuscular radiation, but the mechanism for the Forbush decreases had not as yet been definitely established."

Chalmers (3), finds that --

> "The conductivity of the atmosphere increases with height, and, above heights of about 30 or 40 km, it becomes so great that, from the point of view of atmospheric electricity, the air behaves as a good conductor, and so one is already in the ionosphere, as far as electrical phenomena beneath are concerned, though the term, as applied in radio-wave reflection, refers to levels considerably higher. For most purposes of atmospheric electricity we can imagine the ionosphere to be equivalent to a mettallic conductor."

Recent Evidence

With reference to the work of Landmark, Kane, Nicolet, Aiken, Chadwick, van Zandt, Johnson, e.a., it appears that there is a definite relationship between pressure, density, electron density, composition, temperature and minimum flashover voltage at altitudes from 70,000 to 220,000 feet, which, although dependent upon many factors, such as Zurich sunspot number, latitude, solar declination, Greenwich hour angle of the sun and longitude, has given cause for serious concern. Until more is known about the causes and effects of these atmospheric-electric phenomena through direct observations, Minimum Flashover Voltage vs. Altitude can be expressed as shown in Figure 1. Furthermore, it may well be that there is a connection between these unusal intermediate altitude phenomena and John Glenn's observations of a "haze layer over the horizon" during his historic orbital flight.

