MR B.B. JACKSON

ITT

ELECTRICAL COMMUNICATION

TECHNICAL JOURNAL OF INTERNATIONAL TELEPHONE AND TELEGRAPH CORPORATION

VOLUME 37 **1961** **NUMBER 2**

ELECTRICAL COMMUNICATION
Technical Journal Published Quarterly by
INTERNATIONAL TELEPHONE and TELEGRAPH CORPORATION
320 Park Avenue, New York 22, New York
President: H. S. Geneen
Secretary: J. J. Navin

CONTENTS

Volume 37 **1961** **Number 2**

EDITOR, H. P. Westman

Subscription: $2.00 per year; 50¢ per copy

This Issue in Brief

Transmission of Data over Telephone-Type Circuits—Several of our companies participated in an extensive test programme initiated by Special Committee A on data transmission of the Comité Consultatif International Télégraphique et Téléphonique to ascertain the ability of the European telephone network to transmit data and to determine the most suitable characteristics for data transmission systems using that network.

For the important case of serial binary transmission over switched telephone circuits, which generally are more disturbed than leased circuits, frequency modulation is considered most effective, paying due regard to equipment cost and complexity.

Measurements were made of the errors occurring in thousands of millions of bits of data transmitted over many different telephone connections; although the average error rates were low, occasional bursts of errors warned against accepting average figures as reliable indexes of performance. The received signal level has the most powerful influence on error rate and many measurements were made at limiting values of received signals so as to give the most pessimistic results.

The telephone network is shown to be a suitable medium for transmitting data but there is a need for automatic error correction if error rates satisfactory to customers are to be achieved over most connections. The distribution of errors is such that correction by repetition of blocks containing errors is preferable to correction by error-correction code. Depending on the block size used, the blocks to be repeated may be determined either by means of an error-detection code acting alone or in conjunction with a trouble detector. A fairly complex detection code is necessary and a description is given of work being done to determine the structure of a code having adequate error-detecting capability and involving the minimum of redundant information.

Frequency-Shift Modulation of Binary-Coded Signals for Transmission Over Telephone Circuits—The need to transfer large quantities of engineering, scientific, statistical, and accounting data in short time intervals makes transmission speeds much beyond those of conventional telegraphy necessary. The use of standard telephone channels provides a convenient path for such transmission.

The modem link, consisting of the modulator, line, and demodulator, is considered. Emphasis is placed on frequency-shift modulation, which is compared with phase-shift modulation. The theoretical performance of a phase-shift-modulated system in terms of error rate as a function of signal-to-noise ratio can be calculated and is well known. Unfortunately, frequency-shift demodulation is not so amenable to assessment by calculation, and a measurement program was undertaken to determine what performance such a system could provide in practice. For a given receiver bandwidth, the element error rate was measured for different frequency shifts and modulation rates, against different levels of white noise used as a source of interference.

Analysis of the results shows that a frequency-shift system may be optimized as regards error liability by suitable choice of frequency shift and receiver bandwidth. The practically achieved performance compares very favorably with the theoretical performance of phase-shift systems.

From a large number of error-rate measurements on various loop-connected lines, a relationship between noise performance and the phase nonlinearity of the transmission medium was obtained. White noise injection was also used in this case.

For various applications, different modulation rates and clock-signal frequencies are required. A table covering a suitable range of baud figures is proposed for standardization. A proposed range of modem link terminals is also included.

Practical Evolution of the Commutated-Aerial Direction-Finding System—In the commutated-aerial direction-finder, the individual aerials in a circular array are sequentially connected to a single receiver. Each such sample of the received wave is stored for one aerial-connection period in the receiver for comparison of its phase with that of the succeeding sample from the next aerial. These phase differences stem from the relative difference in path length from each aerial to the source of the waves.

To reduce interaction of the aerials, top loaded thin unipoles, shorter than a quarter wavelength at the highest operating frequency, are disconnected about a third of the distance from the ground plane by a voltage-operated germanium diode. An equivalent dipole can be made of two such units.

