Convention Record of the I·R·E

1953 NATIONAL CONVENTION

Part 1—Radar and Telemetry

SESSIONS ON . . .

Radio Location, Navigation and Airborne Electronics

Significant Trends in Airborne Equipment

Radio Telemetry

Remote Control Systems

SPONSORED BY

IRE PROFESSIONAL GROUPS ON . . .

Airborne Electronics

Radio Telemetry and Remote Control

Presented at the IRE National Convention, New York, N.Y., March 23 - 26, 1953

The Institute of Radio Engineers

CONVENTION RECORD OF THE I.R.E.

1953 NATIONAL CONVENTION

PART 1 - RADAR AND TELEMETRY

TABLE OF CONTENTS

THE TECHNIQUE OF MONOPULSE RADAR

W. Hausz
General Electric Co.,
Syracuse, N.Y.

ABSTRACT

The monopulse technique in radar consists in deriving sufficient information on a single pulse by multiple simultaneously acting receiving channels to determine completely both the angular position and the range of a target or targets. Two commonly used variants are phase comparison and amplitude comparison. The informational aspects of both of these, and of monopulse more generally compared with scanning techniques for angular determination, is given.

The above paper was withdrawn from the Convention program.

REDUCING SKYWAVE ERRORS IN CW TRACKING SYSTEMS

Marvin S. Friedland and Nathan Marchand
Air Force Missile Test Center
Patrick Air Force Base, Florida

Free flight testing of guided missiles is frequently conducted at the Air Force Missile Test Center to determine their performance characteristics from trajectory measurements. One of the means for making this determination is a group of electronic systems employing CW phase-measuring techniques such as the Raydist System. This is one of the systems used at AFMTC to obtain three-dimensional data on flights conducted over certain parts of the test range.

The configuration of the Raydist System is such that a phase comparison is made on a medium high frequency CW signal radiated from the vehicle under test at four or more ground stations. This is accomplished by transmitting to each of the ground receiving stations a reference signal offset from the vehicle transmitter frequency by 400 cycles per second, as shown in Figure 1. At each of these receiving points, both the airborne signal from the vehicle and the reference are received, and a 400-cps heterodyne is produced. Then, these 400-cps tones are relayed to a centrally located master station where a phase comparison is made between them, and is recorded. The trajectory data is extracted by computations made on these phase records.

The data from this type of system is deteriorated by two kinds of interference, atmospheric noise and multipath propagation due to skywave and ground reflections. Of these, the skywave contamination is the most insidious because it cannot be detected until the phase data is reduced; then, the reduced results merely indicate that the data is worthless. Transmissions from the reference transmitter can be protected against these effects by careful geographic location and proper antenna design, but the airborne signal cannot be protected by these techniques. It is the purpose of this paper to present a method of time discrimination which protects this signal from the missile against this type of contamination.

Discrimination is accomplished by pulsing the airborne transmitter to permit the ground receivers to gate out the skywave components. Tests on the system indicate that these components must be kept at least 30 db below the direct wave for the data to be useful; therefore, if possible, the skywave must be reduced by at least this amount. Figure 2 shows the expected delay in skywaves for a missile flying at 40,000 feet and an ionosphere height of 50 and 60 miles. For ease in this computation, the earth was considered flat. An inspection of these curves indicated that a pulse-to-pulse period of 1600 microseconds and a pulse width of 160 microseconds would prove satisfactory so as to permit the system to lock on the direct wave, for ranges out to 100 miles.

Both the missile transmitter and the ground receivers must be modified to apply the time discrimination technique. The transmitter modifications are straightforward in that a pulse modulator is added to the existing CW unit as shown in Figure 3. The power supply and power amplifier are adjusted so the average power remains about the same as for the CW condition. This is not too difficult since the duty factor required is 0.1. Thus, the modulator causes the transmitter to radiate 160 microsecond pulses at the rate of 625 per second. The modifications to the receiving system to match these transmitter pulses, however, must be considerably more complex, and are considered next.

A block diagram of the modified receiver system is shown in Figure 4. It may be seen that the synchronizing circuits are separated from those used for phase measurement, but that they are fed from a common antenna. When the system is in operation, signals from both the CW reference transmitter and the pulse missile transmitter are fed into two tuned gated preamplifiers. These units are arranged so that they are open until the gates are generated in the time discriminator to close them. From these preamplifiers, the signals are fed to each of the receivers for amplification and detection. The output from the pulse synchronizing receiver is sharpened in a differentiator and applied to the time discriminator where it is used to generate a 100-microsecond gate. The termination of this 100-microsecond gate is used to start the 850-microsecond gate which closes the preamplifier feeding the pulse synchronizing receiver. It is also used to control a 625-cycle AFC-sync automatic frequency control unit. The 850-microsecond pulse to the preamplifier prevents the skywave components from entering and confusing the pulse synchronizing receiver, or in the event that the system is started by a skywave pulse, it holds this circuit closed until the next direct pulse arrives.

The output of the AFC-sync unit is used to generate a 1500-microsecond length gate so phased that it allows the first 100 microseconds of the direct wave to enter the phase comparison receiver. All other signals are rejected so as not to deteriorate the signal-to-noise ratio at this point. The output from the phase-measuring receiver is then introduced into a filter whose

center frequency is 400 cps and band width is ± 20 cps. This filter serves as an integrating device to pass the 400-cycle beat and the doppler components due to the velocity of the missile, but rejects the 625-cycle component and its side bands. A 20-cycle band width is enough to accommodate the doppler components of all signals from all types of test vehicles contemplated on the range. The AFC-sync unit serves to stabilize the 1500-microsecond gate during the periods when atmospheric noise and fading cause the time discriminator to drop out. The time constant in this circuit is several seconds since the time variations in the repetition rate of the transmitter are very slow, and change in the doppler effect is small. A dual time constant circuit is used in the control loop to allow for quick lock in, yet reduce the effects of noise and interference.

A time analysis in Figure 5 shows expected time delays between the direct wave and skywave components as they would arrive at a receiver. On the first line, the first pulse received is the direct wave which is then followed by skywave pulses. This illustration shows the earliest and latest delay conditions expected. In the next two lines are shown the gates that must be generated in the time discriminator to allow the direct wave to pass, yet block, the skywaves out of the phase comparison receiver. The first of these gates is 100 microseconds long to permit a portion of the direct pulse to pass into both receivers; the end of this gate is differentiated to start the next of these gates of 850 microseconds length which must be generated to prevent triggering the discriminator by skywave reflections. Another gate of 1500-microseconds length is started at the same time to gate these waves out of the phase comparison receiver. In the event that the discriminator is first started by a skywave, as is shown on line 4, the next direct wave pulse will synchronize the system since the 850-microsecond gate will have opened in time to allow it to pass. Line 5 illustrates the effect of this same skywave effect on the gate to the phase-comparison receiver. The result is that the data from this receiver is not useful until the system locks on the direct wave.

Under ideal conditions where a perfect rectangular pulse could be transmitted, the degree of discrimination could be made infinite; however, the restricted band width available determines the practical limitation of this, the discrimination technique. In practice, the transmitted pulses must be rounded off; therefore, the receiver output may be determined by obtaining the average area under the gated pulse as a function of relative delay. In analysis, the detected pulse may be represented as

$$\text{Detected Pulse} = f_1(t) = E\frac{\tau}{T}\left[1 + 2\sum_{\eta=1}^{K} \frac{\operatorname{Sin}\left(\frac{\eta\pi\tau}{T}\right)}{\frac{\eta\pi\tau}{T}} \cos\frac{2\pi\eta t}{T}\right]$$

where E = Pulse Amplitude
τ = Pulse Width
T = Pulse Repetition Period
The gate can then be stated as

$$\text{Gate} = f_2(t-\delta) = \frac{\tau}{T}\left[1 + 2\sum_{\eta=1}^{\infty} \frac{\operatorname{Sin}\left(\frac{\eta\pi\tau}{T}\right)}{\frac{\eta\pi\tau}{T}} \cos\frac{2\pi\eta(t-\delta)}{T}\right]$$

where δ = an arbitrary time delay between the detected pulse and the gate. Then the detected output from the receiver versus the delay of discrimination function is obtained by integrating the product.

$$G(\delta)\int_0^T f_1(t) f_2(t-\delta)\,dt = \frac{1}{E}\int_0^T f_1(t) f_1(t-\delta)\,dt$$

f_2 can be replaced by f_1 since terms for which $\eta > K$ do not contribute to the integral. This latter integral may be recognized as being proportional to the auto-correlation function of $f_1(t)$. By integrating and normalizing so that $G(o) = 1$, the discrimination function becomes

$$G(\delta) = \frac{1 + 2\sum_{\eta=1}^{K} \frac{\operatorname{Sin}^2\left(\frac{\eta\pi\tau}{T}\right)}{\left(\frac{\eta\pi\tau}{T}\right)^2} \cos\frac{2\pi\eta\delta}{T}}{1 + 2\sum_{\eta=1}^{K} \frac{\operatorname{Sin}^2\frac{\eta\pi\tau}{T}}{\left(\frac{\eta\pi\tau}{T}\right)^2}}$$

Calculations have been made in which the transmitted band width B was restricted to include only that portion of the $\frac{\operatorname{Sin}\frac{\eta\pi\tau}{T}}{\frac{\eta\pi\tau}{T}}$ spectrum which is out to the first null. Figure 6 shows a plot of this function $G(\delta)$ for various values of $\frac{\tau}{T}$, the duty factor. It can be seen here that these curves approximate the triangular discrimination function for the ideal case of infinite band width such as when a full spectrum is transmitted. In this latter case, K becomes infinite (if the duty factor is kept fixed), and the last equation can then be identified as a Fourier series for a triangular pulse whose base is two pulse widths wide, thus making the discrim-ination function zero between $\delta = T$ and $\delta = T-\tau$. Under these conditions K becomes equal to $\frac{\tau}{T}$ and $B = \frac{2K}{T} = \frac{2}{\tau}$.

An examination of the curves in Figure 6 shows that they have finite values near 0.05 at the points just mentioned. If this discrimination were sufficient at a time delay of 200 microseconds, then the pulse width, τ, could be taken as 200 microseconds. Should this not be so, then a shorter pulse width is required. The curve

corresponding to a duty factor ($\frac{T}{PP}$) of 0.1 is the condition in the system as it is described. This appears to give the required discrimination at a delay of 200 microseconds. A calculated value for the attenuation of the skywave components under conditions of normal propagation is 40 db below the ground wave component. This will permit the same frequency to be used for both day and night operation.

The system as described has been tested in breadboard form and found to conform in general to the expected characteristics; however, there are some indications that the 400-cps and 625-cps frequency components may inter-act in the narrow band filter. This will be determined at some future date as the program continues. In the interim, it may be necessary to change the repetition rate somewhat.

In conclusion, it can be stated that by apply-ing time discrimination to the Raydist System, it becomes a much more useful and effective tool for trajectory determination, and that this modification does not make the missile transmitter so complex that its increase in size and weight makes it impractical for use on the range.

It is desired to credit the Sylvania Electric Products Company and the Parsons Aerojet Company, as both have coöperated with AFMTC on this project and made valuable contributions to it.

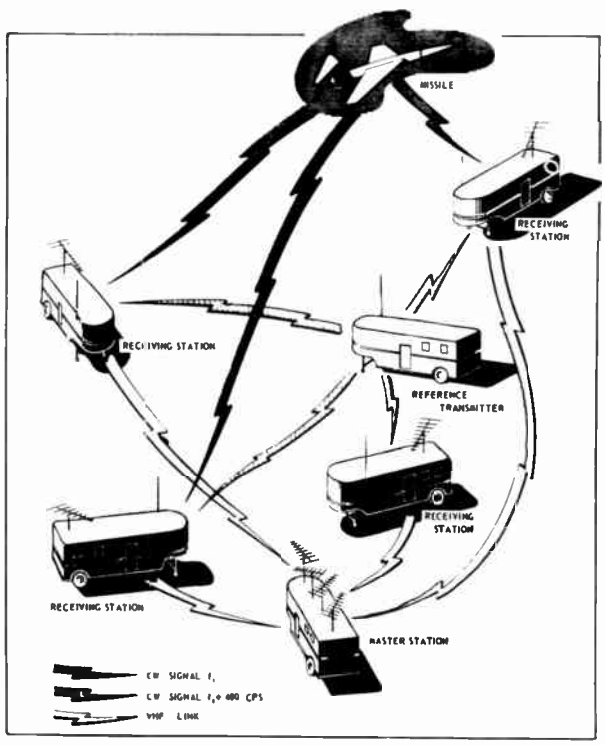


Fig. 1
Raydist system block diagram.

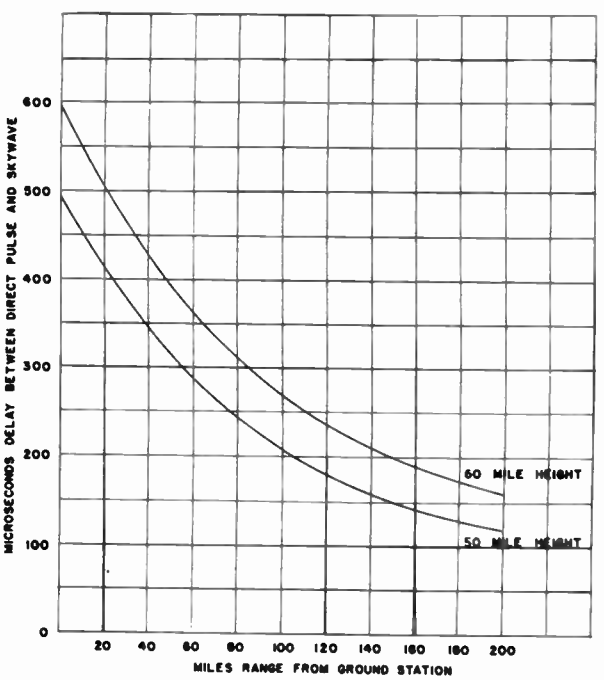


Fig. 2
Skywave delay with respect to direct pulse against range.

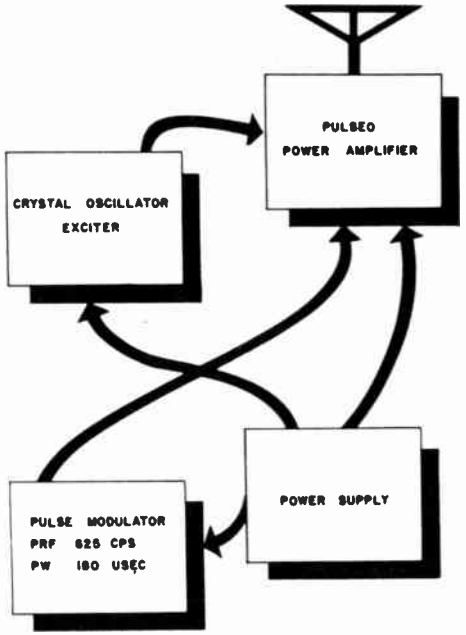


Fig. 3
Missile pulse transmitter.

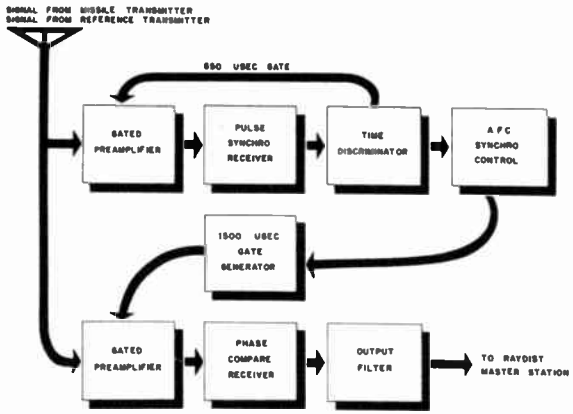


Fig. 4 - Modified receiver system.

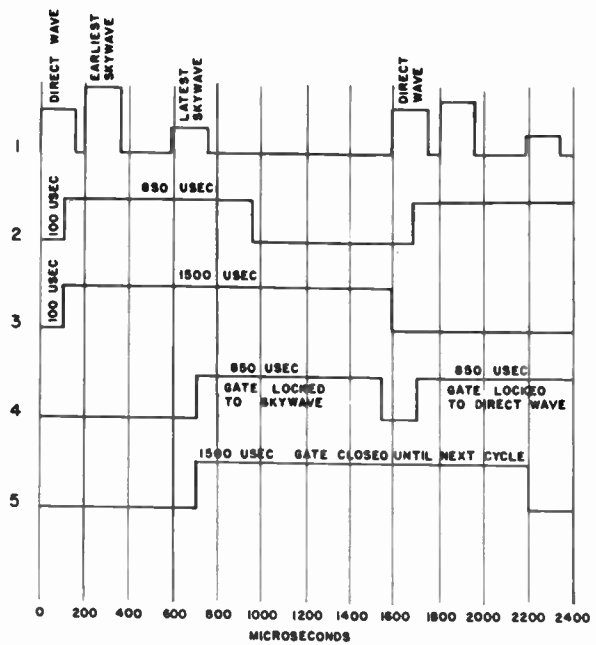


Fig. 5
Time analysis of direct and skywave pulses related to system gates.

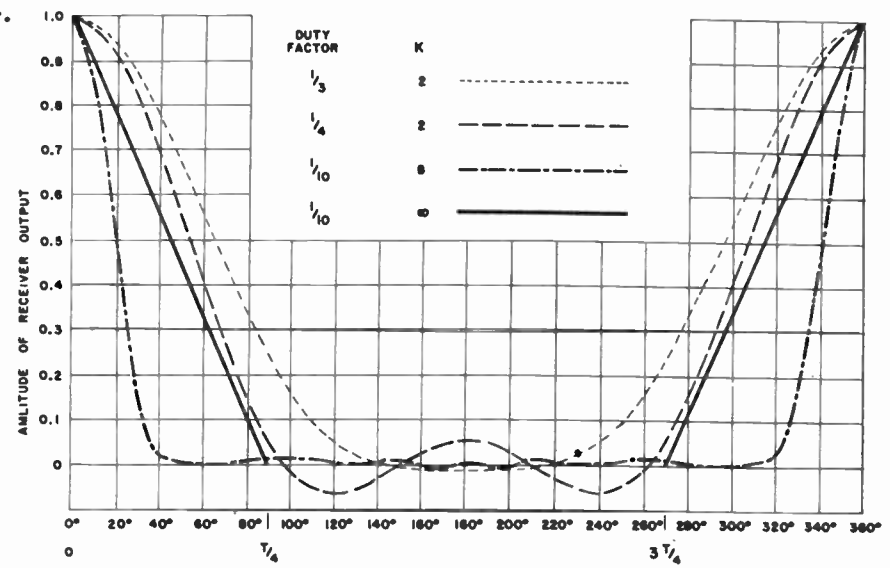


Fig. 6 - Discrimination function $G(\delta)$ for several duty factors.

APPLICATION OF INTEGRATOR-TYPE SIGNAL ENHANCERS TO DIRECTION-FINDING EQUIPMENTS

Charles A. Strom, Jr. and Joseph A. Fantoni
Rome Air Development Center
Rome, New York

INTRODUCTION

Information theory studies made during and shortly after World War II indicated that repetitive-type signals might be greatly enhanced by trading bandwidth for time. An analysis of the output voltage of a radio Direction-Finding system indicated that the Direction-Finding voltages fall into this catagory and that considerable improvement in the Direction-Finding bearing S/N ratio might be achieved by proper operation on the information or output voltage. Such operation, if successful, would go a long way in improving the operational sensitivity of the Direction-Finding equipment and permit usable Direction-Finding bearings on any signal that could be heard by a radio communications station.

While the work done shortly after World War II indicated that both correlation techniques and integration techniques would enable great improvements in S/N ratios, the post detector integration technique was selected for the Direction-Finding studies because of the greater apparent simplicity of the associated equipment. Also, this approach appeared to offer some advantages in simplicity of operation and greater stability. The objectives of the study at the Rome Air Development Center (started at Watson Laboratories, Red Bank, N. J.) was to investigate some of the possible methods of accomplishing a field-type unit which would be useable with existing Direction-Finding equipments and to develop a technique which would be available for use in new Direction-Finding equipment. To further simplify the problem, it was assumed that the signals received by the Direction-Finding station would be from a friendly transmitter and would be of sufficient duration to permit good use of the integration technique. Signals, such as these transmitted from an aircraft under tower or GCA surveillance were assumed as the normal type transmission.

Before proceeding with a description of the post detector integration approaches investigated, it might be well to briefly review the S/N improvement by integration techniques and redefine the terms used throughout this paper.

First, assume a Direction-Finding antenna in azimuth at a constant speed, and connected to a radio receiver. Any RF energy at the input to the receiver will be amplitude-modulated by the azimuth pattern of the antenna. Since we are dealing with a coöperative transmitter, the signal will remain on for several rotations of the antenna and the output wave form from the receiver will have a fundamental period equal to the rotational time of the antenna, even though the Fourier analysis of this signal may show it to be very complex and include many harmonics of the fundamental frequency. Since we are performing post-detector integrations, only the audio (detected) output of the receiver is of interest and will be fed into the integrator. The integration process requires this addition (cycle by cycle in exact phase and amplitude) of the audio output of the receiver. The signal being in phase and amplitude agreement from cycle to cycle should add directly as (n) the number of additions, while noise, being of a random nature, should add only as $\sqrt{n}$.

DESCRIPTION OF TECHNIQUES INVESTIGATED

In order to investigate the applicability of integration as a means of signal enhancement, three separate approaches were undertaken at Rome Air Development Center. They were: (A) Magnesium Delay Line (B) Magnetic Storage Drum and (C) Capacitative Storage Drum. Each was connected to an automatic Direction-Finding station and operated in the system to determine the operational improvement of the station. With the exception of the magnetic drum, these units were experimental models and no attempt was made to miniaturize or package these into a field design. Figure I shows a block diagram of the signal enhancement unit used with a Direction-Finding station.

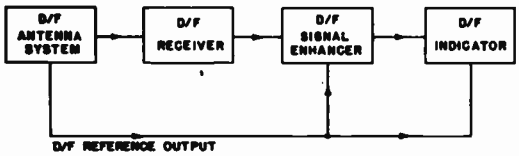


Fig. 1
Signal enhancer unit inserted into a D/F station.

A. One of the first units considered was the delay-line type of signal enhancer. In this unit the signal is delayed in time by multiples of one cycle, and the delayed signals are added in phase

and amplitude coincidence, to form the enhanced signal output. Figure 2a is a block diagram of this system for signal enhancement. Figure 2b shows how one line may be made to do the work of several lines by using different carrier frequencies. The unit tested at Rome Air Development Center was of this type.

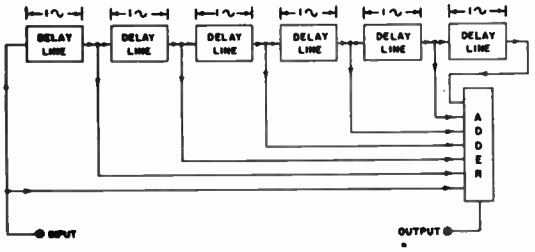


Fig. 2(a) - Multiple delay-line system.

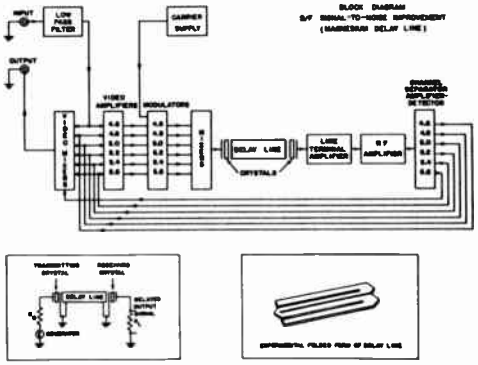


Fig. 2(b)
Single delay-line system employing spaced carrier frequencies.

The experimental model made use of a magnesium supersonic delay line and a carrier frequency of approximately five megacycles. The signal output from the receiver modulates the 5-megacycles carrier and this modulated signal is applied to a crystal at one end of the delay line. The crystal changes the electrical signal into a 5-megacycle mechanical vibration which is propagated down the delay line. At the receiving end of the line, these vibrators excite another crystal which converts the mechanical vibrations back into a 5-megacycle electrical signal. An amplifier is used to raise the signal up to a workable level. The output of several delay lines is then added in phase to produce the enhanced signal.

Figure 3 shows the experimental model of the equipment fabricated for Rome Air Development Center by Crystal Research Laboratories of Hartford, Connecticut.

B. The second approach investigated at Rome Air Development Center for improving the S/N ratio by integration resulted in the fabrication of a magnetic storage drum with multiple read-out heads. The output voltage from the receiver is recorded on the drum and the simultaneous output from all playback heads is used to give an integrated output. Figure 4 shows a block diagram of this equipment.

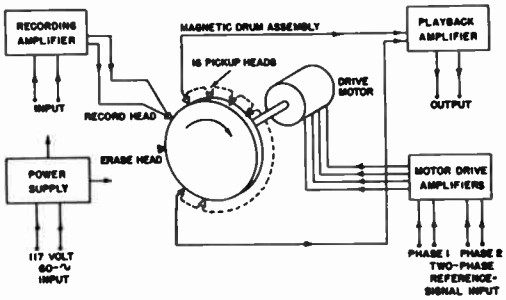


Fig. 4
Magnetic storage drum system with multiple read-cut heads.

The model constructed for Rome Air Development Center makes use of an iron oxide coated drum which is driven at 1/20th of the Direction-Finding rotational speed. The sixteen read-out heads spaced at intervals of one cycle around the periphery of the drum are connected in series and the output voltages which are in phase and amplitude coincidence are the enhanced signal output. An erase head is used ahead of the recording head to remove the old signal prior to recording the new signal. The motor driving the drum is operated in synchronization with the reference voltage and is geared to the drum so that it is driven at exactly 1/20th the speed of the antenna system. Figures 5 and 6 show the engineering mode of this unit fabricated for the Rome Air Development Center by Engineering Research Associated of St. Paul, Minnesota. Excellent phase and amplitude linearity had to be maintained in the recording system to prevent distortion of the Direction-Finding intelligence at the frequencies employed.

C. The third type of post detector-integrator investigated at Rome Air Development Center consists of a capacitor drum as the storage unit. This item was suggested by Mr. H. Busignies and M. Dishal in an article entitled "Signal-to-Random-Noise Radio in Radio Navigation and Direction" PROCEEDINGS OF THE IRE - May 1949.1 In this storage system a drum is made up of several large capacitors and driven in synchronization with the Direction-Finding antenna. The signal output of the receiver is applied directly to the capacitors on the drum through one brush, and the integrated output is read off at a later item through another

brush. Figure 7 shows the circuit and block diagram of this unit.

Each of the capacitors is the same size, and the time constant of the read-out circuit selected so the capacitor will not discharge completely each time it contacts the read-out brush. The voltage assumed by the individual capacitors on the drum will be proportional to the amplitude of the signal at the moment of sampling. As the drum is rotated, the output on the read-out brush will be a series of short pulses equal to the individual charges stored on the capacitors on the drum.

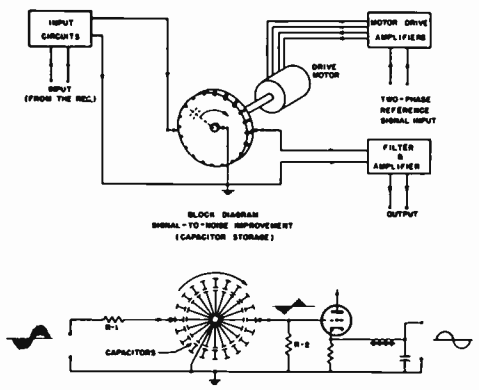


Fig. 7 - Capacitor drum-storage system.

A filter is used to remove the commutating frequency from the resultant wave form. Figure 8 shows the wave forms into, and out of, the capacitor integrator, and filter. The improvement in S/N ratio is limited only by the charge and discharge time of the capacitor-resistor network associated with the integrator. Figure 9 is a photograph of an experimental model capacitor storage integrator built at Rome Air Development Center.

RESULTS

The theoretical improvement in signal-to-noise ratio possible with the post-detector type of signal integrator is shown to be equal to So/No $\times \sqrt{n}$ where (n) is equal to the number of cycles of information over which the integration takes place. Studies made at Air Force Cambridge Research Center by Rogers and Harrington and reported in the previously referred paper,[2] show that in the case of very weak signals, i.e., S/N $\ll 1$ this improvement gradually reduces to So/No $\sqrt[4]{n}$.

In considering the improvement possible with three types of units investigated at Rome Air Development Center, we will consider S/N ratios of 1:1.

The magnesium delay-line unit delivered to Rome Air Development Center permitted a total of seven linear additions giving us a S/N improvement of 20 log $\sqrt{7} = 8.44$ db when positive feed back is employed; the total improvement possible increases, and is limited by the time one is willing to wait and the percentage of enhanced signal that can be fed back without oscillation.

The signal enhancer utilizing the magnetic storage drum permits a total of 16 linear additions, and the theoretical S/N improvement is increased to 20 log $\sqrt{16} = 12$ db. If we add positive feed back (or recirculation) the total improvement may be increased to over 20 db. However, with this much feedback, the information requires several seconds to register and the indications are very sluggish.

Since the capacitative storage drum must follow the variations in phase and respond to new signals in a relatively short time, the circuit must be self discharging, and the number of effective additions will be less than the actual number. The voltage stored may be written as $E_{out} = \sum_{q=1}^{n} E_q e^{-(n-q)\gamma}$ where γ is the discharge factor. In the work of Rogers and Harrington,[2] the signal-to-noise improvements is shown to be equal to $n^{\frac{1}{4}} \left[\frac{\tanh \frac{n\gamma}{2}}{\frac{n\gamma}{2}}\right]^{\frac{1}{4}}$. For 30 additions and assuming $\gamma = 0.1$. The improvement is $4.25 = 12.68$ db. In the case of a 30-cycle per second Direction-Finding antenna rotational frequency, the delay time is equivalent to one second.

Figures 10 and 11 are photographs showing the experimental results obtained at Rome Air Development Center Laboratories when the Radio Direction-Finding equipment was operated with and without the magnetic drum signal integrator. Figure 10a is the bearing presentation on a good signal without the integrator in the circuit. Figure 10b is a bearing presentation on a fair signal without the integrator. Figure 10c shows the bearing on a poor signal presented without the integrator, i.e., the transmitter is near line of sight or the signal is obscured by high noise level. Figure 10d shows the same signal as Figure 10c but with the integrator in the circuit between the receiver and the azimuth indicator. Figure 10e is the bearing presentation on a very poor signal without the integrator in the circuit. Figure 10b is the same as on Figure 10e, but with the integrator in use. Figure 11 shows the effect of employing positive recircuitation in the signal equipment. Figure 11a shows a bearing on a poor signal. Figure 11b shows this same bearing with the signal integrator operating normally. Figure 11c shows the bearing when positive feed back has been used. In Figure 11d we have a bearing presentation on a very poor signal. Figure 11e is this same bearing with normal integration and Figure 11f is this bearing with integration and positive feed back.

The results with the solid delay line were similar to those obtained with the magnetic drum. However, limitations in the length of the line

did not permit the improvement obtainable with the magnetic drum.

Figure 12 summarizes the results of flight tests made with and without the signal integrator. We see from this table that there is no improvement in bearing presentations on good signals. However, the integrator does improve the quality of the bearing presentations for the weak and very weak signals. The net result of these tests were to prove that bearings could be reliably presented on any signal received at the Direction-Finding antenna, and by employing this technique, it was possible to make the Direction-Finding range equal or exceed the communications range.

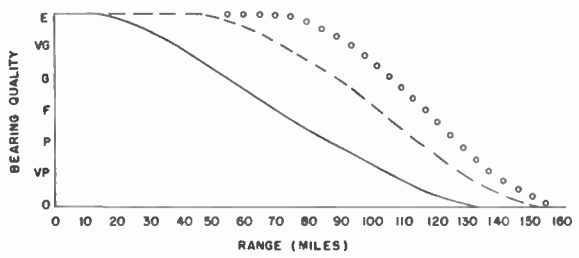


Fig. 12
Bearing quality vs. range with and without the use of enhancer equipment.

In Figure 13, curves are superimposed on a map to show the improvement in range available by the use of signal integration. The inside curve shows the range of a good Direction-Finding station without the integrator. When the integrator is used between the receiver and azimuth indicator, the effective range of the Direction-Finding station is moved out, to approximately line of sight. Additional increase in range obtained by the use of positive feed back.

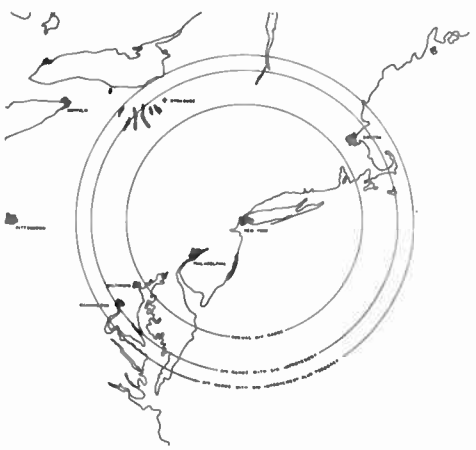


Fig. 13
D/F bearing range with and without signal enhancer.

Another result of this post detector integration technique is to average the bearings received and to reduce the amount of flutter on the bearings caused by propagation difficulties. Azimuth accuracy is maintained only as long as the flutter of the bearings is symmetrical. Since signals which arrive at the Direction-Finding site via a direct path are usually much stronger than reflected signals, the proper signal will usually be favored.

The results of the study show that this technique of improving the Direction-Finding quality and readability by signal integration is fully usable with field-type radio direction-finding station. It is practical to build a simple integrator-type unit which may be inserted between the receiver and azimuth bearing indicator, for enhancing the bearing presented, thereby increasing the readability and extending the range of the station. S/N improvements from 6 to 12db are easily obtainable at the present with the techniques we have described, and greater improvements may be obtained at the expense of a longer integration time. If we arbitrarily accept a one-second time limit as the practical limit of integration, this will restrict us to an improvement in the neighborhood of 12 to 16db overall.

A credit is due to engineers of the Anderson Laboratories and Crystal Research Laboratories, of Hartford, Connecticut for their work on the magnesium delay-line type of signal integrator, and suggestion for improving this technique, and to Mr. Anderson for application of positive feed back. Credit is also due to engineers of Engineering Research Associates at St. Paul, Minnesota who fabricated the engineering model of the magnetic drum type of signal integrator. Credit is also due to engineers of the Air Force Cambridge Research Center who worked with engineers of the Rome Air Development Center in developing the experimental units and applying these techniques to the Radio Direction-Finding Station.

---

[1] H. Busignies and M. Dishal, "Signal-to-Random-Noise Ratio in Radio Navigation and Direction Finding", Proceedings of the IRE; May, 1949.

[2] T. F. Rogers and J. V. Harrington, "Signal-to-Noise Improvement Through Integration in a Storage Tube", Proceedings of the IRE, Vol. 38, No. 10, October, 1950.

Fig. 3
Experimental model magnesium delay-line signal enhancer.

Fig. 5
Engineering model magnetic storage system signal enhancer.

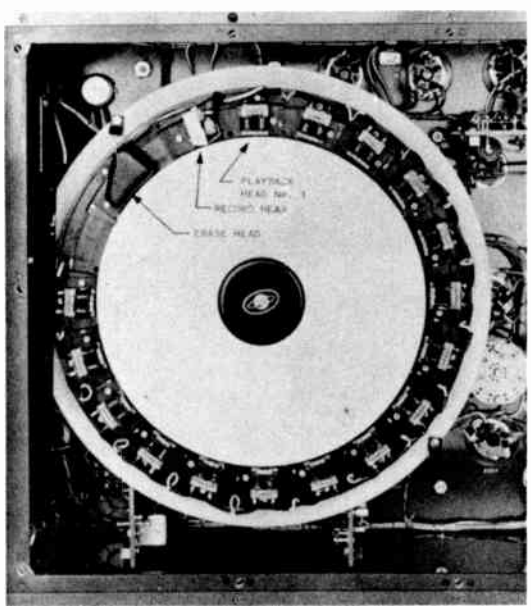


Fig. 6
Side-chassis view of the magnetic storage unit.

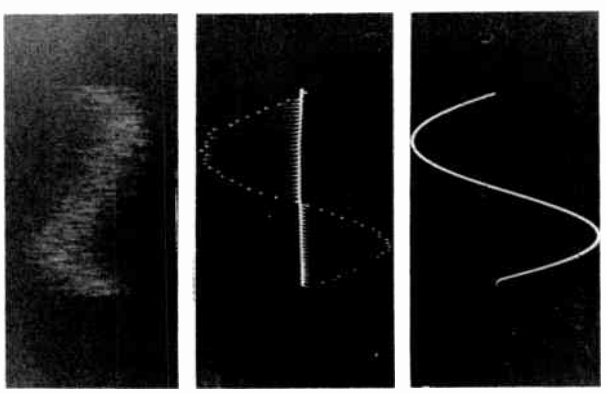


Fig. 8
Input and output waveforms obtained with the capacitor storage unit.

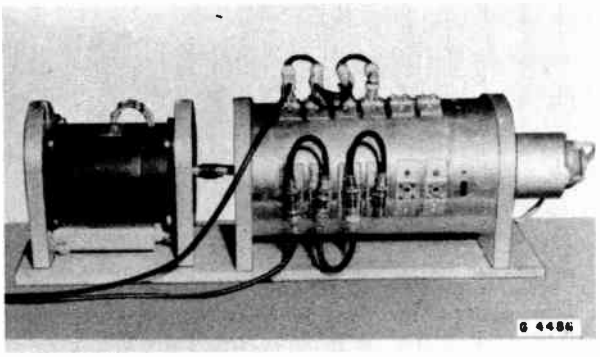


Fig. 9
Experimental model capacitor storage unit.

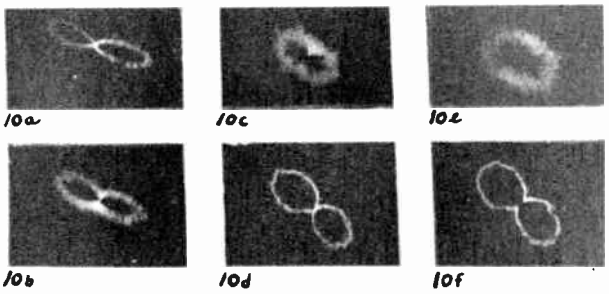


Fig. 10
Experimental results; (a) good bearing; (b) fair bearing; (c) poor bearing; (d) bearing presentation on a poor bearing using the signal enhancer; (e) very poor bearing; (f) bearing presentation on a very poor signal using the signal enhancer.

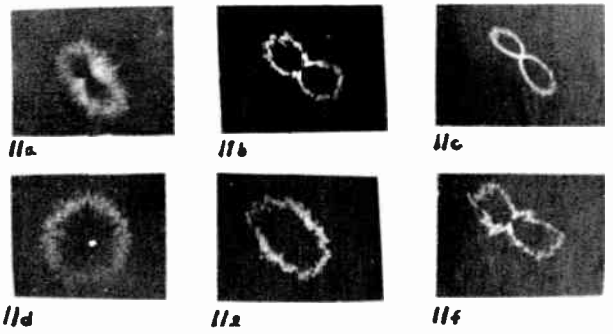


Fig. 11
Experimental results; (a) poor bearing with no enhancement; (b) poor bearing with normal enhancement; (c) poor bearing with enhancement and positive feedback; (d) very poor bearing with no enhancement; (e) very poor bearing with normal enhancement; (f) very poor bearing with enhancement and positive feedback.

A THEORY OF TARGET GLINT OR ANGULAR SCINTILLATION IN RADAR TRACKING

Richard H. DeLano
Hughes Aircraft Company
Culver City, California

Summary

A theory is presented to describe the statistical aspects of tracking a complex isolated structure, such as an aircraft or naval vessel, by radar. The results are expressible in simplest form when the target subtends an angle small compared with the beamwidth. Other situations require special consideration and treatment, but can be attacked by the same general methods. However, when the angle subtended by the target is small, a single description applies to all radar tracking systems. An instantaneous and an effective target displacement from the mean are defined, and their statistical properties derived. Special treatment is given to additional noise arising in conical scanning due to amplitude fluctuations as such. The theory provides information relating to the spectra as well as to the probability densities and rms values of the pertinent quantities. It must be understood that the theory is approximate, is based on a particular model of the target, and leaves the determination of certain critical parameters to experiment in the case of any particular target.

Introduction

The problem of amplitude fluctuations in chaff return, sea return, ground return, ship return, and aircraft return has been the subject of much investigation, effort, and speculation. In many situations the probability density function of the echo amplitudes follows the Rayleigh distribution. In other cases the echo amplitudes follow the probability law of the envelope of a sine wave plus narrow band thermal noise. A summary of much representative work in this field is given in Chapter 6 of Volume 13 of the MIT Radiation Laboratory Series[1] where many additional references can be found. In all such work the Rayleigh distribution of amplitudes is explained by the model of an infinite number of random scatterers with statistically independent amplitudes and phases; in fact, independent phases are enough; the amplitudes can even be constant. The infinity of the number of scatterers is not critical; even 5 or 6 is close to infinity for the purposes of this problem, provided the amplitudes of the scatterers are roughly equal. (See page 554 of Reference 1 regarding this point.)

The assumption of many independent scatterers, or more specifically, of many equivalent point source radiators, to represent the radar target is the basis for the method presented in this paper. Furthermore, it is convenient to consider first the case where all of the equivalent point sources are close to the tracking axis, i.e. where the tracking antenna is not in the process of "resolving" one source from another.

It is assumed in the above model of the target that phase differences are the most important cause of signal fluctuations when observing the target over a small range of aspect angles. The derivations given here are concerned primarily with stationary statistics. Actually, as the target changes aspect by gross amounts, the nature of the target changes also; specifically, the rms amplitudes and the positions of the equivalent point sources change, and as a result the mean radar center, later defined, changes. This change is assumed to be slow and continuous so that in any small angular region the statistics of the target may be regarded as stationary. Multiple reflections are ignored. Polarization is not considered specifically, but is implicitly taken into account in that each polarization contributes separately but differently to the final amplitude and phase of each signal at the output of the antenna.

I. Derivation of Basic Tracking Equation

The amplitude or envelope properties are well known. From N signals of amplitude $a_n(t)$ and RF phase $\theta_n(t)$ measured at the antenna output terminals, the total vector RF signal is

$$V_T = \sum_{n=1}^{N} a_n e^{j\theta_n} = \sum_{n=1}^{N} a_n \cos\theta_n + j \sum_{n=1}^{N} a_n \sin\theta_n$$
$$= \alpha_1 + j\alpha_2 = Ee^{j\theta_r} \qquad (1)$$

As N approaches infinity (assuming all the $\overline{a_n^2}$ roughly equal) $\alpha_1(t)$ and $\alpha_2(t)$ come to have normal or Gaussian distributions and the envelope E(t) has then a Rayleigh probability density. If E(t) is Rayleigh distributed, the standard deviation of E is $\sqrt{(4/\pi) - 1}\ \overline{E} = 0.52\ \overline{E}$ which is the total fluctuation of E about its mean $\overline{E}$, no matter what the spread of this fluctuation in frequency.

In angular tracking, each RF signal $a_n e^{j\theta_n}$ must have an additional part due to the instantaneous tracking error between the antenna tracking axis and the angular position of the equivalent point source radiator. With conical scanning this additional part is an almost sinusoidal amplitude modulation, the amplitude and phase of which modulation are the polar coördinates of the angular error. Thus, for small angular errors the nth signal is

$$a_n\left[1 + b_o(\varepsilon_n)\cos(\omega_s t - \varphi_n')\right]e^{j\theta_n}$$
$$\simeq a_n\left[1 + b_o\varepsilon_n\cos(\omega_s t - \varphi_n')\right]e^{j\theta_n} \qquad (2)$$

$a_n = a_n(t)$, $\varepsilon_n = \varepsilon_n(t)$, $\theta_n = \theta_n(t)$, and $\varphi_n' = \varphi_n'(t)$

where the constant b_o is the fractional modulation per unit of angular error ε and φ_n' is the direction

of angular error. The essential point here is the assumption of linearity of modulation coefficient with angular error. The total RF signal is the sum of all the signals of the type given in Equation 2.

The signal for conical scanning can be considered as broken up into three parts, corresponding to a sum signal and two difference signals. The sum of terms of the form given in Equation 2 can be expressed as

$$V_{TM} = \sum_{n=1}^{N} a_n \left[1 + b_o \varepsilon_n \cos(\omega_s t - \varphi_n')\right] e^{j\theta_n}$$

$$= V_T + \cos \omega_s t \sum_{n=1}^{N} a_n b_o \varepsilon_{1n} e^{j\theta_n} + \sin \omega_s t \sum_{n=1}^{N} a_n b_o \varepsilon_{2n} e^{j\theta_n} \qquad (3)$$

where

$$\varepsilon_{1n} = \varepsilon_n \cos \varphi_n' \qquad (4a)$$

$$\varepsilon_{2n} = \varepsilon_n \sin \varphi_n' . \qquad (4b)$$

The modulations $\sin \omega_s t$ and $\cos \omega_s t$ provide error signals which can be separated out and referred to two perpendicular tracking planes, which can be called plane 1 and plane 2 in agreement with the above notation.

It is convenient to express ε_{1n} and ε_{2n} as the sum of two parts: the first, the deviation of the antenna tracking axis from some fixed point on the target, which is independent of n, and the second, the deviation of the position of the nth equivalent point source radiator from this fixed point. Thus we define

$$\varepsilon_{1n} = -\varepsilon_1 + \delta_{1n} \qquad (5a)$$

$$\varepsilon_{2n} = -\varepsilon_2 + \delta_{2n} \qquad (5b)$$

where

$$\tan \delta_{1n} = \frac{d_n \cos \varphi_n}{R} \simeq \delta_{1n} \qquad (6a)$$

$$\tan \delta_{2n} = \frac{d_n \sin \varphi_n}{R} \simeq \delta_{2n} \qquad (6b)$$

and d_n is the linear distance at range R of the nth equivalent point source radiator from the line through the radar and the fixed point on the target and φ_n is the direction of d_n relative to some reference direction--say, horizontal. Figure 1 illustrates these quantities in a plane perpendicular to the line of sight (LOS). With this substitution Equation 3 becomes

$$V_{TM} = V_T \left[1 - b_o \varepsilon_1 \cos \omega_s t - b_o \varepsilon_2 \sin \omega_s t\right] + \cos \omega_s t \sum_{n=1}^{N} a_n b_o \delta_{1n} e^{j\theta_n} + \sin \omega_s t \sum_{n=1}^{N} a_n b_o \delta_{2n} e^{j\theta_n} \qquad (7)$$

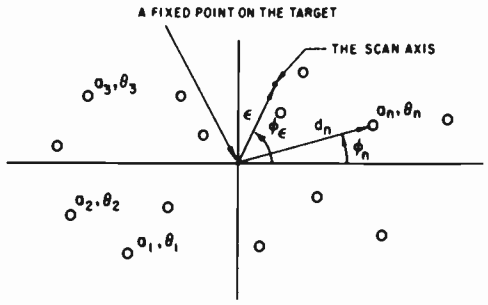


Figure 1. Mathematical Representation of Target in a Plane Perpendicular to Line of Sight

Further discussion will be simplified if the terms with subscript 2 are dropped and only one tracking channel is considered at a time. Under the assumption of small ε_{1n} and ε_{2n}, the interactions between the two channels are negligible. Thus all pertinent information is obtained by considering

$$V_{TM1} = V_T \left[1 - b_o \varepsilon_1 \cos \omega_s t\right] + \cos \omega_s t \sum_{n=1}^{N} a_n b_o \delta_{1n} e^{j\theta_n} \qquad (8)$$

since the subscript 1 is now superfluous, we shall drop it, and the equation can apply to either channel.

Let us now expand the identities of Equation 1 to include the additional modulation terms

$$V_T = \sum_{n=1}^{N} a_n e^{j\theta_n} = E e^{j\theta_r} = \alpha_1 + j\alpha_2 \qquad (9a)$$

$$\sum_{n=1}^{N} a_n b_o \delta_n e^{j\theta_n} = E_s e^{j\theta_s} = \beta_1 + j\beta_2 \qquad (9b)$$

As N approaches infinity α_1 and α_2 become normally or Gaussianly distributed, and so do β_1 and β_2. The asymptotic independence of β_1 and β_2 is easily demonstrated. A particular choice of origin for the δ_n (the fixed point on the target) achieves the independence of the α's and β's, since with normally distributed variables only the mean product has to equal zero to give complete independence

$$\overline{\alpha_1 \beta_1} = b_o \sum_n \sum_m \overline{a_n a_m \delta_m \cos \theta_n \cos \theta_m} = b_o \sum_n \overline{a_n^2} \, \delta_n \, \overline{\cos^2 \theta_n} = \frac{1}{2} b_o \sum_{n=1}^{N} \overline{a_n^2} \, \delta_n = 0 \qquad (10)$$

if the origin is chosen so as to make this mean cross-product zero the asymptotic independence of α_1 and β_1 is established. The condition that $\overline{\alpha_2 \beta_2}$ be zero also results in Equation 10. The statistical independence of the a_n and θ_n was assumed in Equation 10. If the amplitudes a_n are constant with time, the bar over a_n^2 can be removed in Equation 10. With the origin properly chosen, all the quantities in Equations 9a and 9b are asymptotically independent.

With an actual radar target it is well to bear in mind that the point defined by Equation 10 is a slowly varying function of aspect. If this point is a rapidly varying function of aspect, the

theory presented here does not apply.

Linear detection of $V_{TM} = E_T e^{j\theta_T}$ gives the envelope E_T, whereas square-law detection gives the square of the envelope E_T^2. E_T^2 is easily obtained as the dot product or inner product of V_{TM} with itself. This dot product is (assuming ε is small also)

$$E_T^2 = V_{TM}\cdot V_{TM} = \left[\alpha_1(1-b_o\varepsilon\cos\omega_s t) + \beta_1\cos\omega_s t\right]^2 + \left[\alpha_2(1-b_o\varepsilon\cos\omega_s t) + \beta_2\cos\omega_s t\right]^2$$

$$= (\alpha_1^2+\alpha_2^2)(1-2b_o\varepsilon\cos\omega_s t)+2(\alpha_1\beta_1+\alpha_2\beta_2)\cos\omega_s t + \text{terms}$$

$$\simeq E^2(1-2b_o\varepsilon\cos\omega_s t)+2(\alpha_1\beta_1+\alpha_2\beta_2)\cos\omega_s t \qquad (11)$$

and

$$E_T \simeq E(1-b_o\varepsilon\cos\omega_s t) + \frac{\alpha_1\beta_1+\alpha_2\beta_2}{E}\cos\omega_s t$$

$$= E(1-b_o\varepsilon\cos\omega_s t) + \left[\beta_1\cos\theta_r + \beta_2\sin\theta_r\right]\cos\omega_s t$$

$$= E - b_o\varepsilon E\cos\omega_s t + U_a\cos\omega_s t$$

$$= E + \left[U_a - U_b\right]\cos\omega_s t\,. \qquad (12)$$

Equation 12 states the significant result that the total signal component at ω_s (the tracking error signal) is composed of two parts: 1) the useful signal tending to pull the tracking axis back to the geometrical center described by Equation 10 which signal is $-b_o\varepsilon E\cos\omega_s t$, and 2) an independent random error signal $(\beta_1\cos\theta_r + \beta_2\sin\theta_r)\cos\omega_s t$ which can be called the angular scintillation error signal. In addition, the envelope E(t) very often contains components near the frequency ω_s which give error signals of considerable importance. These will be discussed in Section V.

As the number of sources, N, approaches infinity, U_a approaches a Gaussian distribution for the same reason that E becomes Rayleigh distributed. For fixed θ_r the quantity $U_a = \beta_1\cos\theta_r + \beta_2\sin\theta_r$ is Gaussian with zero mean since it is the sum of two Gaussian variables with zero mean. Since the variance of U_a is independent of θ_r,

$$\overline{U_a^2} = \overline{\beta_1^2}\cos^2\theta_r + \overline{\beta_2^2}\sin^2\theta_r = \overline{\beta_1^2} = \overline{\beta_2^2} = \sum_{n=1}^{N}\frac{1}{2}b_o^2\overline{a_n^2}\delta_n^2 \qquad (13)$$

it follows that U_a is Gaussian independent of θ_r as long as β_1 and β_2 are independent of θ_r, which they are. The above expression has been expressed in terms of the properties of the point sources, namely, the mean square amplitudes $\overline{a_n^2}$ and the coordinates δ_n.

If, for example, the $\overline{a_n^2}$ are all equal, the mean square value of U_a is

$$U_a^2 = \frac{1}{2}b_o^2\overline{a_n^2}\sum_{n=1}^{N}\delta_n^2 = \frac{1}{2}b_o^2\overline{Na_n^2}\delta_n^2 = \frac{1}{2}b_o^2\overline{E^2}\delta_n^2$$

$$= \frac{1}{2}b_o^2\overline{E^2}\,\frac{d_n^2}{R^2} \qquad (14)$$

The restoring signal is $U_b = b_o\varepsilon E$. However, on the average only $b_o\varepsilon\bar{E}$ is useful; the remainder, $b_o\varepsilon(E-\bar{E})$ has a mean of zero and is a noise whose rms value is proportional to ε. If ε is small (as assumed), this noise can be neglected, but if a very large steady-state error ε is used experimentally, then this noise will dominate all other sources of error. If U_a is normalized with respect to the mean useful signal per unit angular error, we obtain,

$$\eta_{rms} = \frac{\left[U_a\right]_{rms}}{b_o\bar{E}} = \delta_{rms}\sqrt{\frac{\overline{E^2}}{2\bar{E}^2}} = \sqrt{\frac{2}{\pi}}\,\delta_{rms}\,. \qquad (15)$$

From Equation 11 we can obtain the corresponding ratio when square-law detection is used

$$\eta'_{rms} = \frac{\left[EE_s\cos(\theta_r-\theta_s)\right]_{rms}}{\overline{E^2}} = \frac{1}{\sqrt{2}}\,\delta_{rms}\,. \qquad (16)$$

Square-law detection is observed to be slightly superior as regards its effect on the rms noise angle due to angular scintillation.

Illustrative of Equation 15 is the case where all the radiators are concentrated into two equal groups separated by a length L at range R. This case maximizes η_{rms} for a given L, and the result is obviously

$$\eta_{rms} = \frac{L}{\sqrt{2\pi}\,R}\,. \qquad (16a)$$

If the sources are uniformly spaced along the length L, we have

$$\eta_{rms} = \frac{L}{\sqrt{6\pi}\,R} \qquad (16b)$$

or, if they are concentrated into four equal groups uniformly spaced along L

$$\eta_{rms} = \sqrt{\frac{5}{18\pi}}\,\frac{L}{R}\,. \qquad (16c)$$

Although common sense and speculation can be used to estimate both L and the distribution of sources a priori, experimental measurement is still required in any specific case.

Although these results have been derived specifically for conical scanning, they are equally applicable to any system which develops any kind of modulation on the signal from each source which is linear with angular error and which is in phase with the unmodulated signal from that source (or effectively in phase when used by the system). Essentially all linear modulation systems fall in this category.