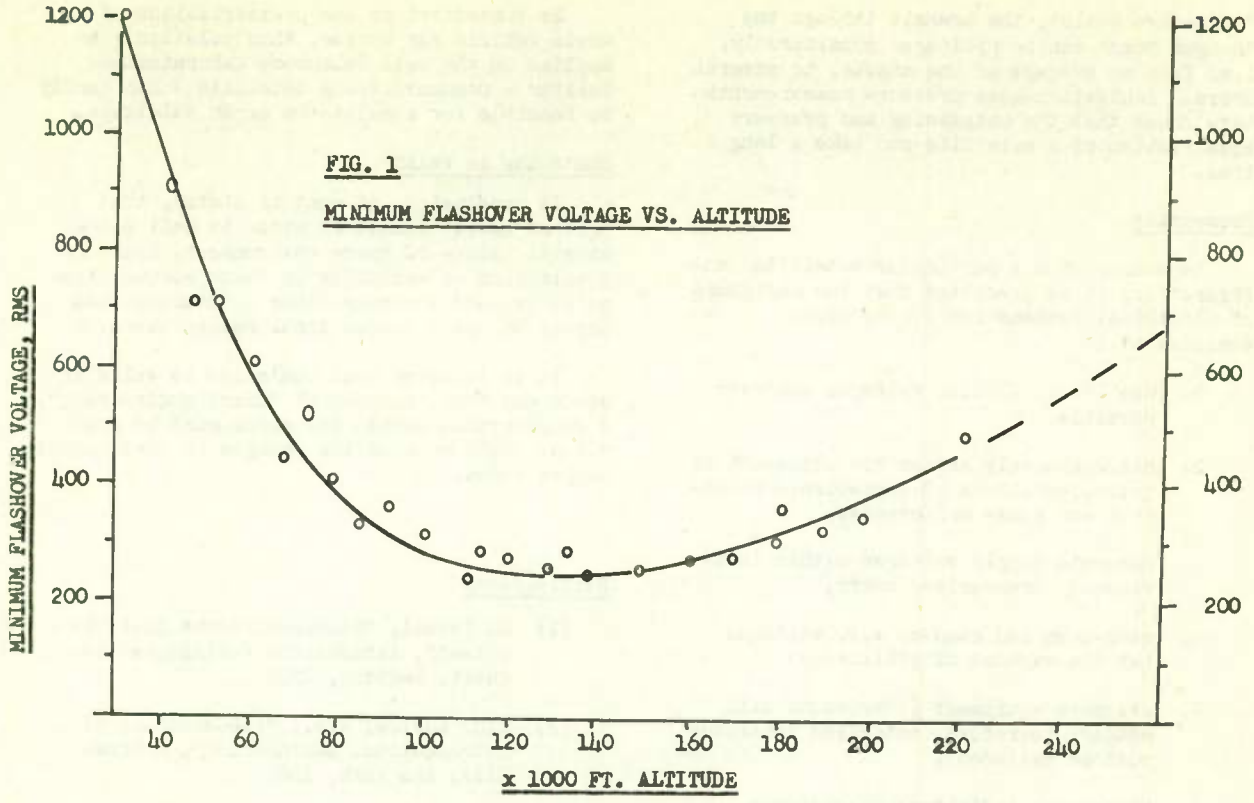


FIG. 1

MINIMUM FLASHOVER VOLTAGE VS. ALTITUDE

A potential flashover exists anywhere in a space vehicle's electrical system during the ascent phase between densely packed connector pins, in terminal boards, across miniature components, on printed circuitry, etc. Once a dependent or independent ignition voltage is reached, corona discharge becomes just as serious as flashover under certain conditions:

> If a connector were to be subjected to altitude and it were effectively sealed, the observed corona incidence for any particular altitude is essentially the same, as would be the flashover voltage for that altitude if the connector were unsealed.

A vacuum-seal exists in commonly used "moisture-proof" connectors.

The Danger Zone

Within the altitudes of 70,000 to 200,000 feet, essentially, the Worst Condition during a satellite's ascent prevails when:

Altitude	110,000 ft (33km)
Minimum Flashover Voltage	230 Volts RMS
Density	10^{-5} g/cm^3
Pressure	7.5 x 10^3 dynes/cm^2
Concentration: O^2/cm^{-3}	4.8 x 10^{16}
N^2/cm^{-3}	1.9 x 10^{17}
Temperature	225°K (-45°C)

Because of the possibility that actual voltages may exceed the minimum flashover voltages, it should be noted that "ideal conditions", e.g. the results of laboratory experiments under a bell-jar, cannot readily be seen compatible with actual atmospheric phenomena, and, therefore,

> "many of the investigations of atmospheric point discharge can only be carried out in the atmosphere, where the conditions are not very much under control." (Chalmers)

It must be assumed that a.c. supply voltages may approach, or even surpass, the recognized minimum flashover voltage limits. This holds true in particular with the addition of out-of/phase voltages, say 115 + 115 = 230 volts a.c. RMS = 324 volts peak-to-peak. At a concurrent presence of a transient, or current spike, from other sources, the "ignition voltage" may reach a momentary level of 1000 or more volts.

Failure?

The abrupt termination of a satellite mission can be foreseen with a momentary overload and breakdown of primary power, subsequent overshooting, saturation of small transformers due to a deviation of 20-30% from nominal, extremely high current flow due to a change from inductive reactance to d.c. resistance in the primary of saturated transformers, and, finally, damage to one or more semiconductors in power conversion equipment.

Without a specific consideration of

"effective seals", the transit through the "Danger Zone" can be prolonged considerably, i.e. from an average of one minute, to several hours. Ionization-gage pressure measurements have shown that the outgassing and pressure equalization of a satellite can take a long time.

Pressurize

Depending upon a particular satellite configuration, it is predicted that the designers of electrical systems for future space vehicles will:

1. use low level d.c. voltages wherever possible,
2. discriminately assign the placement of interconnections of appreciable potential and phase differences,
3. generate supply voltages within individually pressurized users,
4. step-down and step-up a.c. voltages (at the expense of efficiency)
5. evacuate equipment (lubricants will outgas; therefore, motorized apparatus must be excluded),
6. pressurize individual black-boxes,
7. pressurize the vehicle.

As convenient as the pressurization of a whole vehicle may appear, this solution - as applied on the Bell Telephone Laboratories' Telstar - Communications Satellite - can hardly be feasible for a multi-ton earth satellite.

Ohm's Law is Valid

In conclusion, it must be stated, that "ground rules" cannot be given to define the unusual nature of space environment, i.e. the combination of variables in space vacuum, temperature, and the many other conditions that appear to upset conventional recognitions.

It is believed that Ohm's Law is valid in space and that, instead of "tight engineering", a considerable margin for error must be provided, until more of the changes in environment become known.

Bibliography

(1) H. Israel, "Atmosphaerische Elektrizitaet", Akademische Verlagsgesellschaft, Leipzig, 1957

(2) H.H. Koelle, e.a., "Fundamentals of Astronautical Engineering", McGraw-Hill, New York, 1961

(3) J. Alan Chalmers, "Atmospheric Electricity" Pergamon Press, London

1205-5 5
1C-