Commutation of the 12 to 24 aerials usually employed is controlled by a 1-kilocycle-per-second free-running oscillator. Each aerial gate uses 3 diodes in a *T* network and is open for one millisecond. The first part of the gated period may contain switching transients and is rejected, the receiver operating on the last part of each pulse.

Receiver design principles, a linear phase discriminator, and display means are covered. A discussion of system performance includes errors due to site, multi-path, polarization, and interference, as well as the effects of high-elevation sources.

Wide-Base Doppler Very-High-Frequency Direction Finder—Large bearing errors caused by multipath propagation have discredited direction finders in the very- and ultra-high frequencies. Wide-base antenna arrays offer a practical means of substantially reducing these errors.

If a straight line of antennas aligned in the direction of transmission is scanned sequentially by a receiver, a doppler frequency will result from the difference in path length between the source and the several antennas. If the antennas are aligned perpendicularly to the direction to the source, there will be no difference in path lengths and no doppler effect.

To avoid rotating such a line of antennas while they are being scanned, a large number of antennas may be arranged in a circle, a few adjacent antennas forming a relatively short flat arc equivalent to a straight line. If the scanning pattern is rotated around the circle, this effectively straight line of antennas will point progressively in all directions. The scanning of the circle of antennas is accomplished by a uniformly rotating capacitive switch that first scans 6 antennas in the forward direction and then only 5 antennas in the backward direction to progressively cover the entire array.

The bearing is indicated on a simple mechanical display in which a light beam is reflected to a suitable screen from 2 mirrors tilting at right angles to each other under control of the sine and cosine outputs of a two-phase generator mechanically coupled to the capacitive switch. The information from a brief transmission can be retained until a new signal arrives.

Logarithmic Navigation for Precise Guidance of Space Vehicles—The orbital rendezvous of two space vehicles or of a vehicle and a celestial body demands precise control of distance between them and zero relative velocity as they meet. It is assumed that the overtaking vehicle will use deceleration to achieve safe contact with the other body.

After considering control in one dimension, multidimensional control is treated. Instrumentation errors are considered; the effects of bias and dynamic errors are presented in two appendixes.

Seven space guidance missions are treated and include rendezvous with an orbiting satellite, soft landing on a planet, soft landing on a planet while avoiding dangerous terrain, soft landing on a planet in a particular location, soft landing on a planet in a particular location when approaching from a prescribed direction, and matching a desired trajectory precisely in position, velocity, acceleration, and time.