II. Properties of Apparent Radar Center

The next important results to be derived from Equation 12 which gives the total angular error signal in one tracking channel are the properties of the "apparent radar center" whose definition is as follows. The apparent radar center is the position ε of the antenna axis for which the error signal is instantaneously zero; let this value of ε be designated as ε_o. This definition ignores noise

due to components of E(t) near ω_s, which is treated later. Under the assumptions which make Equation 12 valid, ε_o is single valued, or, in other words, ε_o is single valued for a target small compared to a beamwidth and for ε_o small compared to a beamwidth. From this definition and from Equation 12, assuming linear envelope detection, ε_o is given by

$$\varepsilon_o = \frac{U_a}{b_o E} = \frac{\beta_1 \cos\theta_r + \beta_2 \sin\theta_r}{b_o E} = \frac{E_s \cos(\theta_r - \theta_s)}{b_o E} \cdot \quad (17)$$

Since we are at present confining our attention to a single tracking plane, U_a is Gaussianly distributed with zero mean and E is Rayleigh distributed. Since U_a and E are independent it is relatively easy to derive the probability density of their ratio ε_o. The probability distribution of ε_o under these circumstances is a special case of the Student's t distribution in statistics with 2 degrees of freedom. The variable, t, in the Student's t distribution is the normalized ratio of sample mean to sample standard deviation when n independent samples of a Gaussianly distributed random variable are observed.[2] This interesting, but somewhat unrelated fact, provides us with the probability density of ε_o. However, the following results are also derived in Appendix A which considers the general problem of the probability density of the ratio of two independent random variables. In a single tracking channel this probability density is

$$P(\varepsilon_o) = \frac{b_o E_{rms}}{\sqrt{8}\, U_{a\ rms}} \left(1 + \frac{1}{2} \frac{b_o^2 \overline{E^2}}{\overline{U_a^2}} \varepsilon_o^2\right)^{-3/2} \cdot \quad (18)$$

For the second example previously cited—radiators a_n uniformly spaced along a length L (perpendicular to the line of sight from the radar)—this probability density reduces to

$$P(\varepsilon_o) = \sqrt{3}\, \frac{R}{L} \left(1 + \frac{12\, R^2}{L^2} \varepsilon_o^2\right)^{-3/2} \cdot \quad (19)$$

One interesting property of this distribution is the fraction of the time, ε_o, points off the radar target in the tracking plane, i.e. lies outside the angular region of L/R radians, which, in this case, turns out to be

$$2 \int_{1/2}^{\infty} P(\varepsilon_o) d\varepsilon_o = \int_{\sqrt{3}}^{\infty} (1+x^2)^{-3/2} dx = 1 - \sqrt{\frac{3}{4}} = 0.134 \quad (20)$$

or 13.4 percent of the time. Another property of some interest is the second moment, or mean square of ε_o. Since ε_o is <u>not</u> Gaussianly distributed as is U_a for example, the theoretical distribution of ε_o gives infinite mean square or mean fourth power, etc. for ε_o, i.e.

$$\int_{-\infty}^{\infty} \varepsilon_o^2 P(\varepsilon_o) d\varepsilon_o = \frac{1}{2} \int_{-\infty}^{\infty} \frac{x^2\, dx}{(1+x^2)^{3/2}} = \infty \cdot \quad (21)$$

This result violates the initial assumption that ε_o is always small compared to a beamwidth. However, it points out the fact that dividing U_a by E causes the large values of ε_o to be much more probable than if ε_o were Gaussianly distributed.

To the extent that a very rapid AGC divides the total signal of Equation 12 by E(t), these remarks apply also to such an AGC. If the scan modulation components are very small, and if the bandwidth of E(t) is low enough, a rapid AGC acts in approximately this manner.

In considering both tracking channels simultaneously a complete description of the two dimensional statistics of U_a and ε_o will not be presented here. However, it is useful to consider under what circumstances the U_a's are statistically independent in the two perpendicular tracking channels. Examination of Equation 7 shows the two complex vectors which need to be independent. These are

$$\sum_{n=1}^{N} a_n \delta_{1n} e^{j\theta_n} = \beta_1 + j\beta_2 \text{ and } \sum_{n=1}^{N} a_n \delta_{2n} e^{j\theta_n} = \gamma_1 + j\gamma_2 .$$

For simplicity let us assume that all the δ's are small so that Equations 6a and 6b apply. Then it is easy to show that independence of <u>all</u> the components of the above two complex vectors is assured if a single summation vanishes, namely

$$\overline{\beta_1 \gamma_1} = \overline{\beta_2 \gamma_2} = \frac{1}{2} \sum_{n=1}^{N} \overline{a_n^2} \frac{d_n^2 \cos\varphi_n \sin\varphi_n}{R} = 0 \cdot \quad (22)$$

If Equation 22 holds, the error signals in the two channels are statistically independent, and are Gaussian in the case of U_a. Equation 22 holds whenever there exists any one axis through the target plane (or a great circle through the target) with respect to which the equivalent point source radiators are symmetrically located both in angular or linear, distance and in amplitude a_n, assuming always that one of the tracking channel directions coincides with this axis. Equation 22 also implies that if the pattern of point sources is identically symmetrical around any two perpendicular axes, the error signals U_{a1} and U_{a2} are independent with equal standard deviations no matter what the orientation of tracking axes.

As an example of this latter double symmetry, consider equal point sources uniformly spaced over a circular area of radius ρ. This double symmetry allows easy extension of the results for tracking in a single plane to three-dimensional tracking. In the case of the apparent radar center, the probability density is

$$P(\varepsilon_o') = \frac{8R^2}{\rho^2} \frac{\varepsilon_o'}{\left[1 + \frac{4R^2}{\rho^2} \varepsilon_o'^2\right]^2} \quad (23)$$

This distribution also predicts an infinite value for $\varepsilon_o'^2$, as it must. The total probability that ε_o' exceeds ρ/R, i.e., lies outside of the circular target is

$$\int_{\rho/R}^{\infty} P(\varepsilon_o') d\varepsilon_o' = 2 \int_{2}^{\infty} \frac{x\, dx}{(1+x^2)^2} = \frac{1}{5} \cdot \quad (24)$$

III. Statistical Nature of Tracking Transfer Function

Returning to consideration of tracking in a single plane, we can see from Equation 12 that the restoring error signal is proportional to E(t) as well as to ε. It must inevitably follow, unless the AGC is fast, that the antenna tracking equations relating antenna tracking axis to the radar center of the target are not constant coefficient linear differential equations whose properties are easily treated by the use of complex transfer functions Y(ω) or Y(p). The differential equations involved will have coefficients which contain E(t), a statistical variable, in addition to the forcing function U_a, also statistical. The mathematical treatment of such equations is not at present in a state of razor-sharp precision.

Physical reasoning reveals some common situations where the variability of the useful signal due to E(t) will have practically no effect. If the spectrum of E is sufficiently broad, i.e., if E fluctuates up and down fast enough relative to the time constant or averaging time of the tracking servo, the total applied error signal after such averaging will differ only slightly from the ensemble average $-b_o\varepsilon\bar{E}$. In such cases as far as the tracking equation is concerned $-b_o\varepsilon E$ can be replaced by $-b_o\varepsilon\bar{E}$ and constant coefficient equations result. Certain solutions of tracking performance obtained on analog computers at Hughes Aircraft Company have shown that a ratio of, say, 10 to 1 in scintillation to tracking bandwidth is quite adequate for the above substitution of $\bar{E}$ for E. When this substitution is made, the angular error signal is just as though the target were a point source fluctuating in angular position by an amount η_{rms} as given in Equations 16.

IV. Spectra of E and U_a

The spectrum of narrow band noise which has been detected with a linear or square-law detector is easily calculated for a square-law detector and somewhat less easily calculated for a linear detector.[3] Precisely the same situation applies to the random signals which produce E and U_a. Since it has been shown[3] that the shapes of the detected spectra are only very slightly different for square-law and linear detection, we shall consider here only the square-law case. Equation 11 shows the squared envelope to be

$$E^2 = \alpha_1^2 + \alpha_2^2 \tag{25}$$

and the error signal corresponding to U_a to be proportional to $\alpha_1\beta_1 + \alpha_2\beta_2$. The question of the magnitudes of the spectra is not important here, because the spectra can be normalized later, using the well-known relation that the total integral of the power spectrum of any quantity over all frequency is equal to the variance or square of the standard deviation of that quantity, which is known for both E and U_a.

The shape of the spectrum of E^2 (hence of E) can be determined from the spectra of α_1 and α_2. Since α_1 and α_2 have identical spectra, either will do. The power spectrum of α_1^2 is just the convolution of the power spectrum of α_1 on itself.[4] Although the spectrum of E(t) can be measured, there are many simple situations in which it can be calculated as well. For example, if a line of uniformly spaced randomly phased point sources rotates at a constant angular velocity with respect to the line of sight, the spectrum of $\alpha_1(t)$ is rectangular and the spectrum of α_1^2 is triangular, maximum at zero frequency and dropping to zero spectral density at the highest beat frequency, namely that between the two outside point sources.

In like manner the spectrum of $\alpha_1\beta_1 + \alpha_2\beta_2$ can be obtained by taking the convolution of the spectrum of α_1 on the spectrum of β_1. In cases where there is no dependence of the rate of change of phase $\dot{\theta}_n$ on position δ_n or d_n/R, the spectra of β_1 and of α_1 are usually identical. In other cases the spectrum of β_1 is different, but its calculation involves no real problem if the spectrum of α_1 can be determined. In no case will the bandwidth of the amplitude fluctuations E(t) differ significantly from the bandwidth of U_a.

V. Additional Noise Due to Amplitude Fluctuations in E(t)

In addition to signal components at frequency ω_s of amplitude U_a and $-U_b$, the actual envelope E(t) may have noise components close to ω_s. These components are interpreted by the receiver as real signals and, unlike U_a, have no definite relation to the size or geometry of the collection of point sources which we call the target. Any mechanism at all giving rise to fluctuations at this frequency, or close to it, produces an additional noise term. Problems of this sort are considered in Section 6.10 of Vol. 25, MIT Radiation Lab. Series.[5] However, this source[5] states that the rms fluctuations in amplitude about the mean are found in practice to be about 1/4 of the mean, whereas with the Rayleigh distribution, the amount predicted is .52 times the mean.

If the spectral shape for E(t) is known or can be determined, the actual magnitude of the spectral density can be determined through the identity (basic to the definition of power spectrum)

$$\overline{E^2} - \bar{E}^2 = \left(\frac{4}{\pi} - 1\right)\bar{E}^2 = \int_0^\infty \Phi(\omega)\,d\omega \tag{26}$$

where $\Phi(\omega)$ is the power spectrum of E(t), or of $E(t) - \bar{E}$ and E is assumed Rayleigh distributed. The corresponding spectral density after phase detection, i.e., after the error signals have been extracted from the signal E_T of Equation 12, expressed as an equivalent angle, is

$$\Phi_F(\omega) = \frac{2\Phi(\omega_s)}{b_o^2\bar{E}^2} \qquad \text{for } \omega \cong 0 \tag{27}$$

where the factor of 2 arises from the fact that both frequencies above and below ω_s contribute to the output noise at the difference frequency. This spectral density is the same in both tracking channels. The magnitude of this noise in angular units

is independent of range to the target unless the spectrum of E(t) is range-dependent in some way.

VI. Short-Range Theory

When the target is large or the beamwidth is narrow or the range is sufficiently small, it may happen that one of the basic assumptions of the preceding theory is no longer satisfied. This assumption is that the angular error, $\delta_n - \varepsilon$, of each point source radiator is sufficiently small that its individual error signal, or scan-modulation component, is small. When this condition is not satisfied by all the sources simultaneously for all ε's to be expected, the simplicity of the preceding theory is not possible. It is then no longer possible to divide the error signal into two statistically independent parts, one of them proportional to both ε and the signal envelope. In fact, it is no longer possible to consider the two tracking channels separately, as they interact. The error signals in the two channels, and their vector sum, are functions of the orientation of the two tracking axes with respect to the target. The apparent radar center is no longer uniquely defined, and the rms error signal depends on ε which, however, was also true for long-range tracking, but in a different manner. It is possible to derive equations for the mean and rms error signals for fixed ε, but they will not be presented here. Calculations from these equations show how the mean error signal slope decreases rapidly as the target comes to subtend two or three beamwidths.

VII. Finite Number of Sources

The theory presented here has given results for a target whose equivalent point sources are infinite in number. From results in the literature on E, the absolute magnitude of N randomly phased sinusoidal components[1,6] it seems well-established that five or six roughly equal sources is remarkably close to infinity as far as the resulting probability density of E is concerned. An unpublished paper by A. Vazsonyi of Hughes Aircraft Company has shown that the rms value of η, as defined by Equation 15, never exceeds the value given in Equation 15 when the point sources are finite in number and of equal intensity. Even when N is as small as four, the above paper shows the η_{rms} is somewhat greater than 0.9 times the value given by Equation 15. Thus the infinite source theory is slightly pessimistic in all cases but gives answers very close to the more rigorous answer from a finite number of reflectors even so.

Consider briefly a two-source target. For two fixed amplitude signals it is easy to determine one particular property of the apparent radar center, ε_o, and that is the phase and amplitude ratio required to put ε_o outside the two sources. Imagine the scan axis pointed at the larger source, so that its scan modulation is zero. If the resultant of the two complex signal vectors is at an angle of more than 90° with respect to the small vector, its scan modulation will produce a modulation on the resultant out of phase with itself. Such a modulation tells the antenna to move away from the smaller source, i.e., outside the target. From simple geometrical considerations it is apparent in Figure 2 that a smaller signal, added to the larger, whose phase and amplitude are such as to put the resultant completely in the circle fulfills the required qualifications.

When calculations are carried out on the two-radiator case, it is found that for essentially all plausible situations the rms fluctuations in effective radar center, suitably defined in this case, are smaller than those obtained for an infinite number of reflectors. In case the two amplitudes are independent statistical variables, Rayleigh distributed, the results immediately become identical to the infinite radiator case with each source taking the place of a large number of sources. This result seems reasonable at once, since each source can be represented by an infinite number of sources all at the same location, the infinite number of sources producing the Rayleigh distribution of amplitudes.

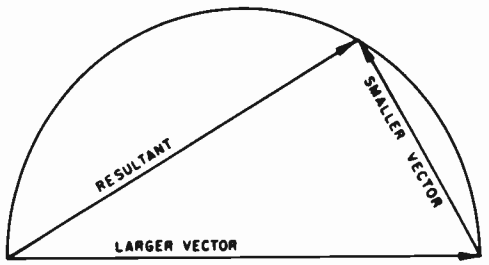


Figure 2. Vector Showing Smaller Vector more than 90° from Resultant

Appendix A. Probability Densities of a Ratio of Random Variables

Consideration of the apparent radar center ε_o, which makes the total error signal instantaneously zero, requires that the probability density of the ratio of two random variables be determined. If we define random variables x, y, and u related according to $u = x/y$ with probability densities $P_1(x)$, $P_2(y)$, and $P_3(u)$, with x and y independent, the probability densities must be related according to the following equation. Assume here for simplicity that x and y assume only positive values, as this assumption is no limitation on the generality of the results but implies that other cases must be treated separately and combined after $P_3(u)$ is determined for each portion of the problem.

$$\int_0^u P_3(u)du = \int_0^\infty P_2(y)dy \int_0^{uy} P_1(x)dx \,. \qquad (28)$$

The integral from 0 to uy of the quantity $P_1(x)dx$ is clearly the total probability that x is small enough so that, for fixed y, $x/y \le u$, and the total equation gives this probability averaged over all values of y. By differentiation of both sides of this equation, we obtain

$$P_3(u) = \int_0^\infty P_2(y)dy \frac{d}{du}\int_0^{uy} P_1(x)dx = \int_0^\infty yP_1(uy)P_2(y)dy. \qquad (29)$$

A useful example is when x is Gaussian and y is Rayleigh distributed, i.e.,

$$P_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x^2/2\sigma^2)} \tag{30a}$$

$$P_2(y) = \frac{2y}{\mu^2} e^{-(y^2/\mu^2)} \qquad y \geq 0 \tag{30b}$$

then

$$P_3(u) = \int_0^\infty \frac{y}{\sqrt{2\pi}\,\sigma} e^{-(u^2y^2/2\sigma^2)} \frac{2y}{\mu^2} e^{-(y^2/\mu^2)} dy$$

$$= \frac{\mu}{\sqrt{8}\,\sigma} \left[1 + u^2 \frac{\mu^2}{2\sigma^2}\right]^{-3/2} \tag{31}$$

due to symmetry considerations, this expression applies to negative u. If both x and y have Rayleigh distributions with mean square values μ_1^2, and μ_2^2, respectively, $P_3(u)$ is

$$P_3(u) = \int_0^\infty \frac{4y^3u}{\mu_1^2\mu_2^2} e^{-(u^2y^2/\mu_1^2)} e^{-(y^2/\mu_2^2)} dy$$

$$= \frac{u\mu_2^2}{\mu_1^2} \left[1 + u^2 \frac{\mu_2^2}{\mu_1^2}\right]^{-2} .$$

References

1. D. E. Kerr, Editor, "Propagation of Short Radio Waves." MIT Radiation Laboratory Series, Vol. 13, Chap. 6. McGraw-Hill; 1951.

2. Paul G. Hoel, "Introduction to Mathematical Statistics." Wiley; 1947.

3. Lawson and Uhlenbeck, "Threshold Signals." MIT Radiation Laboratory Series, Vol. 24, Section 3.8. McGraw-Hill; 1950.

4. S. O. Rice, "Mathematical Analysis of Random Noise." Bell System Technical Journal, Vol. 24, Section 4.5; January 1945.

5. H. M. James, N. B. Nichols, and R. S. Phillips, "Theory of Servomechanisms." MIT Radiation Laboratory Series, Vol. 25. McGraw-Hill; 1947.

6. W. R. Bennett, "Distribution of the Sum of Randomly Phased Components." Quarterly of Applied Mathematics, Vol. 5, pages 385-393; January 1948.

AUTOMATIC DEAD RECKONING NAVIGATION COMPUTERS FOR AIRCRAFT

James L. Dennis
Wright Air Development Center
Dayton, Ohio

High-speed jet aircraft with fewer crew members have made critical the need for increased automaticity in as many navigation and flight control functions as possible. Very little time is available to the busy pilot-navigator of a jet fighter to compute manually dead reckoning position and make plots of position and course on flight charts. Dead reckoning procedure, as applied to air navigation, is a method for establishing the best estimate of aircraft position based on computation of the travel from a known position. Advancing the aircraft position from the known position to a later position is done by computing the travel along a ground track based on the ground speed and time of travel. Rather than require the pilot or navigator to solve this problem by the usual method of using certain manual analog computers, such as the E-6B, and by plotting on flight charts, Air Force contractors have developed computers which provide some automaticity to the solution of the problem.

In dead reckoning air navigation, the navigator customarily deals with the following terms: See figure 1.

Compass Heading (Hc) - The indicated heading of the aircraft with respect to the apparent magnetic pole, uncorrected for local aircraft distortion.

Magnetic Deviation (De) - Distortion of the earth's magnetic field by the aircraft structure. Varies with aircraft orientation.

Magnetic Heading (Hm) - The indicated direction of the aircraft heading with respect to the earth's apparent magnetic pole corrected for distortion of the magnetic field by aircraft structure.

Magnetic Variation (Var) - The angular difference between the apparent direction of the magnetic pole and true geographic north.

Wind Force (Vw) - Expressed in knots.

Wind Direction (Hw) - Expressed as a true direction from which the wind is blowing. Actually, in the A-1 Computer to be described, the wind direction is set in as defined herein but internally Hw is shifted 180 degrees. In the other computer to be described, the AN/APA-58, wind is set in or indicated in north-south and east-west components and represents the true components of wind velocity in those coördinates.

True Air Speed (Vt) - Early air speed indicators gave readings which had to be corrected for temperature and altitude. Indicators are now available which make this correction internally, hence, are called True Air Speed Indicators.

Ground Speed (Vg) - The actual speed of the aircraft over the ground.

Drift Angle (D) - The angle between the true heading of the aircraft and the ground track.

Ground Track Angle (Hg) - The angle with respect to true north which represents the actual direction of travel of the aircraft over the earth.

The first computer to be described is designated as the Ground Position Indicator Type A-1, developed under contract with Ford Instrument Co. This computer indicates present position in latitude and longitude from continuous automatic inputs of true air speed and magnetic heading, and from manually set values of wind direction, wind force, and magnetic variation. The latitude and longitude data can be reset at any time a dependable visual, radar, or celestial fix is obtained. The Type A-1 GPI consists of four units. Figure 2 shows the Indicator Unit. Latitude and longitude counters can be reset by means of the slew switches. The indicator fits in the standard mounting hole for a 3-inch indicator in an aircraft instrument panel and is 8 inches deep. Figure 3 shows the Preset Control Box which has the following dimensions: 5-3/4" x 3-3/8" x 5-3/8". Figure 4 is the hermetically sealed computer unit. It contains no vacuum tubes and is designed with the objective of providing a unit which will give many hours service without a requirement for maintenance. Dimensions are 8-3/4" x 6-13/32" x 9-13/32". Figures 5, 6, 7 and 8, are views of the Amplifier Unit, internal chassis layout and subassembly design. The size of this unit is 8-13/16" x 19-1/4" x 5-7/8". Total weight of the computing system is 46 pounds.

In the A-1 computing system those inputs which are expected to be rapidly varying functions are automatic, whereas those inputs which are slowly varying to an extent that only occasional reset is necessary, are manually

introduced. That is, magnetic heading (Hm) and true air speed (Vt) are automatic inputs, and magnetic variation (Var), wind force (Vw) and wind direction (Hw) are manually set into the Preset Control Box. The problem to be solved is as follows:

(1) Convert magnetic heading to true heading.

(2) Add vectorially the air velocity to the wind velocity.

(3) Resolve the resultant vector into longitude and latitude rates.

In figure 9 is shown a functional diagram of the compass heading circuit. The indication of the compass Hm is transmitted by a synchro from the compass to a synchro control transformer.

Provision is made to correct the magnetic heading Hm for transmission error TE. The voltage corresponding to Hm is followed up by the synchro control transformer and the servo motor through the servo amplifier. The motor positions the transmission error compensator to an angle corresponding to the magnetic heading Hm + TE. The output of the cam follower is proportional to TE, the transmission error for an individual installation of the equipment. This correction is added in the differential mechanically to the output of the servo motor. The sum of the two is fed back to the control transformer. The entire system is in balance and the motor stops turning when the differential output equals Hm and the servo output equals Hm + TE.

Figure 9 also shows how magnetic variation is introduced. This is the other quantity necessary to complete the computation of true heading Ht. The variation control knob positions the rotor of a synchro differential in the Preset Control Box. In this mode of operation a zero set network feeds the differential synchro, and the output is received by a control transformer in the computer section. The variation signal is followed up by the control transformer, amplifier and servo motor. The output of the servo motor is fed to one side of the mechanical differential. The other input to the differential is the quantity Hm + TE, obtained from the servo loop just described. The output of this differential is therefore the true heading. Ht = Hm + TE + Var. Ht is used to position the rotor of an electric resolver which provides an input of E sin Ht to the longitude integrator and E cos Ht to the latitude integrator.

In figure 10 is shown how the second of three voltage functions which are the inputs to the integrators are derived. The rotor of the wind resolver is positioned in accordance with Hw. Outputs are the usual sine and cosine functions. These are then divided by Vt in one set of potentiometers, the arm being positioned by the Vt servo. The resulting voltages are multiplied by Vw in a second set of potentiometers, where the arm is positioned by manual set of wind force, in order to obtain the functions $\frac{Vw}{Vt}$ E sin Hw and $\frac{Vw}{Vt}$ E cos Hw.

Now we have as inputs through summing networks to the latitude integrator the following voltages:

$$E \cos Ht + E\, Vw/Vt \cos Hw .$$

It should be noted that this is the north-south component of ground speed, divided by Vt. Clearly, if one multiplies the above expression by Vt, it becomes

$$E\, Vt \cos Ht + E\, Vw \cos Hw = E\, Vg \cos Hg .$$

However, the purpose of the division by Vt is to yield dimensionless quantities and to produce a variable scale of computations which provides high accuracy at low values of Vt. The effect of this division is subsequently cancelled in the course of integration through multiplication by Vt.

The quantity E cos Ht + E Vw/Vt cos Hw is transformed from voltages to equivalent shaft rotations by a servomechanism shown in figure 11. The servo motor output is transformed into a voltage by the response potentiometer and this voltage is fed back to the summing network, thus completing the servo loop. Instead of being fed directly from the aircraft's power supply, the response potentiometer for the latitude integrator servo loop is excited by a resolver with a locked rotor. This resolver produces a response waveform identical to the waveforms of the outputs from the computing resolvers to yield optimum servo error discrimination in the summing network and in the amplifier.

Also in figure 11 is shown how the output of the integrator servo is multiplied by Vt in order to cancel the earlier division by Vt. This is done by driving the disc of the mechanical ball and disc integrator by the function (Vt. t).

In the A-1 GPI, the mechanical integrator carriage containing the ball is positioned by the velocity component. Since the disc is driven as a function of time, the integrator outputs from the roller are the time integrals of the velocity components.

One other problem exists. Since the linear distance between meridians of longitude decreases as the cosine of latitude, the change is longitude computed from the longitudinal ground velocity, Vg sin Hg, multiplied by the secant of latitude, sec La. Sec La is computed

by inverting cos La. This inversion is produced by utilizing the inverting effect of the feedback loop in a servomechanism.

The output from the ball and disc integrator is therefore longitude rate. That is, the roller will turn through an angle proportional to change in longitude. The same result is obtained in the latitude channel.

The second navigation computer to be described is the Ground Position Computer AN/APA-58(XA-1) developed under Air Force contract with Eclipse-Pioneer Division of Bendix Aviation. This computer provides essentially the same functions as the Type A-1 GPI and additionally provides several other functions. It is designed to be used with an airborne search radar such as the AN/APS-10 (see figure 12). When used in this manner the combination provides a radar position-fixing system and a method of wind determination by memory point tracking of radar reference points. The Computer Unit has provision for computation of Direction to Destination, Course being made good, and True Heading for the aircraft to fly to arrive at the selected destination. Time to go until arrival at the referenced destination is also computed. In one mode of operation a Pilot's Direction Indicator is activated.

The computer unit is a compact assembly 11" x 16" x 19" and weighs approximately 48 pounds. It is shown in figure 13, along with the True Airspeed Meter, one of the inputs to the computer.

Of special note is the mechanical design of this computer. It consists of several functional subassemblies, each being removable from the computer unit frame. Figure 14 shows the True Airspeed and Direction Subassembly removed from the main Computer Unit. In addition to the six removable subassemblies in the front of the unit, there are several removable from the back. Since these subassemblies merely plug in, they can be easily removed and a patch cable can be inserted to facilitate servicing.

Figures 14, 15, 16 and 20 show some of the details of the construction. A metal card is used for mounting of tubes and parts because of its heat conduction characteristics. An added advantage is its shielding effect. Glass bead inserts with feed-through terminals are used extensively.

This computer uses 85 vacuum tubes, 75 of which are subminiature types. It is not considered as conclusive evidence of the superiority of subminiature tubes, but it is encouraging to note that one of these computers has had almost 1000 hours operation with only one tube failure (a miniature type).

Almost all of the transformers in this computer are of special design. They are designed to run at high temperatures and are sealed in aluminum cans filled with inert gas. No difficulty has been experienced with these transformers.

Referring again to figure 13, in the lower center of the panel is a servoed indication of true airspeed. In the upper left-hand corner is a servoed indication of compass heading from the gyrosyn compass input. When the adjacent "Variation" dial is manually set for correct local magnetic variation, the case (and stator winding) of the Synchro from which the "Compass" indication is driven is rotated so that the "Compass" indication becomes true heading.

Plane Position is manually set at the start of the navigation problem and at anytime thereafter that a more reliable position determination can be obtained by any means. In order to do radar position correction, it is first necessary to set the identification counters to the latitude and longitude of a known radar reference point within the range of the radar, but not more than 100 miles away. Also, essential to correct measurement of position is the setting of aircraft altitude on the counters in the lower left of the computer panel. When these things are done, the range circle and azimuth mark, usually referred to as cross hairs, appearing on the radar indicator should intersect on the radar reference point if the Plane Position counters are correct. If the cross hairs do not lie on the proper point, the Plane Position counters can be manually turned until the cross hairs are moved to coincide with the reference point.

Once set correctly, the Plane Position counters will continue to indicate the position of the aircraft except for errors in the input data. The major source of error is the knowledge of wind since it is continually changing in an unpredictable manner. If north-south and east-west wind components were set in initially based on meteorological predictions, it may be that the error due to this cause is not large. In any case, if at any time a reliable setting of Plane Position is made and at any later time as a result of a radar fix it is noted that an error in position has developed, a position correction should be made followed by a Wind Tracking operation. This is done by turning the Tracking clock to 10 minutes, observing the drift of the cross hairs from the radar reference point and repositioning the cross hairs at a suitable time within the ten minutes while the radar target is still in good view. The correct wind (assuming all error is due to wind error) will be computed, indicated and fed to the integrators.

The Function Switch has three operating positions, "Navigate", "Course" and "P.D.I.". All of the operations so far described are done with the switch in the "Navigate" position. When turned to "Course", functions having to do with the Identification point are inoperative and

an azimuth mark signal (representing the ground track angle of the aircraft) is generated for the radar indicator. In the "P.D.I." position , no markers are fed to the radar. Instead, the heading to fly to reach a destination set on the "Identification" counters is computed and displayed on the dial in the lower center of the computed panel. Also, an error signal is generated and fed to the vertical needle of the Indicator ID-48/ARN. The "Time-To-Go" Indicator is actuated during the "P.D.I." function.

Figure 17 is a simplified functional diagram showing the computational arrangement when the Function switch is in the "Navigate" position. The true air speed input is obtained from a special miniature synchro take-off in the True Air Speed Meter ID-205(XA)/APA-58. A position servo operating from the true air speed synchro drives a potentiometer across which a 400-cycle scale voltage appears. The output from the potentiometer is proportional to air speed and is fed through a driver amplifier to the rotor of a synchro resolver.

The other input to the airspeed resolver is true heading. This is obtained in the following manner. A transmitter synchro is located in the magnetic compass. A follow-up synchro in the Compass Subassembly of the Computer Unit generates an error signal when not aligned with the take-off synchro and this signal is fed through a servo amplifier to a low-inertia two-phase motor and a gear reduction box to rotate the follow-up synchro to a null position. Magnetic variation is introduced by rotating the case of the follow-up synchro. The output of this servo system then is true heading and is mechanically coupled to the air-speed resolver. The outputs of the airspeed resolver are two voltages, north-south component of air speed and east-west component of air speed. These voltages are inputs to the latitude and longitude integrator circuits, respectively.

In series with each of the output windings of the air speed resolver is a voltage derived from the corresponding wind potentiometer. For example, the latitude integration circuit has as one input a voltage which is the sum of the north-south component of air speed and the north-south component of wind. Therefore, if the wind setting is correct in the north-south channel, one input to the latitude channel is north-south ground speed. By use of a unique rate servo, which will be described later, the latitude counters are continuously driven to maintain correct indication of latitude within the accuracy of the various inputs. A similar integration takes place in the longitude channel, with the exception that the east-west rate must be modified by the secant of latitude.

Since one of the objectives of this computer design is to allow radar position fixing, it is necessary to generate a range mark and an azimuth mark based on the range and azimuth of some identification point relative to the indicated position of the aircraft. This is accomplished by setting the Identification counters to the latitude and longitude of the reference point. A differential between Plane Position and Identification counters drives a range potentiometer in each of the channels. It should be noted that the computation of range in the longitude channel is a cosine function of latitude, i.e., the higher the latitude the lower the value of range represented by a given number of degrees longitude. The range resolver circuit provides as an output the ground range to the Identification point. In order for this to be usable for presentation on the radar indicator, it must be converted to slant range. This is done in a triangle solver into which is inserted the altitude of the aircraft. The altitude setting on the computer can be based on other sources of altitude data, but may be made by switching to a range mark signal which is a function of altitude only. Then when it is adjusted so that the range mark coincides with the ground return circle on the radar indicator, the correct altitude has been reached.

The Range Mark Generator accepts a synchronizing signal from the radar in order to generate a range mark which will appear at the proper point in range on the radar indicator.

The other signal supplied to the radar set is the azimuth mark representing the direction to the Identification point. The condition necessary for generation of the azimuth mark is that the true heading plus the angle of the radar antenna with respect to the true heading must equal the angle with respect to true north of the Identification point.

When operating in the "Navigate" mode, it is possible to determine the correct wind by memory-point tracking of a radar target. In order to determine wind by this method, the cross hairs are set on a suitable radar reference point. Any clearly definable radar echo may be used, even though its position may not be known. The 10-minute mechanical timing unit is turned on. If the correct wind has been set into the Wind Set dials, then the cross hairs should move synchronously with the radar reference point. If the wind values are wrong, the cross hairs will drift off. At any time during the tracking operation the tracking control or "Joy Stick" can be used to control the positioning of the cross hairs and thereby correct the wind setting automatically.

The wind tracker is illustrated with reference to one coördinate in figure 18. A constant speed timing unit drives the Correction Potentiometer so that the zero correction voltage is available at the start of the period and the maximum at the end. The "Joy Stick" is not operated until it is observed that the cross

hairs have departed from the radar reference point. At this time the "Joy Stick" is operated, clutching in the constant speed drive to the Wind potentiometer, and connecting the correction voltages to the input of the integrator rate servo. The position counters are thereby driven at a higher or lower rate than normally, thus causing the range to the identification point to be corrected. The scale factors chosen are such that the proper change in the setting of the wind potentiometer takes place in the same correction time.

Figure 19 is a block diagram of the integrator circuit. This circuit has certain novel features. In order to avoid the problem of passing through zero speed and the difficulty of obtaining uniform speed characteristics in both directions of rotation, it was decided to design a circuit in which zero is, in effect, 4000 rpm. An induction motor running at 12,000 rpm geared to a reference generator running at 4000 rpm develops a reference voltage. This voltage is fed to the input of the integrator servo amplifier. The servo amplifier drives a variable speed motor coupled to a generator from which a voltage is fed back to the input of the integrator servo to balance the reference voltage. The integrator generator is mechanically coupled to a differential synchro which receives a signal from a transmitter synchro coupled to the reference generator. The operating speed of the transmitter synchro is 48 rpm and the differential synchro varies from 12 to 84 rpm as a function of ground speed in each coördinate. The output from the differential synchro goes to a receiver synchro, servo amplifier and servo motor which mechanically drives through a gear train the receiver synchro over the speed range of 0 ± 36 rpm. The aircraft position counters are also driven from the same servo motor. Reference is made to figure 18 which shows a picture of the Integrator Subassembly.

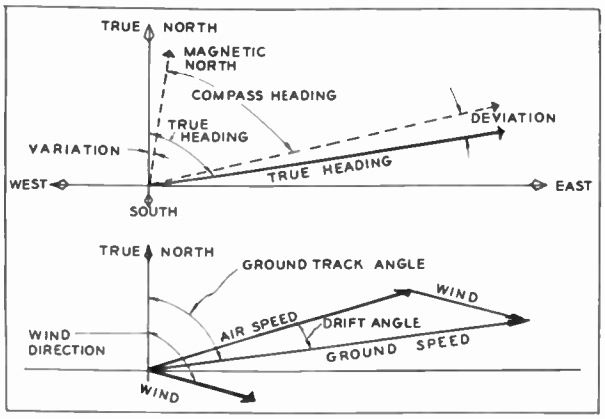


Fig. 1
Illustration of heading and velocity.

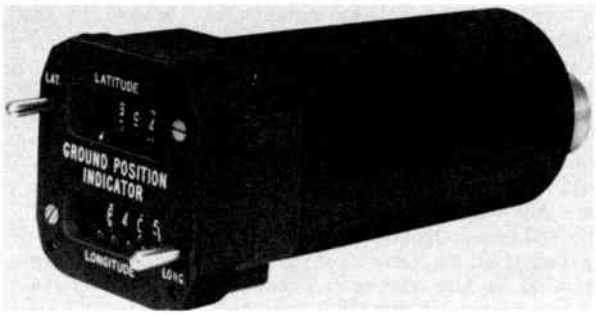


Fig. 2
Indicator unit.

Fig. 3
Preset control box.

Fig. 4
Computer unit.

Fig. 5
Amplifier unit.

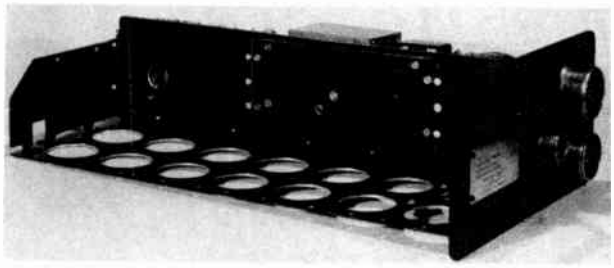


Fig. 7
Amplifier unit chassis
with subassemblies removed.

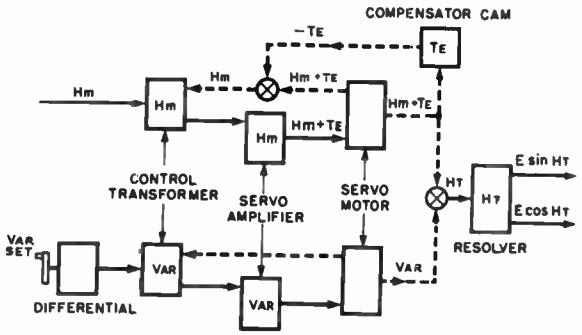


Fig. 9
Functional diagram --
computation of true heading.

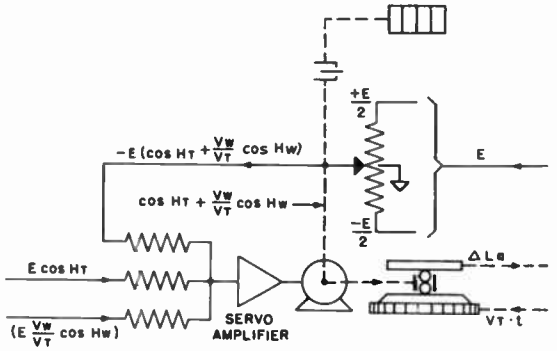


Fig. 11
Computation of change in latitude.

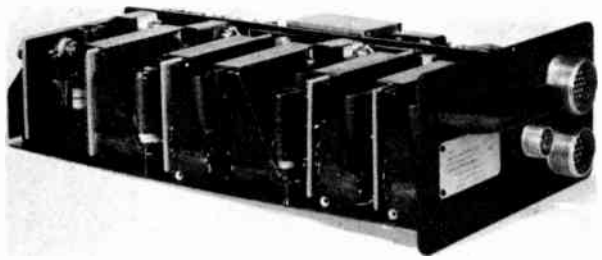


Fig. 6
Amplifier unit with case removed.

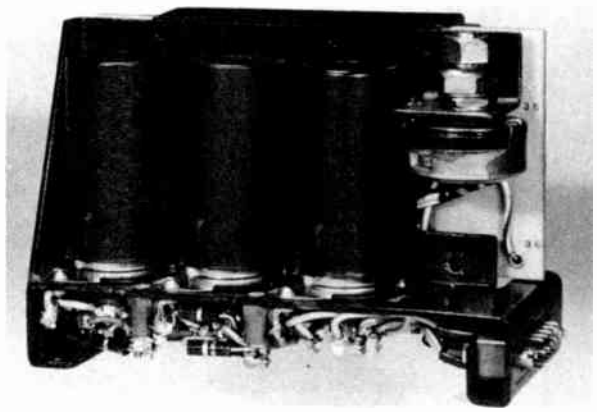


Fig. 8
Typical amplifier subassembly.

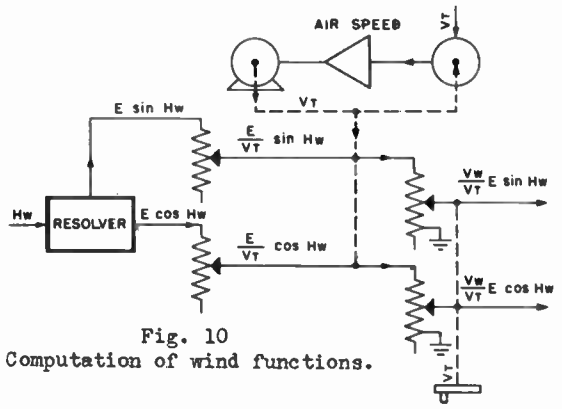


Fig. 10
Computation of wind functions.

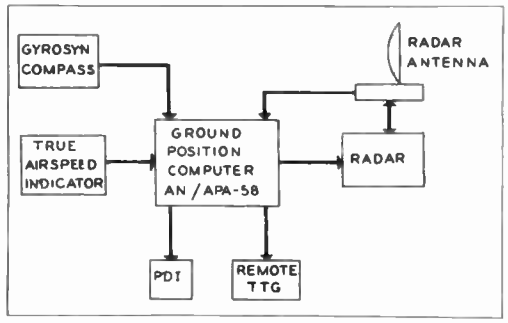


Fig. 12
AN/APA-58.

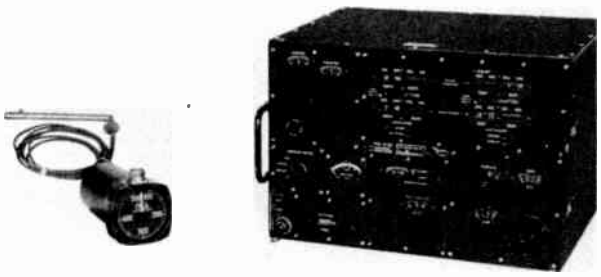


Fig. 13
Computer CP-29(XA)/APA-58 and
true airspeed meter ID-205(XA)/APA-58.

Fig. 14
Computer CP-29(XA-1)APA-58 (front view showing
true air speed and direction unit subassembly removed).

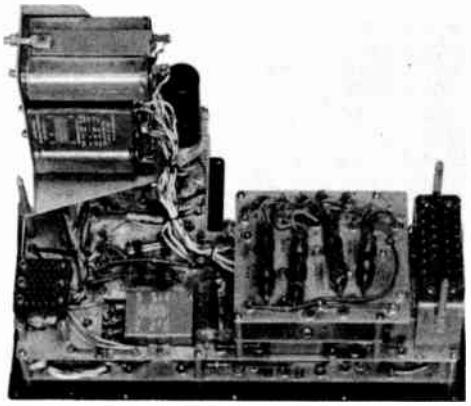


Fig. 15
Calibration unit subassembly
of computer CP(XA)/APA-58 (rear view).

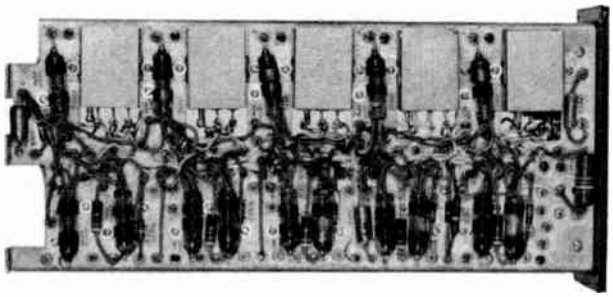


Fig. 16
Calibration unit subassembly
of computer CP(XA)/APA-58 (top view).

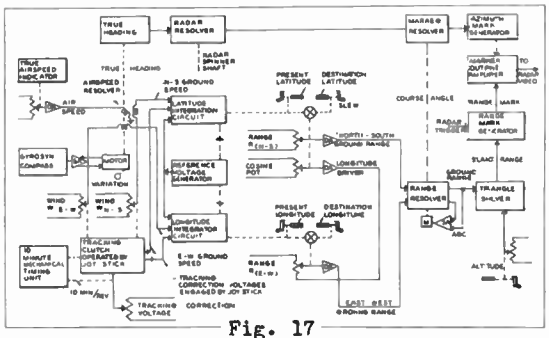


Fig. 17
Simplified block diagram AN/APA-58(XA-1).

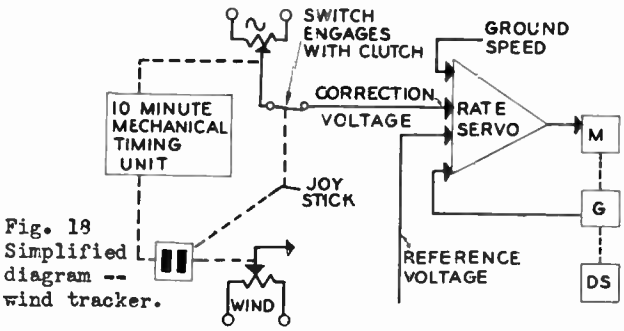


Fig. 18
Simplified
diagram --
wind tracker.

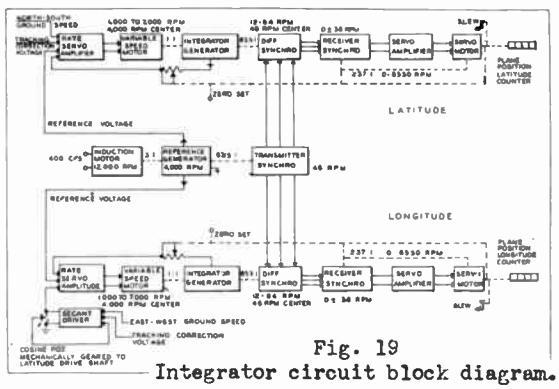


Fig. 19
Integrator circuit block diagram.

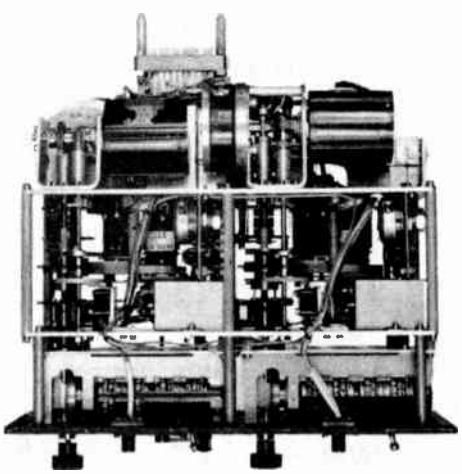


Fig. 20
Position unit subassembly of
computer CP-29(XA)/APA-58 (bottom view).

SOME SYSTEMS CONSIDERATIONS IN FLIGHT CONTROL SERVOMECHANISM DESIGN

Robert J. Bibbero and Roland Grandgent

Republic Aviation Corporation

Guided Missiles Division

New York City, N.Y.

Perhaps the most significant development in airborne electronics today is its tendency to grow away from purely electrical considerations and to merge with other fields of engineering. A radio engineer engaged in the design of airborne electronic systems does not restrict himself by specialization, but, instead broadens his interest to include such widely diversified fields as heat transfer, reliability and noise statistics, shock and vibrational mechanics, and many others. The technology of airborne electronics in many of its aspects is truly one of those borderline regions lying between two or more other fields which, as Norbert Wiener has said, offer the greatest opportunities for advances. [1]

This is nowhere more true than in that phase of airborne electronics which concerns itself with the control and navigation of aircraft. In the design of those elements which make up the steering loop of a directed aircraft, the principles of aerodynamics and flight mechanics must be applied to complete a successful system, to the same extent as those of electronics and of electrical or mechanical engineering.

The reasons for this statement become apparent when it is applied to the technology of aircraft flight control and to the mechanisms which directly steer the aircraft. Although the basic form of the steering intelligence may be something as tenuous as electromagnetic radiation, it must ultimately be converted within the flight control mechanism to a force capable of deflecting a relatively massive control member, whether the latter is a wing or flap, a jet vane, or a gimballed motor. But even if we have sufficient knowledge of the small signal technology, amplification techniques, and mechanics to describe this conversion, the result is meaningless in terms of the overall steering loop unless the response of the airframe to the resultant control deflections can be adequately expressed. Thus, in tracing the flow of information through a steering sequence, we become involved with radio, optical or similar technology; with electrical amplification; the production of mechanical power; and finally with the aerodynamics of the airframe and the response of various flight instruments to aircraft motion.

Occupying a central position in this involved sequence is what we wish to term the flight control servomechanism. We define this device as a machine which accepts at low power levels inputs proportional to some combination of the aircraft steering error and its derivatives and to the airframe motion and its derivatives. Its outputs are control deflections which are directly, or through their derivatives, proportional to a summation of the inputs. The flight control servomechanism has been called a servopilot, since it duplicates many functions of the human pilot. It is similar in many respects to the autopilot, although we prefer to reserve that name for the device intended for use in piloted aircraft.

An illustration may serve to clarify the exact elements which we wish to include in the flight control servo. We can consider a radio controlled aircraft or missile as an example. By means of some intelligent device capable of measuring the steering error, that is the difference between actual and desired aircraft course, a command which is a function of that error is computed and transmitted to the aircraft. The command is conveyed to a servo which deflects a control surface in an amount and direction proportional to the signal. If the steering equation requires that the airframe respond in a particular way to the command, it may be necessary to add other inputs to the servo. For example, if the command is to correspond to airframe rate of turn, the output of a rate gyro should be added. If it is to correspond to airframe lateral acceleration or load factor, the additional input may be a linear accelerometer. All of these devices are integral parts of the flight control servomechanism.

The general scheme just outlined can be expressed in block diagram form. (Fig. 1). The diagram consists of an outer loop and a number of inner loops. The innermost loop and the transducers responding to the airframe dynamics are what we have defined as the flight control servo. The outermost loop is the steering loop.

Although the conventional autopilot using a human pilot as the primary intelligence device is similar to the flight control servo, we prefer to regard it as a separate class for reasons of its performance and historical development. The autopilot was originally developed as a means of relieving the pilot during straight and level flight. As such, its primary duty is to respond to transient loads imposed by gusts and to compensate for steady-state loads developed by cross-winds and control surface malalignment. The piloted aircraft is stable under most flight conditions, hence the autopilot has a comparatively easy task to maintain favorable steady-state conditions.

On the other hand, the flight control servo built into a guided missile or pilotless aircraft is designed to operate primarily in the maneuvering condition. The airframe may be constructed with litte inherent stability, since in that way the highest maneuverability is obtained. Hence, the flight control servo must often provide stability as well as maneuverability to the airframe.

Historically, the first autopilots were designed with no particular airframe in mind, but with regard only to their own performance. The aircraft manufacturer chose from the available models those which best suited his airframe. In a later stage of development, the autopilot manufacturer became forcibly aware of the difference in dynamic response of various airframes and began to design the autopilot specifically for the characteristics of a particular airplane. In fitting the autopilot to a frozen airframe design, inevitable compromises were made and the result was not always as desired.

In the final analysis, the flight control mechanism and the airframe should be designed together and to complement each other. When this is done, the system performance is optimum, regardless of whether the aircraft is partially or totally pilotless. At the risk of oversimplifying, it is our purpose here to delineate this ideal design process.

In the preparation of this paper, the authors have carefully examined the design processes actually employed in various flight control developments of which they have knowledge. Considerable progress has been made in the past ten years.

Early efforts were hampered by great difficulties of component development and manufacture which tended to distract attention from more fundamental problems. All too often the performance of the system as a whole was the last thing to be considered, whereas logically it should be the first. Many guided missiles, for example, evolved around components whose mechanical, electrical, or aerodynamic niceties attracted interest quite apart from the problem to be solved. Frequently, the understandable desire to use familiar techniques and components weakened any attempt to logically relate the design to the task. Today, however, the art has reached a state in which the designer can resist the preoccupation with components per se and focus attention on the system. Such an approach is obviously essential to further progress.

In the initial stages of design, it is necessary to establish the type and source of steering intelligence and the basic geometry of the steering process. These factors are determined from consideration of the task to be accomplished and the tools which come to hand. For example, in the case of a guided missile or perhaps of an aircraft landing system, the source of intelligence will depend on whether the target can best be observed by radar or by other means; whether it is stationary or moving; and many other factors of this nature. In steering geometry or navigation, we have a choice of beam-riding, homing, and precomputed trajectory, as well as variations and combinations of these methods. The location of the intelligence source, whether on or off the airframe, also affects the geometry directly.

The selection of the source of intelligence and the steering geometry are mutually interdependent, and their problems should be considered together.

In a similar manner, the steering geometry and the airframe configuration are interdependent. A configuration entirely satisfactory for beam riding might be very poor for homing, and vice versa.

In considering the major steering loop, it is advisable to determine first the characteristics of the combined control system and airframe, without commitments as to the physical nature of either. (Fig. 2). Here the objective is to find a transfer function for the steering loop such that the system response to inputs and disturbances accomplishes the desired task in the desired time. Knowing the characteristics of the feedback path (the steering geometry) and the closed loop transfer function, the transfer function of the forward path (the control system plus airframe) is determined. The forward transfer function is the ratio of the airframe load factor to the steering error; the feedback circuit or steering geometry converting the airframe load factor to a quantity whose dimensions are the same as those of the input.

Having determined the transfer function of the combined control system and airframe, we can proceed to the manner of achieving the performance in terms of the separate characteristics of the control system and the airframe. At this point, a judicious selection of airframe characteristics may materially simplify the task of the control system.

In general, the design of the airframe is subject to many considerations not directly related to steering. Payload, range, structural requirements, and propulsion method may exert great influence on the configuration. However, the selected intelligence and steering geometry must not be overlooked during the airframe design. For given requirements on range, payload, and so forth, there are available many types of airframe configurations; some are well suited to one purpose, some to another. For example, if the intelligence is a self-contained angular-sensing device, like a homing head, it is important to employ a configuration in which steering is accomplished without excessive angular motions of the airframe axis. Furthermore, in such cases, the feedbacks utilized in the flight control servomechanism must be carefully selected so as to provide whatever static stability or damping may be required to enable the intelligence equipment to operate accurately.

The division of duties between airframe and flight control servomechanism require the utmost in coöperation between the aeronautical designer and the control system designer. Between them, they must arrive at two separate transfer functions that in cascade have the transfer function required of the complete forward path of the steering loop.

When this is done, the major characteristics of the two sub-systems are solidified. On the one hand, we obtain the general aerodynamics and mass parameters of the airframe, and on the other, the nature of the dynamic inputs to the flight control servo, such as transverse acceleration, angular velocity, angle of attack, and so forth. It remains to design the flight control servo to the derived requirements.

This may be done initially by considering the controller and feedback elements as simple gains. Using elementary servo theory, the output of the control servo-airframe cascade may be analyzed in response to some extreme frequency component of the intelligence output. Knowing the phase lags and amplitude ratio for the cascade needed to achieve the necessary following accuracy, the static gain and damping contributed by the servo can be calculated. If the analysis is extended to cover all input frequencies, the usual Bode or Nyquist plots can be drawn and the phase and gain margins measured. However, the value of this type of analysis is somewhat questionable at this point since the computation must be repeated for each point on the flight path where the airframe transfer function parameters undergo a significant change. Furthermore, the traditional frequency analysis tacitly assumes that the transfer function is not time-varying, which is not always the case in the real airframe. The coefficients of the differential equation describing the aerodynamics are not constant, but vary with altitude and velocity. Hence, the frequency analysis is strictly applicable only in the case where these parameters do not vary with time.

Rather than complete the frequency analysis the recommended technique is to set up the equations of motion of the airframe and of the simplified servo on any of the common varieties of analog computer. The aerodynamic parameters can be modified quickly and easily on these machines to correspond to velocity and altitude. The servo gains can be varied about the values found by hand calculation; and finally, the effect of a finite servo time constant can be investigated, starting with a value about one-half to one-fourth that of the airframe. The input to the servo should be simulated by a step-function or combination of a step-function and slow ramp-function. If available, records of actual noise outputs of the intelligence device should be superimposed on the commands. Any of these combinations will prove more realistic than the sinusoidal frequencies assumed in the first analysis.

From the data of these computer runs we can determine a gain setting or gain program for the servo that will yield the required stability and accuracy over the entire flight. In addition, we obtain the value of the largest time constant which can be allowed the servo without adversely affecting overall performance. These figures comprise a set of specifications for the servo design which greatly influence the choice of hardware. However, the specifications are set up on the basis of an ideal linear system. The adverse or beneficial results of non-linearities in the real system are yet to be found.

In order to make final choice of the servo hardware, we must consider the power requirements. The peak power needed to move the controls determines the rating of the actuator and often its generic type, while the average power influences the nature of the power source. To determine the peak power, we must know the velocity and acceleration of the control surfaces and the moments opposing this motion. The control surface motion can be determined from the analog computer records. The opposing moments are due to the inertia, friction and damping, and to the aerodynamic hinge moment, which can be computed from the control surface design.

The hinge moment is nominally considered to be a linear function of control surface position for a given Mach number, although it may often deviate very much from this simple relationship, both by deliberate design and through influence of the airframe angle of attack.

At this point, there exists a good opportunity for the flight control servo design to feed back into the aerodynamic design. If the hinge moments or inertias appear out of line, a substantial saving in servo power or weight can be obtained by redesigning the control surfaces.

It must be noted that the relation between hinge moment and control deflection, even when linear, varies widely with flight Mach number. Hence during a given flight, it may become zero or even negative -- that is, it may tend to increase rather than decrease the control deflection. However, in estimating the peak power, we are interested only in the region where its product with control surface velocity is at a maximum.

Having investigated the peak power, we can often specify not only the kind of actuator, but its size and the size of all elements directly handling the output, such as bell-cranks, gears, or torque tubes.

With the peak power specified, the average power and time of flight together determine the size and weight of the power source and power generating elements, for example, the batteries, hydraulic or pneumatic accumulators, and pumps. The average power will, of course, vary with each flight. It is almost impossible

to determine with high accuracy without a complete program of flight simulation, or experience with actual flight of similar airframes. However, it is possible to make a close guess by estimating the power for a typical maneuver from the computer analysis, estimating the frequency of the maneuvers by reference to the mission of the airframe, and multiplying the resulting average power by a suitable safety factor. In the final design, these estimates must be checked and rechecked by flight simulation or by flight test.

Armed with the estimate of peak power, average power, time constant, and loop gain, the designer is in a position to choose between the various types of servomechanisms for the flight control. The types in general use include hydraulic, pneumatic and electrical, either motor or clutch. Each may be used in conjunction with magnetic or conventional electronic amplifiers. For the range of frequencies and power required in the average high performance missile or aircraft, a hydraulic servo will often prove to be most suitable. Although there are some difficulties associated with hydraulic equipment, they have many advantages over electrical actuators -- primarily in their frequency response and acceleration capabilities. A hydraulic system also has the advantage that less power is required to sustain a large hinge moment. They are preferred over pneumatic systems owing to the time lags and elastance of the latter.