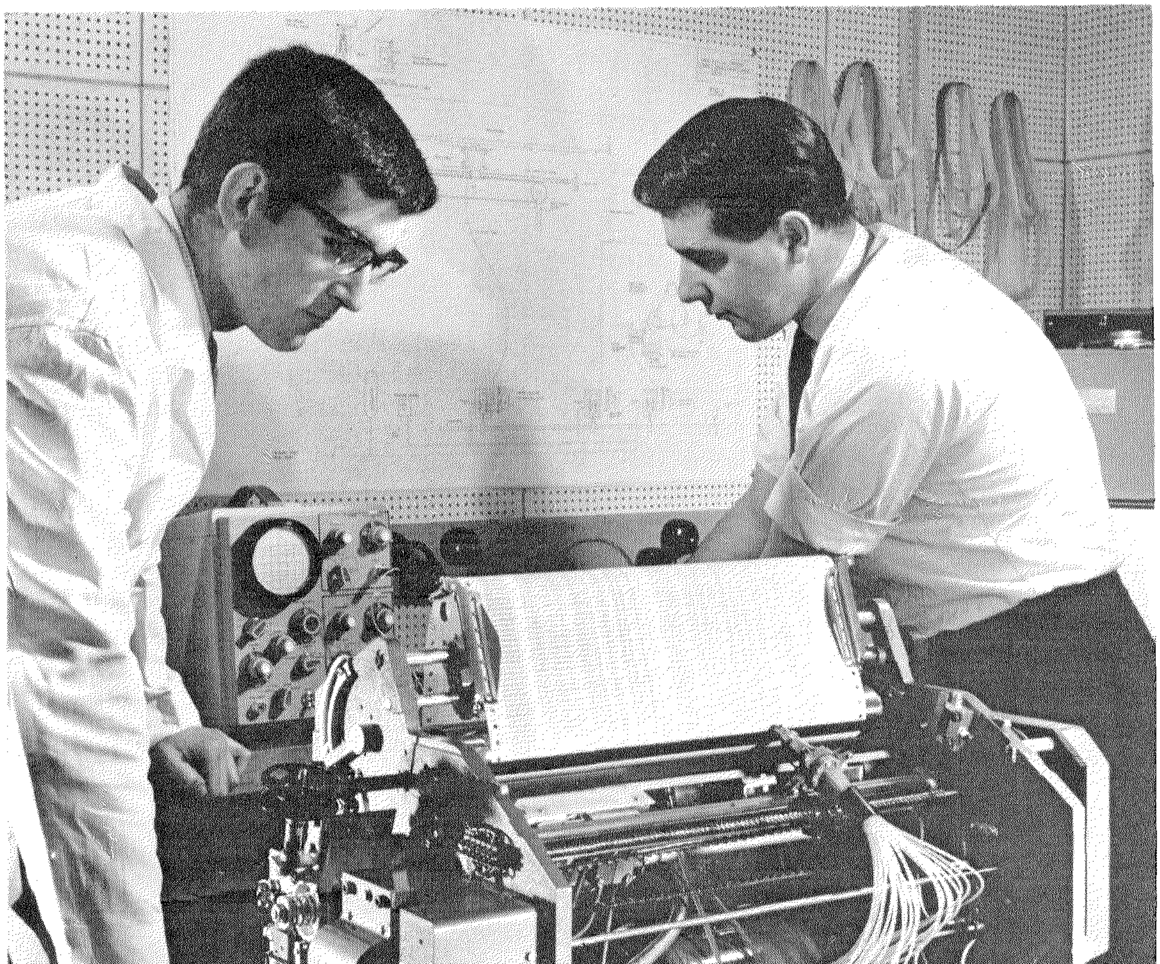


Figure 1—Engineers examine a 1000-word-per-minute page printer designed as one of the output devices for a new order recording and invoicing system.

Recent Engineering Developments

High-Speed Teleprinter—The Model 1000 output printer shown in Figure 1 is a serial (character-by-character) printer designed for use as an economical direct on-line computer output device or as on off-line playback unit controlled by punched paper tape or magnetic tape. Speed of operation is 100 characters per second, which is equivalent to 1000 words per minute. This is several times that of the fastest teleprinters.

Data recording is by an ingenious print-head consisting of 25 hydraulically actuated styli arranged in a compact 5-by-5 grid, the combined operation of which builds up each character in mosaic fashion, permitting an almost infinite number of character patterns to be obtained. Printing is done through a conventional inked ribbon.

Other features include high-speed paper feed controlled by adjustable settings on interchangeable metallic programme discs, variable tabulation control, and adjustable platen tractors to take sprocket-fed stationery of any width between 8·5 and 17 inches (21·6 and 28 centimetres). Single or multi-copy printing is available with a printing line ranging from 10 to 150 characters depending on the width of the stationery used. Signal input arrangements provide for hard valve or transistor circuits.

Creed & Company

Autobanker—A system installed at the First National Bank of Waukesha, Wisconsin, allows the bank teller and the client, although a substantial distance apart, to talk to and see each other and to exchange documents and money. The system is applied to a special drive-in tellers window that is several hundred feet from the banking office position shown in Figure 2. The client remains seated in an automobile during the entire business transaction.

Vision is provided through a two-way closed-circuit television system. Conversation is over a high-fidelity intercommunication system. Papers are carried by a unique pneumatic-tube system that can transport up to 3.5 pounds (1.6 kilograms) at a time at a speed of 25 feet (7.6 meters) per second. The lock on the access door of the pneumatic-tube carrier is under control of the teller, who opens it when the carrier is in proper position before the client. This type of drive-in window provides for every service normally handled by a teller in the bank because that is where the teller is.