The final choice is made not only on the factors mentioned, but on such additional considerations as availability of transducers in the power ranges desired, and availability of prime power from other sources. For example, a missile containing a great deal of electrical equipment may influence choice of an electrical system, while the availability of a source of high-pressure gas may tend to suggest a pneumatic system.

At this point, sufficient information is available to choose detailed components and to construct a laboratory model of the flight control servo. The mechanical model will include many time lags and non-linearities which were not considered in the analytical work. For this reason, a fairly lengthy program of compensation is required to complete the second-order design of the servo amplifier or controller. The method used is a step-by-step investigation of servo and airframe stability for many points on the flight path. The completed servo model is provided with spring and inertial loads simulating the real control surface reaction forces and its output shaft coupled electrically to the computer analog of the airframe. Using a breadboard model of the amplifier, the compensation can proceed under simulated flight conditions utilizing step-function inputs as a test. In general, an attempt is made to find gain settings and compensation networks giving the necessary airframe response without incurring instabilities or exceeding certain overshoot criteria. With a more refined computer set-up it is possible to achieve a better design by taking advantage of a different concept of stability.

If the coefficients of the kinematic and aerodynamic equations are suitably mechanized so as to vary automatically with the changing geometry and flight conditions, it is possible to duplicate realistically and in real time the entire controlled flight. It may be found that some combinations of servo parameters which apparently lead to instability or to inaccurate steering in the step-by-step analysis, now give satisfactory performance. With a higher degree of realism in the simulation, we may find it possible to achieve the desired results with cheaper or lighter components, even though arbitrary stability tests proved them unsuitable.

We have attempted to show how in each stage of the flight control servo design from its inception and specification to the detailed compensation of the components, system considerations - aerodynamic and mechanical - are of equal importance to the electronic aspects. Although the airborne equipment designer is forced to this conclusion by hard facts, his counterpart in industrial controller design can profit by his experience, even when his task is inherently less demanding.

(1) Wiener, N.- "Cybernetics, or Control and Communication in the Animal and the Machine". New York, Wiley; 1948.

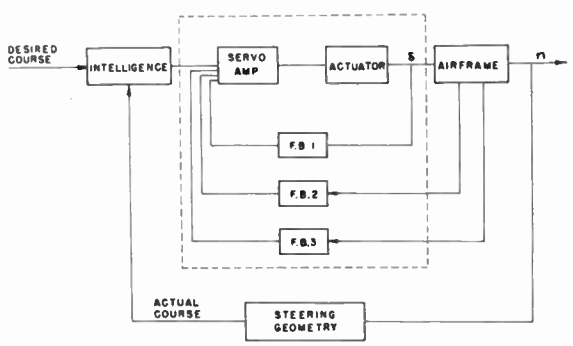


Fig. 1

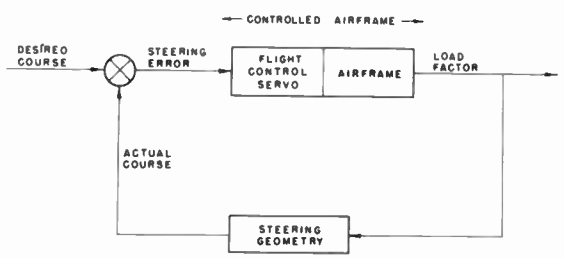


Fig. 2

FAIRED-IN ADF ANTENNAS

By

Louis E. Raburn
Electronics Research, Inc.
Evansville, Indiana

Introduction

During World War II work was initiated on the problem of reducing the size and protrusion of automatic direction-finder antennas (ADF) on military aircraft. This paper describes an extension of this work which was done at Electronics Research, Inc. with the sponsorship of the Navy Department, Bureau of Aeronautics, to obtain basic information and design data on ADF antennas for military aircraft.

Reviewed briefly, the ADF receiver nominally covers the band 100 kc to 1,750 kc. A block diagram of a receiver system is shown in Fig. 1. Two antennas are required: (1) A non-directional vertical (sense) antenna and (2) a loop antenna which is rotated by a servo motor. In operation, the ADF receiver compares the phase of the signal from the sense antenna to the signal from the loop and rotates the loop to a null position. By means of a multiple-synchro system, the position of the loop, and hence the relative bearing of the radio station, are presented on remote indicators in the aircraft.

No specifications on ADF antenna performance were available at the start of the study and the following antenna requirements were formulated to guide the work: First, the strength of the signals picked up by a faired-in loop should be no more than about 3 db below the strength of the signals picked up by a conventional external loop. This assures that signals from fairly distant range stations will be strong enough to override the noise in the input stage of the receiver. Second, the sense signal strength should be at least 6 db greater than the loop signal when the loop is rotated for maximum pickup. This assures quick, definite operation of the loop servo system and an acceptably small amount of modulating buzz in the headphones.

When the loop is mounted on an aircraft fuselage, the distortion of the received signal field in the vicinity of the fuselage causes bearing errors. Time does not permit presenting a concept which has been formulated relating to the distortion caused by the induced currents flowing on the surface of the thick unsymmetrical fuselage. Suffice it to say that errors as high as 60 degrees are encountered in some loop locations, whereas errors greater than 25 degrees cannot be fully corrected by the equipment cam compensating system. In these cases, it is necessary to either move the loop to a new location or to employ some form of electrical compensation such as an inductor-loop or a metal shield.

Preliminary Study of Flush Mounted Loop Parameters

Both iron-core and air-core loops have been developed for use with most ADF sets, and both types were investigated. Preliminary measurements were made of stub and loop antenna installations in the top of the 6' x 6' portable shielded cage shown in Fig. 2. The measurements were made in the center of the Evansville Municipal Airport to minimize effects of re-radiating objects and extraneous signals.

The bearing errors with the iron and air-core loops were investigated by mounting each type of loop antenna alternately in the top center of the cage, as shown in Fig. 2A. The bearing errors were considerable because of the square shape of the cage, but the errors were essentially the same for both types of loop antennas regardless of true bearing or frequency. This was confirmed later by measurements on installations in an SNB-2 aircraft mock-up. The reason is that the effective center of the loop in either case is quite close to the skin of the mock-up and picks up the same distorted field.

Loop sensitivity comparisons were investigated by calibrating the avc voltage of the receiver and measuring this voltage with both types of antennas connected alternately to the receiver. The iron-core loop was found to have substantially the same pickup as the air-core loop, and the receiver loop input stage alignment was the same for both types of loop.

The recess or "well" size required is illustrated in Figs. 2B to 2E, showing the experimental data taken to determine the loss in loop signal as a result of recessing the iron-core loop in various wells. The reference signal in each case is the loop pickup with the loop mounted externally on the top of the shielded cage. Fig. 2D shows that the well must be at least 18 inches in diameter for a loss of less than 6 db. When the air-core loop is mounted in this well, the loss was several db more than for the iron-core loop. The size of the well can be made much smaller if the loop is not required to be completely submerged. Fig. 2B shows that when the iron-core loop is submerged only halfway, a 12" diameter well will cause only 4-db loss.

The shield-detuning effect was determined by aligning the loop input stage of the receiver with the loop normally mounted and realigning the input stage for maximum pickup with each loop mounted in the 18-inch sleeve shown in Fig. 2E. The avc voltage was measured to determine the extent of detuning caused by the shielding effect. The iron-core loop was found to have less than 2 db-detuning, whereas the detuning loss of the air-core loop was much greater. Later measurements on aircraft installations also showed that the iron-core loop requires less retracking than the air-core loop when a flush-mounted installation is made.

Both types of loops show the same bearing errors, and it is possible to use either type of loop to investigate the bearing errors at different antenna locations. The iron-core loop was used for most of the location investigations, as it is more sensitive in recessed locations.

It is structurally desirable to locate the sense antenna in the same space as the flush-mounted loop, but it is usually difficult to get enough sense pickup. As a general rule, if the well is large enough that the sense pickup is less than 6 db below that of a standard vertical wire sense antenna, an iron-core loop can be mounted in this same cavity with a loss in loop pickup of only one or two db.

Bearing Errors of Aircraft Installations

To determine whether the bearing errors of a flush-mounted installation at a given location are comparable to the errors for a surface mounted installation, flight tests were made with an iron-core loop first installed flush in a 17" diameter x 7" deep well in the top astrodome of a SNB-2 (Beechcraft) aircraft and then surface mounted at the same location. The bearing errors for the two installations are quite similar, with perhaps an increase in maximum bearing error of about 2 degrees when the antenna is flush mounted. However, the flush-mounted installation was about 6 db less sensitive than the surface-mounted installation. The same results were obtained when flush and surface-mounted iron and air loops were located in the astrodome of the larger DC-3. These tests, therefore, indicate that various locations for a flush loop on the aircraft could be evaluated by flight tests or compass rose tests of the surface-mounted loop installations.

Next, the bearing errors were measured for a number of loop locations on the SNB and DC-3 aircraft. The measurements were made with the aircraft on the ground, using the south compass rose at the Evansville Airport. Figure 4 shows the bearing error data at 233 kc and 1400 kc for several different locations along the top center line of the SNB with an air-core loop antenna and a 2-foot vertical sense antenna mounted on a small chassis which could be readily moved and bonded to the airframe. These curves indicate that the least bearing error occurs near the nose of the aircraft, and that the error increases as the loop is moved farther back toward the tail of the aircraft. In addition, the dispersal or spread of the error curves at each location for the two frequencies increases as the antenna location is moved back from the nose toward the tail.

Similar bearing error data for two antenna locations along the top center line of the DC-3 aircraft are shown in Fig. 5. The general trends in this case are the same as for the SNB with both the bearing error and dispersal

versus frequency increasing as the antenna location is moved from the top of the nose back toward the tail.

Mock-Up Study of Bearing Errors

The 24'-long full-scale mock-up shown in Fig. 6 was constructed of sheet aluminum and mounted on a rubber-tired dolly so it could be easily turned for measurements related to the installation of ADF sets in fighter type aircraft. A compass rose showing true headings was painted on the decking under the turntable and the power for the ADF receiver was supplied by an adjacent gasoline-engine-driven generator.

This mock-up was built up progressively from the cylindrical fuselage with a loop and sense antenna for the bearing error and sensitivity measurements mounted on the top side at the location as shown. Measurements were made as the mock-up was built up by adding, first, the tapered tail section, next, the hardware-cloth wings, panels, and, finally, the flat nose cover. The data showed that these sections had no appreciable effect compared to the fuselage on the characteristics of this antenna system.

Sense Antennas

An investigation of feasible sense antenna configurations that could be installed with the flush-mounted loop in a single 18"- x 6"-deep cylindrical top-side recess was performed with the antennas shown in Fig. 7. Although these configurations have no vertical element in the conventional sense, the pickup of vertically polarized signals was found to be adequate, due, it is assumed, to the location of the ring an appreciable distance above the aircraft's electrical neutral plane.

The curves in Fig. 7 show the loss in loop sensitivity due to the presence of the sense antenna and the ratio of sense signal to maximum loop signal as functions of frequency for several sense antennas. It is desired that the sense configuration reduce the loop pickup as little as possible, but at the same time pick up a sense signal which exceeds the loop signal by at least 6 db, as previously stated.

A circular plate was the first type of sense antenna investigated, and the best results were obtained with a plate 8 inches in diameter. This antenna causes a loop loss as great as 2.5 db with a 7 db minimum value of sense to loop signal.

An investigation of rings of various proportions indicated that the width of the ring was not critical if it was on the order of 1 inch wide. The last three graphs show the performance for rings with inside diameters of 18, 17, and 15 inches that would just fit inside the cylinder sleeve. These curves show sense signals several db greater than required with no appreciable loss in loop pickup if the outside diameter sense ring is nearly as large as the diameter of the cavity. Additional measurements showed that these sense rings did not cause any noticeable bearing errors in the ADF system at any frequency.

Installation in SNB Waist

An installation of the loop and sense antennas in the top part of the waist of the SNB aircraft was made by cutting away the skin between two bulkheads 33 inches apart. Front and back walls and a bottom shelf of aluminum were added to form the recess as shown in Fig. 8. The top two metal stringers were cut away and replaced by fiberglas plastic bars which, together with the bottom shelf, were designed to carry all the loads normally carried by the metal skin and stringers which were removed. The iron-core loop was mounted on a stand to hold its top end as near as possible to the inner surface of the Plexiglas sheet which covers the recess. The sense antenna is provided by an 8-inch x 24-inch top loading plate and lead-in wire. The wire extends downward and passes through an insulating washer in the bottom of the shelf and then 2 feet forward to the ADF receiver.

As shown in Fig. 9, the maximum bearing error for this flush installation with no mechanical compensation is more than 30 degrees and does not vary greatly with frequency. As the maximum bearing error that can be mechanically compensated is only 25 degrees, additional compensation was needed to make a satisfactory installation at this location. It was found that a length of copper hook-up wire run fore-and-aft just above the loop and connected to both the front

and back walls of the recess resulted in a considerable reduction in bearing error at both frequencies.

Adjustment of Compensating Inductors

An investigation of the following compensating inductor parameters: Numbers of wires, wire cross-section, length, and separations indicated that all of these configurations produced the same type of compensation, and that the amount of compensation increased as the size and number of wires was increased and as the separation of the wires from the loop was reduced. The final choice of inductor configuration was based mainly on mechanical considerations. It consists of two 1/4-inch diameter copper tubes fastened to the skin of the aircraft at the rear wall of the recess, extending forward above the loop, and bent downward to connect to the front of the loop mounting base.

Bending the compensating inductors apart to give maximum clearance from the loop provided minimum compensation with the inductors in place. The bearing error curves in Fig. 9 are without mechanical compensation and show that considerable compensation was produced. The curves for maximum inductor clearance still have the same sign in each main quadrant as the curves without compensation, indicating that the added compensation is less than optimum. The bearing error curves with the inductors straightened so that they run fore-and-aft and moderately close to the top of the loop, show that the sign of the errors in each quadrant has been reversed, indicating over-compensation.

Considerable over-compensation was also tried without any beneficial shaping of the curves or dispersal of the curves with frequency. The final position of the compensating inductors midway between the last two positions shown in Fig. 9 gave optimum compensation at 233 kc with a maximum error of 3 degrees and at 1280 kc a maximum error of 11 degrees with an average displacement of about 8 degrees due to rose errors. This amount of bearing error was considered acceptable without mechanical compensation.

The compensating inductors reduce the loop signal by an amount which varies from 0 up to 15 db, depending upon the bearing and frequency of the received signal, see Fig. 9A. Fortunately, this reduction in loop signal does not appreciably reduce the operational performance of the ADF system as the least sensitivity reduction results in the range frequency band and it occurs at the bearings where the uncompensated loop signal is the least. The minimum loop signal with optimum inductor compensation is only 1 db less at 233 kc and only 6 db less at 1280 kc than the minimum loop signal without compensation.

Installation in Belly of SNB

The SNB belly installation shown in Fig. 10 was made with the iron-core loop and a sense ring mounted in a rectangular aluminum box and fiberglas cover fitted into the space above the tunnel hatch opening. The loop was mounted in an inverted position with the tip of the loop glass cover nearly touching the fiberglas cover sheet. The sense antenna was a rectangular "ring" of 1/2-inch aluminum strap.

The measured belly installation data made on the north compass rose with zero cam compensation is shown in Fig. 11. The maximum error of 20 degrees can be mechanically compensated by the loop cam.

The 12 degrees dispersal versus frequency is comparable to the curves for other installations measured on this same compass rose showing a similar displacement, and was apparently caused by metal fences or buildings near the rose, or by non-uniformity of the ground around the aircraft.

Although this error curve could have been corrected by mechanical compensation, a brief inductor-compensation investigation was performed. Fig. 11B shows typical error curves for an aluminum strip 1/8 inch wide placed athwartship across the box opening just below the loop proper. The bearing error curves are reversed in sign, indicating that in this case the transverse inductor produced considerable over-compensation. This inductor strap caused a reduction in sense signal of about 5 or 6 db. Since inductor compensation was not required, the strap was finally removed and the cams were adjusted to give the curves in Fig. 11C.

Flight Tests

The flight tests of both the top waist and the belly installations were made using the magnetic compass for course reference. During both flight tests the ADF system had good sensitivity, and the resulting error curves are shown in Fig. 12. Most of the random error in these curves at least at the low frequency is a result of the bumpy air during the flight runs which caused the magnetic compass to be very unsteady. The bearing error points in this figure were not joined by a smooth curve since the smooth variation is meaningless. The two flush-mounted antenna systems were satisfactory with regard to sensitivity, bearing errors, and freedom of ambiguities in all normal flight attitudes. For these reasons, it is concluded that faired-in ADF antennas giving fully acceptable performance can be designed, installed and compensated using one or more of the techniques described in this paper.

Other Error Factors

During the compass rose test program it was noted that measured bearing errors were consistently different at the two different compass rose locations at the Evansville Airport. One test was made at both roses with the iron-core loop mounted externally on the top of the aircraft and with the cams adjusted for zero compensation. The bearing error curves at both roses were measured on the same day and comparable curves had the same shapes but appreciable difference in the average displacements. These curves show that the displacement is a function of the frequency and is due to ground variations, metal fences, and railroad tracks in the vicinity. The fences and tracks adjoin the airport and are located many feet away from both rose locations, but in this frequency band these distances are still small in terms of wavelength.

Tests with the loop mounted 12" off the aircraft centerline at the top astrodome location were also made. The loop installations were made on a plate covering the turret compartment opening. In this case the off-center location increased the maximum error a few degrees, but this added error is small compared to the usual fuselage bearing errors. The conclusion is that there is no objection to an off-center mounting of at least 12" if desirable for structural reasons.

Credits

In closing, I wish to give credit to Charles Banks, formerly of BuAer; to Harvey Kees of ERI; and to Les Lorenz, formerly of ERI.

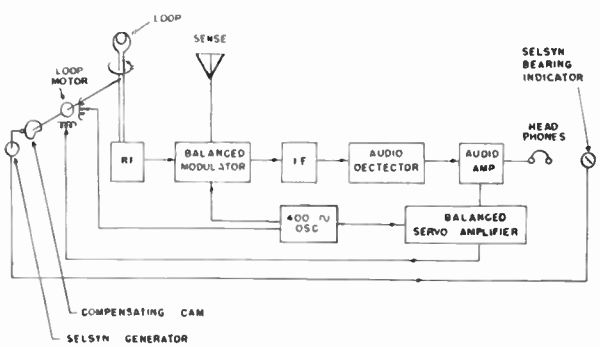


Fig. 1
Block diagram of ADF system.

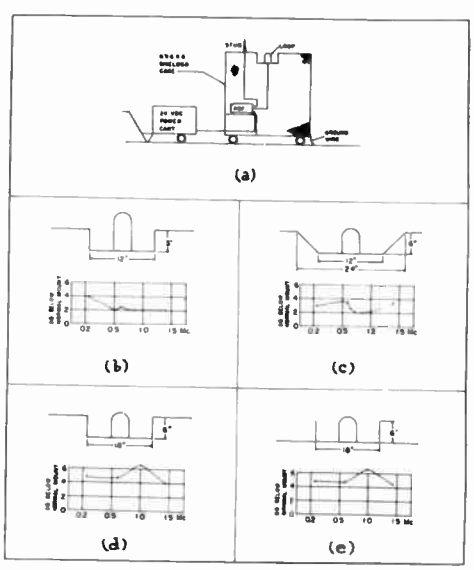


Fig. 2
Recessed loop sensitivity tests:
(a) cage set-up in open field;
(b) well 12" x 3"; (c) well 24" x 6";
(d) well 18" x 6"; (e) sleeve 18" x 6".

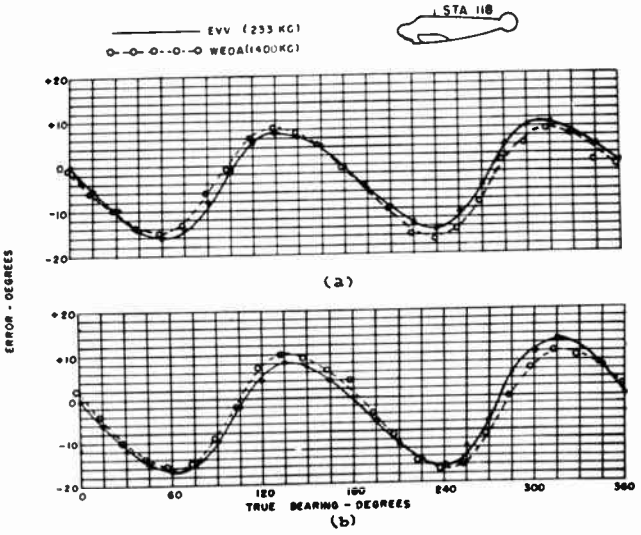


Fig. 3
Loop error vs. loop mounting on SNB-2:
(a) surface-mounted iron-core loop;
(b) flush-mounted iron-core loop.

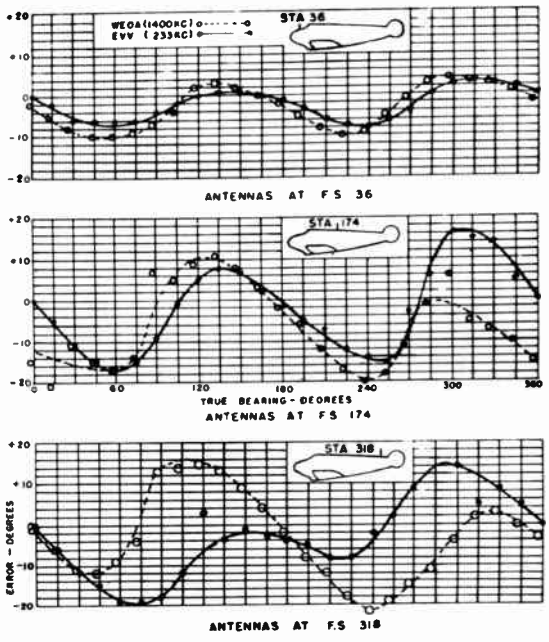


Fig. 4
Bearing errors on SNB-2
at stations 36, 174, and 318.

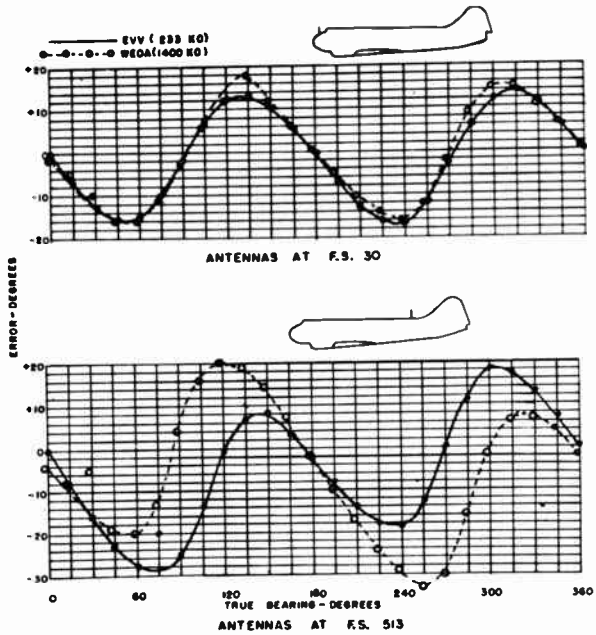


Fig. 5
Bearing errors on R4D at stations 30 and 513.

Fig. 6
Full scale mock-up.

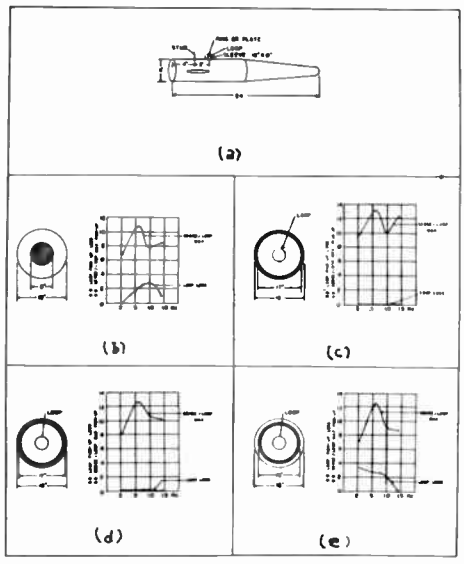


Fig. 7
Recessed sense antenna sensitivity tests.
(a) mock-up for sense tests;
(b) 8" sense plate; (c) 8" sense ring;
(c) 17" sense ring; (e) 15" sense ring.

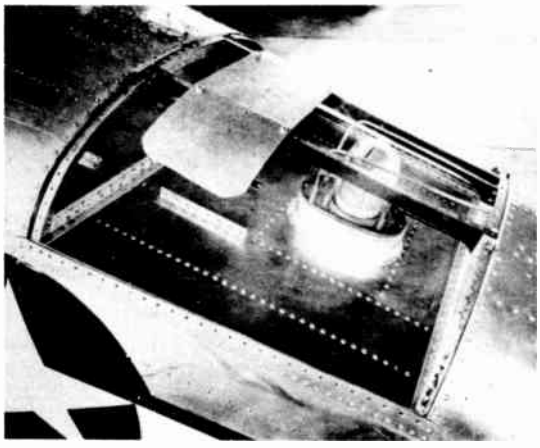


Fig. 8
ADF antenna flush mounted in SNB-2C.

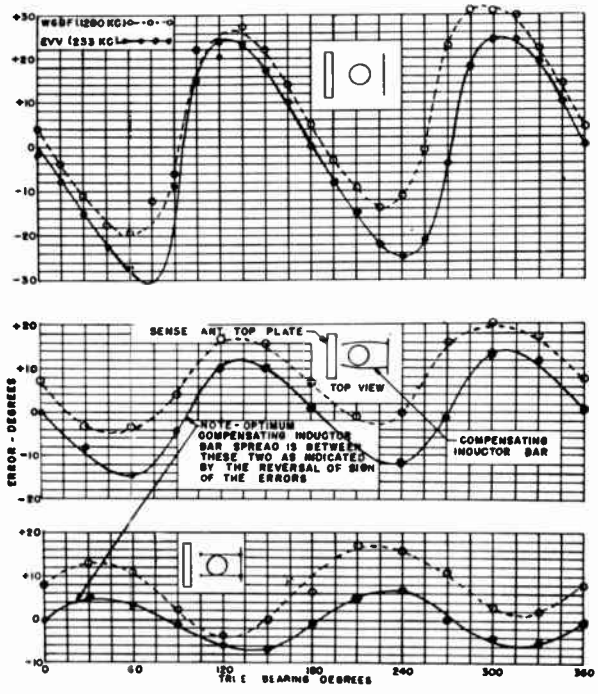


Fig. 9
Bearing errors vs. compensating inductor positions:
(a) no inductor compensation;
(b) inductor compensation less than optimum;
(c) inductor compensation more than optimum.

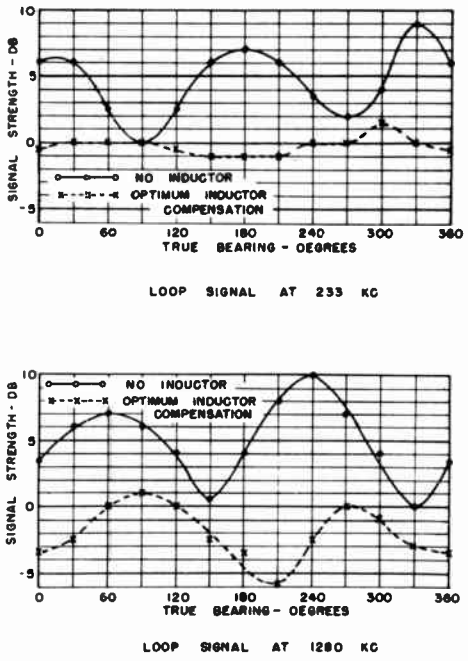


Fig. 9A
Loop signal loss with inductors:
(a) loop signal at 233 kc;
(b) loop signal at 1280 kc.

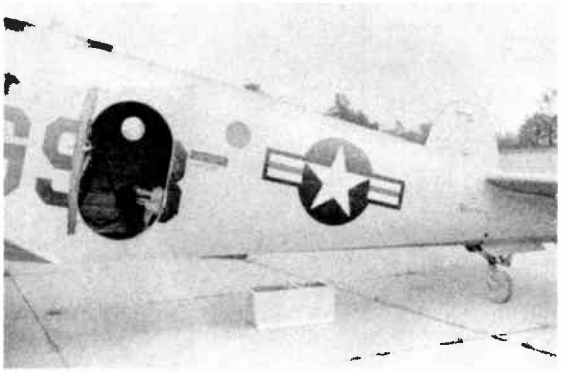


Fig. 10 - Antenna box below belly location.

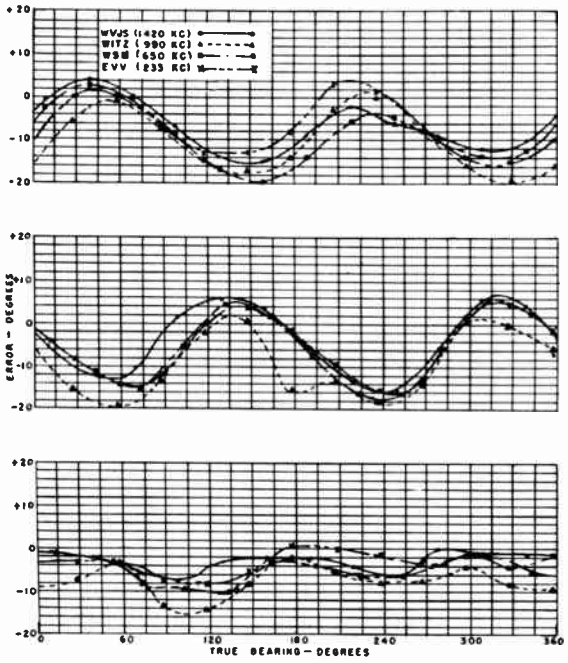


Fig. 11
SNB-1 belly installation errors:
(a) no compensation; (b) inductor
compensation; (c) cam compensation.

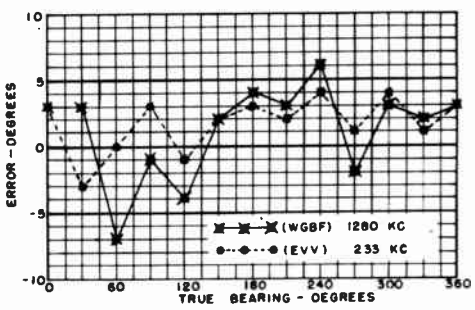


Fig. 12 - Flight test errors of top installation.

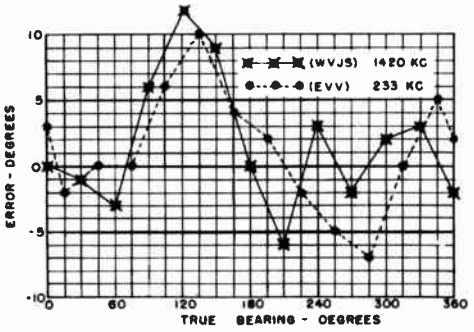


Fig. 13 - Flight test errors of belly installation.

MAGNETIC AMPLIFIERS FOR AIRBORNE APPLICATION

J. K. McKendry

General Precision Laboratory, Inc.

Pleasantville, New York

This paper describes work which was done by the author at General Precision Laboratory, Inc. under an Air Force Contract.

The amplifiers to be described were designed to provide control power for instrument type servo-motors used in an airborne military equipment. They do not embody the most recent advances in the magnetic-amplifier art, principally because it is felt that these have not been reduced to good engineering practice for applications involving the environmental extremes to be encountered by aircraft instruments.

The Position Servo

Several positioning servomechanisms were required, having general similarities but differing widely in some details and specifications. The motor used in all instances is the 2 phase, 400 cps, Mark 7 (Bureau of Ordnance designation). It is rated at 6 watts input per phase, has a stalled torque of 1.5 oz. in., and a free-running speed of 5000 RPM. Ranges of variation in other important servo parameters are:

	Maximum	Minimum
Gear Ratio	3200	533
Control Signal Sensitivity	.39 volts/ degree	.022 volts/ degree
Permissible Error	15 min.	3 min.

In the interests of production economy and in consideration of the spares stocking problem, it was desirable to have a single amplifier design for all these applications. In addition, no factory or field adjustment of the amplifiers was to be required. Fixed attenuators are provided in the system to equalize the variations in control signal sensitivity and gear ratio.

One feature common to all the servos is the assumed presence of coulomb friction in the mechanical loading. A maximum value, based on previous experience with similar mechanisms, was assumed to apply in all cases. Each servo is then required to have a sufficiently high torque-error constant to reduce error to acceptable limits in the presence of this value of load friction. Unfortunately, since the friction term is not controllable except as to its maximum value, it is not possible to rely on it as a stabilizing influence.

The primary system function is that of an analogue computer having slowly varying inputs. Velocity and acceleration errors are accordingly of little significance, and the frequency pass bands of the servomechanisms can be correspondingly restricted. This consideration is, of course, favorable to the use of magnetic amplifiers which are relatively low-pass elements.

It was initially decided that an all-magnetic amplifier would not be feasible, in view of limitations on gain per stage and the anticipated difficulty in stabilizing a servo loop incorporating more than one stage of magnetic amplification. The amplifier, as finally produced, consists of a magnetic amplifier output stage and three dual-triode vacuum tubes of the voltage amplifier class.

The magnetic amplifier, shown schematically in Figure 1, with its associated control tube, is of a conventional type, consisting essentially of two identical saturable core transformers. The primary coils are connected in series across the 115 volt, 400 cps line. The secondary coils are also connected in series, with opposing polarity. With equal currents flowing in the two control windings, the transformer voltages are equal and there is no net output voltage. If the control current of transformer A is increased and/or the control current of transformer B is decreased, the primary impedance of transformer A is reduced relative to that of B, and the primary and secondary voltages of B will be correspondingly greater than those of A. A net voltage of 'B' polarity then appears across the output terminals. Obviously a positive current increment in the control winding of B will reverse the operation and yield an output of 'A' polarity.

Economical design of the magnetic amplifier requires that its internal impedance, as seen at the output terminals, represent a significant fraction of the total load circuit impedance. The voltage delivered to the load is accordingly affected in magnitude and phase by the nature of the load impedance. This effect is usefully applied in the servo-amplifier by trimming the load so that it effectively consists of a resistance in parallel with a capacitor which resonates with the magnetic amplifier inductance at 400 cps. Under these conditions the load voltage

is in exact quadrature with the primary line voltage, which is the desired condition for control of a 2-phase motor. This phase relation holds reasonably constant over the full range of output voltage providing the average of the two control currents is maintained constant. A secondary benefit derived from the load tuning capacitor is the marked improvement in load voltage waveform due to harmonic attenuation. Other, less favorable, aspects of the tuned load circuit will be discussed later.

The preceding electronic stages of the servo-amplifier are shown in the circuit diagram, Figure 2. They consist of an initial two stage, resistance coupled, a-c amplifier, followed by a phase sensitive detector. The detector output is coupled to the grids of the magnetic-amplifier control tube through a resistance-capacitance network which performs the necessary shaping of the amplifier's frequency response.

The necessity for frequency response shaping in this application is largely due to the energy storage effects contributed by the magnetic amplifier, and by the interacting motor and magnetic amplifier. To demonstrate the validity of this statement it is necessary first to examine the frequency dependent transfer function of the motor alone.

The Mark 7 motor, in addition to the characteristics already cited, has a rated rotor inertia of approximately .017 oz. in 2. The effective damping coefficient, computed from the rated stalled torque and the rated free running speed is .0029 in. oz./radian/sec. The ratio of these two quantities in the order given and after conversion to equivalent units, is approximately .014 sec. This is, by definition, the motor time constant. The inverse ratio, sometimes referred to as the motor natural frequency or corner frequency, is approximately 70 radians or 11 cycles per second.

This computed value is somewhat optimistic since the damping coefficient at low control voltages is markedly smaller than at full rated voltage. While the generated motor torque is a reasonably linear function of control voltage, the free-running motor speed at low control voltage levels is considerably higher than proportional. Tests with hard-tube proportional amplifiers, having no energy storage elements effective in the servo frequency range, indicate that the actual corner frequency for low signal levels is approximately 7 cps. With such an amplifier it is possible to achieve a closed loop gain corresponding to a velocity-error constant of approximately 300 radians/sec/radian, a torque-error constant of 0.5 in. oz/radian, and a resonant peak at approximately 20 cps. Such a servo is not well damped but is acceptable for many applications.

The magnetic amplifier has an obvious energy storage mechanism in the inductance of the control windings. In the circuit configuration of Figure 1, using a 12AU7 dual-triode as a control tube, the effect is that of a phase lag network with a corner frequency in the vicinity of 30 cps. This is adjustable, to some extent, by variation of the effective control tube resistance, but practically cannot be increased much above the frequency mentioned.

There is also a second, less obvious, energy storage mechanism associated with the magnetic amplifier load circuit. As previously indicated, the control winding of the motor is paralleled with a condenser which, at 400 cps, creates a condition of parallel resonance with the motor winding effective inductance, plus series resonance with the magnetic amplifier load circuit inductance. When a modulated 400 cps carrier is impressed on this circuit, by the control action of the magnetic amplifier, the higher frequency modulation component lags in phase behind the low frequency component. The net effect is the production of a phase lag in the modulation envelope. In the region of interest this can be satisfactorily represented as a simple phase lag element, operating on the modulation frequency, and having a corner frequency at approximately 30 cps.

The two preceding effects are observable with the servo motor blocked. There is also a third deleterious effect which becomes apparent only when the motor is running free, and results from an interaction between the motor and magnetic amplifier. The stalled motor can be satisfactorily represented (at fixed frequency) by paralleled resistance and inductance. Free running, the motor has an induced internal voltage, which is essentially in series with its equivalent resistive component. The external impedance presented at the motor terminals to this back e.mf. is essentially that of two parallel resonant circuits of quite high "Q". For a fixed value of control signal fed into the magnetic amplifier there will be, then, a considerably higher voltage across the terminals of the free-running motor than across the stalled motor. This increase in voltage with increasing speed constitutes an effective reduction in the motor damping coefficient.

In the closed loop servo equations, reduction of the motor damping coefficient can produce varying results, depending on what compensating changes can be made in other loop parameters. In the absence of all time lags except that of the motor, it must entail reduction of the torque-error constant, reduction of the loop damping ratio or a combination of these two effects. The presence of time lags at higher frequencies furthermore, places an absolute limit on the amount of reduction in damping ratio which can be accepted in lieu of reduced torque-error constant.

The back e.m.f. of the servo-motor is proportionally larger at low control voltages. It

therefore not only reduces the average motor damping coefficient but also enlarges the range of variation of this parameter, to such an extent that it is impossible to compensate the effect adequately by any simple means such as a phase lead network.

It has been found possible, by means of inverse feedback around the magnetic amplifier, to minimize the back e.m.f. effect to a considerable extent. In this particular application use of feedback would have required a large number of additional components, and would have placed some undesirable limits on system functioning. A more acceptable solution was found to be the use of load-damping resistors across the motor terminals. This of course, reduced the maximum useful power output of the magnetic amplifier but not to an intolerable extent. The effective motor corner frequency is, by this means, restricted to a range of variation of approximately 3 to 5 cps. With this modification, the composite open loop transfer function of the motor and magnetic-amplifier can be adequately represented algebraically as a fractional expression having a gain-constant in the numerator and a fourth degree polynomial in jw in the denominator. A graphical plot of the amplitude of the transfer function in decibels against log of frequency can be approximated by three straight-line asymptotes; the first extending from zero to 3 cps with a slope of -6db per octave, the second extending from 3 cps to 30 cps with a slope of -12db per octave, and the third above 30 cps with a slope of -24db per octave.

The phase shift associated with the transfer function reaches 180° (lagging) at approximately 6 cps. For closed loop stability the gain constant must accordingly be adjusted to produce feedback cut-off at or below this frequency. The corresponding velocity-error constant is approximately 75 radians/sec/radian and the torque-error constant is approximately 0.06 in. oz./radian or about one tenth as large as that which can be realized with an uncompensated hard-tube amplifier.

The re-shaping of the transfer function to improve the torque-error constant to a usable level is performed in the resistor-capacitor network shown following the phase detector in the circuit diagram of Figure 2. This is a composite lead-lag network of a common variety sometimes descriptively titled a "notch" network. The straight-line asymptotes of its amplitude response curve on a decibel-log frequency graph are as follows:

1. A straight line of zero slope below .035 cps.

2. A straight line of -6db/octave slope from .035cps to 1.0 cps.

3. A straight line of zero slope from 1.0 cps to 1.7 cps.

4. A straight line of plus 6db/octave slope from 1.7 cps to 48 cps.

5. A straight line of zero slope above 48 cps.

In Figure 3 are shown the amplitude response asymptotes for the composite motor-magnetic amplifier and for the notch network, together with the asymptotes of the curve representing their cascaded response. This over-all response curve has a section of -6db per octave slope between the frequencies of 1 cps and 30 cps (with minor irregularities near the low frequency end). The phase shift corresponding to this amplitude curve reaches 180° at approximately 20 cps. At lower frequencies it remains in the quadrant between 90° and 180° with a maximum lag of 160° in the region of 0.2 cps.

If the loop gain is adjusted to produce feedback cut-off anywhere in the region of 1 to nearly 15 cps, the servo operation will be stable and acceptably damped. With cut-off at 15 cps, the velocity error constant of the servo is approximately 1900 radians/sec/radian. The torque-error constant is 1.4 in. oz./radian for this same gain level.

Computed maximum variation in overall amplifier gain, due to arithmetic addition of all component variations affecting gain, amounts to a range of approximately 5 to 1. In each individual loop a fixed attenuator is included such that the loop gain, with an amplifier having maximum computed gain, will reach cut-off at 15 cps. A minimum gain amplifier will accordingly produce a torque-error constant of 0.28 oz. in./radian. This is sufficient in all cases to meet the static accuracy requirements of the system.

The Rate Servo

The servo-motors used for this application are a 2-phase 400 cps type, similar to Naval Ordnance Mark 8 except that they have 4 poles per phase, yielding a synchronous speed of 12,000 rpm. The stalled torque is 2.5 oz. in. and the free-running speed is 10,000 rpm. The motor damping coefficient, computed from these values is 0.0024 oz. in./radian/sec. Rotor inertia is 0.019 oz. in.^{2} and that of the direct-coupled precision tachometer is 0.57 oz. in.^{2} The effective motor time constant is then 0.64 sec, and the corner frequency is 1.56 radians or 0.25 cycles/sec. As in the previous case, this value is reduced at low signal levels, to approximately 0.15 cps.

The amplifier used is identical in configuration with that of the positioning servo. Differences are in the magnetic amplifier which is slightly larger, as required to furnish the rated motor input power of 16 watts, and in the shaping network values.

The requirements placed on the shaping net-

work are again determined almost entirely by the nature of the transfer function of the composite motor-magnetic amplifier. As in the positioning servo, the phase lag effects due to the control winding inductance of the magnetic amplifier and to the resonant load circuit become effective in the vicinity of 30 cps. There is also a substantial and variable reduction of the effective motor damping coefficient. In the case of the rate servo it was found possible to tolerate this varying effect without direct compensation.

With a varying motor time constant it is not possible to define an exact transfer function in linear terms. However, it is possible to define the limits of the transfer function and determine loop stability for all intervening fixed values. In the lower limit the transfer function can be approximated on a d.b.-log frequency graph by a straight line of zero slope below .025 cps, a straight line of -6db per octave slope from .025 to 30 cps. and a straight line of -18 db per octave slope above 30 cps. In the upper limit the low frequency asymptote extends to approximately 0.15 cps, the upper corner frequency remaining at 30 cps.

With the upper limit transfer function, the maximum stable loop gain would be about 200, increasing to 1200 for a lower limit transfer function. These figures are not acceptable since the integration error permissible is less than one part per thousand.

Figure 4 graphically represents the transfer function of the shaping network, the cascaded shaping network and upper limit motor and magnetic amplifier and the cascaded shaping network and lower limit motor and magnetic amplifier. Allowing for an uncontrolled variation of 6 to 1 in amplifier gain, the minimum loop gain expected is approximately 1000.

Summary

The foregoing description has intentionally emphasized the servo design problems associated with the use of magnetic amplifiers, as a warning against overoptimistic and indiscriminate specification of magnetic amplifiers for this type of application. However, it seems advisable to point out that an equally detailed analysis of any type of amplifier, meeting equivalent performance specifications, would probably disclose design problems of the same order of complexity.

The primary objective of this paper has been to demonstrate the possibility of using magnetic amplifiers in some types of high performance servo-mechanisms. The acknowledged magnetic amplifier advantages in ruggedness and reliability may accordingly be obtained without sacrifice of performance. It is to be anticipated that the newer magnetic amplifier techniques, when applicable, will considerably simplify the design problems.

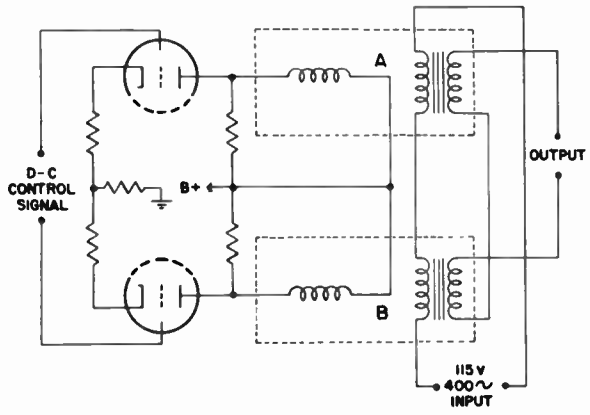


Fig. 1
Magnetic amplifier and control tube.

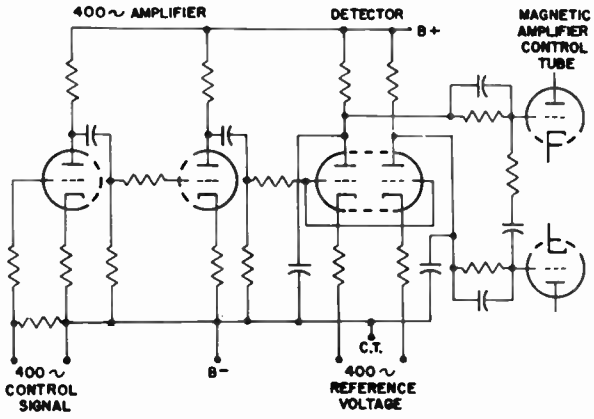


Fig. 2
Electronic amplifier.

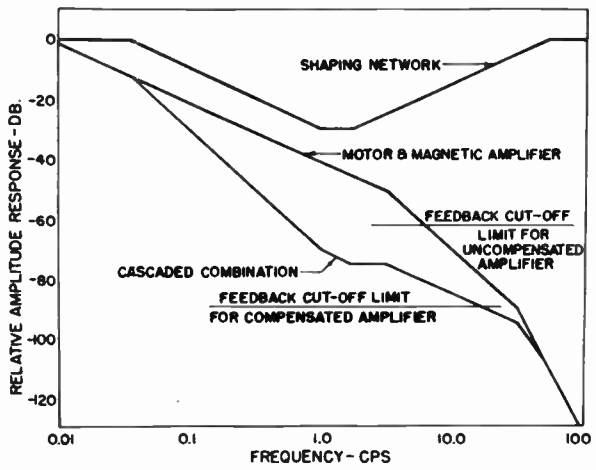


Fig. 3
Position servo transfer functions.

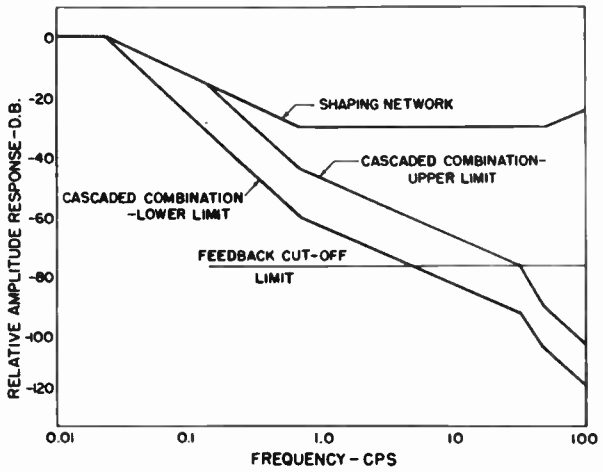


Fig. 4
Rate servo transfer functions.

AIRCRAFT ELECTRICAL POWER

J. C. Dieffenderfer and G. W. Sherman
Wright Air Development Center
Dayton, Ohio

ABSTRACT

Complex aircraft electrical and electronic systems demand that an engineer associated with either system have a working knowledge of the other if the over-all weapons system is to perform its mission effectively. No longer can the electrical power engineer assume that his job is completed when he simply energizes the airplane bus, nor can the electronic engineer assume that he need only to connect the equipment to the bus for satisfactory operation. The power generation and control system in military aircraft is not an infinite source of power nor is it practical to provide the power to classical textbook limits on voltage and frequency regulation, waveform, or harmonic content. It is possible, however, that a fully integrated electrical and electronic system permitting optimum operation of electronic equipment and without unduly compromising either system may be realized if the electronic and electrical engineers approach their problems from the broad-systems concept.

THE EFFECTS OF ELECTRONICS EQUIPMENT STANDARDIZATION ON AIRCRAFT PERFORMANCE

George C. Sumner
Consolidated-Vultee Aircraft Corp.
Fort Worth, Texas

A primary design requirement of the modern aircraft is high performance. Depending upon the particular design, this high performance may be in terms of speed, range, altitude, operating cost per passenger-mile, operating cost per cargo ton-mile, or combinations thereof. To obtain his objective, the aircraft designer has a multitude of considerations, aerodynamic and otherwise. Each structural member, each equipment, and even each rivet affects the performance of the aircraft. The sum total of these effects is significant, indeed, and must be considered early in the aircraft design.

Airborne electronic equipment has been increasing in importance and in magnitude over the past several years. The time is long since past when radio communications equipment constituted the main airborne use of electronics. To-day electronics provides radio navigational aids, instrument landing systems, autopilots, engine controls, electronic fuel gages, intercommunications, and radar sets, as well as many specialized military applications such as fire control systems and navigational computers. Moreover, the trend of assigning more functions to electronics shows no signs of diminishing.

In considering an equipment for airborne use three factors are of primary significance:

1. Does it perform its intended function well, in terms of precision, response, etc?
2. Is it reliable?
3. What is its performance cost to the vehicle in which it is to be installed?

A good engineering design can only result from a judicious balance of these three factors. Functional improvements and simplifications are constantly being made as the state of the art permits. Tremendous effort is being expended in providing better and predictable reliability. The latter factor, equipment cost to aircraft performance, is not receiving such general systematic consideration. It is the purpose here to examine this aspect. Consideration is made from the standpoint of manned airborne vehicles.

Equipment Effects On Aircraft Performance

The most apparent effect of equipment on aircraft performance is the "dead" weight of the equipment. This weight displaces at least an equal amount of payload. Fuel must be expended in transporting it. Every electronics designer who has tackled airborne equipment is well aware of the emphasis placed on weight minimization. There are, however, more subtile and equally as important ways that an installed electronics equipment affects aircraft performance. The performance cost might be expressed by the following equation:

$$P.C. = k_1W + k_2S + k_3P + k_4T + k_5D \qquad (1)$$

where: W is actual installed weight
S is installation space required
P is electrical power input
T is temperature conditioning required
D is direct aerodynamic drag imposed
k's are constants for a particular aircraft

Performance Cost Due to Space

The installation space depends not only upon the actual volume of the equipment but also upon how efficiently the equipment utilizes the available space within the airframe. The value of k_2 depends upon the overall average weight density of the aircraft.

Performance Cost Due to Power Input

Since electronic equipment uses electrical power as input, the aircraft power system must furnish this power to airborne equipment. Fuel must be used to drive the aircraft electrical generators. Additional generator capacity and distribution system capacity is required when an electronic equipment is added. Also, any voltage or frequency conversion device required to adapt a particular electronics equipment further increases the performance cost of that equipment.

Performance Cost Due to Temperature Conditioning

Temperature conditioning, especially cooling, of airborne electronic equipment is becoming a serious problem. The miniaturization trend has increased power densities, i.e. watts per cubic foot. This fact coupled with the increased use of electronics has made temperature conditioning an important consideration. It appears that in the foreseeable future the electronic equipment in some overwater passenger transports may consume as high as six or seven kilowatts. On many military aircraft the total is much higher. This total consumed power less a small percentage which is radiated is dissipated

within the airframe as heat. With such magnitudes of heat, cooling can no longer be obtained by letting "nature take its course". Natural convection or a simple blower will not in general suffice. A conditioning system must be provided to maintain temperatures within allowable limits. Whether conditioning is accomplished by ram air, expanded engine bleed air, evaporation, or whatever means, a performance penalty to the aircraft results. The magnitude of the performance cost for a particular equipment depends upon its dissipation and upon how efficiently it can be conditioned by the aircraft system. The constant k_4 of equation (1) becomes particularly large with high speed aircraft.

Performance Cost Due to Direct Drag

Any protrusion or deviation from ideal aerodynamic lines reduces the obtainable performance. Electronicwise, antennas have been the principal offender in this respect. Fortunately great sucess has been realized in obtaining flush or submerged antennas. There remains, however, an important problem in highly directional antennas, such as for radar use, where physical size is presently dictated by usable beam width. The effect of change in aerodynamic contour may be expressed as added drag. In this case, also, k_5 of equation (1) increases markedly with the aircraft design speed.

Aircraft Growth Factor

A further aid in evaluating performance penalty may be obtained by calulating an equivalent weight for the equipment. This equivalent weight, W_e , is defined as that amount of dead weight which if carried by the aircraft would result in an equal performance penalty. Equation (1) would then become:

$$W_e = W + k_a S + k_b P + k_c T + k_d D \qquad (2)$$

Aircraft performance loss may then be calculated directly from the equivalent weight. The most significant way of considering this performance loss is in terms of how much weight in fuel, engine capacity, etc. must be added to the aircraft to regain the loss. As an aircraft gross-weight change this is given by:

$$\Delta W_g = k_g W_e$$

where k_g is the aircraft growth factor. The value of aircraft growth factor varies among various aircraft, but it may be 10 or more for modern high speed aircraft. As an example, consider an electronic equipment whose actual weight is 50 pounds, whose equivalent weight is 75 pounds, and which is to be installed on an aircraft with a growth factor of 8. The aircraft gross-weight increase necessary to maintain the same performance with the equipment installed is then 600 pounds.

Effects of Present Standardization

For the absolute minimum of performance penalty, each aircraft model should have an electronics system designed especially for it. The constants of equation (1) vary markedly between different aircraft. Therefore an equipment which is a good design for one aircraft type may represent a sizeable penalty for another. This seems to indicate that good aircraft design is inconsistant with equipment standardization. It apparently means that only by specialized equipment can low performance cost be realized. This conclusion is premature, however, as will be seen when standardization is further examined below.

Equipment standardization is highly desirable for many reasons. In times of national emergency it is desirable in order to conserve industrial effort, to simplify logistics, and to effectively use technical talent. At any time it is advantageous in reducing prices and in obtaining other merits of quantity production. It, therefore, behooves the industry to obtain equipment standardization compatible with low aircraft performance cost.

Equipment standardization has three aspects, functional, form, and environmental. Similar usuages, especially for similar classes of aircraft, have resulted in many standard functional requirements. Standardization attempts have brought about common form or packaging schemes such as the ATR system and others. To further obtain standardization it has been the practice in the past to design equipment to operate in a hypotical typical aircraft environment. This hypothetical environment was established to include the environments of almost all conceivable aircraft. Specifications requiring equipments to operate over a wide generalized set of environmental conditions have resulted in both unnecessary and inefficient environmental conditioning for particular installations. In a large measure, most past and present airborne electronic installations fail to measure up to minimum aircraft performance cost mainly because of environmental conditioning difficulties imposed by standardization. The following environmental factors are involved:

1. Power input
2. Pressure conditioning
3. Temperature conditioning
4. Vibration isolation

Power Input

Power input is not generally considered an environmental condition, but the equipment designer is faced with the fact that present aircraft may have one of several primary electrical power systems as regarding voltage, frequency, and regulation. Although attempts are being made to standardize, sufficient merits exist for each of these several systems that it is doubtful if a single common system will be adopted in the foreseeable future. A given equipment with a built-in power supply designed to operate from a certain voltage must be accompanied by an extra power conversion device when installed in an aircraft whose primary power system furnishes a different voltage. This extraneous conversion step costs extra weight, extra space, added heat dissipation due to conversion losses, and fuel loss in the generation of the extra power. These losses can be eliminated and standardization maintained by deleting the built-in power supply and permitting the airframe to provide the required anode and filament voltages in the most efficient manner from its own primary power system. Figure 1A shows the result of installing an equipment with a 28 volts dc input in an aircraft whose primary power system is 115 volts 3-phase. Power efficiency is low because the overall conversion efficiency is the product of the efficiencies of the two conversion devices. Figure 1B shows the same equipment with the built-in supply deleted. Now only one conversion step is required resulting in an appreciable gain in aircraft performance cost. Figures 2A and 2B show a similar condition for an equipment originally standardized of 115 volt input but installed in an aircraft whose primary power system provides 28 volts dc.