ITT Industrial Products Division
Airmatic Systems Corporation

Film Titler—Figure 3 shows a film titler for fast, accurate, durable printing of captions on film negatives. The process is 20 times faster than conventional pen-and-ink methods as it permits film to be titled at speeds up to four

Figure 2—Tellers position for the autobanker. Closed-circuit television, an intercommunication sound system, and a pneumatic-tube to transport papers and money eliminate the effects of distance and permit transaction from a remote drive-in window of all banking services that are normally handled by a teller in the bank.

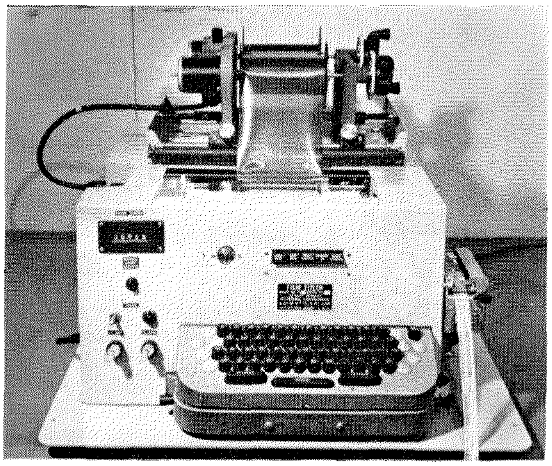


Figure 3—Film titler permits single manual typing of a caption or its repetition under control of a paper tape.

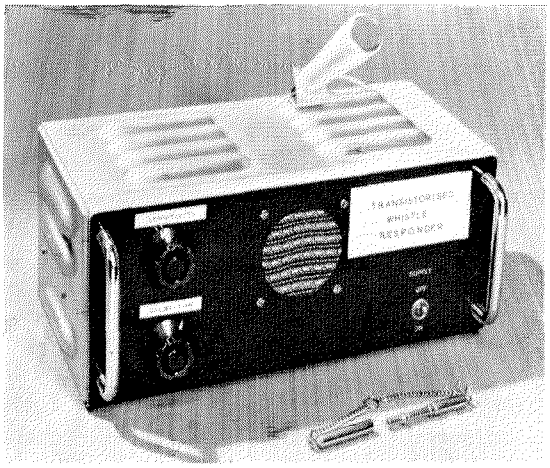


Figure 4—This beacon responds to a superaudible whistle by sounding a bell to inform a blind person of the position of the unit. The blind person is thus enabled to return to the place where he left the beacon.

frames per minute. Any size of roll film up to 9.5 inches (24 centimeters) in width can be handled.

The device is essentially a semiautomatic typewriter having a manual keyboard, as well as a paper-tape mechanism for printing repetitive information. A gold-leaf typewriter ribbon produces lasting golden letters. A person with ordinary typewriting skill can learn to operate the device in two days.

ITT Federal Laboratories

Whistle Responder—A transistor-operated beacon for the blind responds to a small superaudible dog whistle but not to the ordinary sounds we hear. Both the device and one of these whistles appear in Figure 4. The radiation from the whistle is selectively picked up by a microphone on the unit and initiates the momentary operation of a self-contained bell. The blind person can then estimate his position in relation to the known location of the device by judging the direction and intensity of the sound of the bell.

The range of the device is about 60 yards, (55 meters) making it suitable for use within a large area. It could be adapted to other remote-control requirements.

Standard Telephones and Cables

Transmission Of Data Over Telephone-Type Circuits

A review of studies and measurement programmes by companies of ITT Europe in the field of data transmission.

W. T. JONES

Standard Telephones and Cables Limited; London, England

1. Introduction

The use in business and industrial establishments of computers and data-processing systems for the automatic handling and processing of clerical data and also the automatization of methods of measurement, surveillance, and control of manufacturing and other production processes have created a need for the rapid transfer of data between widely separated locations. It has been forecast that in the not-too-distant future data communications may require a network as great as that needed for telephone traffic.

It is inevitable that the national and international switched and leased telephone networks should be considered as possible media for the transmission of these data. The telephone network is widespread, interconnecting many millions of different locations, and facilities exist for its rapid expansion on the basis of well-established standards. The needs of data transmission can be met more rapidly and more economically by using this network jointly for telephone and data communications than by setting up a separate network designed specially for and devoted solely to the transmission of data.