Pressure Conditioning

As aircraft ceilings have risen so have airborne electronic equipment pressure-altitude requirements. Designers have resorted to placing equipments inside pressurized cans or tanks. Such designs result in a very poor installed space efficiency, because the space consumed in the aircraft is usually that of a circumscribed parallelepiped. Often such equipment is installed within pressurized areas for accessability or other reasons with the integral pressure conditioning presenting a sizeable needless increase in weight and volume. This situation can be alleviated by standard pressure specifications which require that the equipment be only as good pressurewise as the men who fly the aircraft. Additional pressure conditioning could then be provided by the airframe as required.

Temperature Conditioning

The temperature environment for airborne equipment depends upon the location within the airframe, e.g. in inhabited areas it is different than in uninhabited areas. It is a function of design speed and altitude of the aircraft. Integral temperature conditioning designed for the worst case will be overdesign for lesser cases. Conversely, built-in conditioning designed for lesser cases requires additional provisions for the worst cases. Moreover, integral means, such as built-in blowers or forced-air heat exchangers, will not in general efficiently match the variables of a particular airframe basic conditioning system. Equipment can be standardized and at the same time aircraft performance not be compromised in this respect, if temperature conditioning is provided by parts non-integral to the equipment which may be varied in accordance with particular aircraft requirements as shown in Figure 2. The recently announced "cold plate" method of cooling is a promising means of accomplishing this.* Other means will probably present themselves.

Vibration Isolation

Vibration conditions are greatly different in different locations within a given aircraft as well as for different aircraft. Some locations present such small vibration that no isolation is required for well designed electronics equipment. Since isolation systems are costly in clearance space, as well as, in weight, vibration isolation should be provided on a location basis rather than on an equipment basis. The aircraft designer can then install standard equipments with isolation as required with no unnecessary performance loss that is attendent to equipments standardized to include vibration isolation, as many present equipments are.

Conclusions

It is seen then that a high degree of equipment standardization can be maintained with a decrease in aircraft performance cost. This may be obtained by changing the level of standardization so that the bulk of environmental conditioning is accomplished by airframe systems. A further requirement is that the aircraft designer be furnished complete and specific environmental limit data for electronic equipment. This change in design concept can prevent aircraft from becoming performance limited by undue burdens of electronic equipment and, at the same time, maintaining the advantages of standardization.

*"Heat Transfer Design Problems in Aircraft Electronic Equipment", Leonhard Katz, National Conference on Airborne Electronics, IRE, 1952.

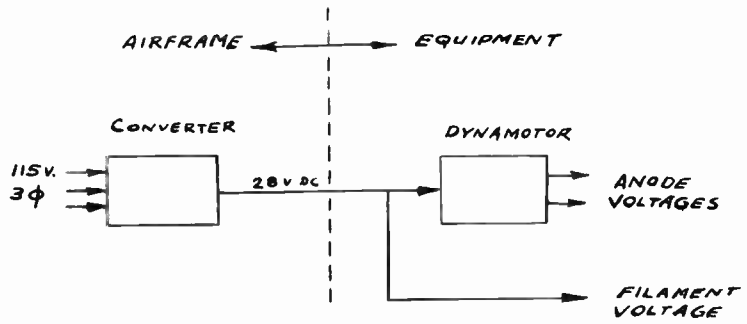


Fig. 1(a)

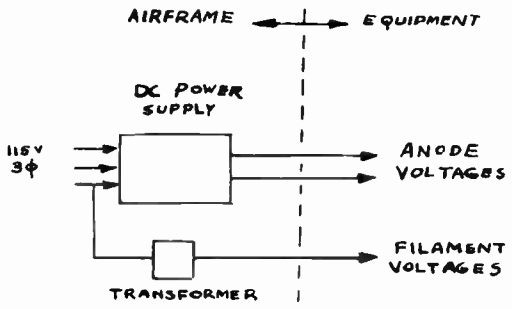


Fig. 1(b)

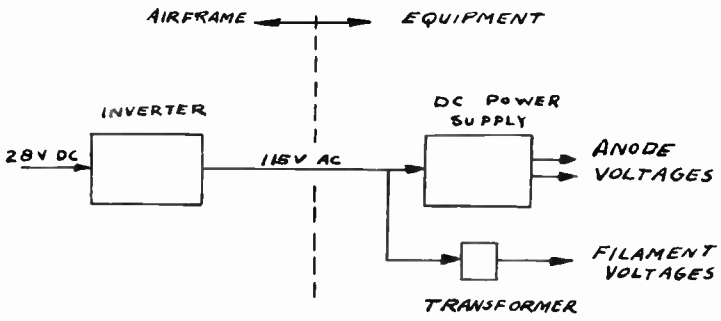


Fig. 2(a)

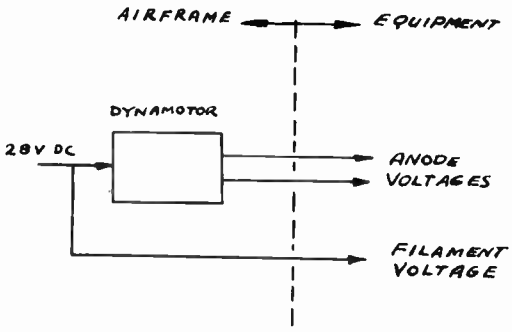


Fig. 2(b)

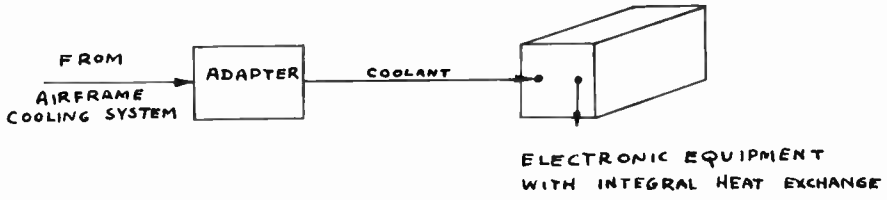


Fig. 3(a)

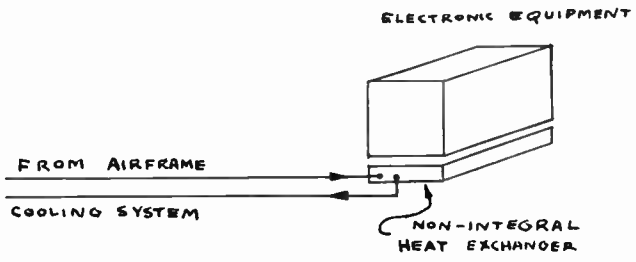


Fig. 3(b)

TELEMETERING REQUIREMENTS FOR UPPER AIR ROCKET RESEARCH

Dr. Marcus O'Day
Chief, Upper Air Laboratory, Geophysics Research Directorate
Air Force Cambridge Research Center, Cambridge 39, Mass.

In the field of upper air research through the medium of rockets, one of the most important problems encountered is that of telemetering. Among the factors to be taken into consideration in a discussion and study of telemetering requirements are weight and size of the airborne units, adaptability to upper air experimental equipment, reliability, number of channels, sampling rate, time resolution, accuracy, and, of course, ease of data reduction.

The airborne telemetering unit is a facility, and, of itself, does not "take" any upper air data; therefore, abnormal size and weight of the equipment may place limitations upon the installation of additional upper atmosphere instrumentation and render infeasible certain experiments which require high altitude. For instance, in one type of Aerobee rocket there is a loss of altitude of of a nautical mile per five lbs. of weight in excess of the standard 150 lbs. of payload. The size of the telemetering unit likewise plays an important role when consideration is being given to the sizes of the equipment required for certain experiments.

Based on experience gained during the last four years, a rough estimate would indicate that the weight and size of the airborne telemetering unit should be less than 25% of the standard payload and volume of the missile. When transistors are available for instrumentation, this percent would undoubtedly grow smaller.

Closely allied to the factors of weight and size is the question of adaptability of the telemeter to experimental equipment and its components. In many guided missile applications it is desirable to have a low impedance input to the telemetering system. However, in practically all upper air experiments vacuum tubes are used so that a high impedance input is desirable and easy to accomplish. The present tendency is to have a high impedance input to the telemetering system of about 100,000 ohms or more, and a maximum signal of 5 volts. These values were used in the original Naval Research Laboratory V—2 telemetering system and have proved very satisfactory.

Though adaptability be achieved, sight must not be lost of the important characteristic of reliability. Upper air instrumentation is expensive and time-consuming even when small rockets are being utilized; consequently, optimum performance and sufficient data should be expected. Usually, data are not obtained over the entire trajectory of the rocket; therefore, it is highly possible that adequate data will not be secured if portions of them are lost during the flight. Gaps in the telemetering response are extremely unfortunate. They are caused, generally, by holes in the antenna pattern, and often occur at the most inopportune times. In the case of the Air Force Aerobee beacon telemetering system, it has been found desirable to have several ground receiving stations in operation so that if data are lost at one station--due to a bad aspect of the missile with respect to that particular point--one of the other stations will receive signals yielding data to supplement those lost.

It is believed that the telemetering system should provide coverage for at least 95% of the time of the useful part of the flight in order to balance the cost and complexity of the scientific instrumentation.

The number of channels necessary in a telemetering system for upper atmosphere research is determined, in general, by the size and characteristics of the rocket available, as well as by the nature and purpose of the experiments themselves. In the case of the V-2 rocket, it was often found that thirty-two (32) channels were not enough; auxiliary systems, therefore, had to be utilized.

When the smaller Aerobee rocket was developed, a preliminary evaluation of its 150 lb. payload, based upon the number of experiments which might be flown, resulted in the estimate that two channels, with a sampling rate of 569 each per second, would suffice. It was believed that mechanical commutation of these two channels would yield adequate data. This estimate, however, has proved to be entirely too low; at times, even one experiment of itself requires more channels. This is due to the fact that the rocket, in general, does not have a steady, stable motion; consequently, when pitch and yaw are present to a high degree, it is necessary that far more data be obtained so that the investigator may be given the opportunity to assay which are significant and which may have been altered so much by the motion of the vehicle as to be worthless.

A relevant example in this situation is found in an experiment which we have flown a number of times with the objective of measuring the intensity of the day airglow. Eight interference filters are rotated in front of a photomultiplier, as shown in Fig. 1, to measure any scattered light from the sun or any other visible radiation which may originate in the upper atmosphere. These interference filters enable us to get a rough spectral distribution of the amount of light. It has been found difficult, however, to obtain all of the data in reducible form by commutating only two channels. If the rocket spins excessively, the commutation rate may even be comparable to the spin of the rocket itself, and when such conditions prevail the data are almost impossible of interpretation. Moreover, data on the performance of the rocket

itself are often necessary. This type of rocket has not been flown enough times to make its characteristics completely predictable, and frequently it has been found necessary to measure certain flight parameters such as chamber pressure, skin temperature, etc. Studies made in the Upper Air Laboratory have indicated that it would be desirable to have ten channels at the above-mentioned sampling rate, half of which could be commutated for data which do not demand so many samples per second.

The Naval Research Laboratory has recently designed a pulse time telemetering system consisting of 15 channels, at a sampling rate of 312 per second, for use in the Aerobee rocket. At the present time, this value seems to be a good compromise between what is desirable and what is absolutely necessary.

If we look ahead to the future, the situation assumes greater significance, because in the next few years transistors will have been developed to the point where the Aerobee rocket may carry two or three times the number of experiments it carries at the present time. The necessity for a larger number of channels in the telemetering requirements will then, of course, be more pronounced than has been indicated here. For the immediate future, the beacon designed for use in the Aerobee during the next two years has been expanded to include six telemetering channels.

Just as the experiments being flown determine, to a large extent, the number of channels needed, so, too, they dictate the sampling rate required. In many experiments a high sampling rate is necessary. In the system developed by the Naval Research Laboratory for the V-2 and Aerobee rockets, a sampling rate of 312 per second per channel was selected. For most missile work a sampling rate of 40 is sufficient. The question, therefore,can well be asked why the experiments designed to study upper atmosphere effects should have a higher sampling rate. If a stable platform on which the equipment could be mounted were available, undoubtedly a lower sampling rate would be acceptable and feasible. Let us look, for example, at one of the rocket flights designed to measure the solar constant without the aid of such a stable platform as a sunseeker. The rocket has a roll rate of from two to three per second. If the optical system receiving the radiation from the sun were to have an aperture of 10° square, it would pass over the sun in 1/72nd of a second or less. With a sampling rate of 312 per second it is obvious that this would permit only about four pieces of data to be obtained. On several occasions it has been necessary to utilize one complete channel of the beacon telemetering system, with its 569 per second sampling rate, for just one part of one of the experiments. Experience has shown, however, that a sampling rate of from 200 to 300 per second will satisfy over 90% of the requirements during a flight.

Another extremely important characteristic of a data transmitting system is the time resolution which it permits. Sometimes high time resolution is synonomous with a high sampling rate, but often the two conditions differ. For instance, two or more phenomena may follow each other in very rapid succession for a very short period of time and yet occur at a slow repetition rate. Two examples will illustrate this point. One of our ionosphere experiments involves the transmitting of pulses on two frequencies simultaneously from the ground to the rocket. One of these pulses is at a frequency of about 4-5 megacycles near the critical one for the E layer. It experiences a delay over the one transmitted at a higher frequency of 500 megacycles. This delay may extend from one to 80 microseconds, depending upon the depth of penetration of the rocket into the E layer. Good data, however, could be obtained at a modest recurrence rate of less than 100/second, yet the time resolution should be a microsecond.

In the second experiment, the velocity of the air stream with respect to the rocket is measured as shown in Fig. 2. A spark serves as the ion source which travels a distance d_s before detection. With the values of d_s, which are practical, and the maximum velocity of the rocket as now flown, the time of passage of the ion cloud will be about 1/4 of a millisecond. These sparks may occur only several times a second, but the time resolution required in the data transmission system is very high, especially when it is borne in mind that an error of even 10% corresponds to 25 microseconds.

In addition to the various telemetering system requirements already discussed, there is the very important factor of accuracy. The upper air physicist is especially concerned with high precision in his data. Some experiments lose significance otherwise.

Let us consider, for example, the measurement of the solar constant at high altitudes. For years the accepted value has been 1.94 gm. calories/cm^2/minute; this quantity is one of our great geophysical constants. It is very desirable to measure it to three significant figures, an accuracy of one part in 200, or 1/2 of 1%. I know of no telemetering system in use which has this accuracy, and I mean over-all accuracy. In the rocket, the value is measured by the use of calibrated thermistors operating in a dynamic circuit, i.e., with a light chopper, so that an a.c. signal is produced. In this case an existing telemetering system was used, and a bias was introduced so that the only values above 1.65 gm. cal/sq. cm. per minute were telemetered, as shown in Fig. 3. The measurements on the ground, however, before the ascent of the rocket, will not give a value above 1.6. Consequently, although the back bias method allows the essential data to be transmitted, it does not enable us to obtain a complete record from the ground to zenith. This situation, naturally, is very undesirable since a check point at ground level is lost. In rocket instrumentation, it is almost axiomatic that values of data obtained from the rocket equipment should check with those obtained from ground-based equipment at the moment of launching, if

possible. Often this procedure, of course, is not possible, but whenever conditions permit, this check point should be obtained.

Some of you may undoubtedly suggest that arrangements can be made to change the level of the signal in flight. While this method is possible, it complicates the equipment, thus introducing a very undesirable feature in a type of apparatus which must perform perfectly without in-flight adjustments. One tiny failure--and the whole flight is lost.

On the other hand, in certain other experiments the bias procedure is impractical. This statement holds true whenever the variations of the signal are comparable to its normal value or where the average signal is difficult to predict.

In any field utilizing telemetering systems, there is always the problem of data reduction. In a flight of a rocket containing upper air research instrumentation, very large amounts of data are obtained. For instance, it has been estimated that in one flight of an Aerobee rocket 600,000 pieces of information are recorded through the telemetering system. In fact, we expect, and hope for, a large amount of pertinent data from every flight. In these days of manpower shortages, however, it is obvious that personnel should not be forced to spend their time--which could be devoted to the more important phases of their research--in deciphering the telemetering records. Moreover, it is always difficult to maintain objectivity in such a process. I have always believed that the data should be reduced automatically and finalized as a set of figures, perhaps from an automatic typewriter.

In the case of the beacon telemetering system used by the Upper Air Laboratory of the Air Force Cambridge Research Center, a device has been developed to accomplish this objective. A rack of equipment which decodes the telemetering signal and operates a bank of neon lights is located at one of the ground stations. These lights, in turn, operate according to a modified binary code so that a photographic record on 35 mm film can be made. The film is then shipped to Boston where companion equipment decodes it, reproducing the data in a set of figures on an automatic typewriter. The amount of data recorded during one flight can keep this machine in operation almost continuously for a period of a week.

It is important that the upper air physicist obtain this reduced data soon after the flight. In a program involving instrumentation of rockets, schedules must be drawn up in advance so that the varied and complex phases involved can be coordinated. The logistics of this type of program take on added difficulty if the experimenter does not have adequate time to analyze his data and engineer any necessary modifications into the equipment before the next flight. A great deal of time is wasted after each flight before the results of the experiment are properly interpreted. The automatic reduction of the telemetered data can materially aid in shortening this time, thereby allowing the upper air physicist to devote more time to perfecting the equipment with the ultimate goal of achieving optimum performance and, with the aid of the telemetering system, more accurate data in the next launching.

In stating these requirements, I have toned down to some extent the desires of the upper air physicist. He would like to have the data transmission system give him data such as he would get in the laboratory. The analysis I have given you is a compromise between what is desired and what I feel is attainable in the comparatively near future with equipment which must be carried in a rocket auxiliary to the main experiments. As techniques improve during the next five years, the requirements may be entirely different.

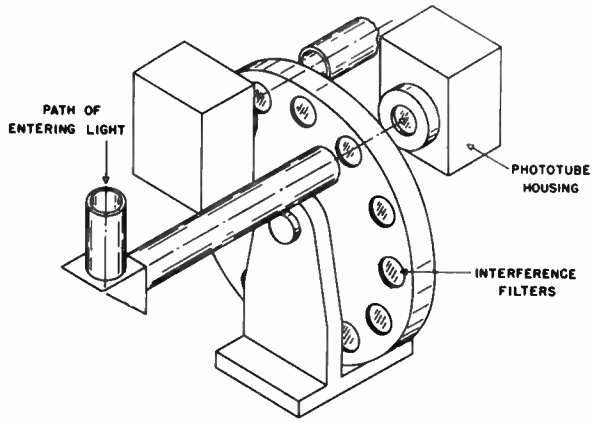


Fig. 1 - Rocket-borne spectrophotometer.

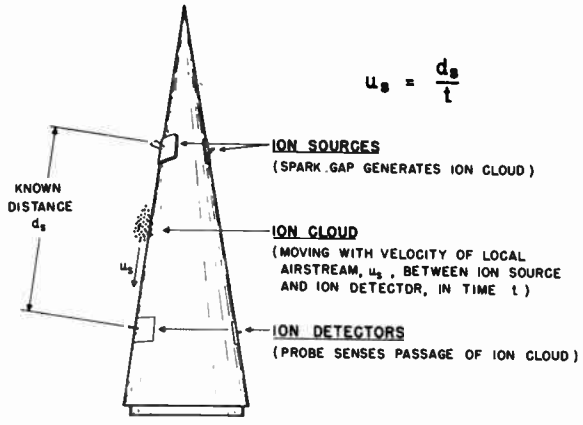


Fig. 2 - Local airstream velocity, u_s.

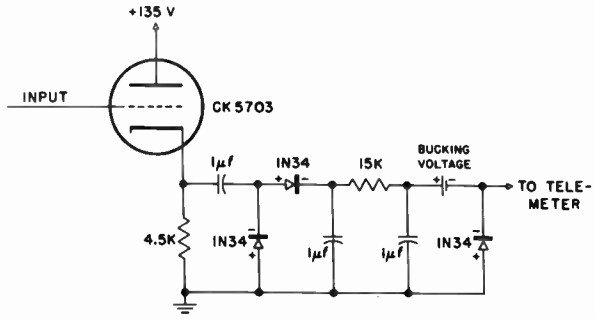


Fig. 3 - Output of solar constant equipment.

TELEMETERING: WIDEBAND ON SHORT ORDER*

Thomas F. Jones, Sc.D.
General Electronic Laboratories, Inc.
18 Tremont Street, Boston, Mass.

Introduction and Abstract

To meet an emergency need of an armed-service research laboratory, a multi-channel telemetering system was developed with unusually wide information bandwidths. The principal problems were systematically reduced to unified practical solutions embodied in an f-m--a-m system which included an airborne relay. Although the equipment design herein described was not seriously constrained by such frequently occurring considerations as size and weight, the designs developed are of general value to the telemetering field because the challenge of the general multi-channel, wideband problem has been successfully met.

The Proposition (or, Whence Came the Order)

The Naval Research Laboratory needed a telemetering system which would provide 24 channels, each having a d-c--10 kc bandwidth. Delivery of the equipment was required in three and one-half months--whence came the title of this paper. Fortunately, the NRL group had laid the ground work of initial study and were thereby able to place their "short order" in terms which obviated the considerable planning period normally associated with so gross a program. In other words, they knew what they wanted and they knew what might be within the realm of possibility in the short period available.

The Proposed System

NRL proposed that the system take approximately the form shown in Figure 1. As proposed, each of three separate systems, identical except for choice of spatial transmission frequencies, provides eight wideband information channels. In each case, eight input signals are converted to frequency modulated subcarriers which are mixed and amplitude modulated upon an r-f carrier operating between 250 and 300 mc. Next the signals are received in an aircraft, demodulated, and then amplitude modulated upon a carrier operating between 350 and 400 mc which in turn is intercepted by a terminal receiver and demodulated into a signal representing the sum of the subcarriers. Subcarrier demodulators then separate out the various signals and demodulate them into the desired information components.

A careful study of the gross problem confirmed the NRL plan as a worthy one and accordingly it was pursued at a rapid pace.

Elementary Blocks and Problems

Frankly, the outstanding problem was the schedule. Even a simple problem is difficult when a dead line is iminent--consider the man on a quiz program--or consider anyone of us trying to fill out a financial questionnaire for the U. S. Government two weeks ago! However, each elementary block presented problems of interest and challenge. Ten basic problems and the elementary blocks in which they occurred are presented in Figure 2. These building blocks will be taken up, each in turn, and some of the principal problems associated with them will be discussed.

The Subcarrier Generators

Reference to Figure 2 reveals a "galaxy" of problems in the subcarrier generator package. The importance of this block of the system justifies a somewhat detailed discussion of the principal problems and the way solutions were achieved. The spectral problem was posed by the requirement for eight f-m subcarriers, each having an information bandwidth of 10 kc. A deviation ratio of five was decided upon as being more than adequate, and a basic study of the theory of frequency modulation indicated both the desirability of accepting frequencies up to twice the deviation and the feasibility of separating contiguous channels of that width. Besides, this choice favored the use of commercially available i-f assemblies. The spectral need resolved itself into a simple equation:

> 8(channels) x 10(kc information bandwidth) x 5(deviation ratio) x 2(to accomodate frequencies up to twice the deviation for higher order sidebands) x 2(upper and lower sidebands) = 1600 kc = 1.6 mc total bandwidth.

Where should this band be positioned? Considerations of linear deviation of a subcarrier generator plus concern over second-order distortion products indicated a lower limit somewhat greater than 1.6 mc. A critical study of third-order

*The work described herein was sponsored by the Office of Naval Research under the technical direction of the Naval Research Laboratory.

distortion products set the specific subcarrier frequencies at 1.975 mc, 2.184 mc, 2.384 mc, 2.604 mc, 2.844 mc, 3.100 mc, 3.392 mc, and 3.664 mc. This choice of band position was also influenced by the approximately 4 mc passband of the "Block III" (CRV-52ABW) transmitters which were made available for this development.

Stability in an f-m subcarrier generator is a serious problem whenever a reactance tube is used. NRL furnished a basic design which met the specifications in several particulars and showed considerable promise. Further development, including such measures as optimum adjustment of the drive and electrode voltages of the reactance tube, increased the deviation sensitivity, linearity, and stability. The final figures were:

Frequency change ÷ heater voltage change = 1 kc/volt

Frequency change ÷ plate supply voltage change ≈ 5 kc/volt

Note: The actual deviation in the plate supply voltage is negligible.

kc deviation ÷ volts to reactance tube grid = 30 kc/volt

Linearity for ± 50 kc deviation = ± 1 per cent of full scale.

The circuit diagram is shown in Figure 3. The resistors in series with each input filter out v-h-f feedback. The diodes are limiters which prevent channel overlapping on greater than full-scale signals. The potentiometer permits setting the over-all deviation sensitivity of each subcarrier generator to the standard value of 10 kc/volt.

The subcarrier generator, being especially sensitive to shock, also presented problems in layout and mechanical designs. The specific applications of the Naval Research Laboratory required acoustical and mechanical isolation of the unit against intense shock. A moderate amount of analysis and a lot of hard work evolved a design which became known as "the derby." (See Figure 4.) By burying the tubes in a heat sink of aluminum and mounting the passive components on a wheel-like chassis thermally isolated from the heat sink by a phenolic disc, several requirements were satisfied simultaneously: temperature rise of temperature-sensitive components was retarded; adequate thermal inertia was provided to permit operation of the unit for significant periods even though all heat was locked in by insulation; adequate heat-exchange area was provided for forced-air cooling; and, an excellent mass-shape for shock mounting inside thick surrounding walls of shock and acoustical (and incidentally but not desirably, thermal) insulation was achieved. (See Figure 5.)

The thermal design permits use without provision for cooling for periods as great as thirty minutes. For longer periods, small blowers connected through acoustically shielded acoustic filters provide stabilization of the subcarrier generator unit at a satisfactory temperature.

The eight subcarrier generators are summed in a resistive network which provides adequate isolation to prevent channel interactions. In the "short order" design, the coaxial cable carrying the subcarrier was fed by the simple vacuum tube circuit shown in the diagram. (See Figure 3.) Since that time, very efficient and effective wide-band matching transformers have been developed. These are admirably suited to the 2-4 mc band. The characteristics of one of these transformers is shown in Figure 6.

Carefully made tests composed of daily checks over a one month period proved the subcarrier generator design to be a good one. The subcarrier frequency error did not exceed 5 per cent of full-scale deviation.

Terminal and Airborne Transmitters

Because of the "short order" feature of the development, the transmitter needs, both at the source and in the airborne relay, were met by modifying "Block III" (CRV-52ABW) transmitters, which were production units of an early airborne television equipment. The four mc video amplifier and the grid modulation system were useful as they were, but in order to permit adjacent channel operation of as many as three equipments, it was necessary to replace the master oscillator of the original MOPA unit with a crystal multiplier "package" which was built into the space formerly occupied by the "synch" generator and the master oscillator.

Serious interferences between the transmitters and the receivers in the airborne relay were prevented by careful choice of carrier frequencies, by double shielding of airborne transmitters, and by spatial separation of transmitting and receiving equipments in the aircraft.

Airborne Receiver

The airborne receiver comprised a low-noise figure r-f amplifier and converter, (See Figure 7.) a crystal oscillator package, and a modified radar i-f strip having a 10 mc bandwidth. The addition of a-g-c to the i-f strip in a particularly satisfactory manner (See Figure 8.) was of considerable importance inasmuch as any overloading of the receiver would have been disastrous to amplitude modulated signals, and the signal strength was expected to vary over wide limits. Furthermore,

these same signals were used immediately to modulate the airborne transmitters which have limitations on dynamic range.

The receiver presented, again, the wideband output problem previously referred to, which was solved quickly by means of vacuum tube circuits. Use of the wideband r-f transformer previously referred to, now available, can improve the efficiency.

Terminal Receiver

The terminal receiver differed from the airborne receiver only in the input stage. This did not require the same consideration for low-noise figure as the input stage of the airborne receiver did because the distance to the airborne relay was to be kept relatively small. A simple first detector-mixer served the purpose. Automatic gain control, crystal oscillator, and output problems were similar to those of the airborne receiver.

Subcarrier Demodulators

Here, too, the "short order" aspect of the problem influenced the development. Commercially available frequency modulation i-f strips were chosen as the nucleus around which the other circuits were designed. A lowpass filter (See Figure 9.) minimized local oscillator interactions among the eight demodulators, while a low-Q parallel resonant circuit provided some of the necessary adjacent channel rejection.

Test points were added to the i-f strips to facilitate alignment, and a zero-center meter connected to the output cathode-follower permitted nulling the cathode-follower to ground--a front-of-panel push button grounds the grid for this operation--and for setting the incoming subcarrier to the null of the discriminator.

Figure 10 is a view of the terminal equipment.

Conclusions

The wideband telemeter described has proved both the feasibility and the applicability of techniques for a multi-channel telemeter carrying 10 kc bandwidth information signals. Although much of the development presented represented compromises dictated by an uncompromisable time schedule, the basic elements of a practical system were proved. The "trail has been blazed" for modification and refinement to meet specific wideband telemetering needs which are arising more and more frequently in our rapidly expanding field.

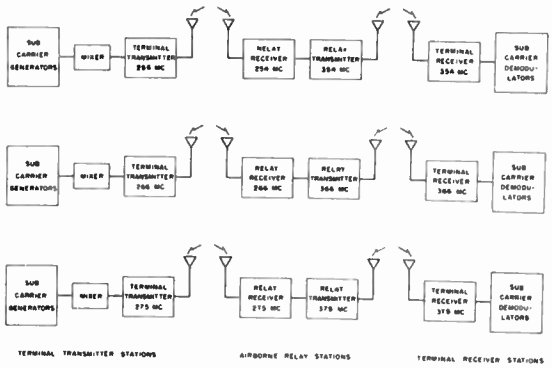
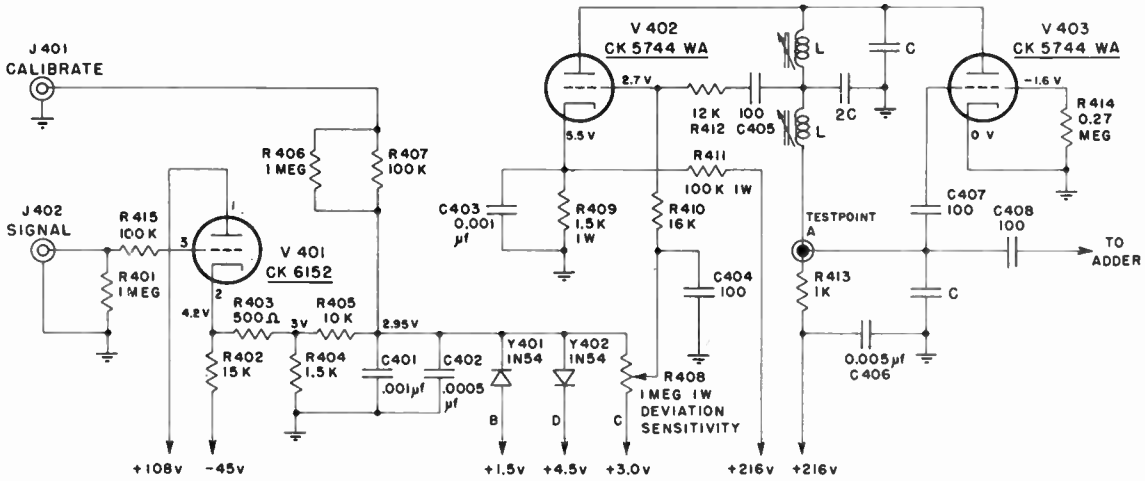


Fig. 1
Functional block diagram
24-channel wideband telemetering system.

	SPECTRUM CHOICE	STABILITY	THERMAL	FILTERING	OVERLOAD	ANTENNAE	LAYOUT	CIRCUIT DESIGN	SHOCK & VIBRATION	POWER SUPPLY
SUBCARRIER GENERATOR	X	X	X	X	X		X	X	X	X
TRANSMITTER AT SOURCE	X	X				X	X	X	X	
AIRBORNE RECEIVER	X	X		X	X	X		X		
AIRBORNE TRANSMITTER	X	X				X	X	X	X	
TERMINAL RECEIVER	X	X			X	X				
SUBCARRIER DEMODULATORS	X	X		X						

Fig. 2
Problem distribution among
basic system blocks.

CHANNEL	1	2	3	4	5	6	7	8
C	120	100	75	75	50	50	50	43
2C	200	200	150	150	100	100	100	100
L	1 MC COIL 110 TURNS REMOVED	1 MC COIL 115 TURNS REMOVED	1 MC COIL 125 TURNS REMOVED	1 MC COIL 125 TURNS REMOVED	1MC COIL 125 TURNS REMOVED	5 MC COIL OK AS IS	5 MC COIL OK AS IS	5 MC COIL OK AS IS



RESISTORS ARE $\frac{1}{2}$ WATT, EXCEPT AS NOTED.
CAPACITORS IN μμf, EXCEPT AS NOTED.
KEEP DEVIATION SENSITIVITY POTENTIOMETER FULLY COUNTERCLOCKWISE (TAP AT TOP).

Fig. 3 - Schematic, subcarrier generator.

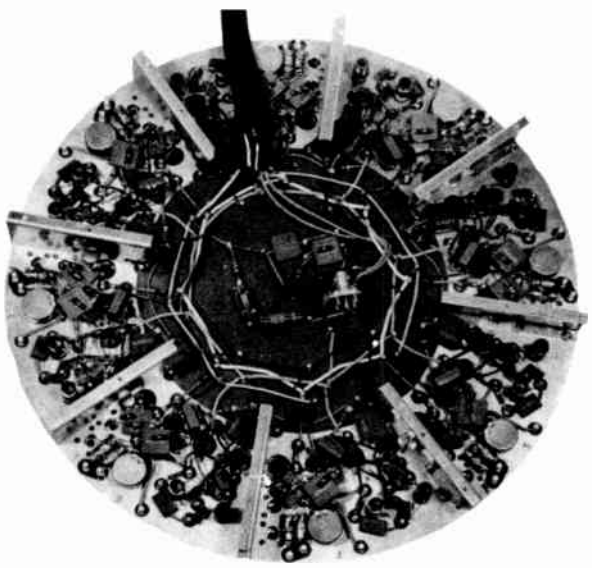


Fig. 4
Subcarrier generator; bottom of chassis.

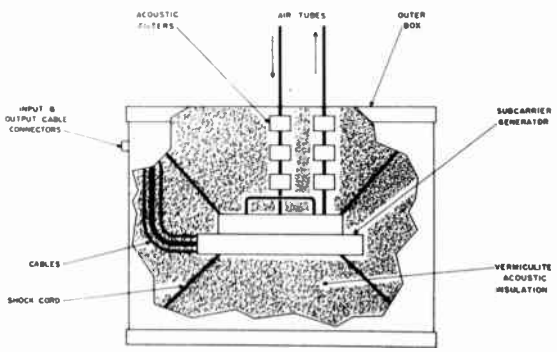


Fig. 5
Suboarrier generator assembly.

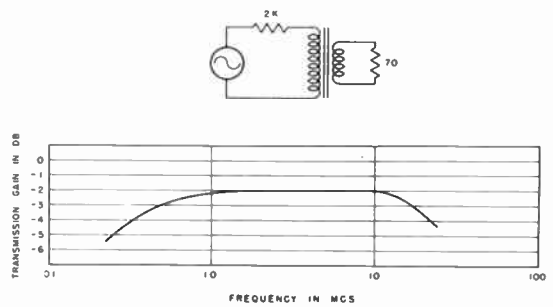


Fig. 6
Wideband R-F transformer characteristics.

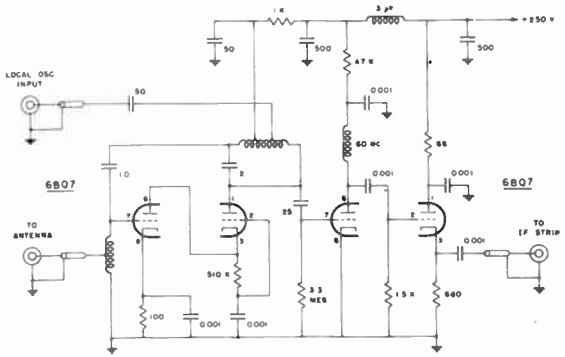


Fig. 7
Airborne receiver, low noise input circuit.

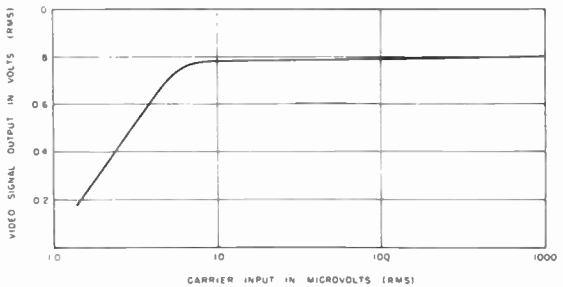


Fig. 8
Receiver AGC characteristics
50 % modulation.

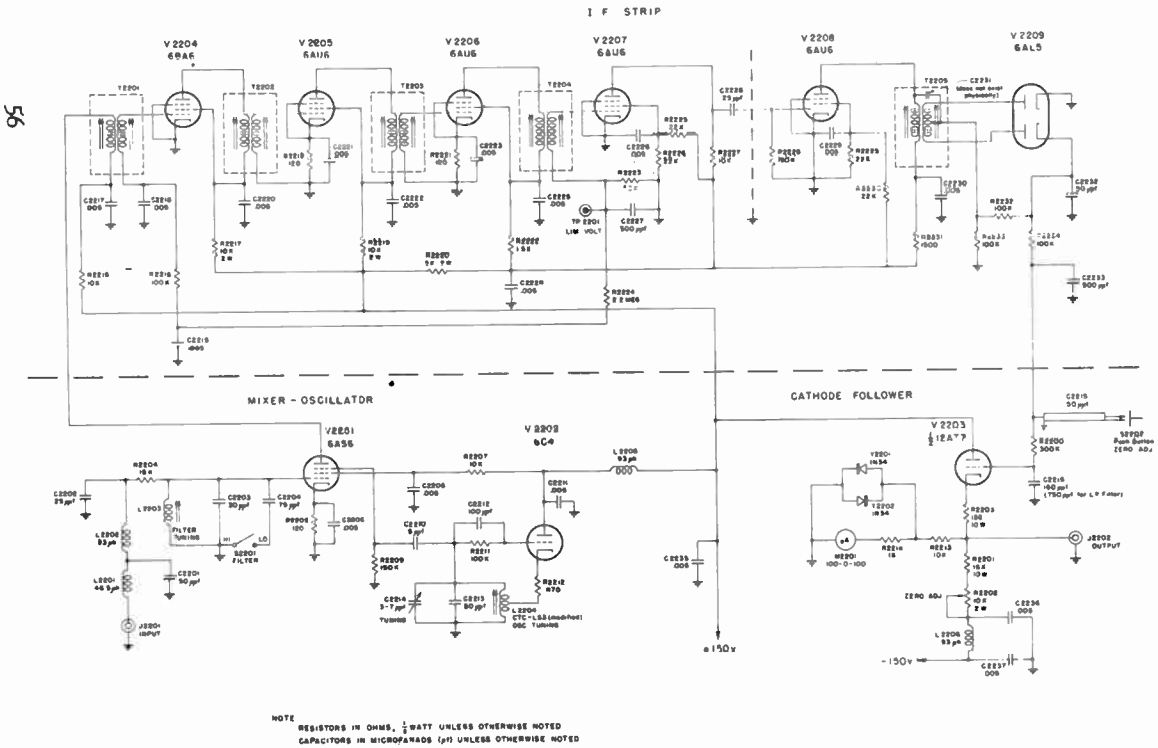


Fig. 9 - Schematic, demodulator.

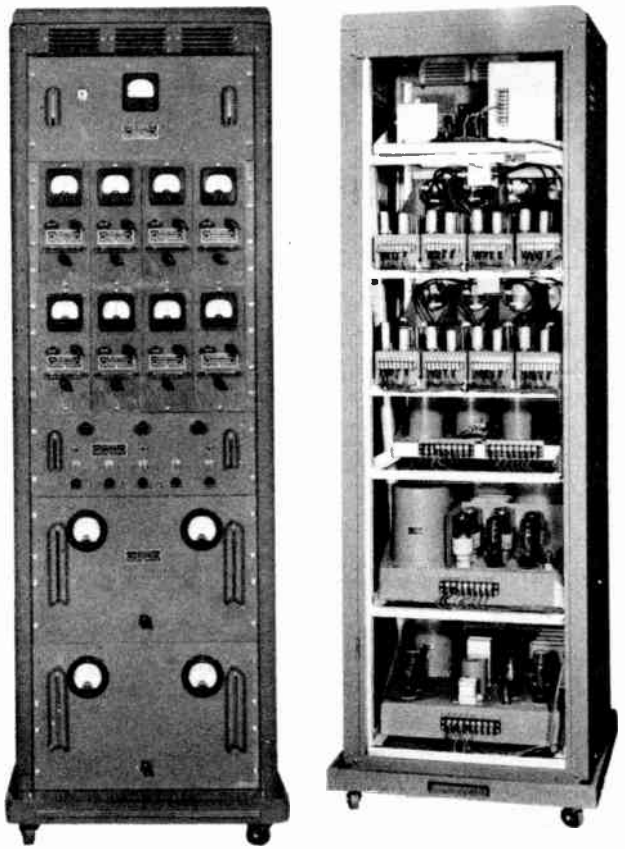


Fig. 10
Terminal receiver station.

FLUTTER COMPENSATOR FOR FM/FM TELEMETERING RECORDER

John T. Mullin
Bing Crosby Enterprises, Inc.
9030 Sunset Boulevard
Los Angeles 46, California

Summary. Records of FM/FM telemetering information reproduced from magnetic tape are restricted in the accuracy with which information may be read because of flutter and DC drift in the tape transport mechanism. A method is described wherein the effects are reduced by electronic compensation during playback. A high frequency pilot tone is added to the FM/FM signals as they are recorded. During playback all channels are modulated by a high frequency carrier. The resulting side bands are demodulated separately for each channel by a carrier whose frequency is controlled in an absolute manner by information derived from the pilot tone. Demodulation thereby restores the channel frequencies to their correct values with variations due to flutter and DC drift greatly reduced.

Introduction

The magnetic tape recorder has become an indispensable tool in many phases of instrumentation work. Its first introduction into this field was by direct application of unmodified sound recorders. Even in the best of such machines a number of deficiencies were obvious from the start, resulting in design modifications which established tape recording on a firm basis for technological work. Speed was increased from 15 inches to 30 and 60 inches per second, and recording and playback circuits were modified to eliminate the pre- and post-emphasis correction employed in sound recorders and to extend the response to as high as 100 kc.

Most of these machines have found application in systems of multi-channel FM telemetering where any individual channel is assigned a center carrier frequency somewhere between 430 and 70,000 cycles. Information is conveyed by a particular channel by deviation of its carrier above and below the center frequency. Any given channel frequency is never shifted up and down in excess of 7-1/2%, or in some cases, 15% of its normal center frequency, since excessive excursion would cause cross talk into adjacent channels.

An available spectrum of roughly 400 to 70,000 cycles permits a total of at least 16 separate channels of the 7-1/2% deviation type to be employed simultaneously. A complex piece of apparatus under test must be observed with the greatest of accuracy, and since a large number of channels may be employed, the difficulty of making simultaneous permanent records from the output of the various channel discriminators can become of unmanageable magnitude. Hence the application of the magnetic tape recorder. The composite signal, consisting of all mixed channels, is received from the apparatus under test and recorded directly on tape. Later, on playback, the discriminators are used to produce final records for analysis. This is a thoroughly workable method which suffers from only one serious drawback.

Wow and Flutter Limitation and DC Shift

Any variation in velocity of the tape, either while it is being recorded or while it is being played back, causes a variation in the carrier frequencies. Since information is conveyed only by frequency modulation of the carriers, tape speed variation manifests itself as extraneous carrier frequency modulation, thereby limiting the signal-to-noise ratio of the system and consequently the accuracy to which information may be determined in the final analysis. Since maximum excursion of a typical carrier is $\pm 7.5\%$ of its center frequency, any speed variation in the tape itself appears to be amplified at the output of the channel discriminator by a factor of 6.67 times. As a typical case, if a tape playback machine shows an overall peak-to-peak speed variation of 0.3%, it can be considered to be a highly acceptable sound recorder, but when used as part of an FM telemetering system the final information can be read only to an accuracy of: $0.3 \times 6.67 = 2.0\%$. Long-term averaging can sometimes be employed during examination and analysis of the string oscillograph records, but even with this aid it often happens that important changes of information, the magnitude of which is somewhat less than the noise limit, cannot be detected. With the exception of one recorder[1] now in production which has been specially designed to assure constancy of tape speed, the problem of speed variation is continually with the user of instrumentation recorders.

Long-standing definitions of the effect of various rates of speed change were derived from their ability to be detected by the human ear as unwanted extraneous effects. Thus, speed variations at cyclic rates of under six cycles are called wow, while above six cycles the applicable term is flutter.

Wow and flutter in a typical tape recorder cannot be fully analyzed in terms of a few strong spectrum lines. While there are always certain prominent lines, particularly in the lower frequencies, due to rotational rates of various shafts, analysis of the flutter frequencies shows a widespread spectrum of noise.

Fig. 1 shows a spectrum analysis of flutter in a typical tape recorder of the form most wide-

ly employed in instrumentation recording. A few prominent rotational lines will be observed, and

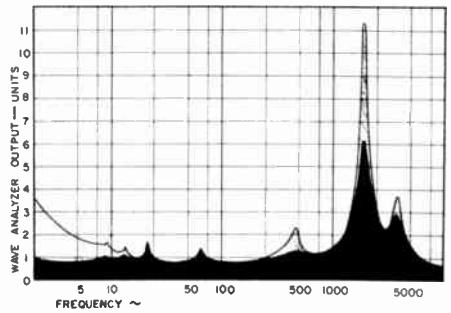


Fig. 1

a widely varying wow spectrum is indicated at the lower end. Such low frequency wow cannot be defined any closer than indicated except at a particular instant, since the rotational rates vary continuously from the start to the end of a reel of tape, and their amplitudes are determined by such factors as the accuracy of alignment of reels on their spindles, care taken in rewinding tape, edge drag on reel flanges, and freedom of shaft bearings.

Of particular interest is the higher frequency flutter spectrum. Here it can be seen that there is an almost uniform distribution of noise up to about 450 cycles where a minor peak is found. There is a very prominent peak at 1800 cycles and another at its second harmonic, 3600 cycles. The amplitude and frequency of such peaks may vary considerably from machine to machine, and with the number of times a reel of tape has been run. Wide variations are also caused by differences in tape tension from beginning to end of the reel. The major peak is very often nearer to 3000 cycles than 1800, as shown here. It is caused by longitudinal vibration of the tape ribbon in passing over heads.

Another deficiency in tape systems arises from the fact that the tape may not travel at exactly the same rate of speed on playback that it did when recorded. This readily happens when AC line frequency drifts or the tape is played back on a machine other than that on which it was recorded, or due to temperature or humidity change of the tape itself which results in an actual change of its physical length. Such variations cause an overall drift or relatively constant shift in the carrier frequencies, thus introducing a DC error in the output of each discriminator. In some ways, this is more intolerable than flutter or wow, since by its relatively steady nature, it is not usually manifest in the final records.

To improve the performance of the recorder which we have been discussing,to the point where none of the above deficiencies would be any longer objectionable would necessitate complete mechanical re-design. A great many such machines are in daily operation, and more are being placed in service regularly. Therefore, a method of overcoming these defects, external to the machine, has long been recognized as being highly desirable.

The device to be described functions in such manner as to reduce the effects of flutter and wow and DC drift by a factor of several times. It is a rack-mounted unit, which is employed on playback only and which is connected between the tape machine and the FM channel units or discriminators. In a normal installation, before installing the compensator, the tape playback is connected to the discriminator units by a single pair of wires (Fig. 2A). Typical discriminator equip-

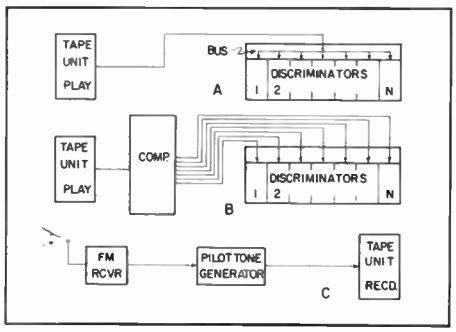


Fig. 2

ment, such as that in the AN-UKR-5 ground station apparatus, takes this signal and by means of an amplifier feeds a low impedance bus across which are connected the various channel units. Each of these responds only to its own frequency, since they each contain a different band-pass filter for channel selection. When compensation is employed, the output of the tape recorder is connected into the compensator, as in Fig. 2B, and after separate correction of each channel, the compensator individually feeds each of the channel units. The compensator operates by reference to a pilot tone which must be put onto the tape records at the time they are recorded. Fig. 2C indicates the method in which this is done, wherein a portable pilot tone generator is employed to add a 50 kc or 100 kc crystal controlled frequency to the other signals which are being recorded from the signal source. More details of this unit will be given later. Suffice to say that the selection of 50 or 100 kc is made on the basis of whether the recorder is to operate at 30 or 60 inches per second. In either event, the pilot chosen lies near the extreme upper end of the recordable spectrum.

The compensator is comprised of frequency multipliers, dividers, adders and subtractors and functions in the following manner, wherein we will follow a typical channel frequency S. Let us assume S is the center frequency of the channel and remains unmodulated by information during the discussion. Coming from the tape playback is the carrier S and a high pilot frequency,P. Initially, we will further assume that there is no flutter in the tape and, consequently, P and S remain

fixed frequencies. Referring to Fig. 3, S and P are separated by high and low-pass filters, and the latter is divided by a frequency divider whose

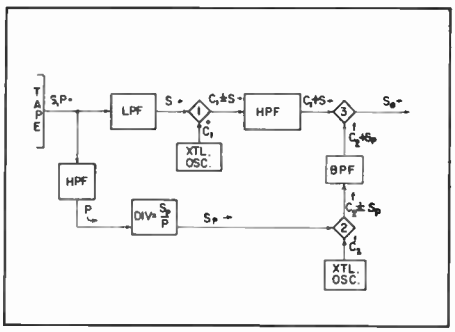


Fig. 3

output is as close to frequency S as it is possible to come. If we assume that P is a simple multiple of S in this initial description, then it is possible for the divider to derive a frequency S_p where:

$$S_p = S$$

Meanwhile, S is modulated by a crystal controlled carrier C_1, in modulator 1, resulting in upper and lower sidebands $C_1 \pm S$. The upper set are selected by a high-pass filter HPF and $C_1 + S$ enter another modulator 3. The output of the divider S_p is added to another crystal frequency C_2 to give sidebands $C_2 \pm S_p$. A band-pass filter BPF selects $C_2 + S_p$, and rejects other unwanted products, and $C_2 + S_p$ is also applied to modulator 3. The output of modulator 3 is an output signal S_o which is the difference of the two input products.

$$S_o = C_1 + S - (C_2 + S_p)$$

C_2 is chosen such that when added to S_p it is equal to C_1, or:

$$C_1 = C_2 + S_p$$

and $$S_o = C_1 + S - C_1$$

or $$S_o = S$$

Thus, the output for a given channel is identically the same frequency as the input to the compensator.

Considering now the action of the compensator in the presence of flutter or drift, we will represent values of flutter by Δf where Δf is an instantaneous percentage of any particular frequency f.

In the above expressions we may substitute $S \pm \Delta S$ for S and $P \pm \Delta P$ for P, and since S_p is related to P by a constant factor of division, we may write

$$S_p \pm \Delta S_p \text{ for } S_p$$

Then: $S_o = C_1 + (S + \Delta S) - [C_2 + (S_p + \Delta S_p)]$
but, $C_1 = C_2 + S_p$
thus, $S_o = S + \Delta S - \Delta S_p$.
Since $\Delta S = \Delta S_p$
then $S_o = S$

And the flutter, wow, or DC drift has been cancelled out.

We have assumed S_p to be equal to the center frequency S of a particular channel. If we are still considering a channel with plus or minus 7.5% deviation from this center, the correction at the edges of the band will be derived. Let the edge frequency be S_e.

$S_e = S + .075S = 1.075S$ for upper edge.
or $S_e = S - .075S = 0.925S$ for lower edge.
thus $\Delta S_e = 1.075\Delta S$ or $0.925\Delta S$.
Since $\Delta S = \Delta S_p$
then $\Delta S_e = 1.075\Delta S_p$ or $0.925\Delta S_p$
and $S_o = S_e + \Delta S_e - \Delta S_p$

For the lower edge:

$S_o = S_e + \Delta S_e - \Delta S_p$
or $S_o = S_e + .925\Delta S_p - \Delta S_p$
$S_o = S_e - .075\Delta S_p$
but $S_p = 1.075\, S_e$
thus $S_o = S_e - 1.075 \times .075 \Delta S_p$
or $S_o = S_e - .081 \Delta S_e$

Whereby ΔS_e is reduced by a factor of 12.3 to 1.

For the upper edge:

$S_o = S_e + \Delta S_e - \Delta S_p$
or $S_o = S_e + 1.075\Delta S_p - \Delta S_p$
$S_o = S_e + .075\Delta S_p$
but $S_p = .925\, S_e$
thus $S_o = S_e + .075 \times .925 \Delta S_e$
or $S_o = S_e + .069\Delta S_e$

Whereby ΔS_e is reduced by a factor of 14.5 to 1.

Should S_p lie at the lower end of a channel, correction will be ideal at that end and will degrade to a minimum improvement at the upper end given by

$S_o = S_e + \Delta S_e - \Delta S_p$
Where $\Delta S_p = .87\Delta S_e$
or $S_o = S_e + \Delta S_e - .87\Delta S_e$
thus $S_o = S_e + 0.13\Delta S_e$

Whereby ΔS_e is reduced by a factor of 7.7 to 1.

When S_p is at the upper end of a channel, correction will be ideal at that end and will degrade to a minimum improvement at the lower end, given by:

$$S_o = S_e + \Delta S_e - \Delta S_p$$

Where $\Delta S_p = 1.15 \Delta S_e$

or $S_o = S_e + \Delta S_e - 1.15 \Delta S_e$

thus $S_o = S_e - 0.15 \Delta S_e$

Whereby ΔS_e is reduced by a factor of 6.6 to 1.

Since it is possible in all cases to derive a frequency S_p that lies within each specific band of frequencies normally employed in these instrumentation systems, the theoretical improvement can always be greater than 6.6 to 1, but in practice this cannot be fully realized for reasons which will be enumerated later. Fig. 4 shows in graphic form the variation in correction to be obtained as the carrier deviates from its lower limit f_L to its upper limit f_H in a typical 7.5% channel.

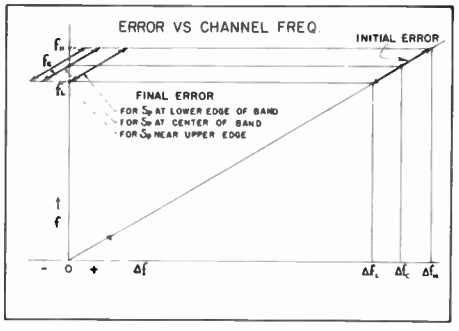


Fig. 4

Design Details

Fig. 5 shows a block diagram of the expanded system as it has been necessary to develop it for a full complement of channels lying within the range of 430 to 70,000 cycles.

Initially, it was hoped that satisfactory sideband separation could be achieved whereby the upper sidebands could be accepted and the lower set rejected after modulation in bridge modulator No. 1. Since it is necessary to employ a carrier frequency of about 100 kc, the separation becomes impractical for channels as low as 430 cycles, since the lower sideband starting at 99,570 cycles must be rejected at least 60 db below the accepted upper band which starts at 100,430 cycles. This becomes an impossible task with anything other than a crystal filter. Even if such a filter were to be employed, any slight drift in the carrier crystal oscillator would produce intolerable shift later in the demodulated lower channels. Again the separation of unwanted products in the addition of divider output plus demodulating crystal oscillator must be thoroughly suppressed. This is achieved by the band-pass filter which feeds modulator No. 3, and if this had to pass 100 kc with 60 db rejection 430 cycles on either side, another major problem would be encountered.

Because of these limitations, the entire system was broken into two similar units. One of these employs a carrier of 100 kc and handles frequencies from 5,000 cycles to over 80,000, while the other employs a carrier of 10 kc and only handles those channels which lie between 430 and 5,000 cycles. This results in more reasonable filter design and fabrication due to the satisfactory separation of upper and lower sidebands.

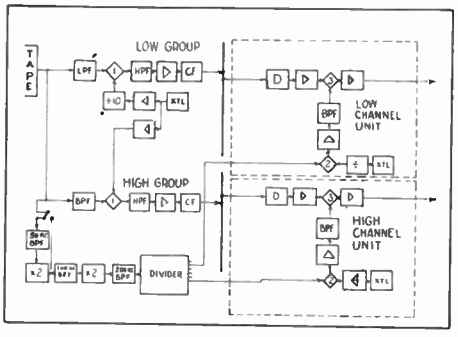


Fig. 5

Examination of the two modulator units shows that they are basically similar. A single 100 kc crystal is used through an isolation amplifier as the carrier source for the high group. Modulator No. 1 is of the ring type, employing a quad of germanium crystals at very low signal level and carefully balanced to assure a minimum of unwanted modulation products.[2] A fairly broad band-pass filter accepts the channel frequencies above 5,000 cycles but cuts drastically before 100 kc to assure removal of the 100 kc pilot tone coming from the tape.

Since the quad is operated at low level, the output of the high-pass filter which selects the upper sidebands must be greatly amplified. A two-stage feedback amplifier drives a cathode follower which provides signal onto a low impedance bus. This bus then becomes the source of signal for all demodulators of the form shown within the dotted area. One such unit is required for each channel, and while they are all fundamentally the same, they all differ slightly depending on the channel they are to serve.

The modulated signal bus input impedance of a typical channel unit is very high and looks into a delay correction network. After the delay network, an isolation amplifier is required which feeds quad 3. Here demodulation takes place and it becomes necessary to employ another amplifier tube to provide sufficient level at 600 ohms to

satisfy the discriminator which will work from this unit.

Meanwhile, coming in from the divider bay is the specific frequency S_p derived by division of the pilot tone. This is far from a pure sine wave, and after modulation by the crystal oscillator via quad 2, a large number of products result which must be suppressed by the band-pass filter before the addition of divider and crystal can be applied to quad 3.

In this process, it becomes necessary to employ some amplification. It has been found that the crystal oscillator itself cannot be applied directly to quad 2 but must be isolated by a buffer amplifier. It then becomes possible to run the oscillator with about 35 volts of plate supply, which assures even greater stability. The low group of frequencies is handled in essentially the same way as those of the high group. The 10 kc carrier is derived from the 100 kc crystal by frequency division, since its absolute frequency must be accurate to the same percentages as that of the 100 kc carrier. After modulation, filtering, and amplification, the upper sidebands appear on a bus of low impedance just as was done for the upper group and individual channel demodulators are similarly employed.

Since the divider output must add to a very stable source of frequency to produce 10 kc, it is necessary to employ crystals for the purpose. Since this frequency must lie between 5,000 cycles and 9,500 cycles, it is most practical to use higher frequency crystals and divide their output by appropriate amounts to provide the frequencies required.

The pilot frequency may be either 50 kc or 100 kc and, consequently, by means of a selector switch, it is passed either directly into a 100 kc band-pass filter or into a 50 kc band-pass filter followed by a frequency doubler. In either case, the signal passed through the 100 kc band-pass filter is doubled to 200 kc where it is cleaned up into a sine wave again by another band-pass filter. Each of these three filters is very wide and contributes a minimum of delay distortion and side band cutting. It is necessary to start the chain of dividers from 200 kc in order to enable all S_p frequencies to fall within the deviation limits of the channels to which they apply. As an example, the S_p for the 70,000 cycle channel is readily derived by dividing 200,000 cycles by 3, giving 66,667 cycles, which lies within the lower limit of the band, the limit for a 7.5% channel being 63,750 cycles. Obviously, no such satisfactory frequency could be derived from 100 kc by division. The dividers themselves are of the blocking oscillator type, with an adjustment in the cathode of each divider to optimize the stability. A total of 21 is required to provide suitable values of S_p for 16 channels.

The delay correction networks are essential to assure satisfactory correction of flutter at the higher rates. In the higher frequency channels the signal arrives at quad 3 over the direct path from quad 1 about 3/4 millisecond earlier than it does via the divider chain and quad 2. Practically all this delay is occasioned by the band-pass filter between quads 2 and 3, and it has been necessary to degrade this as much as possible to minimize such delay. Some additional delay is derived from the frequency doubler filters, of course, but the dividiers themselves contribute a negligible amount. It has therefore been found essential to add delay on the other path of the circuit. These networks are of a relatively simple type, consisting again of band-pass configurations. They are not easy to construct, however, and the lower frequency channels necessitate increasingly difficult corrections.

Physical Arrangement

The entire compensator is contained in one 72-inch rack, as illustrated in Fig. 6. The top of the rack contains the power supplies, of which

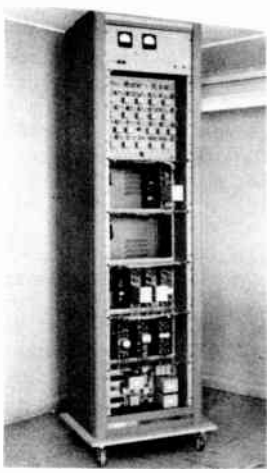


Fig. 6

there are two behind a single panel. One of these is regulated and supplies the divider chain. The other supplies unregulated plate current to the rest of the equipment.

Immediately beneath the power panel are the pilot selector filters and frequency multipliers and dividers.

At the bottom of the rack, mounted on a single panel, are two filament transformers and the two modulator units for the 10 kc and 100 kc groups, including a thermostatically controlled crystal.

In the four rectangular spaces between the dividers and modulator chasses can be seen seven demodulator units for various channel frequencies. One of these units may be examined in detail in Fig. 7. All tubes are available from the front of the unit, and the crystal is mounted as a plug-in unit at the rear. This assures a minimum of frequency drift, since the inside of the cabinet never becomes very warm.

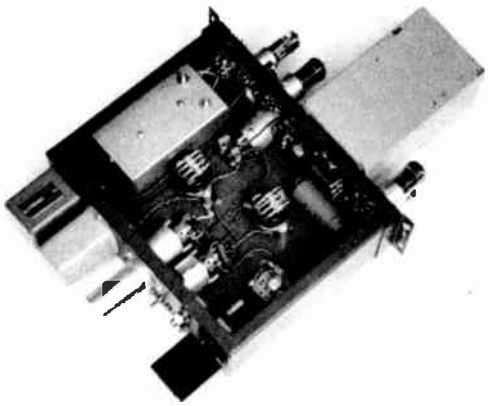


Fig. 7

The pilot generator is shown in Fig. 8. Cannon connectors provide the means of passing all the signals which are to be recorded through the unit before they get to the tape recorder. The unit adds the 50 kc or 100 kc pilot derived from a crystal to the rest of the signals, and the pilot amplitude is variable by one knob, while the other selects the choice of pilot frequency. The levels of the input signals remain unchanged, and they reappear in the output without experiencing distortion.

Since the operation of adjusting delay and maintaining band width is a time consuming one, particularly so when each channel presents a different problem, it must be undertaken separately.

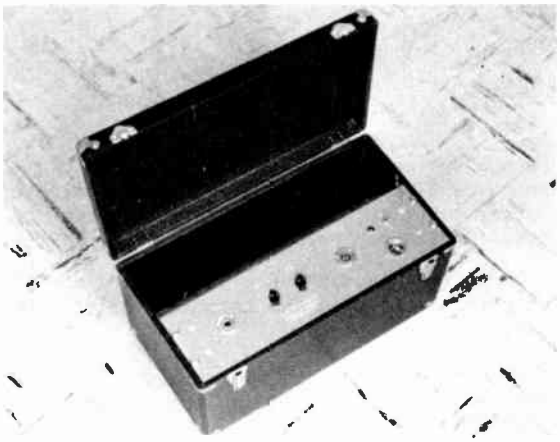


Fig. 8

Performance

Fig. 9 shows the results which can be obtained in compensating various channels. Considerable improvement will be observed in all the traces displayed. While these enlarged sections do not show the improvement which can be gotten on wow due to the long wave lengths involved, they do show the effectiveness of high frequency flutter correction. The 70 kc channel discriminator output was observed through a 4,000 cycle cut-off filter, which was employed to simulate typical service conditions. Thus, it will be observed that all components out to at least 4,000 cycles were greatly reduced in amplitude The oscillogram was recorded from a carrier of exactly 70,000 cycles. This was not the ideal correction frequency, but the residual discrepancies after correction are mostly due to factors which are yet to be discussed. This is true, also, of the traces of the 10,500 cycle and 960 cycle channels. The latter two traces show obvious improvement in wow and low frequency flutter, but the higher frequencies are not corrected as fully as in the case of the 70,000 cycle carrier. This is partly due to modulation products which result in interfering tones within the pass band of the channel itself.

Such interfering frequencies have been a major concern in the development of the compensator, since they can produce serious peak-to-peak deflections of the otherwise corrected trace. We have found that a single frequency lying within the pass band of any channel must be suppressed at least 60 db below the carrier in order to be ineffective as a noise signal in the output. In processing each channel through modulation, demodulation and filtering, it is quite possible for extraneous frequencies to be generated. These must be suppressed by rejection filters to a satisfactory degree. In some channels this has proved to be relatively easy, while in others it has presented a major problem. How some of these products arise will now be shown.

In the first modulation process C_1 plus S results in two sidebands, $C_1 + S$ and $C_1 - S$. The high-pass filter is intended to reject $C_1 - S$. If any of this component gets through, however, it can result in trouble after the demodulator if the carrier $(C_2 + S_p)$ is different in frequency from C_1 by any slight amount. The two demodulated frequencies then have a frequency difference equal to twice the carrier frequency differences and the phase modulation rate will be cyclical at this difference frequency

Again, interference can come about from such products as $C_2 + 2S_p$ or $C_2 + 3S_p$ being demodulated by $C_1 + 2S$ or $C_1 + 3S$ while $C_1 + S$ is being demodulated by $C_2 + S_p$. Whenever C_1 does not exactly equal $C_2 + S_p$, then such form of interference can arise. This can be a very real source of trouble since, for example, it is difficult to make the filter BPF sharp enough to accept $C_2 + S_p$ and reject such unwanted signals as C_2, $C_2 - S_p$ and $C_2 + 2S_p$. S_p itself is strong in harmonics because it is essentially the output wave form of a staircase divider. Any attempt to clean it up by filtering incurs additional delay distortion. BPF itself is limited in its ability to reject some of these products because of a compromise with the delay problem. Under certain conditions of operation the modulators can act like Armstrong Phase modulators and cause further degradation of the signal.

Figs. 10, 11, and 12 show, respectively, the

results obtained when six different channels are simultaneously compensated. In Fig. 10 is shown the record of discriminator outputs bridging the line to the tape recorder as the tape was being recorded. Any fluctuation in these lines is caused by phase modulation in the originating oscillators and by false frequencies generated in the discriminator channel selector filters.

Fig. 11 is the record of playback of these same unmodulated tracks showing wow and flutter introduced by the recorder. In this and in Fig. 12 the 70,000 cycle channel was erroneously not connected, consequently its trace is invalid. Fig. 12 shows the compensated playback. Considerable improvement is evident in all traces, but it can be seen that none of the traces is improved by the factor indicated by the single channel results. Such degradation of performance is due to the large number of side products which arise in the overall process of going through the compensator. There is not a great deal of difference to be observed in dynamic records compared with those shown here for static carrier values. The static condition has been illustrated to facilitate examination of the traces.

Illustrated at "X" in Fig. 12 is another form of error which is actually introduced by the compensator. A spike appears in all traces simultaneously, except for a considerably delayed appearance in the 960 cycle trace. This is caused by a tape "drop-out". Since such tape defects cause a serious loss of playback level for very short wave length signals but seldom affect long waves, they cause loss of count down from the pilot tone. Such discontinuities in the count manifest themselves in all channels as a false correction signal.

In conclusion, a compensator has been developed which provides substantial reduction in error signals caused by flutter and wow and by DC shift. While painstaking means must be employed to reduce unwanted modulation products during the design and construction of the device, its day-to-day operation shows very little tendency to change characteristics. In multi-channel operation its effectiveness is somewhat reduced, and tape drop-outs become more evident in all records than they are without use of the compensator. Research is continuing toward a reduction of drop-out distortion and suppression of interfering modulation products.

References

(1) W. T. Selsted, "Recent Advances in Magnetic Recording for Telemetering Applications". Ampex Electric Corporation, Redwood City, California.

(2) R. S. Caruthers, "Copper Oxide Modulators in Carrier Telephone Systems", Trans. A.I.E.E. vol. 58 pp. 253-260; 1939. Also Bell System Technical Journal, vol. 18, pp. 315-337; 1939.

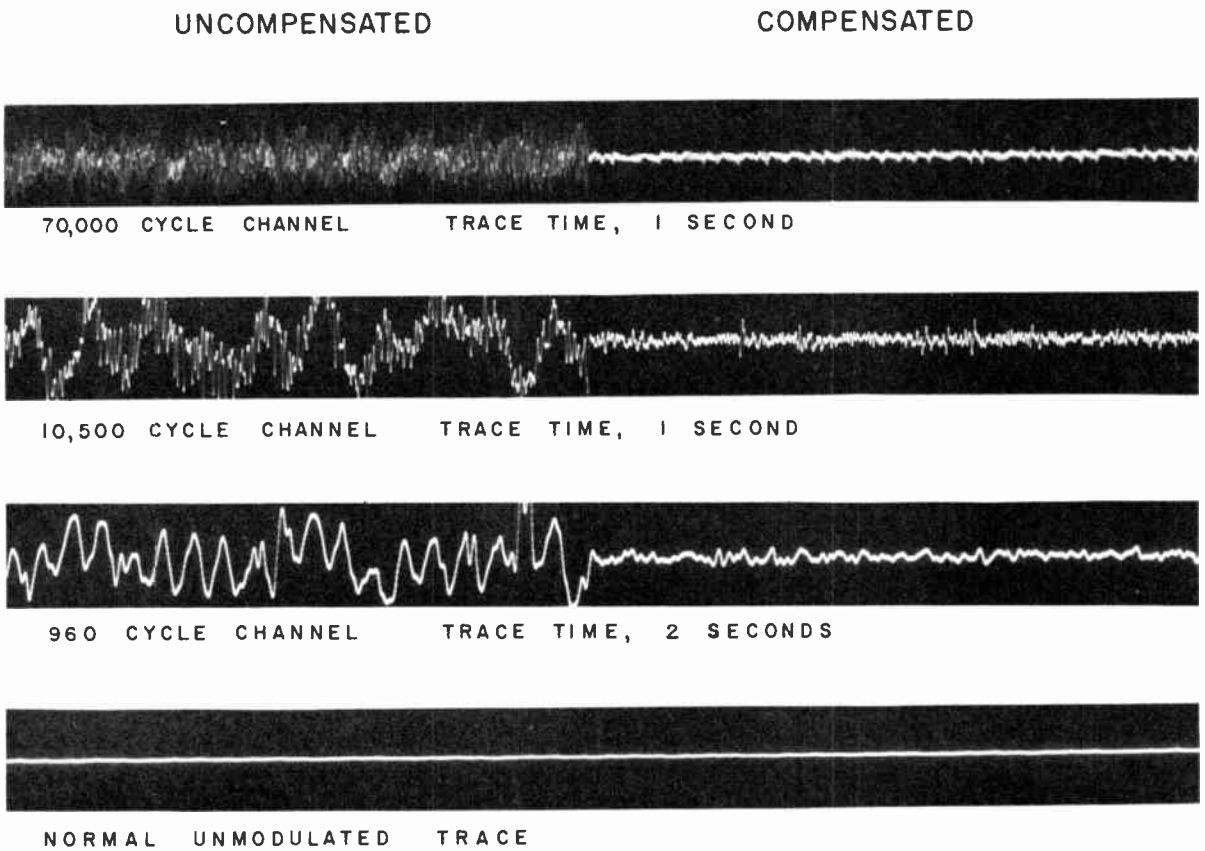


Fig. 9 - Compensated vs. uncompensated playback.

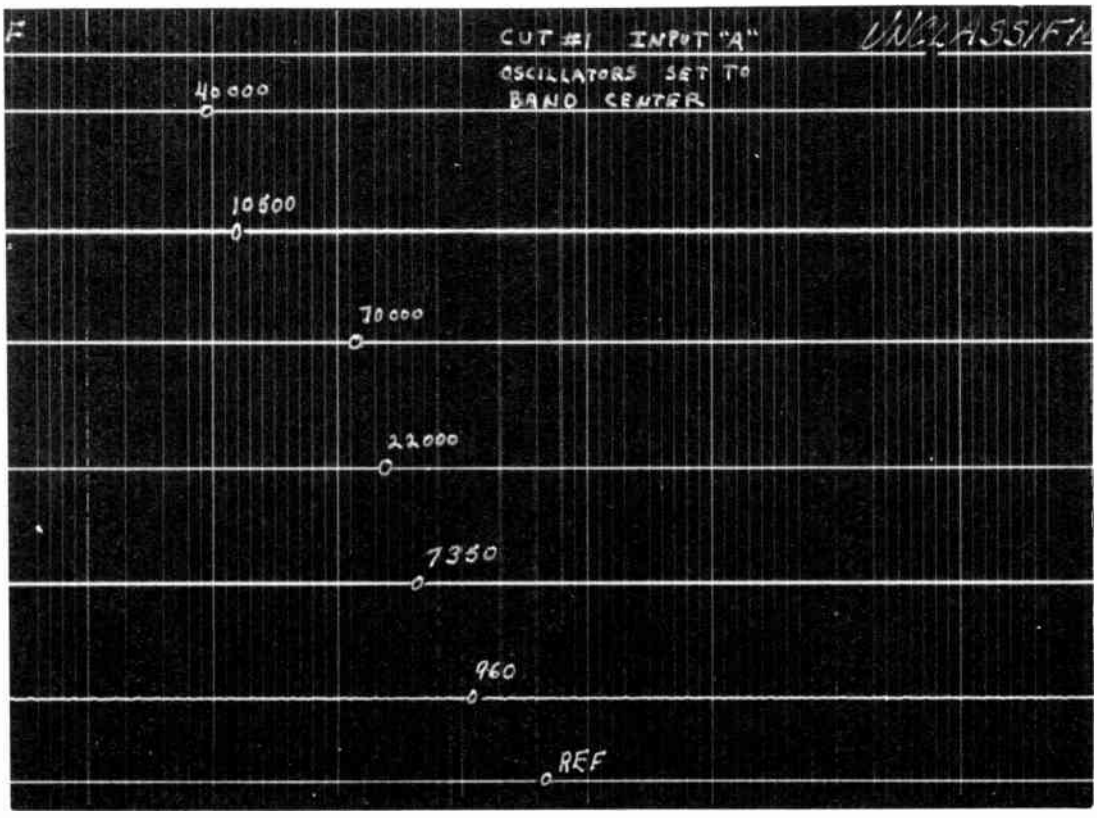


Fig. 10

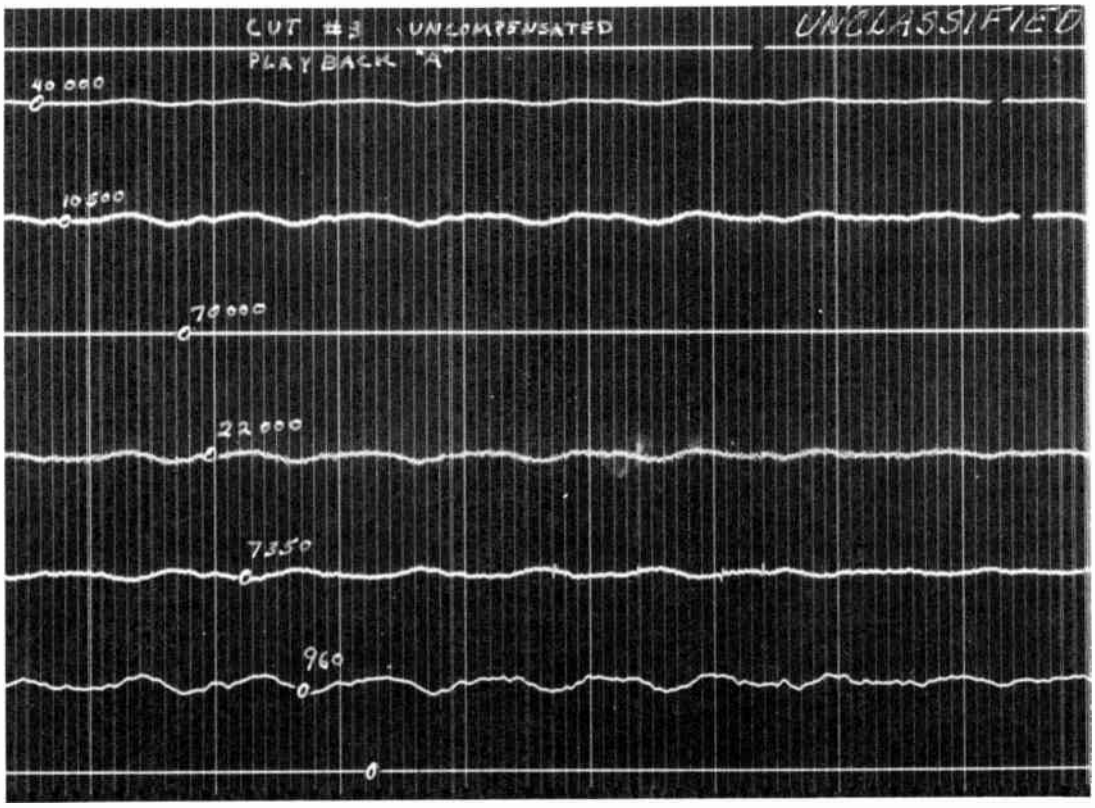


Fig. 11

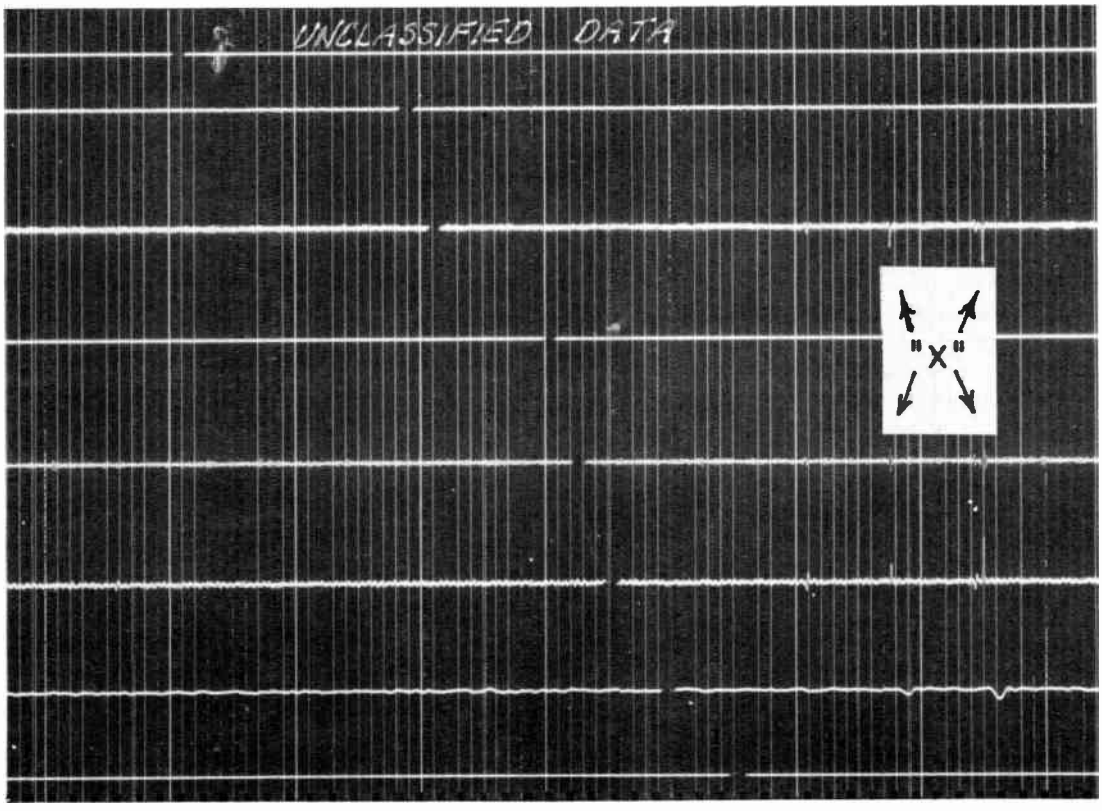


Fig. 12

A MAGNETIC RECORDING SYSTEM FOR PRECISION DATA

Louis L. Fisher
Ampex Electric Corporation
Redwood City, California

INTRODUCTION

Although the principle of magnetic recording was discovered over 50 years ago, it has been only since the end of World War II that this technique has begun to be exploited to its full capabilities. With the development of the use of high-frequency bias, ac erase, and improved magnetic tape materials, the capabilities of this system of recording were found to surpass those of any previously known medium for the recording of sound. Not only was improved performance made possible through greater frequency range, increased dynamic range, and lower distortion, but other advantages of this new medium soon came to light. These include the simplicity of making a recording by eliminating any processing, ability simply to edit and erase, and the permanency of the record which allows countless replays without deterioration. Because of all of these advantages, magnetic tape is now widely recognized as the outstanding medium for recording sound.

These advantages also are desirable in the recording of many other types of information besides audio, so that it was only natural that the technique should be explored further to extend to other fields of application. In the field of instrumentation, magnetic recording offers one other very outstanding feature, particularly in these days when such tremendous quantities of information are being recorded as the result of flight tests, telemetering tests, and similar situations where as much as 100 channels of information must be recorded over periods lasting perhaps several hours. This advantage results from the fact that the reproduced signal is electrical or "live", and, therefore, may be operated upon. This permits the use of automatic devices for translating recorded data, thus saving countless man-hours of work. Also, it permits repetitive and detailed study of any particular signal under a variety of conditions.

Because of the adaptability of magnetically recorded information to automatic data reduction, it has become a necessity in many applications such as the recording of telemetered information from missiles. Extension of the same technique as used in sound recording, that is, the mixing of the signal to be recorded with a high-frequency bias, was achieved to encompass a much wider frequency range, permitting the recording of RDB FM/FM telemetering channels up to 70 KC. Fortunately, the data from the FM/FM telemetering system is in frequency-modulated form, so that some of the characteristics of the direct recording technique which affect instantaneous accuracy are eliminated. However, these limitations seriously affect the results in recording many other types of data, and it was to overcome these limitations that the new FM carrier technique was developed. This new system therefore extends the use of the magnetic recording principle to a wide variety of information, so there remains very little data below 100 KC that cannot be recorded satisfactorily using the magnetic principle.

ADVANTAGES OF THE FM CARRIER SYSTEM

While the direct recording technique of simply mixing the signal with high-frequency bias is applicable to many kinds of information, particularly when the data to be recorded is of a steady state nature and the output will be read on a meter, there are certain kinds of signals on which this system does not yield satisfactory results. Probably the most serious limitation is in the recording of transient information where it is desired to reproduce the instantaneous waveform of the signal. Because the tape coating is not perfectly homogeneous, and the magnetic particles have a tendency to form craters or modules on the surface of the tape, an appreciable variation in instantaneous signal level occurs at the playback head as the passing of a module or defect alters the amount of medium in contact with the playback head gap. Tremendous progress has been made by tape manufacturers in the last few years in making a more uniform coating, and the modules and inconsistencies have been reduced greatly, so that variations in level as read on meters can be held within $\pm$ 1/4 db. However, to reduce such inconsistencies to the extent that the very rapid variations caused by tape nodules can be kept to within the same tolerance poses an extremely difficult requirement, so that even with the best tapes available today, to get accurate instantaneous results requires the use of a different technique.

The effect of such instantaneous amplitude variation is very nicely overcome by the use of an FM carrier. By converting the signal to be recorded into frequency-modulated form, the intelligence is recorded on the tape as a variation in carrier frequency. Therefore, to recover the intelligence on playback, it is necessary only to count the rate at which the signal crosses the zero axis; thus the output signal is independent of any amplitude characteristics of the voltage produced in the playback head.

Another important advantage of the FM carrier system is that it is possible to record low-frequency information down to dc with negligible phase shift. With the conventional high-frequency bias recording system, the output at the playback head is proportional to frequency, so that at low frequencies the output signal from the head falls within the noise level in the playback amplifier, and response to dc is impossible. Also the phase shift produced by the equalizing circuits necessary to achieve flat frequency response introduces serious phase distortion and waveform errors. These effects are, of course, eliminated in the FM system, since the amplitude of the reproduced signal depends only upon the frequency of the recorded signal, and the only phase shift introduced by the system is that of the filter which separates the intelligence from the carrier frequency on playback. It is interesting to note here that the filter characteristic is not a function of the recording, and therefore can be altered as desired on playback.

DESCRIPTION OF THE FM RECORDING SYSTEM

The FM recording system described here has been designed to operate at a tape speed of 30 inches per second. Utilizing a 27 KC carrier, this permits the recording of data from 0 to 5,000 cycles. The principal limitation of the FM carrier system results from variations in tape motion or flutter, which produce changes in the recorded or reproduced frequency, which in turn appear in the output as noise. Since the signal-to-noise ratio of the system is the ratio of the percent flutter to percent deviation for maximum signal, it is obvious that for maximum dynamic range it is desirable to use the widest deviation possible. Therefore, the principal objective in the design of the modulator and demodulator employed in this system was to achieve excellent linearity over the widest possible range of frequency deviation. 40% deviation was found to be the practical limit if overall distortion and linearity were not to exceed 2%.

Because of operating to zero cycles, problems associated with dc drift are introduced. To reduce these effects, an electronically regulated power supply is used, and ballast tubes are provided in the record and playback units where required for heater control. Circuits have been chosen, where necessary, to minimize the effects of drift.

Modulator Unit

A block diagram of the record unit is shown in Figure 1. The input signal required is one volt rms into 100,000 ohms, unbalanced, for 100% modulation (40% frequency deviation). It drives the grid of a cascode stage, 12AX7 modulator. The cascode arrangement uses one-half of a twin-triode as a load for the other half which is operated as a dc amplifier. This method satisfactorily reduces the drift caused by heater voltage changes. In addition, a ballast tube regulator is used for both the amplifier and the multivibrator heaters.

An oscillator suitable for the wide deviations required in this system is the positive-grid, free-running, multivibrator. Its frequency is a linear function of the voltage applied to the grid returns. The grids are returned to the plate of the amplifier; the no-signal voltage is adjusted so that the multivibrator, for the proper R-C time constant, runs at the carrier frequency. The amplified input signal at the amplifier plate varies the voltage on the grid returns, thereby causing the frequency of the multivibrator to change accordingly. An input potentiometer adjusts the degree of deviation.

An amplifier stage following the multivibrator provides isolation and sufficient current for driving the record head to saturate the magnetic tape. The other half of the 12AU7 twin-triode, as a cathode follower, connects to a test jack to provide a suitable check point. By means of a suitable loss pad cable, a signal from the record unit can be connected to a test jack on the playback unit. This is convenient for checking purposes in order to eliminate the need for running tape.

Controls are provided for carrier frequency adjustment and level set. A typical model is set to a carrier frequency of 27 kc. It will deviate plus or minus 11 kc for an input dc level of minus or plus 1.41 volts; the inversion in frequency change is caused by the dc amplifier.

Playback Unit

Figure 2 is a block diagram of the playback unit. The output of the playback head is amplified and severely clipped in three successive

stages, removing all trace of amplitude variation and resulting in square waves, the variation in frequency of which is proportional to the input data. The square waves are then differentiated, converting the leading and trailing edges to positive and negative pulses of identical waveshape. The number of these pulses per time interval is then proportional to the instantaneous value of the recorded input signal.

A cathode-coupled twin-triode amplifies and converts the pulse signals from single-ended to push-pull. In turn, these are rectified by a 6AL5 full-wave rectifier circuit. The pulses, now of a positive polarity and twice the former repetition rate, supply the signal to the cathode follower driving the filter. The filter, in removing the high frequency components, integrates the pulses. This circuit is, in effect, a pulse-counting circuit. Such a pulse-counting type of discriminator is suitable for the wide-deviation signals that are employed in this system. The amplitude and phase characteristics of the overall recorder are determined principally by the filter that is used. In the standard 5,000 cps data recorder, the filter begins to attenuate abruptly at this frequency, and in the region beyond cutoff, has an attenuation of 27 db per octave.

The output of the filter is a replica of the input recorded signal. Being a high-impedance circuit, it connects into the grid of the output cathode follower, which is suitable to drive a 600-ohm load.

The adjustments consist of a level set and a dc zero balance.

The playback output is specified as an unbalanced cathode follower adjusted for normal level to produce a sinusoidal output signal of one volt rms across a 600-ohm termination for 100% modulation.

Construction Features

Since most applications require the recording of a number of events simultaneously, this equipment has been designed in building-block form to allow the convenient assembly of the number of components required by the particular application. The amplifiers are in strip form, such that two strips occupy 5-1/4 inches of rack space in standard 19" relay racks. Each individual strip contains two record amplifiers or one playback amplifier. Front and rear view of the two channel record strip is shown in Figure 3. These views of the single channel playback strip are illustrated in Figure 4. For single channel units, one playback strip, one single channel record strip, and a one-channel power supply are assembled on a single chassis occupying 12-1/4 inches of rack space. For two-track units, two such chassis are employed. For 3 or more channels, record and playback power supplies are supplied, each of which is capable of operating up to 7 channels. In all cases, the electronic assemblies are combined with the Ampex Model 300 Series tape transport mechanism (see figure 5). This tape drive, because of its inherently smooth tape motion, is ideal for use with an FM carrier system. It also has the advantages of readily being adaptable to different tape widths for various numbers of channels, and to different speeds for special applications. Figure 6 illustrates a 14-track record and playback head assembly as used on this transport in order to record 14 separate tracks of information on one inch wide tape. In Figure 7 is shown a completely assembled special 14-track record and playback unit, with Figure 8 illustrating the rear view of this same recorder.

PERFORMANCE CHARACTERISTICS

Overall performance characteristics of the FM carrier recorder are illustrated in Figures 9 through 12. Figure 9 shows typical frequency response and phase shift curves, of a unit operating at the standard speed of 30 inches per second. Production specifications are $\pm$ 1/2 db to 3,000 cycles, with a slight roll-off to 5,000 cycles. As may be seen in Figures 10 and 11, linearity for both ac and dc signals is good. The calibration for dc is slightly different from that for ac because of the characteristic roll-off in low-frequency response.

Transient response is indicated by the results of recording square waves shown in Figure 12. Illustrations (a) and (b) show the overshoot produced by shock-excitation of the filter employed in the playback amplifier. This filter has been chosen to optimize wide-range frequency response, and, with very slight modifications, can be altered to reduce or eliminate overshoot with corresponding sacrifice of high-frequency response. However, these illustrations serve to demonstrate that the recording process has fast enough rise time to transmit a wave front capable of shock-exciting the filter. Since few pickup devices are capable of responding with this speed, the usual results when recording the output of electro-mechanical pickup devices will be more like Figure 12 (c), which is a 50-cycle square wave with the standard 5,000-cycle filter, except that the rise time at the input of the recorder has been increased to 1/4 millisecond.

One of the most important characteristics of performance is signal-to-noise ratio. As already mentioned, this is limited primarily by the uniformity of motion of the tape drive. The principal source of flutter in the tape drive used with this system is a high-frequency flutter caused by oscillation of the tape backing due to friction between the tape and the heads. As the tape is pulled, there is a tendency for it to go into longitudinal oscillation comparable to that produced by pulling a bow over a violin string. This high frequency flutter is broadly tuned to about 3,000 cycles, and is worse with 1/4" wide tape than with wider tapes having correspondingly greater stiffness. Since erase is not used with the FM carrier system, the recorded signal being of sufficient intensity to remove the previous signal, it is possible to install a roller in place of the erase head. This reduces the length of the unsupported section of tape and thereby reduces the amplitude of the frictional flutter. In this way, an rms signal-to-noise ratio of over 40 db may be realized. The roller is not required with tapes over 1/4" wide. If lower frequency response is acceptable, a filter cutting off below 2,000 cycles will improve the signal-to-noise ratio by approximately 6 db.

APPLICATIONS

From the above performance characteristics, it is evident that the recording system described may be used to record a wide variety of data with excellent accuracy. Some of the information for which this system is particularly well suited includes shock and vibration data, stresses and strains in structural members under test, seismic exploration data, temperature, velocity, pressure in flight tests, and countless similar quantities. Some of the advantages that can be realized by recording this information on magnetic tape may be appreciated from the following typical application, that of seismic exploration. Previous techniques consisted of firing a shot in the ground, and recording the reflected energy picked up with a multiplicity of geophones connected to multitrack magnetic oscillograph. In order to obtain the most satisfactory record which will discriminate against unwanted reflection and reverberation and bring out the reflections from the desired subsurface layers, the operator must use his judgment in selecting various filters prior to firing the shot. In recording the geophone output on magnetic tape, not only is the advantage of greater frequency response and dynamic range realized, but such operator judgment is not necessary in the field. By repetitive analysis on playback, any variety of filters may be used to optimize the results. Furthermore, gains in various channels can be altered, the position of channels on the film record made from the magnetic recording can be altered, and a number of analysis techniques can be employed to derive every possible bit of information from the recording.

In many applications, the ability to expand or contract information, time-wise, is of great value in analyzing the data recorded; this operation also is made possible by the FM carrier magnetic tape recorder. For example, very low-frequency seismological information can be recorded at very low tape speeds, so that 24 hours of data may be contained on a single 10-1/2 inch roll of tape. Yet is can be played back at very high tape speeds so that the data can be analyzed in a matter of minutes. Furthermore, frequencies are amplified to a range where separation by filters for analysis purposes is practical. This technique is quite practical with the present state of the art, for although smooth tape motion is difficult to achieve at very low tape speeds, the variations in speed are above the spectrum being recorded, and when speeded up, the flutter frequencies are multiplied outside of the pass band. The converse technique of slowing down information recorded at high tape speed also is very desirable, since it becomes possible in this manner to record high frequency information on direct-writing instruments having limited frequency response. Performance characteristics at low playback speeds are in this case seriously limited by low-speed flutter, so that an appreciable reduction in signal-to-noise ratio occurs. Present developments are directed toward improving the performance of low-speed playback to extend the usefulness of this technique.

CONCLUSIONS

1. An FM carrier system for recording data on magnetic tape has been developed which extends the advantages of the magnetic recording process to applications requiring extreme low-frequency response, excellent transient response, and high degrees of amplitude accuracy.

2. Recording of data on magnetic tape results in an electrically reproduceable signal, which makes possible the use of a great variety of analysis techniques both to reduce the man-hours required in analyzing the data as well as to derive considerably more information from the measured quantities.

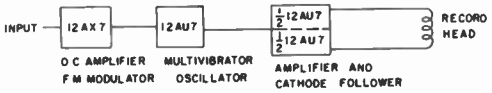


Fig. 1 - Block diagram, record unit, Ampex model 306 FM carrier recorder.

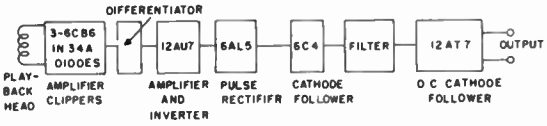


Fig. 2 - Block diagram, playback unit, FM carrier recorder.

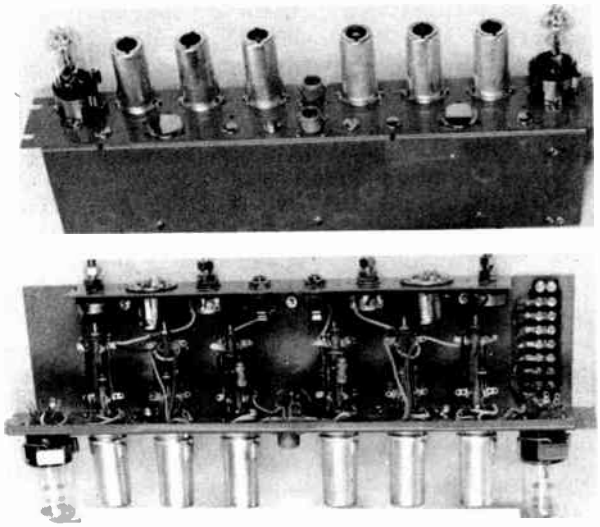


Fig. 3 - Record amplifier, front and rear view.

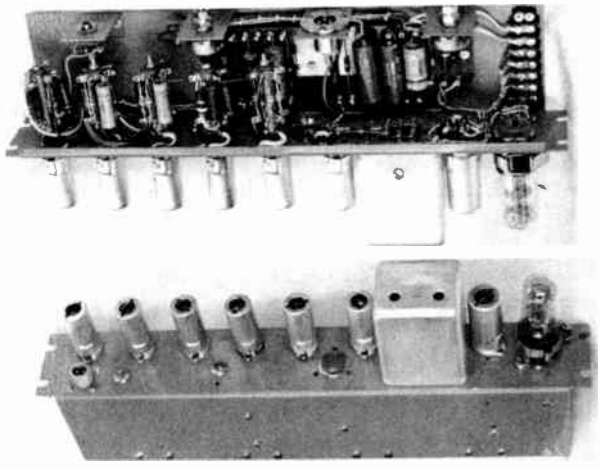


Fig. 4 - Playback amplifier, front and rear view.

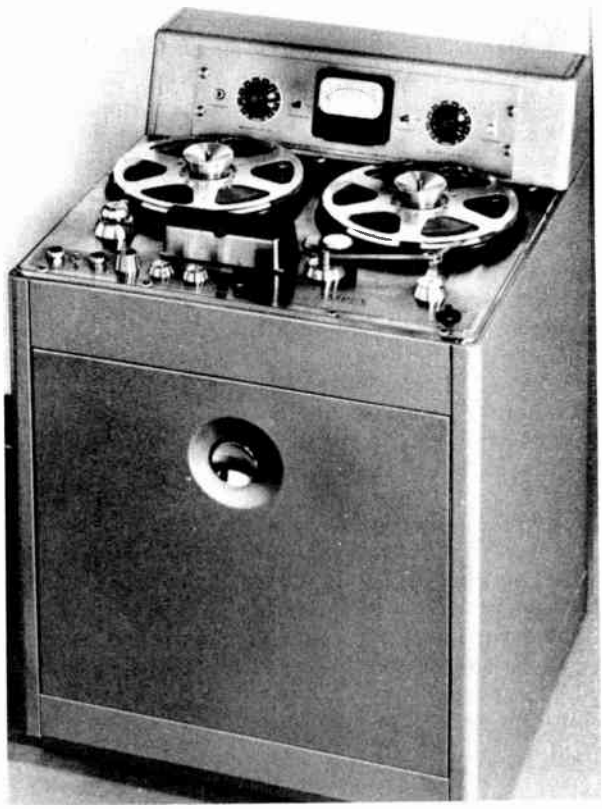


Fig. 5 - Ampex model 300 tape transport.

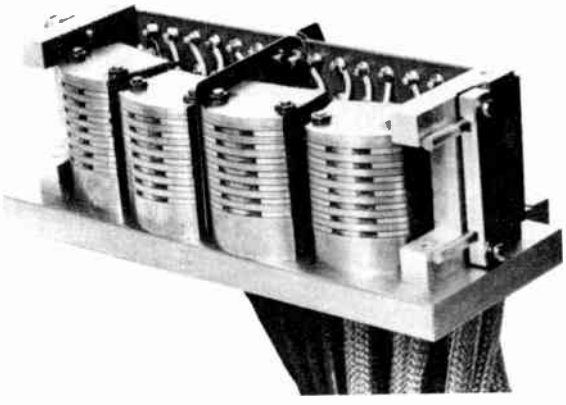


Fig. 6 - Fourteen-track record and playback head assembly.

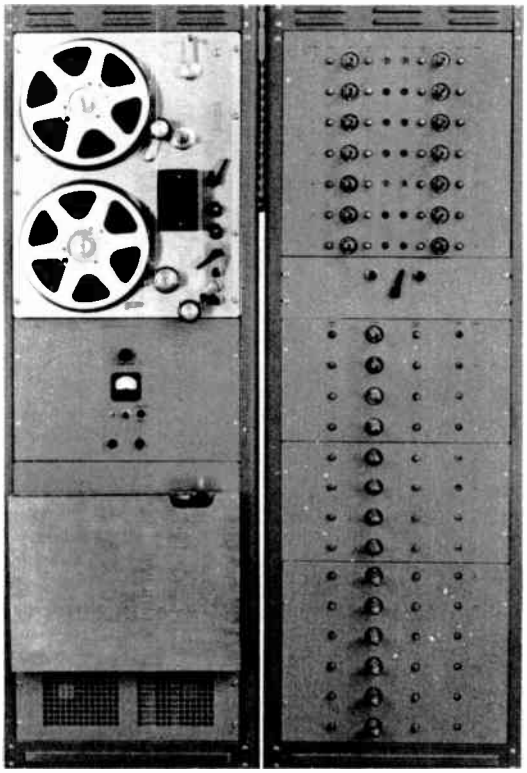


Fig. 7 - Fourteen-channel record-playback unit, front view.

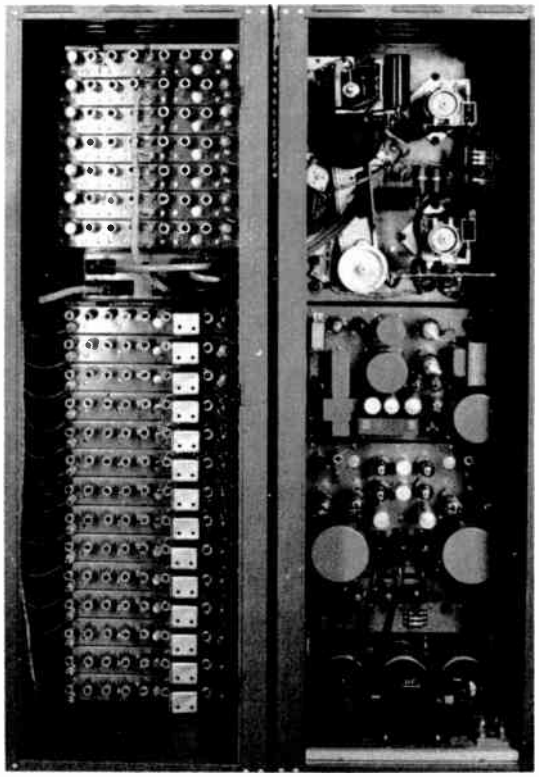


Fig. 8 - Fourteen-channel record-playback unit, rear view.

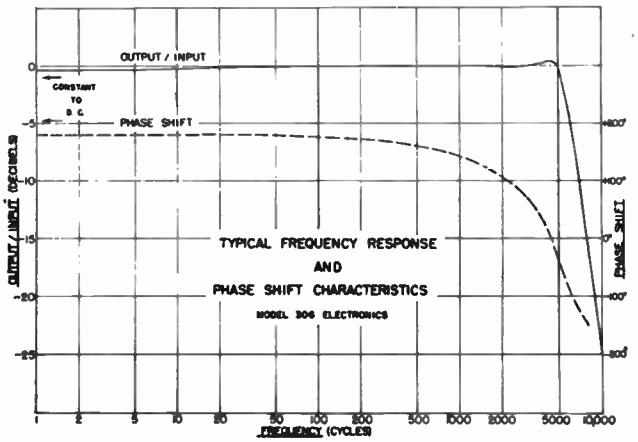


Fig. 9

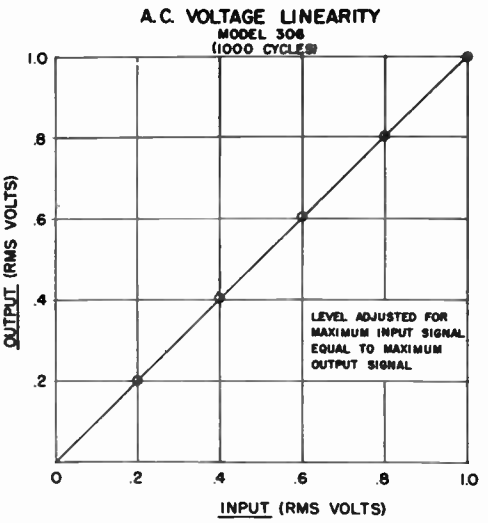


Fig. 10

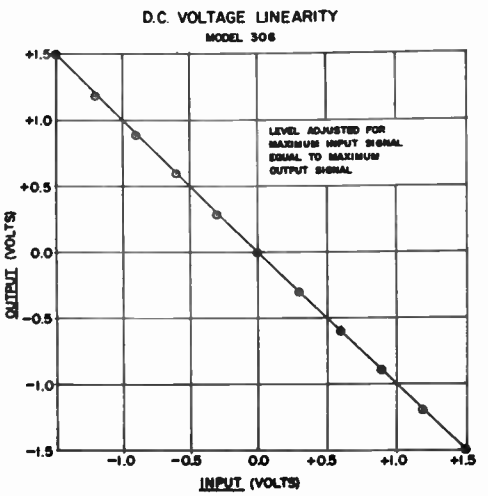


Fig. 11

Fig. 12 - Square wave response

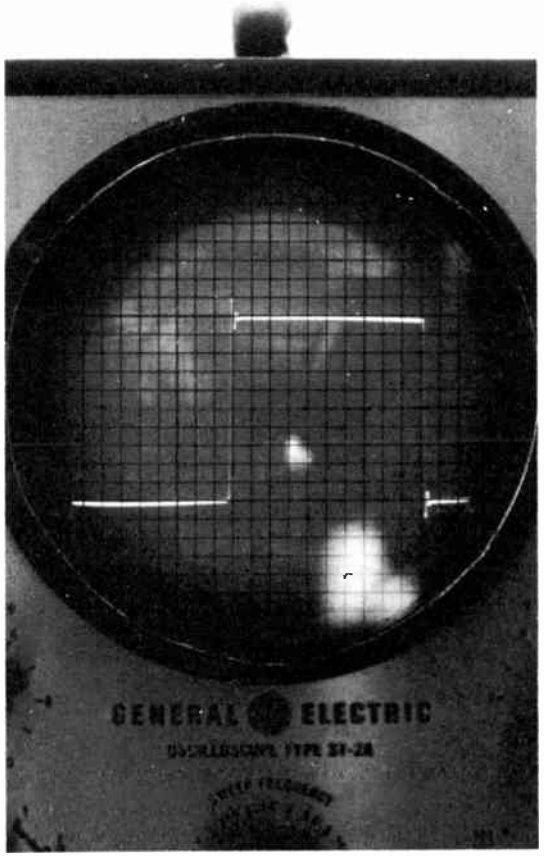


Fig. 12a - 50-cycle square wave with standard filter.

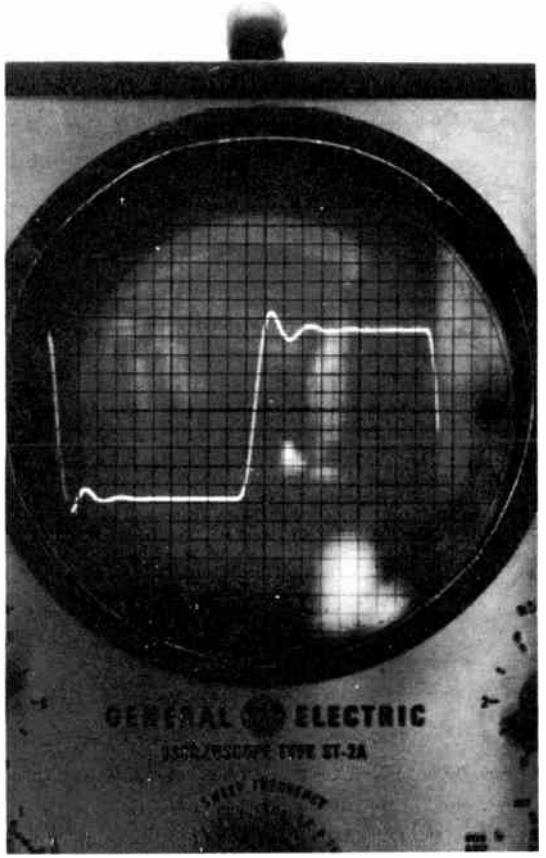


Fig. 12b - 500-cycle square wave with standard filter.

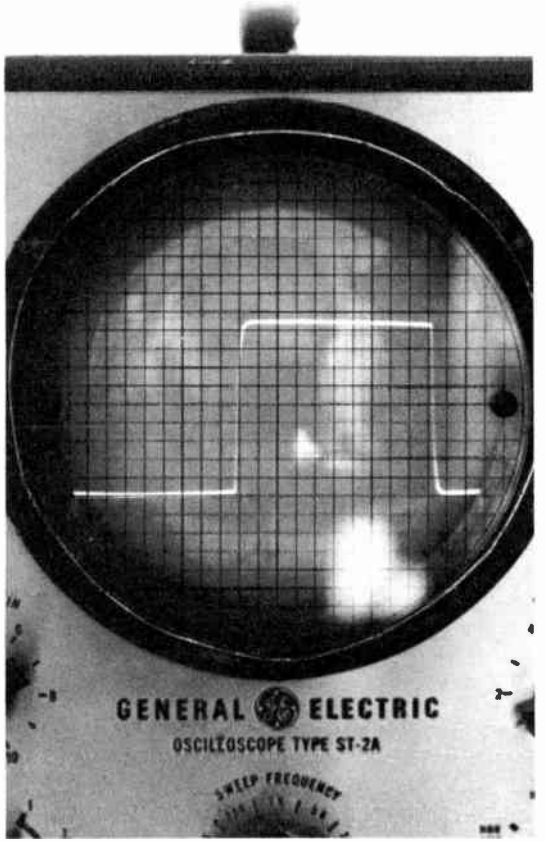


Fig. 12c - 50-cycle square wave with input rise time limited to one-quarter millisecond.

AN IMPROVED FM/FM DECOMMUTATOR GROUND STATION

Foster N. Reynolds,
Chief, Telemetering Section,
The Ralph M. Parsons Company,
Pasadena, California

The basic function of a telemetering decommutation station is to decommutate and demodulate information from a sequential string of pulse amplitude or pulse width modulated data. The particular unit described herein was designed for use with the Research and Development Board's standard telemetering pulse amplitude modulation configuration, which consists of a frame of either 27 or 15 information channels and a 3-channel width synchronizing master pulse.

This system has a long-term accuracy of 1/2 of 1% for a period of eight hours following a one-hour warm up, and a Linearity of 1/4 of 1% or better, which is defined as the deviation of a calibration curve from the best straight line drawn through the calibration points. The response that can be expected from a channel is approximately one-third of the rate at which the end instruments are being sampled. This may be contrasted to the general assumption that with the stair-step type of data a minimum of five samples are required to define a single sine wave of modulation when an operator is employed to reduce data by eye.

The power output of this system is in excess of the amount required to drive to full scale a Brush Pen Motor, a recording galvanometer element and a remote indicating device. Actually, the equipment will develop from each channel ± 32 miliamperes through an 1172-ohm load. During a normal flight test, if one of the loads becomes disconnected or open, the effect on the current flowing through the remaining loads is less than 1/10th of 1%.

Although the equipment was designed to operate with the standard pulse amplitude modulation of an FM/FM system, it is directly adaptable to the PWM or pulse width type of modulation. A ground grating circuit is employed which is triggered by the incoming sequential data and thus eliminates the necessity of using a gating unit in the airborne vehicle.

The output wave form in this equipment is not as is normally experienced with the majority of decommutation systems. Figure I (a) depicts a commutated frame of information. Figure I (b) represents a channel of information after it has been decommutated or stripped from several frames of data. Figure I (c) is typical of the older method of smoothing which results in a series of pulses whose heights are proportional to the initial pulse area. This integration is normally accomplished by recording the data with rather low frequency response galvanometers.

Figure I (d) is the stair-step type of output achieved by a series R-C integrator. In most cases, the discharge spike, occurring at the end of each frame of information is eliminated by a second storage and holding circuit. It should be pointed out that the only voltage of true value in this method of presentation is at the leading edge of each step, since the voltage between the leading and trailing edge is only a memory of the initial value.

Figure I (e) is typical of the output from a demodulator channel of the type used in the system described herein. In all respects the wave form is an exact duplication of the input modulation up to the frequency response limit of the system. Should a sine wave be impressed on a particular contact of an airborne commutator, the output of the associated ground demodulator channel would be an identical sine wave with no spikes, steps or discontinuities not present in the modulation in excess of 1/10 of 1% of its full scale value.