It is inconceivable, in view of the vastness of the telephone network, that it should have to be adapted to the special requirements of data transmission. As the volume of transmitted data grows, it may be expected that the design basis of the network will take data increasingly into account. However, for the present it is reasonable that data-transmission systems should themselves be so designed as to give the desired performance over the network as it exists.

There is need for information on the characteristics of the telephone network relevant to the transmission of data. It was to gather such data that at its meeting of March 1960 Special Committee *A* on Data Transmission (known at the time as Working Party 43) of the Comité Consultatif International Télégraphique et Téléphonique instituted a programme of studies of "the possibilities of telephone-type circuits for data transmission" so that recommendations might be made as regards modulation speeds, modulation methods, error-correction methods, lengths of blocks, et cetera.[1] This paper reviews the studies and measurements undertaken by the European companies of the International Telephone and Telegraph Corporation as part of that programme.

2. Causes of Errors [2]

The transmission properties of a telephone circuit or connection, from the point of view of speech, are determined mainly by the bandwidth and transmission loss, that is, the attenuation–frequency characteristic and the noise level. It is sufficient to think of the noise level as a mean power over a period of time of the order of one minute, the human ear being extremely tolerant of short-term fluctuations in the level of background noise. It is tolerant also of phase distortion, and fairly wide limits can be specified for such distortion in a telephone circuit used for speech.

These properties have to be taken into account in determining the parameters of a data-transmission system intended for operation over telephone circuits. The possible speed of the system will depend on the bandwidth available, as determined from the loss and phase–frequency characteristics, and the frequencies used for the transmission will depend on where that available bandwidth is located in the frequency spectrum; account also has to be taken of any parts of the available spectrum used for in-band signalling systems. Any departure from the ideal loss– or phase–frequency characteristic within the used bandwidth will not in itself be a source of error, although it may reduce the margin of a system against errors from other causes. Likewise, on connections that are considered good for speech, the received signal level and noise level are such that the noise, considered as of constant level, will itself be a negligible source of errors. The properties of a telephone circuit that are of prime interest from the point of view of the error rate in a data-transmission system using the circuit are the probability of occurrence of high noise levels of short duration and short breaks in transmission.

The causes of many of the loud noises liable to occur on telephone circuits are known; probably the most prolific are the high voltages and sudden current changes induced by dialling and relay operations on parallel circuits of the same subscribers' cable. Again, wideband carrier telephone systems are not designed to avoid completely the possibility of overload. During very busy traffic periods momentary overloads will occur in common amplifiers and modulating equipments, resulting in short noise pulses in the derived channels. In telephone exchange equipment the main source of trouble will be fluctuating resistance of switch contacts, brought about by vibration caused by road traffic or the operation of nearby relays or selectors. If the contacts are "wetted" spurious signals can be generated by modulation of the current passing through them; when "dry" the resistance changes are usually greater, leading to intermittent interruption of the circuit. Short breaks in telephone connections are also caused by accidental short circuits and the disturbances of weak joints, fuse mountings, et cetera on distribution frames during maintenance operations and by routine changeovers of power and carrier supplies on wideband systems.

When data are transmitted serially over telephone circuits the duration of each signal element will be of the order of a millisecond, depending on the speed of transmission. Disturbances such as those described, although short from the standpoint of speech, may well extend over a number of data signal elements, many of which would then be received incorrectly; because of the highly redundant nature of human speech, they are of much less significance when the connection is used for telephony.

As neither the probable mean error rate nor the distribution of errors in time is deducible from the known characteristics of telephone circuits, the Study Group on Data Transmission of the Comité Consultatif International Télégraphique et Téléphonique included in its future studies a programme of error measurements designed to gather information on these points. The measurements were recommended to be taken over leased and switched circuits that are representative of the facilities likely to be available at the speeds and transmission levels envisaged for data-transmission systems. The measurement programmes undertaken by the companies of the International Telephone and Telegraph Corporation were on the basis of the guidance so laid down by the Comité Consultatif International Télégraphique et Téléphonique.