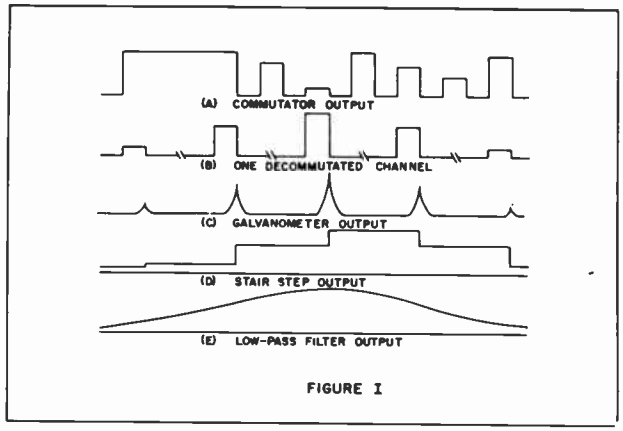


FIGURE I

In general, decommutation systems normally use the following method of extracting information from the amplitude modulated pulses:

> Some types of series resistance and capacitive integrating network is used to produce a voltage proportional to the pulse height or area assuming a constant width. This voltage is stored either in the integrating capacitor or in a following storage condenser, the voltage across the condenser being changed at the occurrence of each channel pulse.
>
> From a signal-to-noise standpoint, it is interesting to note that in some cases full advantage of this type of integration is not taken, due to the fact that the time constant of the integrating network is sufficiently short to produce almost a peak reading type of system. This, of course, decreases the approximate 6 db signal-to-noise improvement ratio that can be realized with perfect integration.

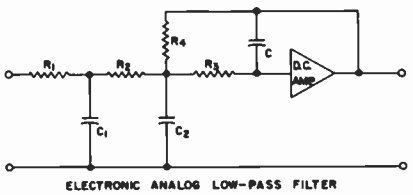
The decommutation system described herein accomplished both integration and storage through the medium of an electronic-analog low-pass filter. A low-pass filter, if it could be designed to be theoretically perfect, would be the ideal device to use where data modulation is to be extracted from either pulse amplitude or pulse width carriers. In theory, the low-pass filter should have an infinite rate of attenuation and its band-pass characteristics should be flat up to the cutoff frequency. If such a filter could be designed and constructed, one would make the cutoff equal to the commutation rate divided by 2. This would permit data to be demodulated from the pulse train up to a frequency equal to F/2, where F is the commutation rate.

Unfortunately the perfect filter cannot be realized in practice, but with cascaded LC sections it is possible to achieve an extremely high rate of attenuation and thus approximate the curve of the perfect device. At higher frequencies this is practical, but when the cutoff frequencies required by the standard FM/FM telemetry system are considered, the lowest being 1.25 cycles per second, it is obvious that, to achieve the needed attenuation, the physical size of the inductors and capacitors becomes prohibitively large.

A solution to the problem of generating the characteristic curve of a pi-section m-derived LC low-pass filter is to construct an electrical analog, which has as its solution exactly the same differential equation but which has no inductance and no limitation on the low frequency cutoff point. Very low frequency low-pass filters can be obtained by passive networks of resistors and capacitors without inductors, if a sufficiently high impedance level is maintained, but such filters have poor uniformity of response in the pass band and lack sharp cutoff characteristics. Uniformity of pass band and sharpness of cutoff are desirable, for example, in cases where a filter is used to smooth a discrete time series into a continuous time series.

The electronic-analog low-pass filter, as developed by Dr. L. L. Rauch of the University of Michigan, in reality is a device providing an amplitude response characteristic which results from the sinusoidal solution of a third-order differential equation. A circuit diagram for this device is shown in Figure II. It is easy to assign values, through the equations describing the analog, to R and C, so a curve, which has a rate of attenuation of 18 db per octave, after the cutoff frequency has been reached, is obtained.

A more desirable response (more uniform pass band) with the same high frequency roll off is shown in equation (1). This is made possible by combining two slightly different filters of Figure II so that their characteristics are within a 0.1 db of equation (1), Figure II. It is also possible, by the choice of the feedback network around the dc amplifier, to produce any positive zero frequency voltage gain within the unit, thus overcoming the normal problem of filter insertion loss. A point of interest is that the voltage across C_2 is exactly the derivative of e_o and may be picked off by a second amplifier. The voltage e_o is, of course, an efficiently smoothed version of e_i.



ELECTRONIC ANALOG LOW-PASS FILTER

TWO FILTERS IN CASCADE HAVE THE RESPONSE:

$$\frac{a^2}{\sqrt{1+\left(\frac{\omega}{\omega_c}\right)^{12}}}$$

FIGURE II

The time delay of this filter or any device, expressed in t_d seconds, is related to the frequency and phase delay by the following equation:

$$\theta = t_d \Omega$$

where the phase delay θ is in radians and Ω is the frequency in radians per second. Any filter having a linear proportionality between phase delay and frequency has a constant time delay. A plot of the equation relating phase delay and time delay for these filter results in a fairly linear relation up to the cutoff frequency, which means that such a filter will provide a result not far from an ideal constant time delay within the pass band. It is easy to compute that, for example, a filter whose cutoff frequency is 10 cycles per second has a time delay of 1/30 of a second and in general a time delay of $1/2\pi f_c$ seconds.

In this equipment, the low-pass filter is used in the following manner: Two filters are placed in series, the amplitude modulated pulses being fed to the input circuit of the first filter, and the recording equipments being connected to the low impedance output of the second. The cutoff frequency, or 3 db point, for the system is nominally chosen at about 1/3 of the commutation rate, or F/3. With this cutoff frequency, examination of a typical attenuation curve for two such units will show that, at the commutation rate F, the filters are approximately 60 db down, indicating that frequency components of F are attenuated at the output to 1/10th of 1% of the full-scale value.

Each individual filter is designed to have a positive zero frequency voltage gain of 10, giving an over-all channel gain of 100. This is necessary due to the fact that the filters act as a true integrator and produce a voltage proportional to the total area occupied by one pulse and the following off period. If it is assumed that this duty cycle is 1 to 100, in other words the pulse is on 1/100th of the time, and if the modulation frequency is kept below the cutoff of the filter, it becomes obvious that for every volt of pulse applied to the input of a channel, one volt may be expected at the output circuit, which, of course, would not be possible if an LC type of filter construction were used due to its insertion loss and the small duty cycle, unless the filters were followed by appropriate DC amplifiers.

The output circuit of the second filter is composed of three tubes, two connected in parallel making up the control element of a cathode follower, and a third used to replace the normal cathode resistor. This produces a non-linear element which alleviates the power handling requirements of the top two tubes under certain operating conditions. The open loop gain of the DC amplifiers in each filter is approximately 10,000 with their roll-off starting at 100 cycles per second so that, by the time the open loop gain is equal to 10, Bode's theorem of a maximum of 12 db per octave attenuation has not exceeded. It may thus be seen that, with feed-back the amplifier circuits have been linearized to a point somewhat better than 1/10th of 1%.

In order to insure minimum drift in the amplifiers, the power supplies for the high gain stages are well regulated. The negative 250 volt supply is chopper stabilized with reference to a temperature controlled subminiature regulator tube. In practice, the tube voltage drift is approximately 400 microvolts about an operating potential of 87 volts. It is easily shown that this method of deriving a standard voltage for high-voltage supplies is superior to that obtainable with a standard cell. The positive supply is then chopper stabilized with respect to the negative, and a resulting differential drift measured to ground for the two is in the order of 700 microvolts.

The equipment is designed to operate under the standard RDB specifications for tolerances in airborne equipment, and it will generate five false switching pulses which are used to maintain sequential counting, should a burst of noise obliviate a portion of a frame of commutation. It will also operate under the condition of decreased sampling rate from a value of full speed to full speed minus 20%.

Due to the fact that the filters are true integrators, the output voltage is a function of the average DC value of the decommutated pulses and, therefore a change in commutation speed appears as a change in output level. A solution to this problem is to employ a closed loop feedback system to maintain a constant duty cycle ratio, regardless of the commutation rate. Figure III is a block diagram of the component parts of this loop.

The first differential amplifier not only raises the input level, but also performs the function of a gate which, when driven by the servo loop, maintains a constant duty cycle ratio. This loop consists of a synchronized free running multivibrator which is triggered by each leading and trailing edge of the data pulses. The multivibrator drives a constant pulse height generator, which in turn provides amplitude limited pulses for an electronic analog low-pass filter.

The output of this filter is actually an error signal which is the difference between a standard reference voltage and a voltage proportional to the duty cycle of the incoming

data pulses. This error voltage is then amplified by an integrating amplifier and used as the clamping voltage for a phantastron, which is triggered by the trailing edge of the raw data, and its delay is a function of the clamping voltage level. A primary gate is also closed at this time, and is opened at the end of the delay period. The closed time is actually the off period of the duty cycle which is adjusted to approximately 40 per cent on and 60 per cent off. A change therefore, in commutation speed results in a change in duty cycle which produces an error signal at the output of the low-pass filter. The amplified error signal changes clamping level on the phantastron with consequent change in delay in such a direction that is necessary to maintain a constant duty cycle. The high open loop gain of this degenerative feed back system results in less than 1/10 of one per cent error with a 20 per cent reduction in commutation speed.

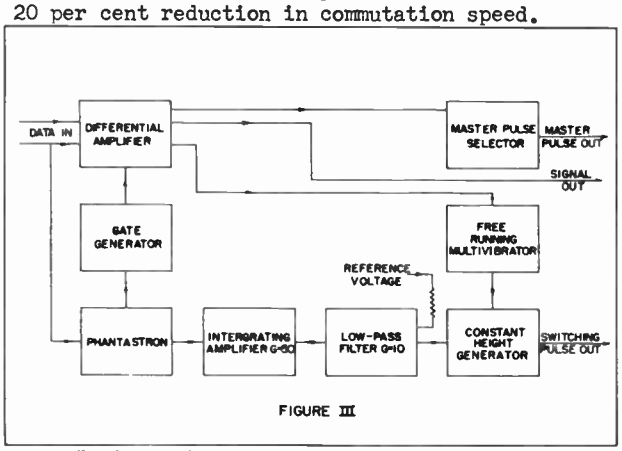


FIGURE III

It is in this manner that the system generates its own ground gating, thus eliminating the necessity of large, power-consuming airborne gating units. In practice, the system will work on any signal which has a duty cycle ranging from 40-60 to 95-5, i.e., 40 per cent of a channel width is on time and 60 per cent off.

The count down circuit, which is used to decommutate the individual channels, is a sequential open loop counter and is started by the trailing edge of the master synchronizing pulse. The outputs of this counter are fed to the individual demodulator chassis through low-impedance lines to the gate tube of each filter unit. The counter tubes and circuits, which are constructed in individual plug-in cans for ease in servicing, are located on a separate chassis from the demodulator channels so that, during an operation, any one of the channels may be removed without affecting the counting or stability of channels following the particular one being investigated or serviced.

The mechanical design of the equipment takes into consideration component cooling and ease of maintenance. The most critical chassis, as far as heat is concerned, are the filter units. In the particular section of the rack where these are housed three vertical unobstructed airducts are provided. This is achieved by the design of the chassis so that they mount in a vertical position with the tubes extending horizontal. All major units are either of the plug-in type, or mounted on slides which permit complete rotation of the chassis when in the extended position, so the equipment may be serviced in an operating condition.

Summarizing the foregoing, the following are the outstanding characteristics of the telemetering decommutation system.

The over-all long term system accuracy is 1/2 of 1% for a period of eight hours. The linearity is better than 1/4 of 1%, and the output wave form generated is an exact duplication of the input modulation.

The response is appreciably better than can normally be achieved with a stair-step type of output function.

The power output is sufficient for practically every application, providing a maximum of 75 volts peak to peak across an 1172 ohm load.

Disconnection of loads has no material effect on the current flowing through other parallel recording devices.

The system is readily adaptable to pulse-width type of telemetering modulation.

The equipment incorporates its own ground gating system, thus eliminating the added burden of airborne gating in the missile.

SOME INDUSTRIAL APPLICATIONS OF TELEMETRY

by

Harold R. Hoyt
Great Lakes Pipe Line Company
Kansas City, Missouri

and

John H. Van Horn
Midwest Research Institute
Kansas City, Missouri

Abstract

Some current techniques are presented for telemetering process variables where operations are widely distributed as in the petroleum industry. Telemetry systems employing carrier telephone and microwave techniques in the transmission of data on such variables as pressure, flow rate, level, position, etc. are described. Also discussed are transmitters and receivers used to convert electrical magnitudes to pulse-duration signals for observation and recording. A method of telemetering by teletype signals is shown. A unique development is described which makes use of a talking signal from a magnetic recording. A total of 16 spoken words and numbers is stored and chosen in the proper sequence to convey the desired information to the listener.

Introduction

As defined by the American Standards Association, "Telemetering is the indicating, recording or integrating of a quantity at a distance by electrical translating means[1]. Although there have been some translating means that were non-electrical, these were limited to such short distances to be of little use in most industrial applications. Early industrial applications of any appreciable scope concerned the generation and transmission of electrical power but other industries soon made use of telemetering principles[2]. At first these involved comparatively short distances, but modern telemetering systems provide indicating and control functions over hundreds of miles. The only limitation appears to be that present with modern electrical communication equipment. Telemetering and control systems are now used in such diverse fields as electrical transmission, the pipe line industry, oil refineries, meteorology, on board ship, nuclear installations, steel mills and guided missiles. In general, telemetering principles are applied whenever it is more convenient, economical, safe or impossible to use information or provide control at some distance from the source.

The information to be telemetered depends upon the associated industry. The transmission and use of electrical power require information regarding kilowatts, kilovolt-amperes, volts, amperes, transformer tap position and transferred trip. Pipe lines associated with the petroleum industry require flow rate, pressure, level and position. The field of meteorology makes use of air pressure, temperature, relative humidity, wind velocity and wind direction. To transmit such diverse information it is necessary to translate it into an electrical signal which can be sent over the distance involved. This requires a transducer or pickup which converts the quantity indicated into electrical energy, then a transmitter provides a signal that can be sent over some distance. Often the transmitter involves more than is normally present in a simple communications system, since a dc signal can seldom be transported very far. After transmission, a receiver reconverts the signal into usable information which is then observed or recorded. Transmission methods can make use of the following information: current, voltage, frequency, position or impulse characteristics. Considering impulses one can count the number, determine the frequency, or observe the spacing or duration. The last one named is the most commonly used at present.

It should be noted that there is a close relation between telemetering and remote control or regulation. In many industries it is only a matter of direction and end result. For example, in the pipe line industry it is desirable to control flow and pressure and to switch paths

in order to separate various fluids. These functions are closely allied to the telemetered information.

Telemetering by Microwave

It is the intention of this paper to concentrate on telemetering applications as used in the transmission and storage of petroleum products. Since the operations of this industry are widely distributed and the control points are kept to a minimum, there is considerable dependence on such long distance systems that make use of carrier telephone and microwave relay networks. Many types of carriers have been used with most of the information being transported by means of tones or frequency-shift signals. However, since microwave transmission is of more current interest the applications of this communication medium to the pipe line industry will be considered briefly here.

It is believed that the first microwave link utilized by a pipe line company was in 1949 and was a comparatively short system. Now some systems extend over 1000 miles and require a large number of relay stations of various characteristics. It is generally thought that microwave will pay its way if the distance involved is more than 25 miles and 4 or more voice circuits or the equivalent are required. The following types of communication facilities are often provided in a microwave network: party line and private line dial telephone circuits for dispatching, maintenance, and executive use; party line ringdown telephone circuits for general use; party line teletype available at all drop-out stations; vhf 2-way mobile communication to aircraft and ground vehicles; telemetering circuits such as remote indication, fault location and alarms; and remote control of many functions from a number of stations in the network. As many as 27 channels of 3000 cps bandwidth may be provided in each direction. The 3000 cps bandwidth per channel is dictated as the normal band passed by the ordinary telephone line, which is used to feed telemetered information and communication circuits into the relay station from the nearest pumping station. The microwave stations are of three types: terminal, through repeater, or drop-out repeater. Any or all of the channels may be available at the terminal and drop-out stations.

Telemetering and control signals often consist of dc pulses of variable time duration. These are used to control audio oscillators so that a single frequency is used for each function to be transported. These frequencies lie in the 3000 cps audio band and are spaced 100-150 cps apart. More than one audio channel may be used if necessary. The audio channels are separated by frequency-division multiplex or time-division multiplex and the multiplex signal is then used to frequency modulate the klystron oscillator for transmission over the microwave link.

A simplified diagram of the frequency division multiplex system is shown in Figs. 1 and 2 for 9 channels. Bridge or lattice modulators and demodulators are used with single side band transmission. If more than 9 channels are required, as many as three groups may be group modulated by a similar system before applying the signal to the klystron circuit. In the receiver the various channels are separated by band pass filters and demodulated using the same oscillator frequency as provided for the original modulation of the corresponding channel.

A simplified time division multiplex system is shown in Fig. 3. This can be used with 4-24 channels and a pulse amplitude modulation circuit is used to scan the audio channels. Other types of pulse modulation can be used equally well. The time multiplex signal is then used to frequency modulate the klystron oscillator to provide the transmitted microwave signal. The reverse action takes place in the receiver to obtain the original audio frequency channels at the terminal or dropout stations.

A duplex frequency system is used to provide for two-way communication without interference between the two signals. A typical frequency allocation is shown in Fig. 4. A bandwidth of 500-700 kilocycles is available in each direction to accommodate the FM signal.

Since there are many manufacturers producing microwave relay equipment there are several variations on the systems described here. However, this one is typical of the services supplied and the equipment involved in an installation serving an extensive pipe line facility.

Transmitters and Receivers

In addition to the normal transmitting and receiving functions familiar to all communications engineers, it is necessary in a telemetering system to convert the output of the pickup into electrical information which can be used to modulate the carrier energy. After demodulation the transported information must be reconverted into a usable form for observation and recording. Until very recently industry has used electromechanical devices for instrumentation, transmission,

and reception of measurement and control signals. However electronic devices are rapidly becoming very common in this field.

Impulse System

One of the oldest and simplest methods of transmitting information is by the impulse system. This system is ideal for the type of information that can be readily expressed in digital form, for example, liquid flow. A simple interrupter device driven by the meter counter shaft is the transmitter. A magnetically operated counter is the receiver. This method has been used in the pipe line industry for transmitting meter readings from the remote corners of a tank farm to central control points. Electronic counters may be used in place of relay operated counters. However the impulse system has some disadvantages. Transients on the line can cause erroneous pulse counts and power outages result in loss of counts during the outage period. This method is not widely used at present.

Variable Frequency System

The variable frequency system is more dependable and is used for a number of applications at the present time. It employs electronic circuits and devices and is considered one of the more modern systems. In general, a primary measuring element is used which produces a dc millivolt output proportional to the measured quantity. Most electrical measurements are made by a thermal converter. Established designs are available for practically any mechanical quantity. Transducers, tachometers or slide wires may be used to provide an electrical signal which is a function of some mechanical action. The conversion from millivolts to frequency is performed by the transmitter. This is an electronic generator which has an output frequency proportional to an applied voltage. Figure 5 shows a block diagram of the transmitter. A fixed bias, E_b, is added to the signal, E_m, from the primary measuring device in order to obtain a voltage, E_o, which will give the base frequency necessary for a range of 15-35 cycles. A voltage, E_o, from the oscillator is directly proportional to the frequency of oscillation and is balanced against E_o, the sum of the voltage from the primary element and the bias voltage. If these two voltages are not equal, the difference is applied to the amplifier and the output of the amplifier in turn governs the oscillator to increase or decrease the frequency as necessary to bring the two voltages into balance. It is evident that this is a null balance system. The receiver is an electronic converter which receives the transmitted signal and converts it back to a dc voltage proportional to the applied frequency. This system is used extensively by power companies. Carrier systems over power transmission lines are utilized for the transmission of signals.

Pulse-Duration System

One of the most versatile and widely used systems of telemetering employs the timed pulse or duration of time principle. The principle is simple and the equipment dependable. Variations in voltage and line resistance have no effect. Indication may be continuous or periodic as desired. Transmission may be over telephone lines, carrier systems or microwave. Two general types of equipment are available. One uses electromechanical equipment for transmitting and receiving and the other employs photoelectric and electronic instruments. The primary measuring element may be a bourdon tube type instrument, mercury manometer, millivolt meter or any mechanical system for positioning a pointer.

Figure 6 shows schematically one type of transmitter. The primary measuring element positions a pointer (A) on a scale. Immediately behind the pointer is a spiral shaped cam (B) driven by a synchronous motor. Attached to the pointer is a small projection or rider which contacts the surface of the cam and lifts the pointer a fraction of an inch when it is on the surface of the cam. When the pointer is lifted, the bar (C) is moved forward actuating the lever arm and opening the switch (S). Thus it can be seen that if the cam is rotated at constant speed, the on and off periods of the switch (S) depend on the position of the pointer (A). At low scale readings the switch is closed a very short period of time; at high scale reading it is closed for a long period. The cam is generally rotated at a speed of 1 revolution each 15 seconds. Thus the signal cycle is 15 seconds with the duration of the impulse varying from approximately 2 seconds at minimum indication to 13 seconds at maximum indication.

The receiver shown in Fig. 7 usually includes a circular-chart recording-type instrument fitted with a conventional chart and pen arm assembly. In the drawing the recording equipment is represented by an arm and indicating scale. The receiver mechanism consists of two differential gear trains driven continuously in opposite direction by a synchronous motor. The gear train design is such that if disc (A) is held stationary, arm (C) will rotate in a counterclockwise direction. In so doing, it will pick up arm (D) which

is attached to floating pinion (E) and cause the pinion to rotate engaging quadrant (F). This in turn will move the pen arm up scale on the chart. The shaft-carrying arm (C) is spring loaded in such a manner that when the large disc (A) is released, the arm (C) will return to its starting position leaving the pen arm at the point to which it was moved. Likewise, when the disc (B) is stopped, its part of the mechanism acts to move the pen arm down scale. The electromagnet (M) receives the impulse signal from the transmitter. When it is energized the disc (A) is locked and the pen is moved up scale during the time the signal is on. During the off portion of the signal, the magnet being de-energized locks disc (B) and the arm (D) tends to move the pen down scale starting at the maximum scale reading. For steady state signals, the pen arm is neither moved up or down scale from its position. A change in measured variable effects the proportion of on and off time of the signal and the pen arm will be repositioned at a new point on the scale corresponding to the new reading.

A very ingenious pulse-duration system makes use of photoelectric scanning to generate the impulse signal. Figure 8 illustrates the transmitter. Light from the bulb (1) is reflected from the conical reflector (2) and focused by lenses (3) through slots (4 and 5) onto the double end prisms (6). The prisms in turn reflect the light into the phototubes (10 and 11). Phototube (10) is a scanning tube and (11) is a trigger tube. The scanning disc (7) is rotated at 6 rpm by the synchronous motor (8). The stationary disc (9) has a long scanning slot (5) located behind the instrument pointer and a short trigger slot (4). The pointer (12) is the indicating needle of the primary measuring element. Its position is indicative of the value to be transmitted.

In operation, the scanning disc rotates as indicated by the arrow. When a lens passes slot (4) a beam of light falls on the trigger tube (11) which sets up a relay circuit making it possible for the telemeter signal to be transmitted. As the other light beam falls through slot (5) phototube (10) is energized and the telemeter signal begins. It continues until the light beam is broken by the pointer (12). The length of the impulse depends on the position of the pointer. The next impulse will start after the trigger tube has again been illuminated through slot (4). The signal transmitted by this device is of the same type as that transmitted by the electromechanical device already described.

The receiver for this system may be of the electromechanical type already described. Another type of receiver converts the signal into a direct current which magnitude is proportional to the duration of the impulses. This current is read on conventional milliampmeter instruments. The receiver uses the principle of charging a condenser through a resistor from a constant dc voltage source. The duration of the signal determines the charging time. Thus the condenser voltage is proportional to the impulse duration. This voltage is applied to the grid of a triode to control the plate current flowing through the indicating meter.

The duration of impulse system is used extensively in the pipe line industry for the transmission of pressures, flow meter readings and power readings. Water supply systems utilize it to telemeter reservoir levels, gate positions and flow meter readings. Normally a 60-cycle power supply is required for the transmitter and receiver to drive the synchronous motors. The signal can be transmitted over 2-wire metallic circuits, one wire and ground, by carrier or microwave. One west coast gas company which desired to telemeter flow meter readings from a remote point had no power source at the remote location. The duration of impulse method using electromechanical instruments was desirable. A transmitter was modified to operate on a chronometrically controlled dc instrument-type motor powered by batteries. The arrangement proved quite successful, thus demonstrating the versatility of this system.

Telemetering by Teletype

Another telemetering system utilizes teletype circuits and teletype equipment for the transmission of pressure and ammeter readings. The system is in use by a major pipe line company. It is part of a remote control system for the automatic pump stations. The system may be classed as a coded pulse system and the fundamental principles are simple. A transducer or local transmitter is used that gives a timed pulse which duration is proportional to the measured variable. This pulse controls the long distance transmitter which generates teletype signals.

The letter D is used for the transmission of pump-discharge pressure and each D represents 20 pounds per square inch. A standard teletypewriter is used for a receiver. If the measured pressure is 800 psi, for example, the transmitter will send 40 D's and the receiving teletype machine will print them out. The operator need only count the number of characters to determine the value of the

measured variable.

To check the operation of any station on the system, the pipe line dispatcher dials a number using a standard telephone-type dial. The selector at the desired station sets the telemetering equipment into operation. The transmitting equipment sends first discharge pressure as described, then S's for suction pressure and finally A's for motor current. Each S represents 10 psi. Each A represents 2.5 amps.

Figure 9 shows a simplified schematic drawing of the various components of a transmitting system. A is a bourdon tube pressure pickup connected to an adjustable core transformer. The output voltage of the transformer is proportional to the measured pressure. B is a scanning device consisting of a linear rise cam which positions the core of a second adjustable core transformer. Since the cam is driven at constant speed (1 revolution each 15 seconds), the voltage output of this transformer varies proportionally with time. The outer cam mounted integrally with the scanning cam actuates a microswitch. The setting of the cams is such that the microswitch closes when the transformer output is zero and opens when the transformer output has reached maximum. The output of pickup A and scanner B is fed into an electronic relay. The output of the electronic relay keeps its magnetic relay coil energized and the circuit to the teletype transmitter closed as long as the pickup voltage is greater than the scanner voltage. When the scanner voltage equals the pickup voltage, the electronic relay de-energizes its magnetic relay and the circuit to the teletype transmitter sends characters for a period of time which is proportional to the pressure pickup voltage and therefore to the measured pressure. Thus, the number of characters transmitted is an indication of pressure.

The full scale value of any measured variable is set at 60 characters. The transmitting speed of the teletype equipment is 368 characters per minute so that 60 characters require 9.78 seconds. The remaining time between 9.78 and 15 seconds is used to switch to the various pickups. Any variable which can be expressed as a voltage can be transmitted by this method.

Telemetering by Voice

A new development in telemetering employs a magnetic storage drum containing recorded words and numbers, and transmits information over standard voice channels by means of spoken words. The present application is the telemetering over a telephone line of liquid levels in a gasoline storage tank. This application is in connection with the operation of a remote controlled pipe line pumping station. The problem is to measure the liquid level in a 40-foot high tank to the nearest one-eighth inch and transmit this information to the point of control, in this case, 30 miles distant.

Figure 10 is a diagram showing the principle components of the system. The primary measuring element is a float member inside the tank. As the float moves up and down, it drives a transmitter by means of a perforated tape pulled over a sprocket. This part of the equipment, along with the indicating panel, is commercially available equipment. The float position is transmitted over a multipair cable to the indicator panel. This arrangement, in itself a telemetering system, is limited in operation to a distance of a few thousand feet. The transmitter is a selector switch device which applies power sequentially to four groups of wires (35 in all) as the float moves up or down rotating the drive sprocket. The 35 wires terminate in an indicator panel having four columns of lights as shown. The four columns reading from the left represent tens-of-feet, feet, inches and fractions-of-inches. Those wires of the 35 which are energized by the selector switch at the tank light their corresponding lights to indicate the position of the float in feet, inches and fractions. One indicator panel can serve an entire tank farm. The multipair cable is common to all tank transmitters. An extra wire to each tank carries power to the individual transmitters and serves as a selector wire to bring in the reading from the desired tank.

The long distance telemetering equipment is designed to work in conjunction with the local telemetering system just described. This equipment is shown schematically in Fig. 11. The drum A has 16 voice channels recorded on its magnetic surface and 16 pickup heads permanently mounted as shown. The recorded words are the numbers ZERO to ELEVEN inclusive, the words FEET, AND, EIGHTHS, and INCHES. Any tank level to the nearest eighth inch can be given by using eight of these sixteen words. In transmitting a tank gauge, the drum makes eight revolutions playing one word each revolution. A typical tank gauge would be a spoken sentence such as this "THREE ONE FEET SIX AND FIVE-EIGHTHS INCHES" to indicate a liquid level of 31' 6-5/8". The output from the pickup heads is amplified and can then be transmitted by any method by which voice can be transmitted.

Figure 11 also shows the various circuits involved in transmitting the words corresponding to the lights displayed on the indicator panel. A relay is used in conjunction with each lamp, each relay being connected in parallel to its corresponding lamp. The relay contacts are in series with the voice circuits from the pickup heads on the drum. The relays are in groups corresponding to the columns of lamps. During the first revolution of the drum, one side of all contacts at the first bank of relays is connected to the amplifier by the stepping switch. Only one relay will be closed (that one connected to the lighted lamp) and the pickup head for that relay will be connected to the amplifier. There is a dead spot on the drum at the end of each word. While the drum is rotating through this dead spot, the stepper switch is advanced one step by the cam actuated microswitch. This connects the second group of relays to the amplifier. Again that relay which is closed routes the output from its pickup to the amplifier and the second word is transmitted. This is repeated for each column of light. At the appropriate places in the message, the stepper switch selects the pickups for the words FEET, AND, EIGHTHS, and INCHES. The drum is 9 inches in diameter and 10 inches high. It is driven at 72 rpm through suitable gearing by an 1800 rpm synchronous motor. Each word occupies approximately 270° of drum surface. The remaining 90° of rotation is used for switching. The complete liquid level is transmitted in approximately 6 seconds.

The system may at first glance seem unnecessarily complex to accomplish the desired results. However, it has some definite advantages over other conventional methods which might be adapted to the application. One transmitter can serve any number of tanks at any given location. It will transmit digital information without the necessity of converting to analog information, which action at least tends toward greater accuracy. Transmission is accomplished over existing voice channels, telephone or otherwise. The information may be received at any point having telephone communication without any special receiving equipment whatever. If desired it would be entirely possible to assign each tank in a system a telephone number. By dialing the desired number, a dispatching office could obtain tank gauges from any location in the system.

Conclusions

It was pointed out that telemetering is the practice of obtaining measurement data at a more convenient location than the source. It is apparent that this field is of considerable interest to the electronic engineer because of the variety of electronic equipment involved in a telemetering system. Some component types included are microwave, carrier, other communications equipment, selsyns or the equivalent, servo mechanisms, magnetic recorders, multiplexing systems, various types of modulation and demodulation, photoelectric equipment, and many others. When combinations of these various equipments are considered there are many possibilities in this field. In addition to application of known techniques there are many basic problems still to be solved. Two of these in the pipe line industry are cited. One is to provide an automatic method of distinguishing products so that switching can be made at the proper time in order to separate the fluids. Another is to provide an automatic or semi automatic means of controlling safety valves in the case of pipe leaks or breaks. Undoubtedly there are many more in the petroleum and other industries which will attract the attention of the radio and electronic engineer in the future.

Acknowledgments

The authors are indebted to the Philco Corporation and the Southwestern Bell Telephone Company for the assistance of their engineers in answering technical questions regarding various equipment.

References and Bibliography

1. A.S.A. Standards C42-1932, 30.40.150.
2. "Principles and Methods of Telemetering" by Perry A. Borden and Gustave M. Thynell. Chapt. 1, Reinhold Publishing Corp., New York.
3. "Multipoint Telemetering System Using Teletype Transmission" by A. J. Hornfeck and G. R. Markow, Electrical Engineering, October 1952.
4. Bulletin EA-2. Control Corporation, Minneapolis, Minnesota.
5. "The Platte Story-Communications System" by Allan R. Heidebrecht. p. D-31, Petroleum Engineer, December 1952.
6. "Microwave Communication Links" by Joseph Racker. Petroleum Engineer, March 1952.
7. Bulletin 41-810, "Freq-O-Tron Telemetering Equipment", Westinghouse Electric Company.

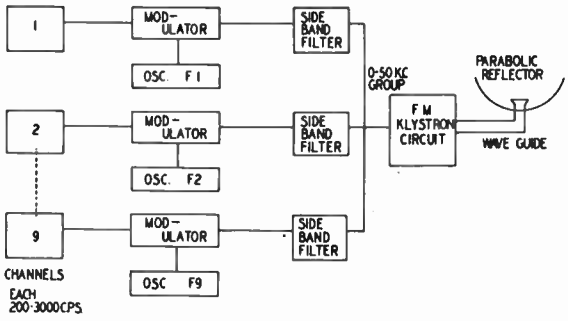


Fig. 1
Simplified diagram of frequency-division multiplex transmitter.

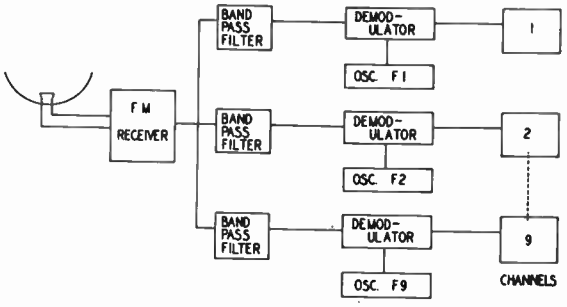


Fig. 2
Simplified diagram of frequency-division multiplex receiver.

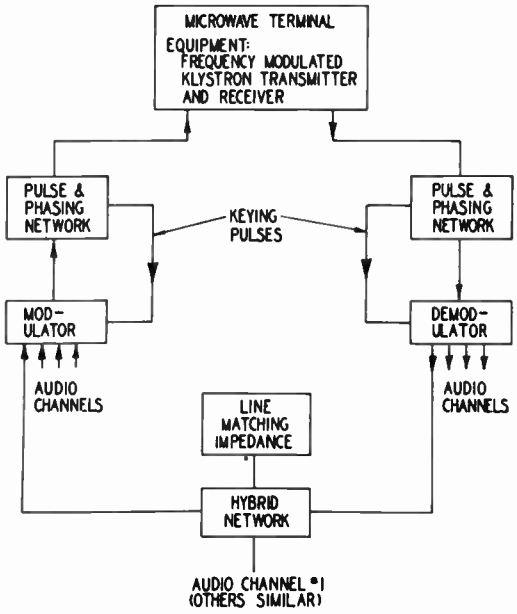


Fig. 3
Simplified diagram of time-division multiplex system.

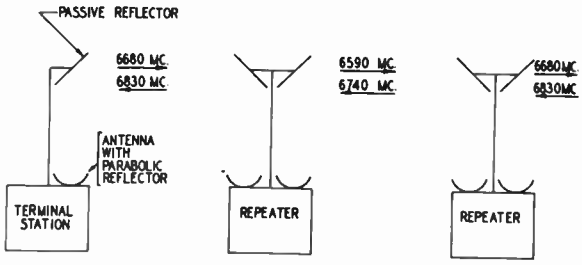


Fig. 4
Typical frequencies used in two-way microwave relay system.

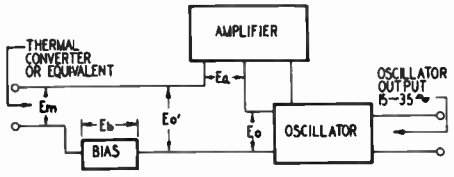


Fig. 5
Variable frequency transmitter.

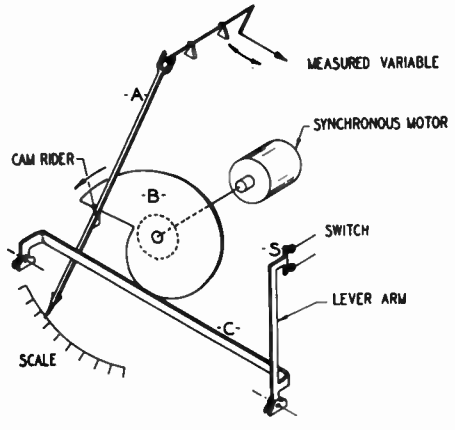


Fig. 6
Duration of impulse transmitter.

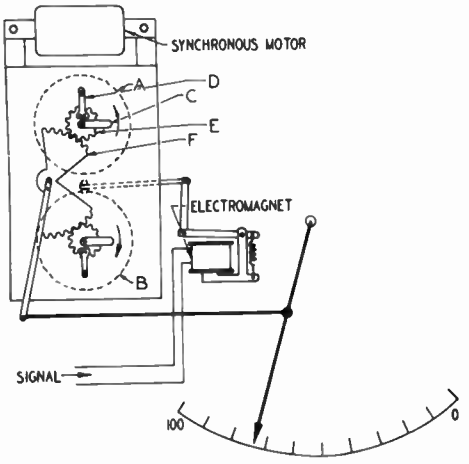


Fig. 7
Duration of impulse receiver.

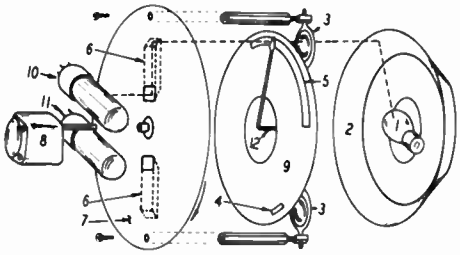


Fig. 8
Photoelectric scanner.

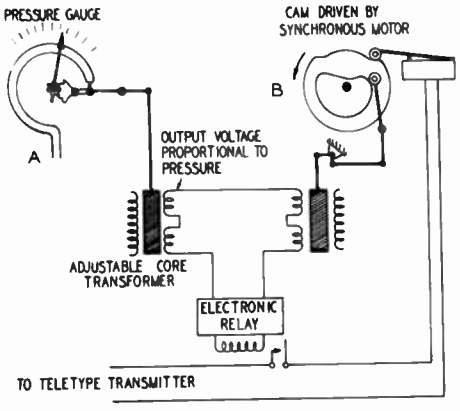


Fig. 9
Impulse generator for teletype telemetering.

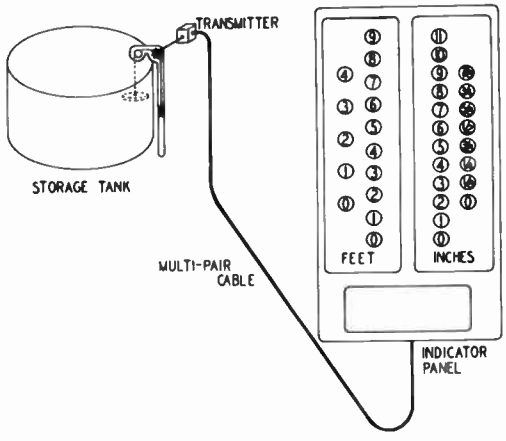


Fig. 10
Local telemetering system for liquid level indication.

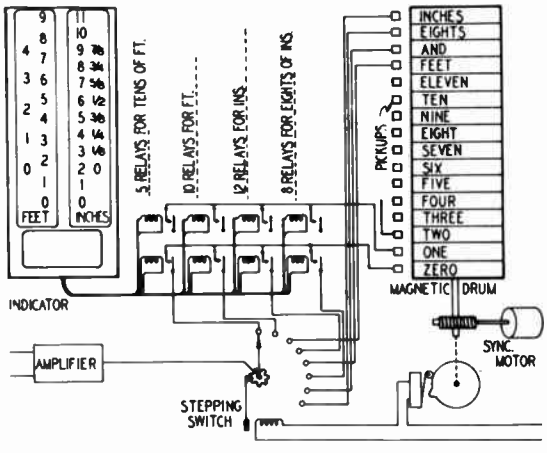


Fig. 11
Voice telemetering transmitter.

THE ORGANIZATION OF A DIGITAL REAL-TIME SIMULATOR

Harry J. Gray, Jr.
University of Pennsylvania
Moore School of Electrical Engineering
Philadelphia, Pennsylvania

I. Introduction

Heretofore, most real-time simulators of dynamical systems have been analogue devices; but with the advent of high-speed electronic digital computers, the possibility arises of using them in simulators, their flexibility being a great advantage.

Under the sponsorship of the Office of Naval Research, Special Devices Center, the University of Pennsylvania conducted a study to determine the feasibility of building a digital real-time simulator for a system described by thirteen first-order non-linear differential equations.

A preliminary study indicated that the computation time for one integration interval took about 0.22 seconds for a computer similar to the Raytheon Hurricane. Since less than 0.1 seconds were permitted an investigation, the results of which follow, was conducted to find a faster, realizable machine organization using existing techniques.

II Serial versus Parallel Machines

Serial type machines were considered in preference to parallel type machines because (a) existing serial memories of the size contemplated are as fast as or faster than existing reliable parallel memories and (b) the arithmetic unit in a serial machine is appreciably smaller (see sections V and VII).

The superficial organization was assumed to be similar to that introduced in the EDVAC, and now in common use: representation of numbers as signed, binary magnitudes, automatic sequencing by instructions stored in the memory, a central control unit, and an algebraic unit. Time was divided into pulse times (the reciprocal of the clock frequency), minor cycles (the duration of a word) and major cycles (number of words in a mercury delay line times duration of one minor cycle).

III n-Address Codes, $n \leq 4$

In the four-address code, an order type specifies the operation and three addresses specify the memory locations of the two operands and the result. A fourth-address, omitted in the three-address and one-address codes, specifies the location of the next order. In the three-address and one-address codes, the location of the next order is specified by an order counter. In the one-address code, the result remains in the arithmetic unit until called therefrom by an order

A one-address machine (machine #3 table 1) using an 18 digit word and suited to the application was set up using the same orders that were proposed for the Institute of Advanced Study machine.[1] Three-address and four-address codes (machines #2 and #1 respectively) were similarly set up and time estimates for one integration interval with the given equations were made. Improvement over the aforementioned figure of 0.22 seconds is evident even for the four-address machine. Selection of order types more suitable for the given equations resulted in even greater speed (machine #4).

IV Partitioning of the Memory

Since, except for Compare (see Appendix), adresses refer only to the number memory positions of which there were relatively few (about 500) it was felt that a possible saving in address length (and computing time) could be achieved by partitioning the memory into separate order and number sections. The word structure is given in Table 2.

A further saving occurs in the one-address code because switching time is not needed. For the three-address and four-address codes, a two pulse time switching blank was allotted for extracting numbers. None is needed for extraction of orders since this is provided by the time taken for the arithmetic operations. Assuming synchronism of order and number memories yields the values for the word length in table 3. A definite improvement in computing speed is evident (machines #5, #6, and #7).

V Whipple-tree Multiplier

Multiplication by repeated additions takes about 15 minor cycles for a 15 bit plus sign word. Even if three times as many other operations taking one minor cycle apiece are performed in one integration interval, use of a multiplier which can multiply in two minor cycles results in over a three-fold increase in speed.

Knowing that two-cycle multipliers, such as the "whipple-tree" multiplier, exist; but leaving the choice of the best multiplier to a later date, time estimates for one-, three-, and four-address codes (machines #5, #6, and #7) were made to see if computing time for the given problem were greatly reduced.

Incidentally, the superiority of the one-address code is increasingly evident.

VI Auxiliary Storage Registers

One advantage of the one-address code is that it reduces the need for returning information to the memory unless it needs storage for use later on, but the reduction is not complete. Since an order always carries an address which specifies a

number memory location, it would be desirable to extrac an order from the memory only when access to the numbe- memory is desired. However, for example, a number may be removed from the memory, multiply a number stored in the arithmetic unit, and the product be added to the result of the preceding computation, the result to remain in the arithmetic unit. To do this with one order requires a "composite" order and auxiliary registers in the arithmetic unit. Such a composite multiply-add order was set up and times were estimated (machine #8). A small improvement is evident.

If a complete set of composite orders is utilized, the times become those for machine #9.

VII Whipple-tree Divider

Division, as performed in the conventional non-restoring process,[2] takes roughly as many word times as there are digits in the quotient, because the decision on whether to add or subtract depends on the sign of the preceding partial remainder. If for n partial quotient digits, the 2^n partial remainders corresponding to the 2^n possible partial quotients are computed in one word time as in the structure of figure 2, with the adders being those used in the whipple-tree multiplier, then for a quotient of N digits, N/n word times are needed. If A and B are respectively the magnitudes of the dividend and divisor then only one of the two partial remainders, $R_0=2A$ and $R_1=2A-B$ corresponding to assumed quotient digits 0 and 1 respectively is correct. Rules for determining the correct remainder and establishing that its properties are the same as the dividend have been obtained. However, the divider of fig. 1 with machine #8 gives a 5% increase in cost for a 5% increase in speed.

VIII Double-Bus Computer

In machine #8, which now combines the greatest speed with the greatest economy, numbers and orders are extracted from the memory in alternate word times (minor cycles).

In all operations taking one cycle (all operations except multiplication and division) the flow of information on the memory busses assuming separate order and memory busses can be as shown diagrammatically in figure 2.

Curiously, in order to mechanize a double-bus computer, less equipment is likely to be required because the same tank selector switches would be used as before and the output busses would no longer need a combining unit.

It is now necessary to include a switching blank in the word. A two pulse time blank has been selected and time estimates have been made for machine #10 which includes the features described in the next section.

IX Minimum Time Programming

In a serial machine using mercury delay line memories having for reasons of economy n words of duration T stored in a tank (delay line)each word is accessible at time instants nT apart (once a major cycle).

All previous times have been estimated assuming that each memory position called upon is accessible without delay (minimum time programming), a condition attainable if each memory position is specified by an address.

In the proposed machine, the number memory positions are specified by an address while the order memory positions are specified by a counter. Rather than using the costly method of having the counter call on the order memory positions in a nonsequential but predetermined manner, every operation time is reduced to either one major cycle or one minor cycle, automatically achieving minimum time programming for the order memory (except for compare). All operations except multiplication and division take one cycle, however a study of the time relationships in multiplication show that the significant part of the product becomes available in one minor cycle and is ready to be operated on further, so that multiplication takes effectively one cycle. However, the multiplier is occupied for two minor cycles for a product, so that to avoid losing a major cycle when a multiply order follows a multiply order, one order memory position is wasted by introducing a mark-time order (W) between successive multiplications.

In division, it would not be too expensive to use a one or two digit whipple-tree divider to permit division to take not more than one major cycle. For example, if a tank contains 16 words, a 16 binary digit quotient could be obtained using either a one digit whipple-tree or a conventional divider, but for more than about 16 and less than about 32 quotient digits a two digit whipple-tree divider is needed. (See machine #10)

X Conclusions

The characteristics of the "optimum" computer utilizing existing computer components are as follows.

(a) It is serial and uses a binary number representation.

(b) The number and order words will be at least 16 bits in length with a 2 pulse time blank.

(c) It will use a one address code.

(d) It will have separate order and number memories having separate input and output busses.

(e) It will use serial memory units.

(f) It will have a single arithmetic unit incorporating a high-speed multiplier, a one or two digit whipple-tree divider, and a multiply-add composite order.

(g) It will have auxiliary registers in the arithmetic unit.

(h) A revaluation of memory requirements for the given problem indicate that a total memory capacity of at least 2000 words is needed. Sufficient reserve capacity will be included to enable the simulator to be used for more complex problems.

Acknowledgements

The author wishes to acknowledge with thanks the assistance and the many valuable suggestions given by Mr. Puzant V. Levonian and Dr. Morris Rubinoff. This work was done under contracts

N6onr-24913 and N6onr-24915 sponsored by the Office of Naval Research, Special Devices Center, Port Washington, New York.

Appendix - List of Order Types

(A+) Add number coming from memory position specified by address into arithmetic unit.

(A-) Same as (A+) except change sign of number in arithmetic unit before adding.

(Ac) Same as (A+) except clear arithmetic unit before adding.

(S+) Subtract number coming from memory position specified by address from contents of arithmetic unit.

(S-) Same as (S+) except change sign of contents of arithmetic unit before subtraction.

(Sc) Same as (S+) except clear arithmetic unit before subtracting.

(AM+),(AM-),(AMc),(SM+),(SM-),(SMc) These orders correspond to the above except that the sign of the number coming from the memory is deleted.

(M) Multiply number coming from memory position specified by address by number in arithmetic unit.

(Da) Divide number in arithmetic unit by number coming from memory position specified by address.

(Dm) Same as (Da) except that number in arithmetic unit is the divisor.

In all the above orders the result of the operation remains in the arithmetic unit.

(T) Transfer number from arithmetic unit to memory position specified by address.

(S) Shift number in arithmetic unit by an amount specified by address. Sign is not shifted.

*(Cc) Examine sign of number in arithmetic unit. If positive (zero is positive by definition) sequence or orders is not disturbed. If negative, order counter is set to value specified by address.

*(Cu) Order counter is set to value specified by address.

(L) Number in arithmetic unit is logically multiplied by number coming from memory position specified by address. Result returns to arithmetic unit.

(W) Order counter is advanced one count at normal rate. No other operations are performed.

(T'+) Number is transferred from memory position specified by address to a special register in arithmetic unit.

(T'-) Same as (T'+) except that sign of number is reversed.

(M') Number in special register is multiplied by number coming from memory position specified by an address. Result is added into arithmetic unit.

References

1. "Preliminary Discussion of the Logical Design of an Electronic Computing Instrument", Burks, Goldstine, von Neumann, IAS, Princeton, N.J., 1946.

2. "A Functional Description of the EDVAC", EDVAC Staff, University of Pennsylvania, Moore School of Electrical Engineering, Research Division, Report 50-9, November 1, 1949, Volume I, Section 4.4.2.4, 4.4.2.5.

Footnote

*It is recognized that the address in these compare orders may not have sufficient digits to permit entering all the order memory position. This is not expected to be serious since the given problem involved less than 30 compare orders.

MACHINE NO.	PARTITION MEMORY	WHIPPLETREE MULTIPLIER?	COMPOSITE ORDERS?	DOUBLE BUS?	CODE	COMPUTING TIME SECONDS*			COMMENTS
						1 MC.	2.25 MC.	4 MC.	
1	NO	NO	NO	NO	4A	.237	.106	.059	SEE SECTION III
2	NO	NO	NO	NO	3A	.218	.097	.054	SEE SECTION III
3	NO	NO	NO	NO	1A	.206	.091	.050	IAS CODE (CODE ONLY)
4	NO	NO	NO	NO	1A	.176	.078	.043	MODIFIED CODE
5	YES	YES	NO	NO	4A	.115	.051	.029	SEE SECTION IX, X
6	YES	YES	NO	NO	3A	.098	.044	.025	IX, X
7	YES	YES	NO	NO	1A	.069	.031	.017	IX, X
8	YES	YES	MULT-ADD	NO	1A	.062	.028	.016	SEE SECTION XI
9	YES	YES	FULL USE	NO	1A	.056	.025	.014	SEE SECTION XI
10	YES	YES	MULT-ADD	YES	1A	.040	.018	.010	SEE SECTION XIII, IX

Table I
Computing times for all machines.

*The 1-mc rate refers to a computer similar to SEAC or the Moore School MSAC; the 2.25 and 4-mc rates to the Univac and Hurricane, respectively. These items are the minimum possible.

CODE	OPERAND 1	OPERAND 2	RESULT	NEXT ORDER	ORDER TYPE	TOTAL
1A	10	x	x	x	5	15
3A	10	10	10	x	4	34
4A	10	10	10	13	4	47

Table II
Word structure.
Number length: 15 bits and sign = 16 bits.

CODE	NUMBER WORD	ORDER WORD
1A	16	16
3A	18	36
4A	18	54

Table III
Word lengths in pulse times.

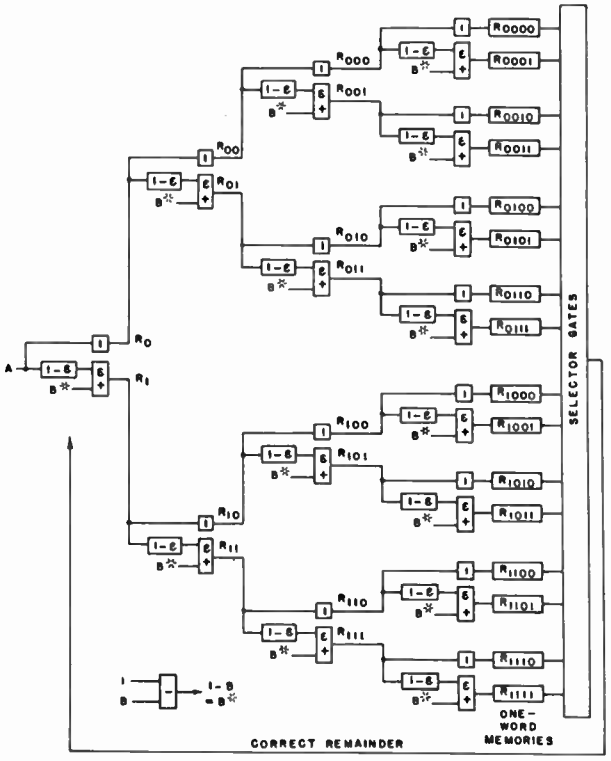


Fig. 1
Whipple-tree divider.

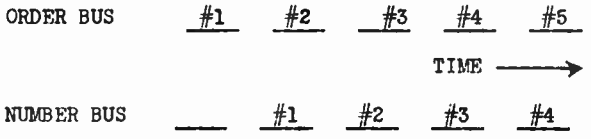


Fig. 2
Illustrating double bus timing.

CONTROL SYSTEMS ENGINEERING APPLIED TO AUTOMOBILE SUSPENSIONS SYSTEMS

Gunther J. Martin — Research Engineer - Ford Motor Company*
Robert D. Jeska — Research Engineer - Ford Motor Company*
Dearborn, Michigan

Introduction

Stability and ride characteristics of automobile suspension systems have been the subject of an intense investigation for a number of years. Until recently, however, the methods of analysis of this problem have consisted mainly of finding solutions of the differential equations of motion, and data taken in the field or on a bump rig. With the advent of the electronic analog computer, a somewhat different approach to this problem is being developed. Although the method is still mainly one of finding solutions of the differential equations of motions for various values of physical constants and geometrical arrangements of suspension members, a different insight to the problem is gained by using the language and the representations used by the control systems engineer. The advantage of this method is mainly in its approach. Emphasis is not placed on the method of solution, but rather on the physical arrangement and its equivalent block diagram representation. It is this representation which stresses a need for a revision in the velocity sensitive components of a suspension so that good vehicle stability and ride characteristics need not be opposing fsctors.

Part 2 of this paper deals with the simulation of an automobile suspension in which a different approach to the velocity sensitive element problem is tried.

1. Simulation of Dynamic Systems

A simple dynamic system and its equivalent block diagram are shown in Figures 1 and 2.

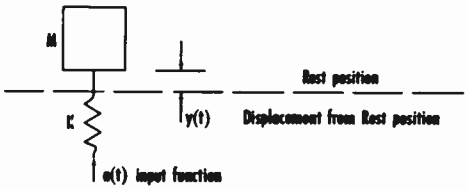


Fig. 1

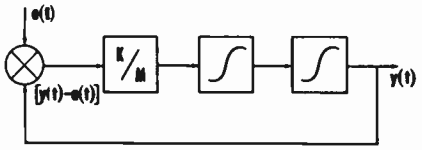


Fig. 2

The transformed equations for these two systems are stated as follows:

$$K/M \quad E(p)\frac{1}{p^2} = Y(p) \qquad (1)$$

$$E(p) = [A(p)-Y(p)] \qquad (2)$$

$$[p^2 + K/M]\, Y(p) = K/M\, A(p) \qquad (3)$$

An examination of these equations suggests that the system shown in Figure 1 has a force feedback like its equivalent in Figure 2, and in addition they both have the same transfer characteristics. It should be noted that the position error term, E(p), of the system in Figure 1 is exactly the same as its classical equivalent in the servo-mechanism and that this error is acted on by the system gain, K/M, until it becomes nearly zero. In this example, the error cannot become zero since it will oscillate indefinitely.

Since the physical system exhibits a positive definite potential energy function so that the roots of its characteristic equation will appear only in the left hand p-plane, the usual criterion for stability need not be invoked here.

The system response and its associated error, however, have very important significance. In general, a very small response to the input a(t) is desirable. Other factors of course must enter into this problem such as the form of the function a(t), static loads, and coupling between modes of vibration in the automobile. These factors may change the requirements of the performance, but in general, for driving over a road consisting of washboard and pot hole type inputs, we desire small response or large errors. This is the opposite of the view taken by the control engineer, who usually desires small errors in control systems. An automobile suspension may be likened to a closed loop control device, but one in which the fidelity is low. It is noted that this low fidelity is accomplished in part by making the system gain, K/M, small, which means weak springs and a large mass.

Velocity sensitive elements are the most troublesome suspension members; these are usually characterized by the so-called shock absor-

*Formerly of Willow Run Research Center, University of Michigan.

bers and may be of varying types such as viscous or dry friction. Their main function is one of damping; however, damping can have adverse as well as helpful effects on the desired response depending on whether or not the mass is subjected to a force equal to $C[\dot{Y}(t) - \dot{a}(t)]$ or $C\dot{Y}(t)$. Figures 3 and 4 show two equivalent systems in which damping is included.

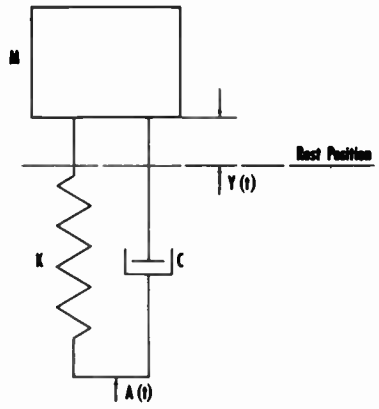


Fig. 3

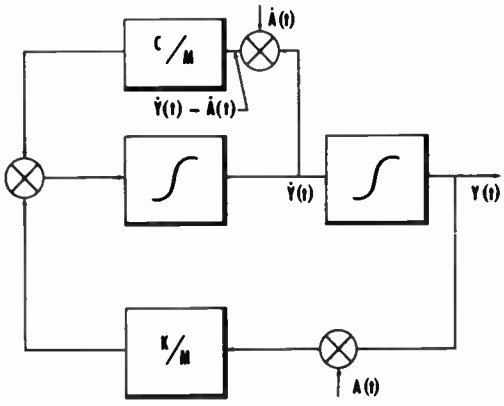


Fig. 4

The system shown in Figure 4 shows how large values of $\ddot{a}(t)$ can introduce large accelerations into the system. These large values of $\ddot{a}(t)$ are caused by driving over a washboard type road at high speed or hitting pot holes. Varying the values of C/M alone, cannot be depended on to give successful results, i.e., weak response to input functions, since large values of C/M introduce large accelerations due to the input $\ddot{a}(t)$, and small values of C/M cause the system to be underdamped. These results make a compromise choice of C/M necessary.

The system shown in Figure 5 overcomes, in part, the need for a rigid compromise value of C/M; here the damping value C is controlled by the wheel acceleration. This arrangement essentially prevents high damping forces from getting into the system during the time when a large input acceleration is acting.

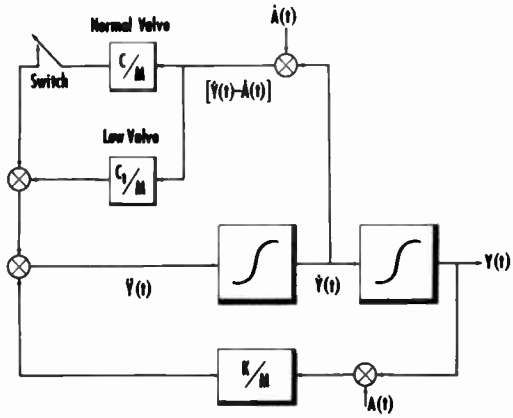


Fig. 5

The switch shown in Figure 5 is normally in the closed position and it is automatically opened only when the wheel acceleration exceeds a predetermined value.

Other arrangements for the purpose of producing weak responses to input functions are of course possible. The one shown in Figure 5 illustrates how a particular portion of a new suspension system may first be tentatively designed. Ultimately, of course, one must design equivalent mechanical elements which have similar transfer characteristics to those selected in the block diagram representation.

2. Computer Solution of the Suspension Problem

This section illustrates how an actual problem dealing with velocity sensitive elements of a suspension system was solved. A simulation of an automobile suspension in translation and pitch was first set up on an electronic analog computer. Figure 6 shows the system which was used.

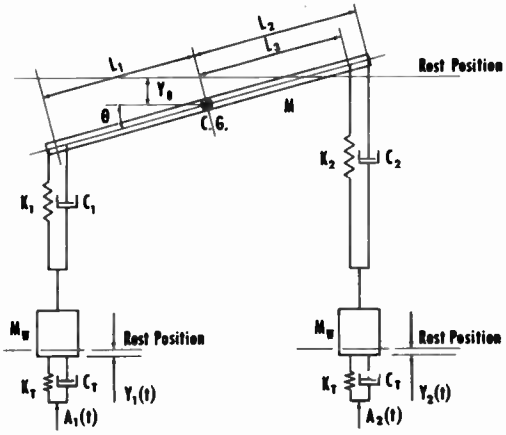


Fig. 6

The equations of motion are obtained by writing equilibrium equations of forces and moments about the center of gravity and the two wheel masses, M_w. They are written in matrix form as follows:

$$A\ddot{Y} + B\dot{Y} + CY = X \qquad (4)$$

Where

$$Y = \begin{pmatrix} Y_o \\ \theta \\ Y_1 \\ Y_2 \end{pmatrix}, \quad \dot{Y} = \begin{pmatrix} \dot{Y}_o \\ \dot{\theta} \\ \dot{Y}_1 \\ \dot{Y}_2 \end{pmatrix}, \quad \ddot{Y} = \begin{pmatrix} \ddot{Y}_o \\ \ddot{\theta} \\ \ddot{Y}_1 \\ \ddot{Y}_2 \end{pmatrix}, \quad X = \begin{pmatrix} 0 \\ 0 \\ a_1(t) \\ a_2(t) \end{pmatrix}$$

and

$$A = \begin{pmatrix} -(K_1+K_2) & -(K_1L_1-K_2L_2) & K_1 & K_2 \\ -(K_1L_1-K_2L_2) & -(K_1L_1^2+K_2L_2^2) & K_1L_1 & -K_2L_2 \\ K_1 & K_1L_1 & -(K_1+K_t) & \\ K_2 & -K_2L_2 & & -(K_2+K_t) \end{pmatrix}$$

$$B = \begin{pmatrix} -(C_1+C_2) & -(C_1L_1-C_2L_2) & C_1 & C_2 \\ -(C_1L_1-C_2L_2) & -(C_1L_1^2+C_2L_2^2) & C_1L_1 & -C_2L_2 \\ C_1 & C_1L_1 & -(C_1+C_t) & \\ C_2 & -C_2L_2 & & -(C_2+C_t) \end{pmatrix}$$

$$C = \begin{pmatrix} M \\ Mk^2 \\ Mw \\ Mw \end{pmatrix}$$

The solution of equation 4 can be carried out on an analog computer as shown in Figure 7. It should be noted that the system consists of four degrees of freedom, each degree is characterized by a resonator having a transfer function:

$$\frac{Y_o(p)}{Y_1(p)} \quad \frac{1}{p_2 + a_{11}p + b_{11}}$$

The matrices B and C show the coupling that exists between the various resonators, and this coupling appears on the computer as the gain in the closed loops around the same resonators as in the actual system. The transfer functions for the coefficient blocks a_{ij}, b_{ij} are given in equation 5. The suspension elements are in general non-linear; however, for the sake of simplicity, neither the nonlinearities nor their accompanying circuits are shown. The computer "setup" then becomes one of making closed loop patterns around transfer functions which are equivalent to those found in the original system. This equivalence is again pointed out by the fact that the matrices B and C show no direct coupling between the wheel masses, and the fact that the closed loop between wheel transfer functions are absent.

$$a_{ij} = a_{ji} \qquad b_{ij} = b_{ji}$$

$$a_{11} = -(C_1+C_2) \qquad b_{11} = -(K_1+K_2)$$

$$a_{12} = -(C_1L_1-C_2L_3) \qquad b_{12} = -(K_1L_1-K_2L_2)$$

$$a_{13} = C_1 \qquad b_{13} = K_1 \qquad (5)$$

$$a_{14} = C_2 \qquad b_{14} = K_2$$

$$a_{22} = -(C_1L_1^2+C_2L_3^2) \qquad b_{22} = -(K_1L_1^2+K_2L_2^2)$$

$$a_{33} = -(C_1+C_t) \qquad b_{33} = -(K_1+K_t)$$

$$C_{44} = -(C_2+C_t) \qquad b_{44} = -(K_2+K_t)$$

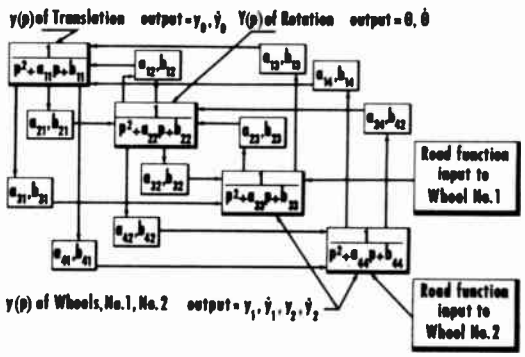


Fig. 7 - Computer diagram of suspension simulation.

The input functions for this system were obtained by surveying the road over which the actual vehicle was driven. A wheel cut to the proper rolling radius, and having a small light at its center, was rolled over the road at night, and a time exposure photograph was taken. The result was transferred to two function tables shown in Figure 8. The drag cards were properly phased for the wheelbase and swept across the road functions.

Translational and angular acceleration records, at the C.G., were taken in the actual

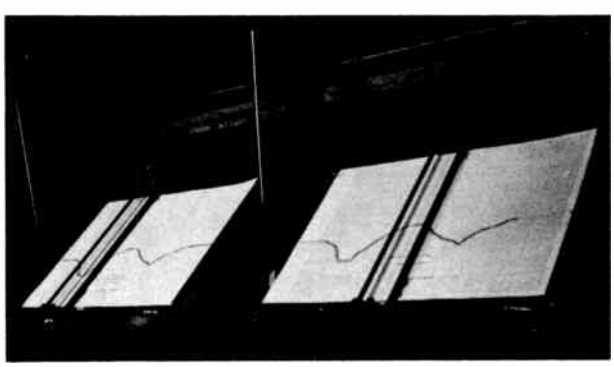


Fig. 8

vehicle and these were compared with the computer solutions. Figure 9 shows the excellent comparison between the computer runs and the field tests. These results indicate significantly that comparisons between computer solutions and field tests can be in good agreement if careful attention is given to details.

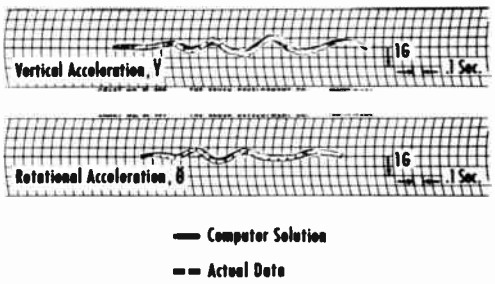


Fig. 9
Comparison of computer results with actual data.

The simulation of the automobile suspensions, having been felt to be reliable and quite representative of the characteristics of its physical counterpart, was next incorporated in the system shown in Figure 5. The problem is restated in somewhat more detail as follows: The forces acting at the shock absorber will be, in general, $F = f[\dot{y}_c - \dot{y}_w]$ where $\dot{y}_c$ equals the vertical velocity of car body and $\dot{y}_w$ equals the vertical velocity of the wheel mass with respect to their rest positions. Since this is an error rate damped system, large damping forces can be introduced by large values of $C\dot{Y}_w$. Since it is not feasible to lessen this force for long periods of time by making the damping force smaller, an alternative is proposed which effectively utilizes two values of the damping coefficient and sets up a bi-modal action. It is, therefore, desired to study the effect of selecting one value, depending on wheel acceleration, so that the shock absorber is effectively by-passed during the transient periods of high acceleration. The elements of such a shock absorber are shown in Figure 10.

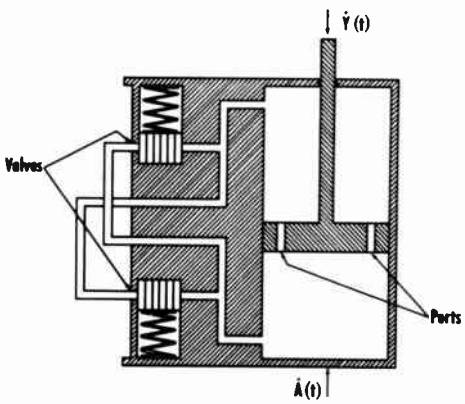


Fig. 10

Prior to putting this shock absorber in the simulation, the system was subjected to a square wave input shown in Figure 11A. Figures 11B, 11C, and 11D show the position response, translational and angular acceleration at the C.G., respectively. The position response is shown to follow the input with a fair degree of faithfulness, which points out that the fidelity of this system is rather high, and, as expected, the acceleration responses at the beginning and end of the input function predominate. There is little doubt that passengers would sense this particular input with some degree of harshness. With the acceleration modulated shock absorber in the simulation, the system was again subjected to the same square wave input shown in Figures 11A and 12A. Figures 12B, 12C, and 12D show the new position response, translational and angular acceleration respectively. The position response now shows that the fidelity of the system is much

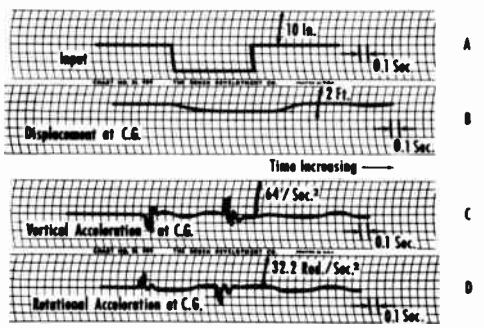


Fig. 11
Response of system with conventional shock absorber to square wave input.

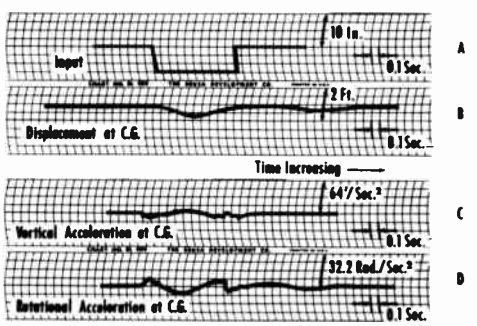


Fig. 12
Response of system with modulated shock absorber tc square wave input.

lower, which means in this case that the vehicle eases into the bump more gradually. The accelerations at the beginning and end of the input function are now significantly less. The results indicate that this acceleration modulated shock absorber has given the system a "softer" response during the time that high wheel accelerations are acting, but it should not be taken to mean that

some compromise is no longer necessary since relieving the harshness is gained at the expense of some overshoot in the position response. Unlike the normal case however, less overshoot is now present.

3. Conclusion

The response curves point out that motions in automobile suspension systems closely resemble those which occur in some closed loop control devices. This is so because dynamic systems are fundamentally force feedback systems. The use of the principles of force feedback enables us to make a more basic approach to the problem, and it is hoped that it will ultimately permit the direct application of the theory of best automatic control.

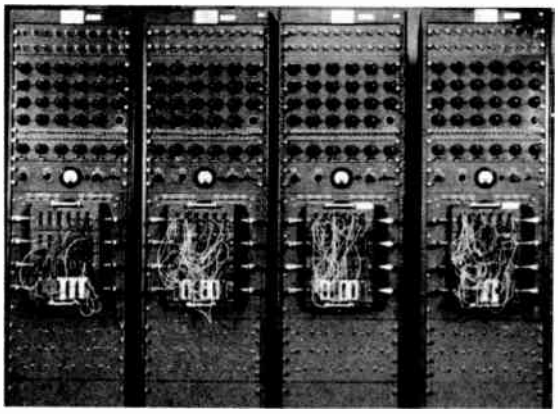


Fig. 13

EXPERIMENTAL EVALUATION OF CONTROL SYSTEMS BY RANDOM-SIGNAL MEASUREMENTS

William W. Seifert
Dynamic Analysis and Control Laboratory
Massachusetts Institute of Technology
Cambridge, Mass.

Abstract

Theoretical advances in the past decade indicated the advantages of random-signal calculations over sinusoidal and step-response methods in control-system design, but practical application of random-signal measurements to experimental evaluation of control systems was relatively limited. At M.I.T. the Dynamic Analysis and Control Laboratory has developed equipment and techniques for generating and monitoring random signals. One recent application of this technique, the experimental evaluation of a rate servomechanism, is presented, and the experimental results are correlated with analogue-computer and approximate theoretical results. The practical advantages of random-signal measurements are emphasized and special attention given to the study of nonlinear systems under realistic operating conditions.

Introduction

Sinusoidal or frequency-response measurements and step responses have for some time been the accepted methods for testing and describing the performance of servomechanisms. These methods are adequate for handling the majority of the servo problems but show serious shortcomings when truly high performance systems are considered. If an ideal linear system is assumed, its response to sinusoids or steps can be calculated relatively easily. Physical systems, however, are not linear, and as better and better performance is demanded, more and more attention must be given to the nonlinearities in the system. At very low output levels backlash and stiction become significant, while at high levels velocity and acceleration limiting present difficulties. Consequently, with step-response testing, the transient response actually observed becomes a function of the step level, while in sine-wave testing, the frequency response of the system depends upon the amplitude of the input signal used during the test. In fact, one step response or frequency response, in itself, does not describe the system adequately. Furthermore, in actual operation, a servo follows some randomly varying signal rather than either steps or sine waves.

Ten or more years ago Wiener and others developed analytical methods for the statistical analysis of linear systems, and many people have used these methods to a limited degree in the design of servomechanisms. In spite of this, very little attempt was made to employ random-signal methods for the experimental evaluation of servos. Over the past four years, however, in operating the M.I.T. Flight Simulator, the Dynamic Analysis and Control Laboratory has developed equipment for generating and monitoring several types of random signals. Because a number of high-performance servos are used in the Simulator as well as in the control systems that have been studied, the analysis of these servos has received special attention at the D.A.C.L. Gradually the techniques for random-signal testing[1] have been developed, and the advantages of this method of testing are being appreciated as more and more experience is gained.

The basic mathematical expression to be used in making random-signal calculations for a linear system relates the power spectra at the input and output of the system. If the system function is specified as $H(i\omega)$, the power spectra of the input $\Phi_{in}(\omega)$ and of the output $\Phi_{out}(\omega)$ are related by the expression

$$\Phi_{out}(\omega) = \left|H(i\omega)\right|^2 \Phi_{in}(\omega). \qquad (1)$$

Furthermore, if the mean-square value of the output, whether it be a mechanical motion or a voltage, is denoted by $\overline{E^2_{out}}$, then

$$\overline{E^2_{out}} = \int_{-\infty}^{+\infty} \Phi_{out}(\omega)\,d\omega. \qquad (2)$$

Consequently,

$$\overline{E^2_{out}} = \int_{-\infty}^{+\infty} \left|H(i\omega)\right|^2 \Phi_{in}(\omega)\,d\omega. \qquad (3)$$

A proper choice of the function $H(i\omega)$ allows assignment of various meanings to the quantity E_{out}. For example, E_{out} may be the error signal, the output voltage or shaft motion, or the velocity or acceleration at the output shaft. Once $H(i\omega)$ and $\Phi_{in}(\omega)$ are known, the calculation of E_{out} becomes a relatively simple matter. The inverse problem of determining $H(i\omega)$ from measurements of $\overline{E^2_{out}}$ made with several input power spectra is a more difficult problem since it involves solution of an integral equation. Significant progress has been made, however, in handling this problem, and the method shows distinct advantages, particularly as regards the ease with which the experimental data are collected.

The mathematical methods presently available are not adequate to analyze the mathematical problems which arise in the analysis of many control systems. When the methods fail, considerable information can be obtained by studying the system on an analogue computer. If the results from the computer setup agree with those derived from the actual system, it is probable that the mathematical statement of the problem on which the computer setup is based adequately represents the system. A combination of analytic techniques, computer studies, and tests of the actual system has been found to be very effective in evaluating complex control systems.

Random-Signal Generating and Monitoring Equipment

The availability of suitable equipment for generating and monitoring random signals is the first requirement if experimental random-signal measurements are to be made either on the actual control system or on its analogue representation. Several basically different approaches to these problems are available, but the present discussion will be limited to consideration of the methods which have been found to meet most nearly the needs which have arisen at the D.A.C.L.

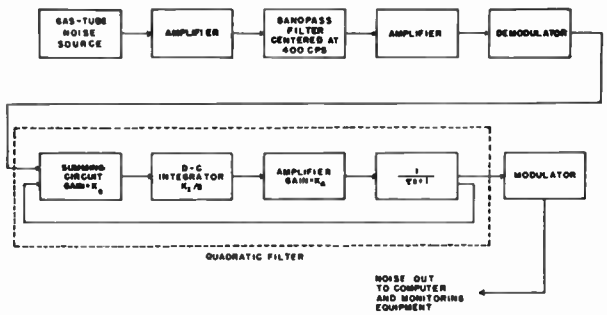


Fig. 1—Noise-generating system.

Figure 1 shows a block diagram of the random-signal or noise-generating system which has been developed. The basic noise-generating source is a gas-filled diode. The noise generated by such a tube extends over a wide range of frequencies and has properties that are relatively well defined and stationary except in the vicinity of zero frequency. Since, for the type of work considered here, primary interest is centered on the characteristics of the random-signal spectrum in the range from zero to 20 or 30 cps, the noise generator should produce signals that are well defined in this region. In order to overcome the basic uncertainties around zero frequency in the signal generated by the gas tube, this noise-generating system first selects by means of a bandpass filter a rather broad band of frequencies and then demodulates this signal, employing as the carrier a signal in the center of the filter pass band. The resultant signal, which is now translated to the zero-frequency end of the spectrum, is essentially flat over the half bandwidth of a bandpass filter and can be shaped with either an active or passive low-pass filter so as to produce the desired power spectrum. This signal exhibits a Gaussian amplitude probability distribution. Unfortunately the amplitude of the output of the gas-tube noise source exhibits an inherent instability with distinct jumps in level occurring at random times separated by intervals of a few minutes to an hour or more. Consequently some means must be provided for controlling the output of the gas tube. A very satisfactory solution to this problem has been obtained by recording the signal appearing at the output of the bandpass filter on magnetic tape, monitoring it carefully during the process, and making suitable gain corrections for these large jumps. When a noise tape an hour or two long has been recorded properly, it makes a very stable and convenient noise source.

The problem of monitoring the resultant noise signal deserves more attention than might appear necessary at first. If the signal is read by an ordinary voltmeter, even of the thermal type, the averaging time provided by the meter is not sufficient to damp out the fluctuations resulting from attempting to follow the variations in signal amplitude. In addition, although thermal meters may measure sinusoidal voltages accurately, when Gaussian voltages are measured, difficulties arise from the inherently limited square-law range of the thermocouple. If the meter is used near the lower part of the scale, the accuracy of reading is poor. On the other hand, if the meter is used near full scale, the peaks of the signal exceed the accurate square-law range of the thermocouple. As an alternate monitoring method, the noise signal could be recorded and the rms value obtained by reading a sufficiently large number of points from the record, but this method requires much effort. A very simple device developed for monitoring the noise is shown in Fig. 2 and consists of a squaring device

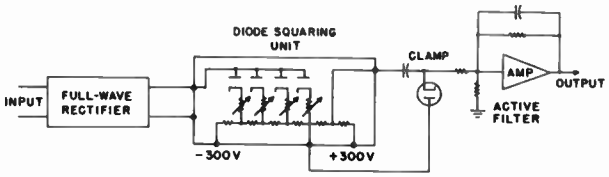


Fig. 2—Noise monitor.

followed by a filter with a time constant of the order of two minutes. The squaring device developed for this purpose is a series of diodes which are biased for successively high voltages. The signal is applied in parallel to the group of diodes and the outputs are summed. The design of the resistance network employed determines the function generated (in this case, a square law). To obtain a filter with a long time constant, resistance-capacitance feedback is employed around a high-gain amplifier.

Evaluation of a Rate Servo

One of the principal computing components in the generalized a-c section of the Simulator at the D.A.C.L. is the electromechanical rate servo which is employed as an integrator. The integrator employs a 2-phase 400-cps motor to drive a high-precision tachometer. A velocity servo is formed by subtracting the tachometer output from the input voltage and using this difference, after electronic compensation, as the input to the amplifier which drives the motor. Various output units are connected through appropriate gearing to the motor shaft. Figure 3 shows a block diagram of the servo. Because the demands placed upon the unit

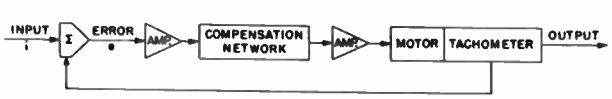


Fig. 3—Block diagram of integrator servo.

are very high, great effort has gone into developing a high-performance system. In the course of this work the experimental difficulties associated with obtaining frequency-response characteristics for such systems and the problems in determining whether or not the system attained is adequate have appeared repeatedly.

The conclusion was reached that a better understanding of the operation of the integrator could be obtained if the input signals used in testing it more nearly duplicated the inputs which the integrator is called upon to handle in actual operation. A random signal with an appropriately chosen power spectrum and amplitude probability distribution appeared to meet the requirements both of realism and experimental practicality. Furthermore, it was felt that experimental studies of the actual system combined with analytic investigations and analogue-computer studies of the system would prove most effective both in arriving at an adequate understanding of this particular servo and in seeing how the results of this study would apply in other cases.

The first step in this program consequently consisted in obtaining some experimental data showing how the system operated when random-signal inputs were applied to it. The criterion to be used in judging the performance of the integrator servo is the rms value of the error signal, which is the difference between the input and the derivative of the output. Figure 4 shows a family of curves of the rms error signal plotted as a function of the rms value of the input signal for

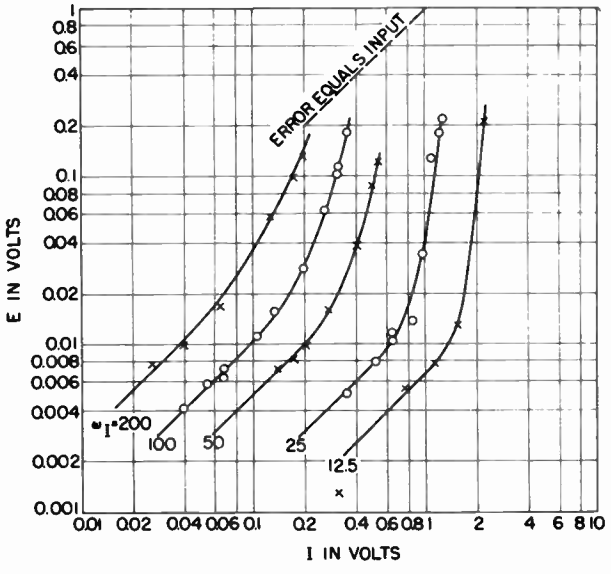


Fig. 4—Curves of rms error as a function of rms input to the servo.

input-signal power spectra of various bandwidths. The random signals employed were formed by passing white noise through a quadratic filter of the desired bandwidth, but always maintaining a damping ratio of 0.7. If the system were linear, the error signal would increase linearly with an increase in the rms value of the input signal. The actual system deviates from linearity at the low-level end because of backlash and stiction, and at the high end because of acceleration or velocity limiting. The present discussion deals primarily with the effects of limiting. In Fig. 4 the error depends on the bandwidth of the input signal as well as on its rms value because the torque developed at the output shaft depends on the error signal, and a greater torque is required to produce a given rms output as the bandwidth of the input signal is increased. Figure 4 also shows that as the input is increased, the error voltage exhibits a rather sharp, large increase, and with a further increase in output the error again shows a steady increase. For extremely large inputs the error equals the input for inputs of all bandwidths. The operation is, in effect, bimodal with an abrupt transition region between the two modes.

In order to see whether this performance could be predicted mathematically, an analytic study of the system was made. Figure 5 shows a mathematical representation of the servo. In the

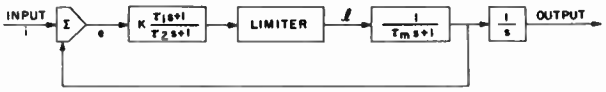


Fig. 5—Mathematical representation of the servo.

linear region, theoretical error curves can be calculated from (3). Calculation of the performance for inputs that require peak accelerations greater than are available at the motor requires inclusion of the effect of the limiter in the system. Booton[2] has developed a quasi-linearization procedure for handling such a case analytically. There the limiter is replaced by a device which inserts a pure attenuation. The attenuation is determined by the rms value of the input to the limiter. If the input is stationary, the determination of the error signal and of the attenuation which mutually satisfy the system becomes relatively easy. Figure 6 shows a set of theoretical curves obtained

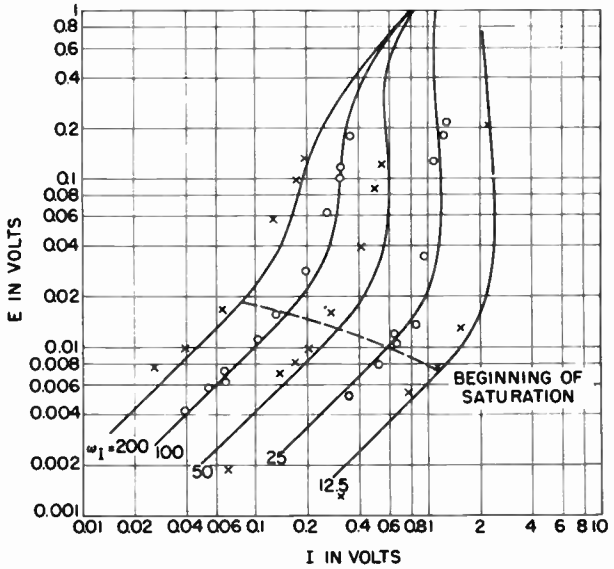


Fig. 6—Curves of theoretical performance and experimental points from servo.

in this manner, with a group of points indicating the results obtained experimentally by applying random inputs to the actual servo. The dotted curve labeled "beginning of saturation" was computed on the basis of a linear system and shows the input level for which the input to the limiter begins to exceed the limiting level a significant part of the time. For the Gaussian signals employed, the peaks rarely exceed three times the rms value of the signal; consequently, the beginning of saturation was calculated on the basis of an rms signal at the limiter equal to one-third the limiting level. The corresponding frequency of the input signal is indicated by the intersections with the other family of curves.

In order to establish further the validity of the approximation method used in obtaining the theoretical curves, the system shown in Fig. 5 was set up on an analogue computer and studied in some detail. Figure 7 shows a set of curves obtained from an experimental study of the analogue system

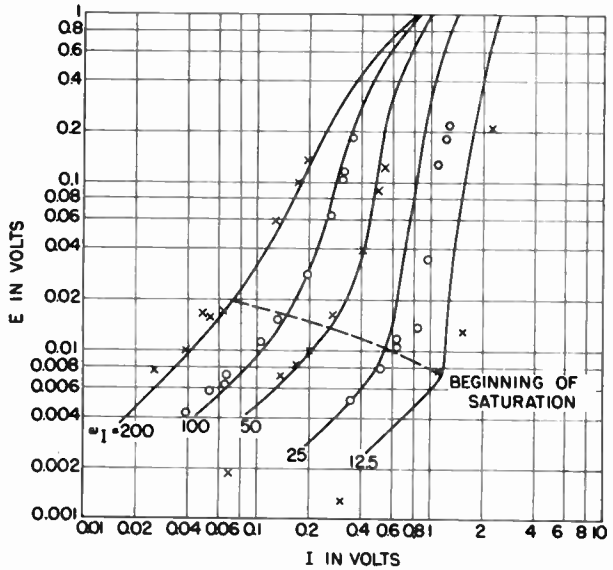


Fig. 7—Curves from analogue system and experimental points from servo.

together with the points obtained from the actual servo. While the correspondence of data for these various cases is not perfect, it is very good for the nonlinear problem involved and certainly adequate for servo design.

The fact that the d-c analogue curves do not match the servo merely means that the description of the servo as set up on the computer was not exact. This description was based on sinusoidal measurements which were rather difficult to make. It is possible to work backwards from the random-signal measurements to a transfer function, although there are problems connected with solving the integral equation involved. These curves merely demonstrate that the results obtained in different ways agree relatively well. Actually, the principal interest would probably center on adjusting the analogue of the system so as to secure the closest agreement between the results from the servo itself and those from its analogue.

A more complete understanding of the performance of this system can be gained by examining the signals within the system. Figure 8 shows the input signal, the error signal, and the output of the limiter for a case in which the signal at the

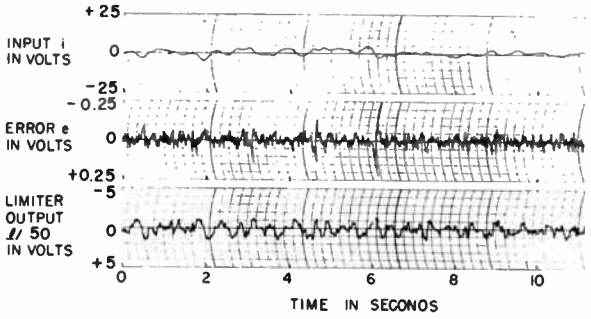


Fig. 8—Servo performance for a small random input.

limiter is usually below the limiting level. Here the error signal is very low except for one or two large spikes where the limiting occurred.

Figure 9 illustrates conditions in the system for a somewhat larger input voltage. In this case the error signal exhibits a definitely bimodal characteristic. For a time the error voltage remains very small. Then limiting occurs in the

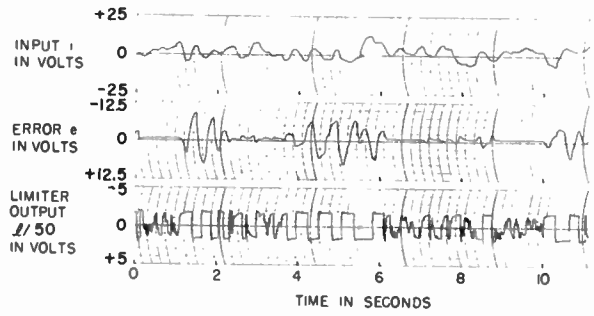


Fig. 9—Servo performance for inputs causing moderate limiting.

system and persists for several seconds with the limiter output jumping from a maximum of one polarity to the maximum of the other polarity through several cycles. During this time the error signal is very high. Finally, the system stabilizes again on a low error voltage for a few seconds. Then limiting reoccurs. An examination of the input signal reveals very little change between the times when the error is small and when the error is large. Figure 10 shows the condition for a still larger input voltage. Here the system is in one limit or the other nearly all the time, and the error voltage is nearly the same as the input.

Some additional tests with a sine-wave input rather than a random input were made on the analogue representation of the system. In these tests the error voltage stays very small so long as the peaks of the voltage at the limiter are below

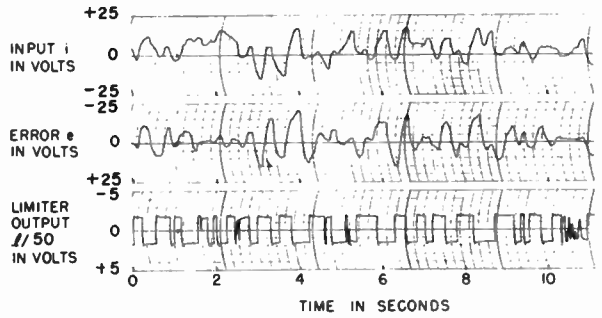


Fig. 10—Servo performance for inputs causing severe limiting.

the limiting level. Figure 11 indicates conditions in the system when the input is such as to cause a small amount of limiting. With a sinusoidal input, limiting occurs on each half cycle, and curves of error as a function of input show a very sudden transition from the linear mode of operation to the

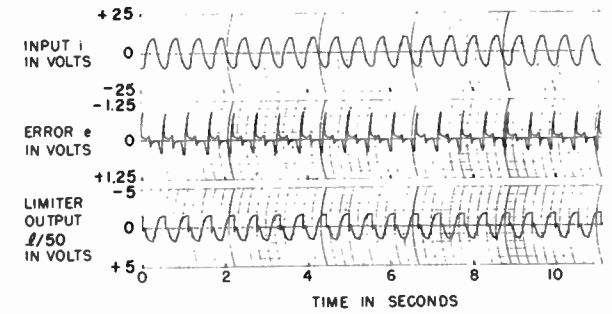


Fig. 11—Servo performance for sinusoidal input causing a small degree of limiting.

saturated mode. The system also exhibits some hyteresis effect. When the input signal is reduced, the system returns from the saturated mode to the linear region at a different input from that which drove the system from the linear mode to the saturated mode.

Conclusions

Experience at the Dynamic Analysis and Control Laboratory has shown conclusively that data of much better consistency can be obtained on high-performance servo systems if random-signal inputs are employed than if sinusoidal inputs are employed. Relatively simple equipment has been built for generating these random signals and measuring their mean-square values. The older sinusoidal and step-input techniques were more or less adequate for evaluating the relatively low performance servos in use ten or fifteen years ago. In high-performance servos, nonlinearities such as backlash, stiction, and velocity and acceleration limiting cannot be overlooked. If an experimental technique for evaluating the system is to give realistic results, the signal must operate under conditions that closely resemble those which it encounters in practice.

In this paper principal emphasis has been on the evaluation of an existing servo system. The use of random-signal techniques coupled with an analogue representation of the system, however, can provide a powerful approach to the servo design problem, since the computer provides a convenient way for changing the parameters of a system.

Acknowledgment

The author gratefully acknowledges the cooperation afforded him by other members of the Dynamic Analysis and Control Laboratory during the course of this investigation. Particular mention is made of the work of Dr. R. C. Booton, Jr., in developing approximate analytic techniques for analyzing nonlinear control systems and to Mr. M. V. Mathews for his efforts in developing experimental techniques and gathering the experimental data for this particular system.

[1] R. C. Booton, Jr., W. W. Seifert, and M. V. Mathews, "Nonlinear Servomechanisms with Random Inputs," Dynamic Analysis and Control Laboratory Report No. 70, Massachusetts Institute of Technology, Cambridge, Mass., forthcoming.

[2] R. C. Booton, Jr., "Nonlinear Control Systems with Statistical Inputs," Dynamic Analysis and Control Laboratory Report No. 61, Massachusetts Institute of Technology, Cambridge, Mass., March, 1952.

EXTENSION OF CONVENTIONAL TECHNIQUES TO THE DESIGN OF SAMPLED-DATA SYSTEMS

W. K. Linvill & R. W. Sittler
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

Introduction

The analysis or design of a sampled-data system at a first glance often appears to be a discouraging problem. The engineer perceives that in addition to considering familiar system components he must take into account the action of new component types. The presence in the sampled-data system of sampling operations and of linear filters whose input and output signals have a sampled nature seems to suggest to him that he is faced with a totally different problem than those to which he is accustomed. This first impression of difficulty is still further intensified when the engineer turns to the literature on the subject of sampled-data systems. Quite naturally such literature is often preoccupied with emphasizing the differences between sampled-data systems and conventional systems. The engineer is confused by the use of apparently new or unconventional mathematical methods. Under all these adverse impressions it is no wonder that a belief arises that sampled-data systems have an exotic nature which the conventional methods of systems analysis cannot penetrate.

This first impression is totally wrong. It is a fact that no new methods of analysis or design are required. What are really required are the old conventional methods extended and supplemented to fit the new situations. The similarities between sampled-data and conventional systems are, in fact, much more striking and important than their differences.

In this paper we cannot hope to prove in detail this thesis of overwhelming similarity. It is possible here only to set forth a philosophy for treating sampled-data systems through conventional techniques and to illustrate this philosophy by using these techniques in a logical succession of example problems. The reader should consult references 1 and 2 for a more complete and systematic account.

The main advantage from realizing the similarity between continuous-data systems and sampled-data systems comes in design problems. System design is the art of compromising between what is desired in a certain system and what is physically realizable. Whereas a specific analysis problem has a unique solution, very often no specific design problem exists for a given situation and even if it did there would exist many solutions rather than only one. Accordingly, the technique of system design is usually a matter of cut-and-try wherein certain designs are visualized, analyzed, and modified. To be workable, this cut-and-try process must have short steps based on simple approximate analysis procedures. In this paper analysis will be discussed before design. First the components in a sampled-data system will be described. Second, techniques for flow graph (block diagram) manipulations will be illustrated to reduce a complicated block diagram to a simple cascade equivalent. Third, responses of a cascade of sampled-data system components will be characterized by pole-zero locations and partial-fraction expansions in the s-plane, the ε^{-sT}-plane and the $\varepsilon^{-sT/2}$ plane. Design techniques will be discussed after analysis.

Analysis

System Components

The major underlying reason why conventional methods may be used in sampled-data system problems is the linearity of the sampled-data system. This linearity follows from the linearity of the system components. The linearity of the system components follows from the way in which they are defined.

We recognize three basic types of components. The first is the ordinary continuous-signal linear filter, the basic building block of conventional systems. Its operation is described by the usual transfer function, K(s). The second is the discrete-signal linear filter whose only claim to unconventionality is that it is particular what kind of signals it accepts and delivers. The discrete-signal filter operates linearly on an input consisting of a sequence of equally spaced impulses to produce an output of equally spaced impulses in time synchronism with the input pulses. Its operation is described by a transfer function, $D(\varepsilon^{-sT})$, which is most conveniently written in terms of a variable ε^{-sT} (T is the interval between impulses) instead of frequency, s. The third basic building block and the only one with a strong claim to unconventionality is the sampling device, a device which samples its input every T seconds and delivers only the samples to the following components. Although such a device is often visualized as a switch, it is better to view it as a modulator, whose constant carrier waves are unit impulses spaced by T seconds. The samples delivered at the output are impulses of areas equal to the input wave at the sampling times. Hence the sampling device, the impulse modulator, is linear since doubling the size of the input leads to double the output and so forth. Therefore all sampled-data system components are linear as is the overall system.

Although the modulation (sampling) is linear, it is the type of linearity called time-varying. Signals applied at one time do not invoke the same response as those applied at another time. A conventional transfer function cannot therefore be applied in general to describe the sampling operation. However the modulation process, being a simple multiplication, is time varying without any complicated dynamical features. By tying down all modulators to synchronism and phasing the carrier so that one impulse of it occurs at t=0 all impulse-modulated signals have simple transforms which are rational functions of ε^{-sT} and all discrete-signal filters have transfer functions which are rational in ε^{-sT}. The time variability of the sampled-data system is associated only with modulation, where it is handled very simply and hence the time variability is only an incidental detail rather than an important factor.

Flow-Graph Manipulation

The first step in an analysis problem is to reduce the given system to a form which permits simple computation. In the first example, therefore, we show the reduction of a sampled-data feedback system of some complexity to a simple equivalent system without feedback. An extension of conventional flow-graph reduction procedures is used (see reference 3).

Consider the system block diagram of figure 1 a. Figure 1 b shows the same system diagram translated into flow-graph form. (M = modulation (sampling) operation, D = discrete-signal filter operation, K = continuous-signal filter operation). Now flow-graph reduction procedures are based on the linearity of the system together with the possibility of finding the equivalent transfer function for a system with a single feedback loop when the feedback and open-loop transmissions are described by transfer functions (the familiar $\frac{K}{1-\beta K}$ formula). In sampled-data systems it is not so obvious that such equivalences exist since the loops in general contain modulators whose operations are not describable by conventional transfer functions. However if we make the restrictions previously mentioned on modulator synchronism, simple equivalences do exist. The result is that any sampled-data feedback system can be reduced to an equivalent system without feedback.

Following through our example in 1 c, we pick residual modes to cut feedback in such a way that we cut as much as possible by choosing nodes at which the signal is purely sampled. Here it is necessary to pick only one such node. We then finish the job by picking an extra residual node where necessary. No further choices are necessary in our example. In 1 d only residual nodes are retained with the operations between them written in symbolic form. This new graph is reduced by removing the self-loops around nodes in reverse order to which the nodes were chosen above. The only residual node which we must consider in the example (except the input and output nodes) is a node at which the signal is purely sampled. Because of the syncronization previously mentioned the self-loop operation of this node has an equivalent transfer-function description and may be removed by the $\frac{K}{1-\beta K}$ formula. Figure 1 e, 1 f, show the results.

A central feature in these reductions is the basic transformation which allows us to replace the operation on sampled inputs of a continuous signal filter, K (s), followed by a modulator, M, with an equivalent transfer function $K^*(s)$. $K^*(s)$, since it relates sample impulses to sample impulses, is really the transfer function of a discrete-signal filter and is most easily written as $K^*(\epsilon^{-sT})$. We can find K^* by applying an impulse to K, finding the time output in the filter in terms of exponential components, finding their sample sequences as delivered by the modulator, finding the Laplace transforms of these component sequences in series form using delay factors ϵ^{-sT}, and rolling up the geometrical series obtained in closed form.

As a result of the reduction procedures we can always, as in figure 1 f, write the equivalent system in the form of a parallel set of branches, each branch containing a standard cascade sequence of a continuous-signal filter followed by a modulator, a discrete signal filter and another continuous signal filter (degenerate cases occur). The analysis of a specific system response hinges therefore on the analysis of the response of the basic $K_a M D K_b$ cascade sequence.

Response of K_a M D K_b Cascade

The second example is the problem of computing the response of the $K_a M D K_b$ sequence of figure 2 a ($K_a = 1$). Laplace transforms and pole-zero patterns are used. A trick is used to estimate the output ripple due to sampling.

A unit ramp $f_i(t)$, is applied at an instant of sampling. We can think of this ramp as being generated by a unit impulse applied at a sampling time to a prefilter with transfer function $F_i(s)$. The overall transfer function of an equivalent continuous filter preceeding the modulator is $F_i(s)\,K_a(s)$. Now we can connect the unit impulse with the modulator output through the basic transformation giving an equivalent discrete signal filter $(F_i K_a)^*$. This is the transform of signals leaving the modulator. The transform of the output signal $f_o(t)$ is then $F_o = (F_i K_a)^* D K_b$. In the example

$$(F_i K_a)^* = \left(\frac{1}{s^2}\right)^* = 0 + T\epsilon^{-sT} + 2T\epsilon^{-2sT} + 3T\epsilon^{-3sT} + \cdots$$

$$= \frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})^2}.$$

$$F_o = (F_i K_a)^* D K_b = \frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})^2}\,(1-\epsilon^{-sT})\,\frac{a}{s(s+a)}$$

$$= \frac{T\epsilon^{-sT}}{1-\epsilon^{-sT}} \cdot \frac{a}{s(s+a)}.$$

Figure 2 b shows a pole-zero diagram of F_o in the s plane. We note two basic regions of the diagram, a central region whose poles represent low frequency response components, and the remainder of the plane whose poles represent ripple components in the output. The s-plane diagram of $F_o(s)$ tells us the qualitative nature of the response. In this example the diagram shows that the output signal has a ramp component (double pole at s=0), a possible step component (obscured by the double pole), a decaying exponential (pole at s = -a) and a ripple of constant magnitude (simple poles at $s = \pm j n\Omega$, $n = 1, 2, 3\cdots$ where Ω is the sampling frequency).

The reason we do not go any further and use the s-plane diagram to quantitatively determine response is that there are an infinite number of poles on the diagram in general. We would like to calculate only a finite number of components. We turn now to an approximate calculation of the response, still using nothing but Laplace transforms.

Suppose we only desire the output at the times of sampling. We find this output by appending to the original system a second synchronized modulator. We evoke our basic transformation, relating by means of an equivalent discrete signal filter, K_b^*, the

sampled signals delivered by the discrete signal filter D and the output samples. The transform of the output samples is therefore $F_o^* = (F_i K_a)^* D K_b^*$. In the example system of figure 2 a, we have:

$$F_o^* = \left(\frac{1}{s}\right)^* (1-\epsilon^{-sT}) \left[\frac{a}{s(s+a)}\right]^*$$

$$= \frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})^2} (1-\epsilon^{-sT}) \left[\frac{1}{s} - \frac{1}{s+a}\right]^*$$

$$= \frac{T\epsilon^{-sT}}{1-\epsilon^{-sT}} \left[\frac{1}{1-\epsilon^{-sT}} - \frac{1}{1-\epsilon^{-aT}\epsilon^{-sT}}\right]$$

$$= \frac{T(1-\epsilon^{-aT})(\epsilon^{-sT})^2}{(1-\epsilon^{-sT})^2(1-\epsilon^{-aT}\epsilon^{-sT})} .$$

We see that Fo^* differs from F_o in that Fo^* is not a mixed expression in ϵ^{-sT} and s is but a rational function of ϵ^{-sT}. F_o^*, therefore, has only a finite number of poles in the ϵ^{-sT} plane, each one corresponding to a component of response. These components relate only to low frequency response, the primary response for which the system is designed, since the intersample behavior of the output which includes ripple behavior is not detected. The response can be resolved into components by resolving Fo^* into a partial fraction expansion and identifying component ϵ^{-sT} functions with component time functions. The ϵ^{-sT}-plane pattern of F_o^* provides a quantitative indication of the primary response.

The ϵ^{-sT}-plane pattern of F_o for the example is shown in figure 2 c. The pole at $\epsilon^{-sT} = 1$ indicates sampled ramp and step components, the pole at $\epsilon^{-sT} = \epsilon^{aT}$ indicates a sampled exponential decay. The quantitative calculations are

$$F_o^* = \frac{T(1-\epsilon^{-aT})(\epsilon^{-sT})^2}{(1-\epsilon^{-sT})^2(1-\epsilon^{-aT}\epsilon^{-sT})}$$

$$= \frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})^2} - \frac{T}{1-\epsilon^{-aT}} \frac{1}{1-\epsilon^{-sT}} + \frac{T}{1-\epsilon^{-aT}} \cdot \frac{1}{1-\epsilon^{-aT}\epsilon^{-sT}} .$$

The $\frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})^2} = 0 + T\epsilon^{-sT} + 2T\epsilon^{-2sT} + \cdots$ represents the transform of a sampled unit ramp. Similarly $\frac{1}{1-\epsilon^{-sT}} = 1 + \epsilon^{-sT} + \epsilon^{-2sT} + \cdots$ is the transform of a sampled step and $\frac{1}{1-\epsilon^{-aT}\epsilon^{-sT}} = 1 + \epsilon^{-aT}\epsilon^{-sT} + \epsilon^{-2aT}\epsilon^{-2sT} + \cdots$ is the transform of a sampled exponential decay. By inspection we can find the envelope curve of the sampled response. In the example $f_e(t) = t - \frac{T}{1-\epsilon^{-aT}}(1-\epsilon^{-at})$ for $t \geq 0$. This envelope is an approximation of the actual response neglecting ripple.

To find the ripple by transform methods we resort to the trick of sampling the output every T/2 seconds instead of every T seconds. Although the filter D delivers samples every T seconds we may nevertheless regard them as occurring every T/2 seconds with alternate samples having zero amplitude. The basic transformation again applies and we can find an equivalent discrete signal filter, K_b^{**}, connecting signals delivered by D to the doubly sampled output. Now, however, the variable appearing in K_b^{**} is $\epsilon^{-T/2}$ instead of ϵ^{-sT} as found in K_b^* and s as found in K_b. We write for the transform of the doubly sampled output $F_o^{**} = (F_i K_a)^* D K_b^{**}$ In our example:

$$F_o^{**} = \frac{T\epsilon^{-sT}}{1-\epsilon^{-sT}} \cdot \frac{\epsilon^{-s\frac{T}{2}}(1-\epsilon^{-aT/2})}{(1-\epsilon^{-sT/2})(1-\epsilon^{-aT/2}\epsilon^{-sT/2})}$$

$$= \frac{T(1-\epsilon^{-aT/2})(\epsilon^{-sT/2})^3}{(1-\epsilon^{-sT/2})^2(1+\epsilon^{-sT/2})(1-\epsilon^{-aT/2}\epsilon^{-sT/2})} .$$

The whole expression has been written so that it is rational in $\epsilon^{-sT/2}$. The pole-zero pattern of F_o^{**} may be plotted in the $\epsilon^{-sT/2}$ plane as shown in figure 2 d. Comparing with 2c we see that the only new component indicated is the pole at $\epsilon^{-sT/2} = -1$ representing the term $\frac{1}{1+\epsilon^{-sT/2}} = 1 - \epsilon^{-sT/2} + \epsilon^{-2sT/2} - \epsilon^{-3sT/2} \cdots$, an oscillating component of the output. This is the steady state ripple. Our double sampling has caught the ripple! The $\epsilon^{-sT/2}$ plane pattern of F_o^{**} provides a quantitative estimation of ripple as well as a more accurate estimation of primary response.

If we follow the procedure used previously of finding the magnitudes of these components we can get an envelope curve for the double frequency response samples and with it an improved approximation to the actual response. The envelope turns out to be $f_e = t - \frac{T}{1-\epsilon^{-aT}}(1-\epsilon^{-aT}) + \frac{T}{4}\frac{1-\epsilon^{-aT/2}}{1+\epsilon^{-aT/2}}\left(1-\cos\frac{2\pi}{T}t\right)$. (Figure 2 e shows the input, output, and the two sample envelope approximations). If necessary the approximations can be still further improved by still more rapid sampling of the output wave so that in essence the complete response picture is given through the use of Laplace transform methods.

System Design

Response Requirements Versus Realizability Conditions

The picture of system design has two aspects. One aspect describes what sort of response charcteristics are wanted. The other aspect describes what sort of system is realizable. Each aspect can be seen in proper perspective only with the other aspect as a background. In servo systems design a suitable output member is usually chosen at the outset on the basis of power considerations and this member is fitted into a system by compensating filters to provide proper response characteristics.

First, desirable response characteristics should be described. Usually the responses of a sampled-data system will be acceptable if transients are short and well-damped, if disturbances in the feed-forward section are adequately suppressed, if errors in following smooth inputs are small, and if sampling ripple is small at the output. Most of these desired response characteristics can be described for a specific system from a simplified block diagram study and location of the poles and zeros of the overall system transfer function in the ϵ^{-sT} and $\epsilon^{-sT/2}$ planes. The quality with which a single-loop servo suppresses disturbances and follows smooth inputs can be predicted readily from error coefficients (reference 4).

Second, what modifications in response characteristics are realizable by use of compensating filters? The realizability conditions arise

from the fact that a given compensating filter, either discrete-signal or continuous-signal, can't respond before it is excited. How the physical realizability conditions for a given filter affect physical realizability conditions on the overall system transfer characteristics depends upon the system configuration chosen but usually a given overall system transfer function for a one-loop ysstem with conventional output member will be realizable if it has a zero at the origin of the ϵ^{-sT}-plane. (This zero allows for delay in the power member.) As with continuous servo design it is not feasible to give more than realizability conditions for individual filters used to compensate sampled-data servos, but it should be observed that realizability of compensation is usually less a problem in sampled-data servos than it is in continuous servos. Physically, this fact arises because sampling in a system reduces the frequency horizon to half the sampling frequency and one can usually handle that limited frequency range very well with realizable compensating filters.

Example: Compensation of a Single-Loop Servo

The system to be compensated is shown in figure 3 a. The equivalent system without feedback for the uncompensated system turns out to be just the single cascade sequence of figure 2 a whose ramp response has already been determined. Figure 2 e shows the response. As a design problem we desire to remove the steady-state lag in the ramp response by adding compensation in the feed-forward section. The velocity error coefficient is to be made zero.

The uncompensated system should first be analyzed. It is shown in references 1 and 2 and easily verified by using the flow-graph (or block diagram) reduction technique previously illustrated that a system consisting of a single loop of negative unity feedback around a modulator-filter feed forward section reduces to an equivalent system of the following type. If DK is the transfer function of the filter consisting of a discrete signal part, D, and a continuous signal part, K, then the equivalent system is a cascade of components in the order: modulator, filter with transfer function $\frac{1}{1+DK^*}$, filter with transfer function DK. In the example:

$$D = (1-\epsilon^{-aT}\epsilon^{-sT}) \; ; \; K = \frac{a}{s(s+a)} \; ;$$

$$DK^* = \frac{(1-\epsilon^{-aT}\epsilon^{-sT})(1-\epsilon^{-aT})\epsilon^{-sT}}{(1-\epsilon^{-sT})(1-\epsilon^{-aT}\epsilon^{-sT})} = \frac{(1-\epsilon^{-aT})\epsilon^{-sT}}{1-\epsilon^{-sT}} .$$

$$\frac{1}{1+DK^*} = \frac{1-\epsilon^{-sT}}{1-\epsilon^{-aT}\epsilon^{-sT}} \; ; \; \frac{DK}{1+DK^*} = \frac{(1-\epsilon^{-sT})a}{s(s+a)}$$

Since the equivalent filter $\frac{1}{1+DK^*}$ delivers signals to DK, the feed-forward filter, the output of $\frac{1}{1+DK^*}$ is the sampled error of the system. The respective error coefficients are the respective magnitudes of the error samples at t = 0 to a set of unit input impulses, a ramp sequence of impulses, etc. applied at t = $-\infty$. The procedure for finding error coefficients is completely analogous to that used in conventional systems except that we work with discrete signals. In general, for a sampled-data feedback system with unity feedback, the kth error coefficient can be shown to be

$$\epsilon_k = \frac{1}{k!}\frac{d^k}{ds^k}\left[\frac{1}{1+DK^*}\right]\Bigg|_{\substack{s=0 \text{ or} \\ \epsilon^{-sT}=1.}}$$

A system with zero positional error coefficient and zero velocity error coefficient should have zero steady-state error to a ramp set of pulses. In the example if we apply a ramp set of pulses, whose transform is $\frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})^2}$, the sampled error transform is $\frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})^2}\cdot\frac{1}{1+DK^*} = \frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})(1-\epsilon^{-aT}\epsilon^{-sT})}$. The $(1-\epsilon^{-sT})$ factor shows the presence of a steady state component. We see that in order to prevent this factor from arising we need at least a factor $(1-\epsilon^{-sT})^2$ in the numerator of $\frac{1}{1+DK^*}$. In conventional systems with unity feedback we talk about the need for extra integration. Here $(1-\epsilon^{-sT})$ plays a role analogous to the s factor in conventional systems. Note that $\lim_{T\to 0}\frac{1}{T}(1-\epsilon^{-sT}) = s$.