3. Transmission Methods

A number of different signal forms may be considered for the representation of data in course of transmission over telephone circuits. In serial or time-division transmission, the data are sent in a continuous stream, each bit being represented by a signal condition that persists for a certain time, the element period. In parallel or frequency-division transmission, a number of bits are sent simultaneously along relatively narrow-band channels occupying different parts of the frequency bandwidth of the telephone circuit. With suitable adaptive equipment the parallel method may be used whether the data are presented to the transmit terminal in parallel form or serially. Both for the single channel of serial transmission or the several channels of parallel transmission, different modulation methods are available; the method best suited to data transmission is one of the points currently under study.

3.1 Modulation

An important case is the modulating method for serial transmission over the entire bandwidth of the telephone-type circuit. The main possibilities are amplitude (double and vestigial sideband), frequency or phase modulation of a carrier of suitable frequency, and direct transmission of the binary signal without any modulating process. The judgment criteria are relative costs, efficient use of available bandwidth from the point of view of transmission speed, and susceptibility to errors in the presence of the kinds of disturbance likely to appear on telephone circuits.

As the binary signal will contain a direct-current component that is not transmissible over most modern telephone circuits, the last method (direct transmission of the binary signal) cannot be considered as it stands. However, an adaptation of this method, in which each binary digit is represented by two signal elements, eliminates the direct-current component. A system using this principle has been developed by Compagnie Générale de Constructions Téléphoniques in Paris for transmitting telemetering data and remote-control instructions over physical cable circuit private networks.[3] The terminal costs of such a system are low but, unfortunately, even small frequency shifts such as must be expected when transmitting over carrier-derived circuits give rise to errors so it is limited in its application to leased physical circuits. For general application over the switched and leased telephone networks, it is necessary to choose one of the modulating methods.

Amplitude modulation requires only simple terminating equipment but its performance in the presence of noise or when there are signal-level variations is known to be inferior to that of frequency or phase modulation. Of the modulating methods, vestigial sideband is the most economical of bandwidth but as part of the signal energy is taken up by the transmitted carrier and because of square-law distortion its error performance against noise is the worst. The choice seems to lie, therefore, between frequency and phase modulation; these methods are comparable as regards economy of bandwidth utilization, but phase modulation, which requires more complex equipment, usually has attributed to it a small theoretical advantage in terms of error rate for a given signal-to-noise ratio. The object of a measurement programme undertaken by Standard Radio & Telefon AB in Bromma, Sweden was to determine whether such advantage as may be realizable in practice justifies the extra equipment complications.[4]

The theoretical performance of a phase-modulation system in terms of error rate as a function of signal-to-noise ratio can be calculated and is well known. Unfortunately, frequency modulation is not so amenable to assessment by calculation. The measurement programme consisted, therefore, of measurements of element error rate for frequency-modulation systems of different frequency shifts, modulating rates, operating in different bandwidths, and against different levels of background noise. White noise was used as the source of interference. It seemed reasonable that a system optimized on the basis of white-noise performance would also be the optimum under actual telephone-line conditions. Also, this method facilitated comparison of the results with the theoretical predictions for phase modulation.

An analysis of the results shows that the performance of a frequency-modulation system may be optimized as regards error liability by suitable choice of frequency shift and modulating rate. For a given receiver bandwidth there is an optimum value of frequency shift that, for the same signal-to-noise ratio, gives a lower error rate than any other value. Further, there is an optimum modulating rate above which a penalty is incurred in the form of rapidly increasing error rate but below which rate the error rate is practically independent of speed.

In Figure 1 a curve relating error probability and signal-to-noise ratio for a frequency-modulation system, optimized in this way as regards modulating rate and frequency shift, is compared with the well-established theoretical curves for a phase-reversal system using an ideal phase reference and a differential coherent phase-modulation system in which the phase of the received signal during each unit interval is

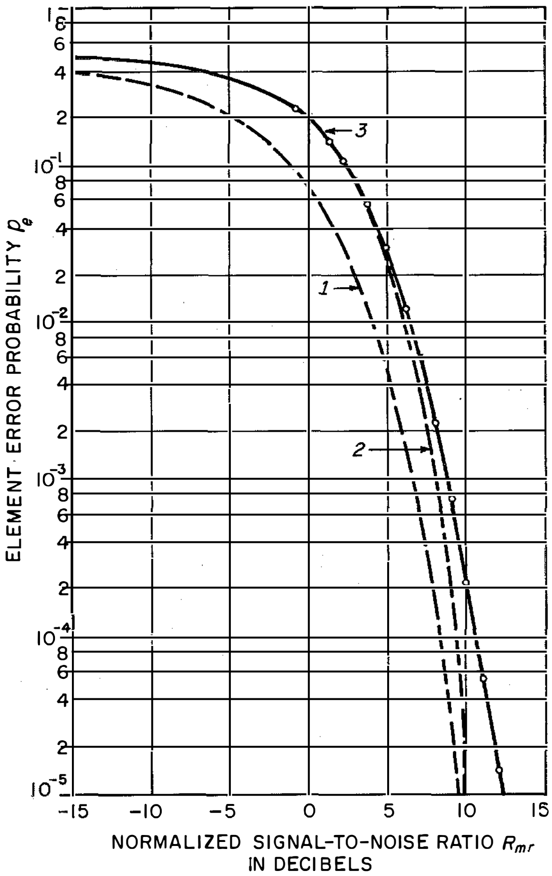


Figure 1—Probability of element error p_e versus normalized signal-to-noise ratio R_{mr}. Curve 1 is the theoretical optimum for phase-shift modulation with ideal phase reference. Curve 2 is for the theoretical optimum case of phase-shift modulation with differential coherent detection. Curve 3 gives the measured values for frequency-shift modulation.

compared with the phase during the preceding interval.

The practically achieved performance of the frequency-modulation system at low signal-to-noise ratios equals the theoretical performance of the differential phase-coherent system and is only some 3 decibels inferior to the theoretical performance of the phase-reversal system with ideal phase reference. In practice, of course, this theoretical performance is unattainable because of unavoidable disturbances of the phase reference and it becomes doubtful whether the theoretical superiority of the phase-modulation method is in fact realizable. Bearing in mind also the added cost and complexity of the phase-modulation system receiver, in which a phase reference has to be derived from the measuring signal, frequency modulation has been preferred and has been used in all the error measurement programmes described later.

3.2 Biternary Detection

"Biternary" detection is a method of detection for synchronous systems under development by Standard Telephones and Cable in London, which increases the optimum signalling speed by using two threshold signal values instead of one, the particular threshold used at any one examination instant depending on the previous signal condition. Three separate conditions are detected, mark, space, and zero, hence the name; a zero condition is interpreted as a mark if the previous signal condition was space and vice versa. The speed increase arises from the shorter interval required for a signal transition from a mark (or space) condition to zero than from mark to space. Biternary detection, at a conservative assessment, gives a speed increase of 50 per cent with substantially no change in error rate, that is, if, for a switched connection and a leased circuit, the transmission speeds using normal detection are 1200 and 2400 bauds the use of biternary detection should permit speeds of 1800 and 3600 bauds, respectively. Alternatively, it could be used to give effectively the same speed with a superior error rate resulting from the better signal-to-noise ratio of a narrow bandwidth.

3.3 Parallel Arrangements

For most applications of data transmission over telephone circuits the serial mode (time division) is to be preferred. The terminal equipment is relatively simple and there is greater flexibility both as regards speed adjustments and the form of binary code that may be used. However, the parallel mode has some advantages that make it more suitable for some kinds of applications.

The complexity of the parallel mode is primarily in the receiving equipment. In the data-gathering type of system the flow of data is from a large number of out stations, often in small quantities, into a central computing or data-processing centre. If the data are transmitted in a highly redundant code, permitting automatic error correction at the centre, or if backward repetition of received data for checking purposes is done verbally, from suitable recordings, the transmission of binary data in the outward direction can be avoided altogether. The cost of the one receiving equipment is shared among many stations and a very simple design of outstation equipment using a parallel mode of transmission becomes possible. Such a system is the "Digitel" developed by Standard Telecommunication Laboratories at Harlow.