Having analyzed the uncompensated system, we now turn to the problem of deciding what sort of response characteristics would be desirable and realizable. Since the sampled input to sampled error transfer function is simple and describes many important response characteristics of the system we select an input to error transfer function which would be acceptable and work backwards to see if the compensating filter used to get this form would be physically realizable. In general compensating filters could be placed either in the feed forward or feedback section or could be employed in minor loops. In this example we will assume that the filter is cascaded in the feed-forward section.

The compensated sampled feed-forward transfer function is a rational function of ϵ^{-sT} and can be written as the quotient of two polynomicals in ϵ^{-sT} as $\frac{P}{Q}$. The sampled input to sampled error transfer function is given by $\frac{1}{1+P/Q} = \frac{Q}{P+Q}$. The zeros of this transfer function correspond to poles of $\frac{P}{Q}$, the poles of this transfer function correspond to natural frequencies of the system, and $\frac{Q}{P+Q}$ can be used to evaluate error coefficients. The question now is can it be realized? Realizability can be best discussed in terms of P/Q. The straight-forward way to test realizability of P/Q is to work backwards to find the compensating filter. For continuous compensation, K_c would be evaluated as follows: First: divide out the discrete filter transfer function from P/Q. Next find an unsampled transfer function of the quotient. K_c is the transfer function which must multiply the transfer function of the continuous fixed part. $\frac{P}{Q}\cdot\frac{1}{D} = \left[K_c\cdot\frac{a}{s(s+a)}\right]^*$.

Given $\left[K_c\frac{a}{s(s+a)}\right]^*$ there are many possible values of $K_c\frac{a}{s(s+a)}$. A realizable K_c can always be chosen if $\left[K_c\frac{a}{s(s+a)}\right]^*$ has a zero at the origin of the ϵ^{-sT}-plane. Usually the low-pass K_c is used if it is realizable. Note that the uncompensated system K^*D has a zero at the origin of the ϵ^{-sT}-plane. This zero corresponds to the fact that the impulse response of the output member starts off at zero because its transfer function has two more poles than zeros. To remove this zero at the origin would require a feed-forward section with a non-zero value of impulse response at t=0

or the motor would have to receive its signal before t=0. This is clearly impossible because the compensating filter would have to respond before it was excited in order to "warm the motor". (Refer to reference 2 for detailed realizability conditions.)

In our design we choose $\frac{1}{1+DK^*_{new}} = \frac{Q}{P+Q} = \left(\frac{1-\epsilon^{-sT}}{1}\right)^2$ This choice yields: a) zero velocity error, b) no transient poles. From it we intend to find a new K and determine what compensation must be added to the old K to get the new one. Since $\frac{Q}{P+Q}$ has been given we find immediately $P = 1-(1-\epsilon^{-sT})^2 = \epsilon^{-sT}(2-\epsilon^{-sT})$ and $\frac{P}{DQ} = \frac{\epsilon^{-sT}(2-\epsilon^{-sT})}{(1-\epsilon^{-aT}\epsilon^{-sT})(1-\epsilon^{-sT})^2} = K^*_{new}$.

$$K^*_{new} = \frac{AT\epsilon^{-sT}}{(1-\epsilon^{-sT})^2} + B\left(\frac{1}{1-\epsilon^{-sT}} - \frac{1}{1-\epsilon^{-aT}\epsilon^{-sT}}\right).$$

$$K_{new} = \frac{A}{s^2} + B\left(\frac{1}{s} - \frac{1}{s+a}\right)$$

$$= \frac{(A+aB)s + aA}{s^2(s+a)}$$

where $A = \frac{1}{T(1-\epsilon^{-aT})}$; $B = \frac{1-2\epsilon^{-aT}}{(1-\epsilon^{-aT})^2}$.

Therefore K_c is: $K_c = \frac{K_{new}}{K_{old}} = \frac{s(s+a)}{a} \cdot K_{new} = \frac{\left(\frac{A}{a}+B\right)s + A}{s}$.

K_c is part integration and part straight amplification and is obviously realizable.

To see what we have done we note that this compensation yields the system M followed by $\frac{DK_{new}}{1+DK^*_{new}} = (1-\epsilon^{-sT})^2(1-\epsilon^{-aT}\epsilon^{-sT})\frac{(A+aB)s+aA}{s^2(s+a)}$.
If a ramp is applied we get the output transform $F_o = (F_i)^*\frac{DK_{new}}{1+DK^*_{new}} = T\epsilon^{-sT}(1-\epsilon^{-aT}\epsilon^{-sT})\frac{(A+aB)s+aA}{s^2(s+a)}$.
A diagram of F_o in the s-plane (figure 3 b) shows that there are poles only at s=0 and s= -a. Hence there is no ripple. The compensation has removed the ripple. We have designed the system to eliminate velocity error so we know it is absent. To get a further quantitative idea of the output we determine the output samples. We get
$F_o^* = (F_i)^*\frac{DK_{new}}{1+DK^*_{new}} = \frac{T(\epsilon^{-sT})^2(2-\epsilon^{-sT})}{(1-\epsilon^{-sT})^2} = \frac{T\epsilon^{-sT}}{(1-\epsilon^{-sT})^2} - T\epsilon^{-sT}$
This result shows that the first two samples are zero and then the output follows the input exactly. A sketch of the time responses of the compensated and uncompensated systems is given in figure 3 d. Note the great improvement.

It may be noted that if for some reason we had desired to use some transient poles instead of just 1 in the denominator of $\frac{1}{1+DK^*_{new}}$ then ϵ^{-sT}-plane diagram of $\frac{1}{1+DK^*_{new}}$ shows the damping characteristic of these poles. (See figure 3 c) In the diagram we may draw ζ contours showing relative damping as reference lines (mapped over from the s plane). Root-locus methods can also be applied to the pole-zero pattern of $\frac{1}{1+DK^*_{new}}$ in the ϵ^{-sT} plane to find how changes in gain affect the damping.

It is also possible to compensate the above system by an alternative method using discrete-signal filter compensation. It is true, however, that such compensation does not improve the ripple characteristics of the system.

Summary

The main point of this paper has been the assertion that sampled-data system problems can be handled by anyone familiar with conventional system analysis and design methods. Once the modification introduced by the presence of sampling operations are determined, both analysis and design problems can be handled in a strightforward manner. It is true, however, that such compensation does not improve the ripple characteristics of the system.

References

1. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulations, W. K. Linvill, AIEE Transactions Vol. 70, 1951.

2. Design of Sampled-Data Systems by Extension of Conventional Techniques, W. K. Linvill and R. W. Sittler, Project Whirlwind Report R-222, Mass. Inst. of Tech., 1953.

3. Some Properties of Signal Flow Graphs, S. J. Mason, (To be published in IRE Proceedings.).

4. Theory of Servomechanisms (book), James, Nichols, Phillips; McGraw-Hill Book Company, New York, New York, Vol. V., 1947.

5. The Analysis of Sampled-Data Systems, J. R. Ragazzini and L. A. Zadeh, AIEE Transactions, Vol. 71, 1952.

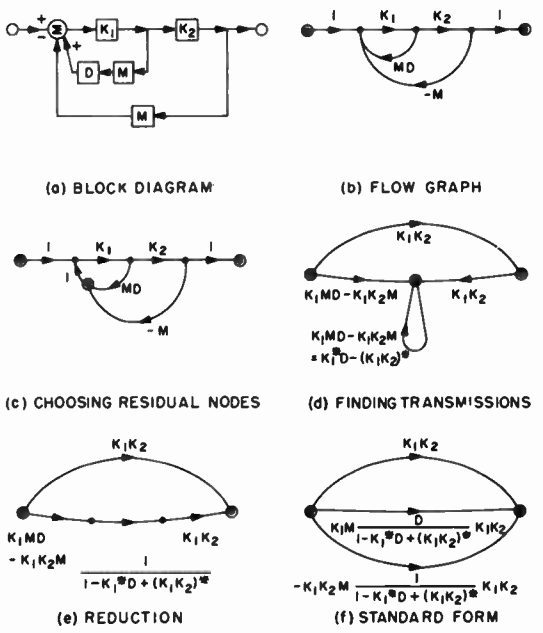


Fig. 1
Flow graph reduction of a sampled data feedback system to an equivalent standard form.

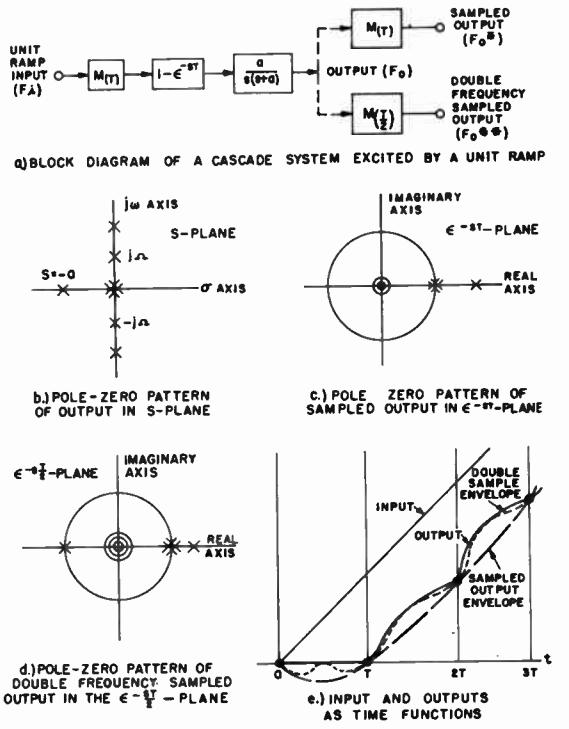


Fig. 2
Transient response analysis of a cascade system.

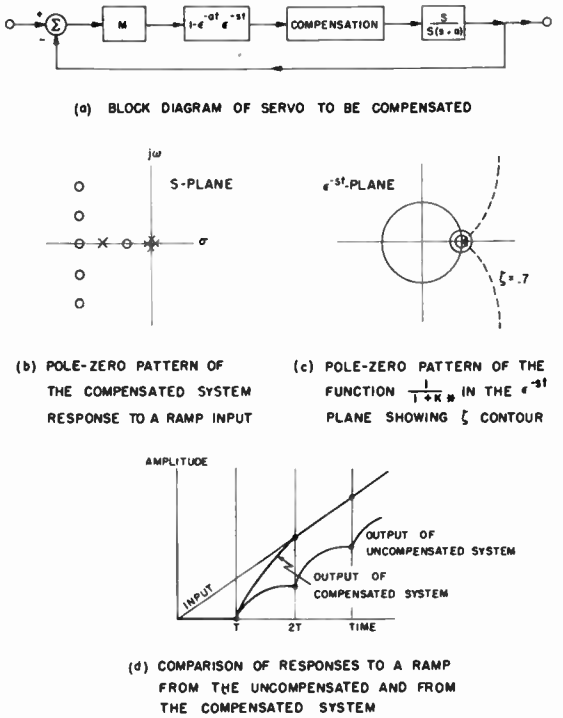


Fig. 3
Compensation design for a simple sampled-data servo.

GENERALIZED SERVOMECHANISM EVALUATION

Dr. Wm. P. Caywood, Jr., and Mr. William Kaufman

Introduction

Every year more automatic devices are being designed and installed by industry. Human labor is being supplanted increasingly by machines. The wide spread acceptance of servomechanisms and other automatic control devices has posed the problem of predicting an evaluation of system performance as integrated over its useful life in its expected application.

The first analyses made of servomechanism systems, the present day classical methods, primarily allow for the calculation of system stability, and secondarily allow for the calculation of its response to an input signal, either of which may be considered to be a Heaviside step function, or a known arrangement of them, or may be considered to be a periodic function describable as a series of Sine and Cosine functions. The calculation of performance usually comprises determining the amount of the first overshoot and the free resonant frequency in response to a Heaviside step, or the amplitude and phase response to a sinusoidal signal. In instances when the system is of complexity greater than the most simple, the response calculated may be simply rule-of-thumb inference to the Heaviside response of the closed loop system, obtained by the sinusoidal relation holding for the system with its feedback path opened. (See Chestnut and Mayer, Servomechanisms and Regulating System Design, for example.)

As a means of evaluating a servomechanism the classical analysis appears to have a rather severe limitation in accuracy. First there can be no apriori knowledge of the exact input signal if the servomechanism is to serve as a control device, rather than a mere generator of waveforms. Instead, there must be a certain repertory of possible messages, and the occurrences of the individual possibilities can be stated only in terms of probabilities. Additional uncertainty in the functional expressions for the input signal result from the inevitable contamination of the desired signal with an undesired one, called noise. The noise comprises another repertory of possible values the occurrence of which is in part, if not wholly, unrelated to the occurrence of specific, possible messages of the signal and again is given only in terms of probabilities. The response to the combination of signal and noise generally is the one of importance in performance evaluation or synthesis problems.

The limitation of classical analysis lies directly in its inability to effectively evaluate the performance of a servomechanism because of the inherent approximation of the input signal. A survey of the literature on the classical approach is indicative; different authors give, as the optimum, different amounts of overshoot in response to a step function, none of the authors relating his choice with the application at hand.

The present paper is divided into two parts. The first part comprises a review of the use of correlation methods to evaluate the performance of servomechanism systems for cases in which the input signal is treated statistically. Certain restrictions are tolerated:

1. That the servomechanism systems perform linear operations, invariant with respect to the choice of an origin in time.

2. That the convolution integral given by the expression:

$$\int_{-\infty}^{\infty} a(\tau)\, H(t-\tau)\, dt \qquad (1)$$

possesses mean square or mean "nth" integrability at least, where $a(\tau)$, $H(t-\tau)$ and mean "nth" integrability are more completely defined below.

The second part of this report is an extension of the applicability of the methods reviewed in the first part which is effected by considering numerical measurers of error, or norms, that are more general than the quadratic. That is to say, the error that may exist between the output and input at any instant is weighted according to an appropriate weighting function that is tailored to the application being considered. For input signals that are typical the average value of the weighted error is calculated, and it serves as a single number that is a measure of the "poorness" of performance. A synthesis of the best system for any particular application is simply a matter of minimizing the mean, weighted error by proper choice of the system parameters.

Error, as it applies to communications and servomechanism theory, is the difference between the actual output of a system and the desired or ideal output. In symbols,

$$e = f_o - f_d \qquad (2)$$

where e = error; f_o = actual output; f_d = desired output.

If, for example, f_o and f_d are voltages as functions of time, e is also a voltage as a func-

tion of time, and the mean square value of error is a measure of power developed by the error that must be overcome.

In any linear system, the output is the result of some linear operation on the input. In symbols,

$$f_o(t) = A(t)\, f_i(t) \tag{3}$$

where A(t) is some linear operator; $f_i(t)$ is the input. This could also include a time delay, and it also can be written as the convolution integral as follows:

$$A(t) f_i(t) = \int_{-\infty}^{\infty} a(t)\, f_i(t-\tau)\, dt \tag{4}$$

$a(t)$ is the derivative of the system response to a unit step and upon integration becomes $A(t) f_i(t)$ The desired output can be written similarly

$$f_d(t) = B(t)\, f_i(t) \tag{5}$$

For the B operator there is the choice of desiring the output to be identical with the signal component of the input, in which case B is unity. On the other hand there may be conditions such as the desirability of limiting the power rating of motors that suggest other choices of the B operator.

If the actual input is the sum of a signal input function $f_i(t)$ and some random noise function N(t), both resulting from a stationary random process, then the output is

$$f_o(t) = A(t)\left[f_i(t) + N(t)\right] \tag{6}$$

The error is the difference between desired and actual output, and can be expressed:

$$e(t) = A(t) f_i(t) + A(t) N(t) - B(t) f_i(t) \tag{7}$$

The quadratic norm can be further expanded:

$$e^2(t) = \left\{\left[A(t) - B(t)\right] f_i(t) + A(t)\, N(t)\right\}^2 \tag{8}$$

$$e^2(t) = \left\{\left[A(t) - B(t)\right] f_i(t)\right\}^2 + \left\{A(t) N(t)\right\}^2 + 2A(t)\left[A(t) - B(t)\right] N(t)\, f_i(t) \tag{9}$$

Employing the convolution integral, as in equation 4 above, and determining the average value:

$$\overline{e^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left\{ \int_{-\infty}^{\infty} \left[a(\tau_1) - b(\tau_1)\right] f_i(t-\tau_1)\, d\tau_1 \int_{-\infty}^{\infty} \left[a(\tau_2) - b(\tau_2)\right] f_i(t-\tau_2)\, d\tau_2 - T \right.$$

$$+ \int_{-\infty}^{\infty} a(\tau_1) N(t_1 - \tau_1)\, d\tau_1 \int_{-\infty}^{\infty} a(\tau_2) N(t-\tau_2)\, d\tau_2$$

$$\left. + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(\tau_1) N(t-\tau_1) \left[a(\tau_2) - b(\tau_2)\right] f_i(t-\tau_2)\, d\tau_1\, d\tau_2 \right\} dt \tag{10}$$

Assuming mean square integrability (if uniform convergence is not complete) the order of integration may be interchanged and:

$$\overline{e^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[a(\tau_1) - b(\tau_1)\right]\left[a(\tau_2) - b(\tau_2)\right] \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f_i(t-\tau_1) f_i(t-\tau_2)\, dt\, d\tau_1\, d\tau_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(\tau_1) a(\tau_2) \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} N(t-\tau_1) N(t-\tau_2)\, dt\, d\tau_1\, d\tau_2$$

$$+ 2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(\tau_1) \left[a(\tau_2) - b(\tau_2)\right] \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f_i(t-\tau_1) N(t-\tau_2)\, dt\, d\tau_2\, d\tau_1 \tag{11}$$

The three limiting processes can be seen to define in equation 11 the two auto-correlation functions for the signal and the contaminating noise, and the cross correlation between signal and noise respectively. Expressing the auto correlations as

$$\Phi_{ii}(\tau_2 - \tau_1) \quad ; \quad \Phi_{NN}(\tau_2 - \tau_1)$$

and the cross correlation by $\Phi_{Ni}(\tau_2 - \tau_1)$:

$$\overline{e^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[a(\tau_1)-b(\tau_1)][a(\tau_2)-b(\tau_2)]\phi_{ii}(\tau_2-\tau_1)d\tau_1\,d\tau_2$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}a(\tau_1)a(\tau_2)\phi_{NN}(\tau_2-\tau_1)d\tau_2 d\tau_1 \qquad (12)$$
$$+2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}a(\tau_1)[a(\tau_2)-b(\tau_2)]\phi_{Ni}(\tau_2-\tau_1)d\tau_2 d\tau_1$$

In many cases the noise is independent of the signal and the cross correlation between the two, ϕ_{Ni}, is zero. In this case the last term of equation 12 is removed.

Determining the best solution to the synthesis problem is done by finding a stationary value of mean squared error which represents a minimum. It is treated in the literature by Wiener, by Levinson and by others and will not be pursued further here.

The More General Criterion of Error

The quadratic norm has a certain realism in some applications, such as when the power or energy of the error is important. However, its mathematical convenience may not be supported by a very useful applicability in many situations. Consider fire control systems as an example. Rather than an error being "bad" in proportion to its size, the target is either hit or not. And, if it is missed completely, it makes no difference by how much.

As a general means of treating the problem of the error criterion, let us assume that whatever the realistic relation is, it can be expanded as a power series:

$$e_w = C_0 + C_1 e + C_2 e^2 + C_3 e^3 + \cdots \qquad (13)$$

In cases of symmetry the odd terms are absent of course, and in many cases there will be no need to move the origin by use of C_0. The average value of weighted error:

$$\overline{e_w} = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e_w\,dt \qquad (14)$$

Substituting the power series approximation for e_w:

$$\overline{e_w} = \lim_{T\to\infty}\frac{1}{2T}\left\{\int_{-T}^{T}C_0\,dt + \int_{-T}^{T}C_1 e\,dt + \int_{-T}^{T}C_2 e^2 dt + \cdots\right\} \qquad (15)$$

Considering the separate integrals:

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}C_0\,dt = C_0 \qquad (16)$$

which contributes a shift in the origin. The second:

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}C_1 e(t)\,dt = C_1\overline{e}^{av} \qquad (17)$$

With e given as the difference of the system operator A operating on message plus noise and the idealistic operator B on message:

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}C_1\left\{A(t)f_i(t)+A(t)N(t)-B(t)f_i(t)\right\}dt \qquad (18)$$

Applying the convolution integral as in equation 1 above:

$$\overline{e_w}^{av} = \lim_{T\to\infty}\frac{C_1}{2T}\int_{-T}^{T}\int_{-\infty}^{\infty}[a(t)f_i(t-\tau)+a(t)N(t-\tau)$$
$$-b(t)f_i(t-\tau)]d\tau\,dt \qquad (19)$$

Interchanging the orders of integration

$$= C_1\int_{-\infty}^{\infty}[a(\tau)-b(\tau)]\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f_i(t-\tau)dt\right]d\tau$$
$$+C_1\int_{-\infty}^{\infty}a(\tau)\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}N(t-\tau)dt\right]d\tau \qquad (20)$$

which is zero if the average overall time of both $f_i(t)$ and $N(t)$ is zero. The third term represents the quadratic norm that is described above. The fourth is zero in cases of criteria that are even functions. The next term:

$$C_4\overline{e^4} = C_4\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left\{\int_{-\infty}^{\infty}[a(\tau)-b(\tau)]f_i(t-\tau)dt\right.$$
$$\left.+\int_{-\infty}^{\infty}a(\tau)N(t-\tau)\,d\tau\right\}^4 \qquad (21)$$

To reduce the complexity of the expression consider cases in which there is no correlation between signal and noise. Then changing the order of integration and proceeding as for the quadratic case above:

$$C_4\overline{e^4} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{[a(\tau_1)-b(\tau_1)][a(\tau_2)-b(\tau_2)][a(\tau_3)-b(\tau_3)]\right.$$
$$\phi_{iii}(\tau_3,\tau_2,\tau_1)$$
$$\left.+a(\tau_1)a(\tau_2)a(\tau_3)\phi_{NNN}(\tau_3,\tau_2,\tau_1)\right\}d\tau_1 d\tau_2 d\tau_3 \qquad (22)$$

where: $\phi_{DDD}(\tau_3,\tau_2,\tau_1) =$

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}D(t-\tau_1)D(t-\tau_2)D(t-\tau_3)\,dt$$

$\phi_{DDD}(\tau_3, \tau_2, \tau_1)$ is a function that depends on the third moment probabilities, and hence is termed the third moment correlation function. It can be calculated, or measured on a machine, in a manner analagous to the auto correlation functions. One can proceed to the general case which is given by:

$$\overline{e^n} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Big\{ [a(\tau_1) - b(\tau_1)] \cdots [a(\tau_n) - b(\tau_n)] \times \phi_{i \cdots i}(\tau_1, \tau_2, \cdots \tau_n) + a(\tau_1) \cdots a(\tau_n) \phi_{N \cdots N}(\tau_1, \tau_2 \cdots \tau_n) \Big\} \times d\tau_1 \, d\tau_2 \cdots d\tau_n \qquad (23)$$

where

$$\phi_{D \cdots D}(\tau_1, \tau_2 \cdots \tau_n) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} D(t - \tau_1) D(t - \tau_2) \cdots D(t - \tau_n) \, d\tau$$

The correlation function $\phi_{D \cdots D}$ is again a higher moment one, in general of nth moment where n is also the degree of the term in the power series being considered.

REDUCTION OF FORCED ERROR IN CLOSED-LOOP SYSTEMS

Leonard H. King
Instrumentation Laboratory
Massachusetts Institute of Technology
Cambridge, Mass.

Summary – This paper illustrates a method for reducing the forced dynamic error in servomechanisms by design based on error coefficients. After a brief review of the dependence of the forced dynamic error upon error coefficients, the relationship between error coefficients and the parameters of a servomechanism is derived. This relationship is then used to show how closed-loop systems can be modified to obtain favorable error coefficients that reduce the forced dynamic error. The method has been tested by simulation, and photographs of simulator response show how the effect of additional integrations can be achieved by error coefficient adjustment.

I. Introduction

In a perfect positional servomechanism, the output position would always be aligned with the input command position regardless of noise, disturbing torques, or variations in the input position. As it is impossible to build a perfect positional servomechanism, good design practice consists largely of trying to reduce the following components of error:

1. Noise error, due to the transmission of input and internal noise to the output.
2. Position error, caused by disturbing torques.
3. Transient error, due to initial conditions or input discontinuities. By definition, this error disappears in one response time after which the output remains within five percent of the correct output for a step change in the input.[1]
4. Forced dynamic error, which exists after one response time and is due to variations in the input command position.

This paper illustrates a method for reducing the forced dynamic component of error by design based on error coefficients.[2,3,4,5] After a brief review of the dependence of the forced dynamic error upon error coefficients, the relationship between error coefficients and the parameters of a servomechanism is derived. This relationship is then used to show how closed-loop systems can be modified to obtain favorable error coefficients that reduce the forced dynamic error. The method has been tested by simulation, and photographs of simulator response show how the effect of additional integrations can be achieved by error coefficient adjustment. Specifically, it is shown (1) how a loop with no integrations can be made to have zero positional and velocity error and (2) how a loop with one integration can be made to have zero velocity and acceleration error.

II. Relationship between Forced Dynamic Error and Error Coefficients

If the input to a servo system varies continuously with time, then there will be a forced dynamic error E(t) between the desired input $q_{in}(t)$ and the actual system output $q_{out}(t)$. For finite derivatives of the input function, the forced dynamic error can generally be expanded in a series as

$$E(t) = a_o q_{in}(t) + a_1 \frac{d}{dt} q_{in}(t) + a_2 \frac{d^2}{dt^2} q_{in}(t) + \ldots + a_n \frac{d^n}{dt^n} q_{in}(t) + \ldots \quad (1)$$

where:

$E(t)$ is the forced dynamic error

$q_{in}(t)$ is the input function

a_o = zero-order (positional) error coefficient

a_1 = first-derivative (velocity) error coefficient

a_2 = second-derivative (acceleration) error coefficient

a_n = n^{th}-derivative error coefficient

For example, if $q_{in}(t)$ can be represented as $1 + 0.1t + 0.01t^2$, the error as a function of time, after initial conditions have disappeared, can be evaluated completely as

$$E(t) = a_o(1 + 0.1t + 0.01t^2) + a_1(0.1 + 0.02t) + a_2(0.02)$$

It can be seen that the error coefficients lead directly to the evaluation of the resulting error for all values of time rather than only for time equal to infinity, as with error constants, which have been treated elsewhere in the literature.[2,5,6]

III. Evaluation of Error Coefficients

A perfect servo is defined as one having all error coefficients equal to zero. In practice, it is desired to make as many of the lower-order error coefficients zero as possible. This is equivalent to making the more familiar error constants K_p, K_v, K_a, etc., infinite.

The error coefficients can be calculated by considering the Laplace transform of the error between the input and the output of the system.

Let

$$E(s) = q_{out}(s) - q_{in}(s) \tag{2}$$

where E(s) is the Laplace transform of the error E(t) between the output and input quantities $q_{out}(t)$ and $q_{in}(t)$. Dividing both sides of Equation (2) by $q_{in}(s)$ yields

$$\frac{E(s)}{q_{in}(s)} = \left[\frac{q_{out}(s)}{q_{in}(s)} - 1\right] \tag{3}$$

where $q_{out}(s)/q_{in}(s)$ is the closed-loop transfer function if zero initial conditions are assumed. Equation (3) can be expanded in a Maclaurin series for small values of s as

$$\frac{E}{q_{in}}(s) = a_o + a_1 s + a_2 s^2 + \ldots + a_n s^n + \ldots \tag{4}$$

From the similarity between Equations (1) and (4), it can be seen that the a's are the so-called error coefficients.

It is possible to evaluate successive error coefficients from the following relationships derived from Equation (4)

$$a_o = \lim_{s \to o} \left[\frac{E(s)}{q_{in}(s)}\right] \tag{5a}$$

$$a_1 = \lim_{s \to o} \frac{1}{s} \left[\frac{E(s)}{q_{in}(s)} - a_o\right] \tag{5b}$$

$$a_n = \lim_{s \to o} \frac{1}{s^n} \left[\frac{E(s)}{q_{in}(s)} - \sum_{k=0}^{n-1} a_k s^k\right] \tag{5c}$$

Instead of calculating error coefficients from the Maclaurin expansion of $[(q_{out}/q_{in})(s) - 1]$, it is easier and more instructive to calculate the values from $(q_{out}/q_{in})(s)$, which is the familiar closed-loop transfer function. From Equations (3) and (4), it can be seen that

$$\frac{q_{out}}{q_{in}}(s) = 1 + a_o + a_1 s + a_2 s^2 + \ldots + a_n s^n + \ldots \tag{6}$$

Equation (6) is used to calculate the error coefficients. Let

$$a'_o = 1 + a_o \tag{7}$$

and

$$\frac{q_{out}}{q_{in}}(s) = \frac{n_o + n_1 s + n_2 s^2 + \ldots + n_n s^n}{1 + d_1 s + d_2 s^2 + \ldots + d_m s^m} \tag{8}$$

Then, from Equations (6), (7) and (8):

$$a'_o = \lim_{s \to o} \left[\frac{q_{out}}{q_{in}}(s)\right] = n_o \tag{9}$$

$$a_1 = \lim_{s \to o} \frac{1}{s} \left[\frac{q_{out}}{q_{in}}(s) - a'_o\right] = n_1 - a'_o d_1 \tag{10}$$

$$a_2 = \lim_{s \to o} \frac{1}{s^2} \left[\frac{q_{out}}{q_{in}}(s) - a'_o - a_1 s\right] = n_2 - a'_o d_2 - a_1 d_1 \tag{11}$$

$$a_n = \lim_{s \to o} \frac{1}{s^n} \left[\frac{q_{out}}{q_{in}}(s) - a'_o - \sum_{k=1}^{n-1} a_k s^k\right] = n_n - a'_o d_n - \sum_{y=1}^{n-1} a_y d_{n-y} \tag{12}$$

Equation (12) provides a simple method of computing higher-order error coefficients in terms of lower-order error coefficients. Moreover, Equations (8) and (12) give an insight into how $(q_{out}/q_{in})(s)$ should be modified to provide the desired error coefficients. For example, with all the lower-order error coefficients up to a_n equal to zero, the coefficient a_n is made zero by making n_n equal to d_n.

IV. Methods of Obtaining Zero Error Coefficients

To correlate the desired error coefficients with the parameters of $(q_{out}/q_{in})(s)$, two block diagrams are next examined in detail in order to see how error coefficients of order n and n + 1 can be made equal to zero in single loops containing integrations of order n. Photographs of test results are shown.

A. Loop with No Integrations

First, it is shown how the system of Fig. 1, which contains no integrations, can be modified to obtain zero positional and zero velocity error coefficients. With K_2 gain compensation inserted into the feedback loop of Fig. 1, $(q_{out}/q_{in})(s)$ can be written as follows:

$$\frac{q_{out}}{q_{in}}(s) = \frac{n_o + n_1 s + n_2 s^2}{1 + d_1 s + d_2 s^2} \tag{13}$$

where

$$n_o = \frac{K}{1 + KK_2}$$

$$n_1 = \frac{KT_1 + KT_3}{1 + KK_2}$$

$$n_2 = \frac{K_1 T_1 T_3}{1 + KK_2}$$

$$d_1 = \frac{T_2 + T_4 + KK_2 T_1 + KK_2 T_3}{1 + KK_2}$$

$$d_2 = \frac{T_2 T_4 + KK_2 T_1 T_3}{1 + KK_2}$$

Equations (7) and (9) show that for a_o to be equal to zero, $(1 + KK_2)$ must equal K. Therefore K_2 must equal $(1 - (1/K))$, and with this value the system of Fig. 1 now can be classified as a positional servomechanism since it has zero positional error. The same system configuration with unity feedback, however, does not meet with the zero positional error requirement of a servo.

To determine a_1, Equation (10) is now used:

$$a_1 = \frac{KT_1 + KT_3}{1 + KK_2} - \frac{T_2 + T_4 + KK_2 T_1 + KK_2 T_3}{1 + KK_2} \tag{14}$$

But

$$KK_2 = K - 1 \tag{15}$$

Therefore

$$a_1 = \frac{K(T_1 + T_3) - T_2 - T_4 - (K-1)(T_1 + T_3)}{K} \tag{16}$$

$$a_1 = \frac{T_1 + T_3 - T_2 - T_4}{K} \tag{17}$$

By adding a network to the forward loop so that the sum of the time constants in the numerator equals the sum of the time constants in the denominator, the servo will now have a zero velocity error coefficient.

In order to illustrate the effectiveness of the proposed method of compensation, the block diagram of Fig. 2 was examined on the General Purpose Simulator at the Instrumentation Laboratory, Massachusetts Institute of Technology, and photographs were taken of the transient response as seen on an oscilloscope. The pictures of Fig. 3 show that the forced dynamic error of this specific system has been reduced by using eight-tenths feedback instead of unity feedback. Furthermore, the gain of the feedback loop K is not critical in the system. As long as K is between six-tenths and unity, the forced dynamic error of the system would be less than that present with unity feedback.

It can also be seen from the pictures of Fig.3 that the forced dynamic error for a ramp input has been reduced to zero by insertion of a passive network in the forward path. As there is a wide choice in the parameters of the actual network used, the transient error responses are shown for several different networks.

B. Single Loop with One Integration

It will now be shown how the system of Fig. 4, which contains a single integration, can be modified to obtain zero velocity and acceleration coefficients.

As a first modification, the network $\frac{(T_5 s + 1)}{(T_6 s + 1)}$ is placed in the feedback path. Then

$$\frac{q_{out}}{q_{in}}(s) = \frac{\frac{T_1 s + 1}{T_2 s + 1} \times \frac{T_3 s + 1}{T_4 s + 1} \times \frac{K}{s}}{1 + \frac{T_1 s + 1}{T_2 s + 1} \times \frac{T_3 s + 1}{T_4 s + 1} \times \frac{T_5 s + 1}{T_6 s + 1} \times \frac{K}{s}} \tag{18}$$

$$\frac{q_{out}}{q_{in}}(s) = \frac{n_o + n_1 s + n_2 s^2 + n_3 s^3}{1 + d_1 s + d_2 s^2 + d_3 s^3 + d_4 s^4} \tag{19}$$

where:

$$\begin{aligned} n_o &= 1 \\ n_1 &= T_1 + T_3 + T_6 \\ n_2 &= T_1 T_3 + T_1 T_6 + T_3 T_6 \\ d_1 &= T_1 + T_3 + T_5 + \frac{1}{K} \\ d_2 &= T_1 T_3 + T_3 T_5 + T_1 T_5 + \frac{1}{K}(T_2 + T_4 + T_6) \end{aligned} \tag{20}$$

From these equations, used in conjunction with Equation (12), it is seen that

$$a_o = 0$$

and

$$\begin{aligned} a_1 &= T_1 + T_3 + T_6 - (T_1 + T_3 + T_5 + \frac{1}{K}) \\ &= T_6 - (T_5 + \frac{1}{K}) \end{aligned} \tag{21}$$

Therefore, a_1 is dependent only upon the parameters of the feedback path and the gain of the forward path. Specifically, in order for a_1 to be equal to zero, T_6 must equal $T_5 + (1/K)$. With T_5 equal to zero and a_1 made equal to zero by

the proper choice of T_6, a_2 is now given by the equation

$$a_2 = T_1T_3 + T_1T_6 + T_3T_6 - T_1T_3 - \frac{1}{K}(T_2+T_4+T_6) \tag{22}$$

Because $T_6 = 1/K$, then

$$a_2 = \frac{T_1}{K} + \frac{T_3}{K} - \frac{T_2}{K} - \frac{T_4}{K} - \frac{1}{K^2} \tag{23}$$

With a_1 equal to zero, a_2 can therefore be made equal to zero by ensuring that in the forward part of the loop the sum of the first-order time constants in the numerator is equal to the sum of the first-order time constants in the denominator plus 1/K.

To illustrate further the results obtainable by the use of error coefficients, the block diagram of Fig. 5 was also examined on the General Purpose Simulator. The results, summarized in Fig. 6, show that it is possible to obtain zero velocity and zero acceleration error coefficients without appreciably changing the stability or response time of the system.

C. Generalized Technique

The effect of a servo in the feedback loop, such as in Fig. 7, is considered next. For a constant angular velocity input q_{in}, the following steady-state equations can be written:

$$q_1 = q_{in} - \frac{\dot{q}_{in}}{K} \tag{24}$$

$$q_1 = q_{out} - \frac{\dot{q}_{out}}{K_A} \tag{25}$$

$$\dot{q}_{in} = \dot{q}_{out} \tag{26}$$

For q_{out} to be equal to q_{in}, K_A must therefore equal K. However, from the viewpoint of the over-all loop, the servo in the feedback path is equal to the passive network $1/[1 + (1/K)s]$, which is identical with the modifying network that was calculated for the servo in Fig. 4. Consequently, the modifying passive network acts as though it were a servo having the proper velocity constant. In fact, any passive network which has a transfer function equal to the closed-loop transfer function of a servo having the correct velocity constant can be used in the feedback path of Fig. 4 to obtain a zero velocity error coefficient. In general, this analog method can be used for selecting the network necessary to make a_n equal to zero in a loop containing n integrations where n is 1 or larger.

D. Alternate Compensation Techniques

Hitherto in this paper, emphasis has been placed on using a network in the feedback path in order to make the a_n error coefficient equal to zero in an n integration loop. In many cases, it may be impossible or inconvenient to put a network in the feedback path, due to the nature of the system components. However, even with a unity feedback path, it may still be possible to have the desired error coefficient. Figure 8 shows two possible ways in which this may be done. In multiloop systems more ways are possible and the equivalent method of compensation will depend to a large extent upon the ingenuity of the designer.

E. Use of Additional Integrations

In a single-loop system with unity feedback, higher-order error coefficients can be made equal to zero by the insertion of additional integrations in the forward path. Thus, the compensation techniques developed in this paper are mathematically equivalent to the use of extra integrations in the forward path of a unity feedback system. For example, the system of Fig. 5, which contains only one integration and two passive compensating networks, is mathematically equivalent to a system with unity feedback and three integrations in the forward path. In general, each additional passive compensation network in a single-loop system can be replaced by an active integrator. However, in multiloop systems, it may be possible to have one network take the place of several integrators.

V. Effect of Zero Error Coefficient Modifications upon System Response

A. Stability and Band Width

It has been shown in the preceding section that the passive network used in the feedback loop is equivalent to a servo. If the band width of the equivalent servo is sufficiently larger than the band width of the original loop, then the over-all loop will see unity feedback inside its band width. Therefore, the passive network will not appreciably affect the stability or band width of the over-all loop. Consequently, it is generally possible to design the passive network used in the feedback path so that it does not adversely affect the stability or band width of the over-all loop. Moreover, if the original loop has marginal stability as a result of trying to achieve a large error constant, then both the stability and error constants of the loop can be increased by reducing the forward loop gain and using a network in the feedback loop.

The modifying network used in the forward loop has the function of making the a_{n+1} error coefficient equal to zero in an n integration loop. In a zero integration loop, this network should be such that the sum of the numerator time constants is equal to the sum of the denominator time constants. Since only the over-all sum is important, the networks shown below can be considered equivalent in the steady state.

$$\frac{5s+1}{7s+1} \text{ or } \frac{1s+1}{3s+1} \text{ or } \frac{3s+1}{6s+1} \times \frac{2s+1}{1s+1}$$

Consequently, there is a wide latitude in the choice of the modifying network, and it may be possible to pick the forward loop network so that

it does not affect the stability or band width. Likewise, in higher-order systems there is a wide latitude in the choice of the modifying network.

B. Reshaping of Transient Response

It is known that the set of linear differential operators with constant coefficients is commutative.[7] Consequently, since a ramp function is the integral of a step function, the error response to a ramp input of a linear system with constant coefficients can be found by integrating the error response of the system for a step input. Similarly, the error response for a constant acceleration input is the double integral of the error response for a step input. Making higher-order error coefficients equal to zero therefore means that the error response to a step input is modified so that higher integrals of this response become equal to zero. This is illustrated in the pictures of Fig. 9.

VI. Conclusion

A method has been presented for reducing the forced dynamic error in a closed-loop system. In general, there is a wide latitude in the specific parameters of the networks used to accomplish this effect, and it is possible to design a network that does not appreciably change the stability or response time of the system.

Furthermore, even though it is a matching method of compensation that assumes the system parameters do not change, significant reduction of forced dynamic error can often be obtained without the network parameters having to be set precisely. For example, if a network $1/(T_7 s + 1)$ is used instead of the theoretically correct network $1/(T_6 s + 1)$ in the feedback path of Fig. 4, then the forced dynamic error for a ramp input will be less than would be present with unity feedback as long as $0 < T_7 < 2T_6$. The choice of the network can be facilitated by the following general rule:

In a single loop containing n integrations, the error coefficient a_n is independent of the dynamic lags of the forward path. It does, however, depend upon the gain of the forward path and the parameters of the feedback path. Specifically, in a loop containing no integrations, a zero positional error-coefficient can be obtained by using a feedback path gain of one minus the reciprocal of the gain of the forward path.

In a loop containing one integration, a zero velocity error coefficient can be obtained by using the feedback network $(T_5 s + 1)/(T_6 s + 1)$ where T_6 equals T_5 plus the reciprocal of the gain of the forward path.

Although this paper is restricted to the consideration of positional servomechanisms, the techniques developed herein are applicable to a larger class of systems. If it is desired that the output q_{out} be some function of the input $f(q_{in})$, then the system error may be defined as $q_{out} - f(q_{in})$. With this definition of error, a new set of error coefficients may be formulated in terms of the system parameters. The system parameters may then be modified intelligently to reduce the forced dynamic error. Finally, since the compensation used is contingent upon modifying the feedback loop, it has been shown conclusively that servo performance can be improved by using other than unity feedback.

REFERENCES

1. Draper, C. S., McKay, W., and Lees, S., "Instrument Engineering," vol. I, McGraw-Hill Book Co., Inc., New York, N. Y., 1952.
2. James, H. M., Nichols, N. B., and Phillips, R. S., "Theory of Servomechanisms," vol. 25, Radiation Laboratory Series, McGraw-Hill Book Co., Inc., New York, N. Y. 1947.
3. Floyd, G. H., Jr., and Eisengrein, R. H., "Design of fire-control systems," Servomechanisms Laboratory, M. I. T., Cambridge, Mass., July, 1950, (Confidential).
4. Bower, J. L., "A note on the error coefficients of a servomechanism," Jour. Appl. Phys., vol. 21, July, 1950.
5. Chestnut, H., and Mayer, R. W., "Servomechanisms and Regulating System Design," vol. I, John Wiley & Sons, Inc., New York, N. Y., 1951.
6. Brown, G. S., and Campbell, D. P., "Principles of Servomechanisms," John Wiley & Sons, Inc., New York, N. Y., 1948.
7. Hildebrand, F. B., "Advanced Calculus for Engineers," Prentice-Hall, Inc., New York, N. Y., 1949.

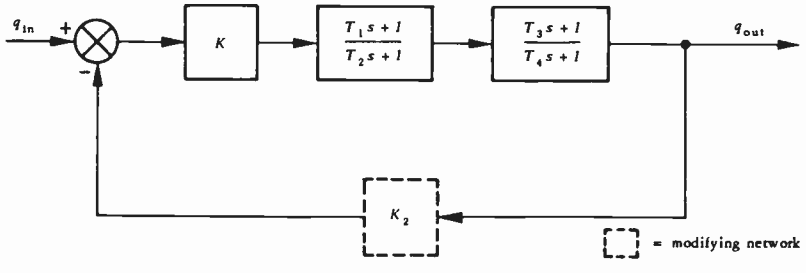


Fig. 1
General closed-loop system containing no integration.

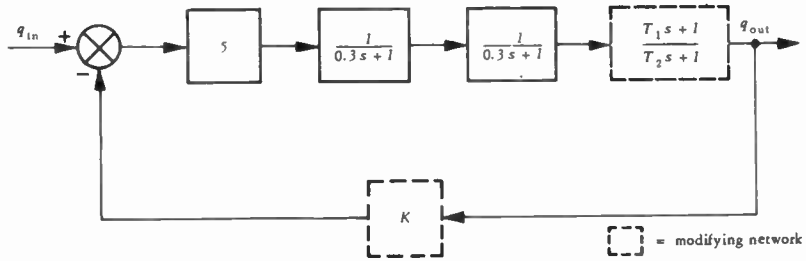


Fig. 2

Sample loop containing no integrations.

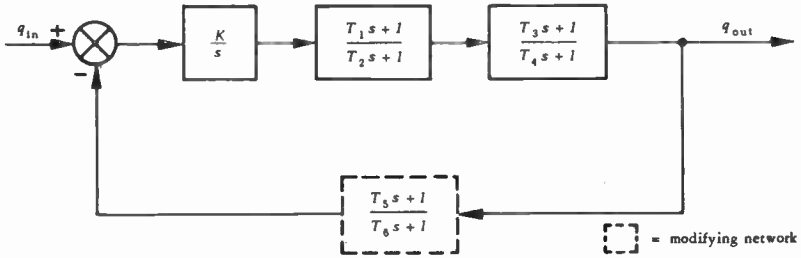


Fig. 4

General closed-loop system containing one integration.

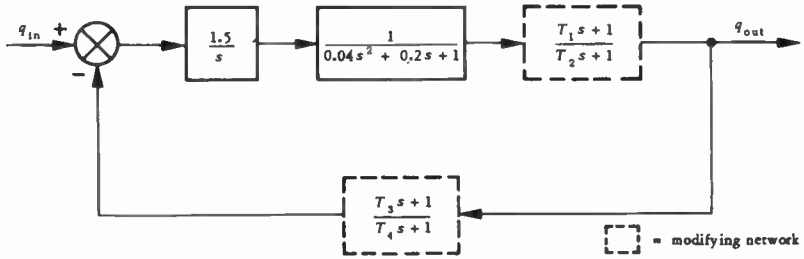


Fig. 5

Sample loop containing one integration.

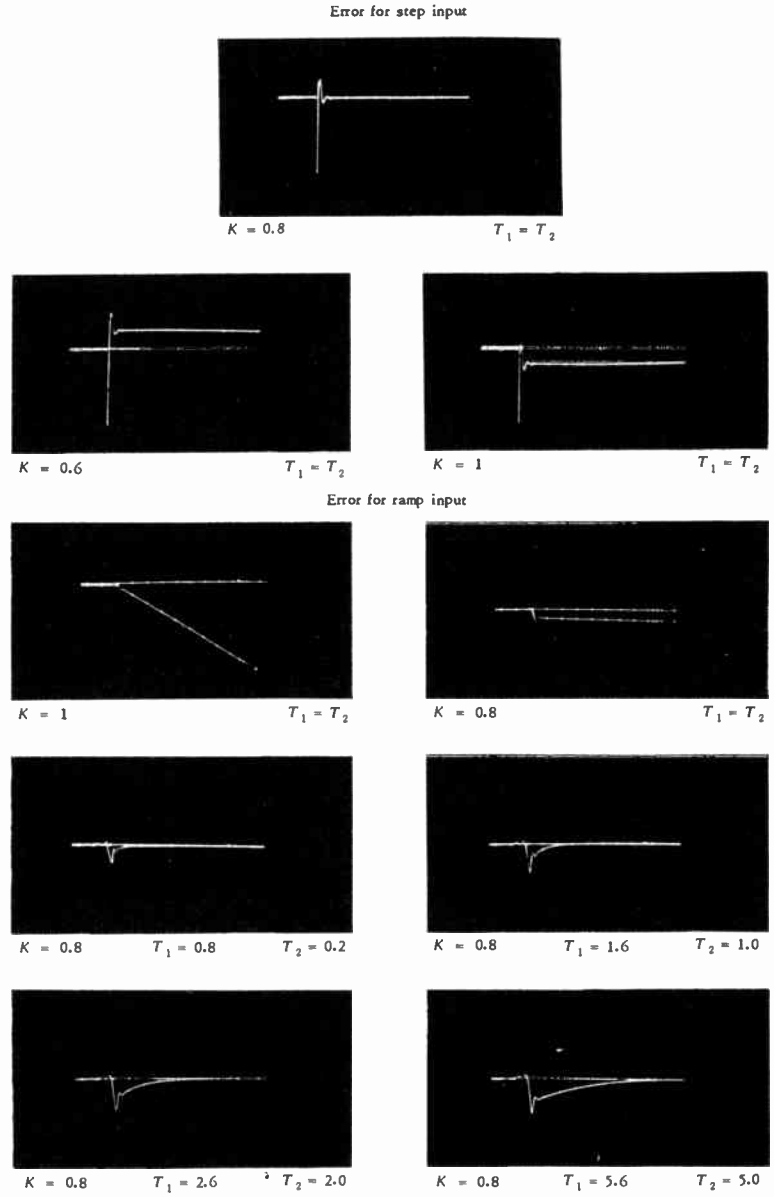


Fig. 3

Simulator results obtained for system of Fig. 2.

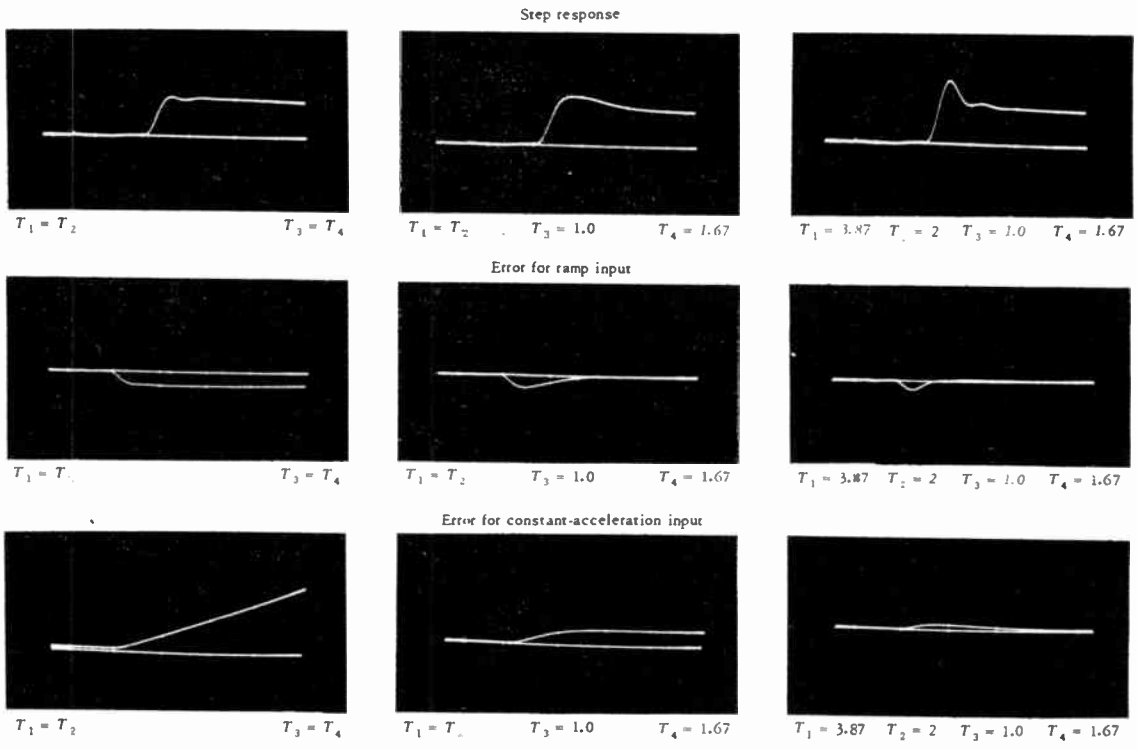


Fig. 6

Simulator results obtained for system of Fig. 5.

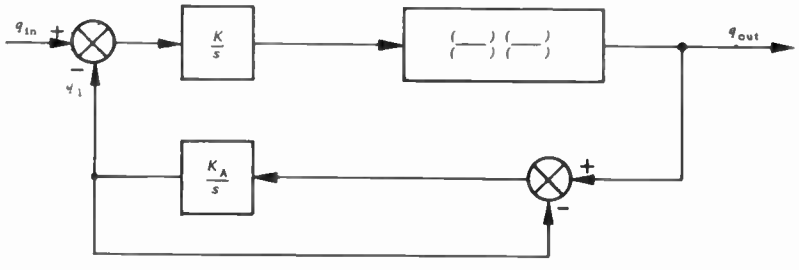
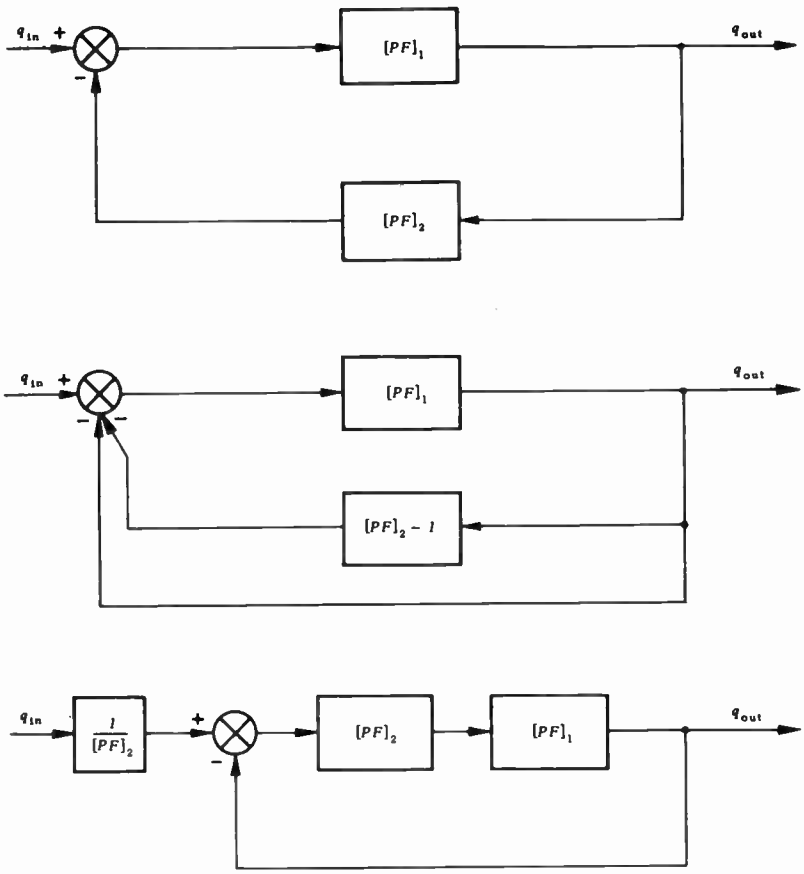


Fig. 7

Example of system containing a servo in the feedback path.



$[PF]_1$ = performance function of forward path

$[PF]_2$ = performance function of modifying network

Fig. 8

Equivalent methods of compensation.

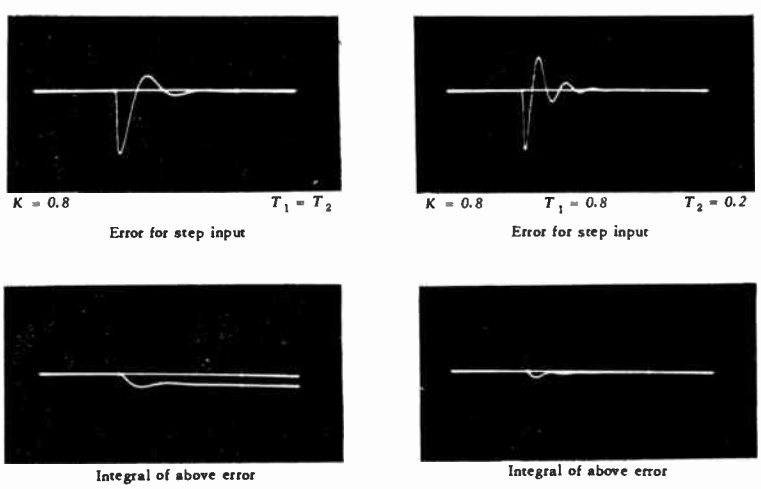


Fig. 9

Effect of modifications on transient response of system of Fig. 2.