Again, when data signals are passed over several relatively slow channels arranged side by side in a telephone-circuit bandwidth, nonlinearity in the wideband phase characteristic can be considered separately over each of the narrow-band channels in turn. For this reason a parallel mode of transmission can use effectively a larger part of the bandwidth of a connection, and give a higher speed in consequence, than a serial system. Because of the narrower filters the liability to error from impulse noise is less and the longer signal elements mean that reflections are less troublesome. Parallel systems, using perhaps six to twelve channels, each of several hundred bauds, are likely to find their best application over leased circuits.

4. Error-Measurement Programmes

4.1 Standard Telecommunication Laboratories[5,6]

The measurements by Standard Telecommunication Laboratories at Harlow were made over lines made available by the British Post Office, the test equipment, both transmitting and receiving, being installed at the Harlow Laboratories. Each connection measured was a loop which began and terminated at Harlow, an arrangement which avoided the difficulties of transporting equipment and personnel. The connections were established at British Post Office Laboratories at Crucifix Lane, London, and extended to the Harlow Laboratories over a four-wire circuit of zero equivalent in each direction. For all practical purposes these extensions were error free; in a test of the extensions alone and involving the transmission of nearly 15 million bits no errors began to appear until the transmitted signal level was reduced to -40 decibels referred to 1 milliwatt; even then with an element error rate of little more than 1 in 10^6.

The tests were carried out with data-transmission equipment using frequency modulation. The very earliest tests, those which took place in 1959, were at 250 bauds using a mean carrier frequency of 1800 cycles per second and a frequency shift of ± 100 cycles per second. Subsequent tests were at speeds of 800 and 1000 bauds, a mean carrier frequency of 1500 cycles per second and a frequency shift of ± 400 cycles per second.

The test message, which would be repeated a number of times in a single test, consisted of 2000 bits. One part of the message consisted of consecutive space elements, another of reversals, and another a random message.

Up to 12 May 1961 a total of 478 measurements over 119 different connections had been carried out, the number of bits transmitted exceeding $1 \cdot 4 \times 10^9$. The tests covered local connections within the London area, toll connections between 200 and 600 miles (322 and 965 kilometres) in length, and international connections, some to Paris, others to Stockholm. The majority of the connections were switched, but a certain number of leased (unswitched) circuits were included. The signal was transmitted at different levels both to ascertain the relation between error rate and level and also to produce at the last switch in the connection the sort of level that will apply in some of the near-limiting conditions expected to be encountered in practice. The maxima recommended by Study Group A (1 milliwatt at the input of the connection or −6dbm0 (decibels referred to 1 milliwatt measured at a point of zero relative level) were not exceeded in any instance. On a number of tests equalizing networks were used either in the sending (pre-emphasis) or receiving (post-emphasis) equipments to ascertain whether compensation for the loss–frequency characteristic of the connection improved the error rate.

The main purpose of the programme, however, was to gather information on the distribution-of-errors characteristic of telephone-type circuits. For this reason, errors were not only counted but the precise position in the message of each error was recorded in a way suitable for subsequent analysis by a computer. As each message was received it was compared, element by element, with the transmitted message and errors were recorded by punching holes in a separate error tape. The production of voluminous quantities of blank tape was avoided by arranging for this tape to step only from the first error of a cycle until the end of that cycle, when a number identifying that cycle was also punched. Nearly 70 000 errors were recorded in this way. A part of the error tape record is illustrated in Figure 2.

4.2 Standard Elektrik Lorenz[7]

The programme undertaken by Standard Elektrik Lorenz was the measurement of error rates and error distribution on switched connections in the local area of Stuttgart; its object was to ascertain whether data transmission is practicable over such circuits and, if so, the degree of error correction required.

The test equipment was a prototype system transmitting data in blocks of 63 bits at 800 bauds over a frequency-modulated channel of a mean frequency of 1800 cycles per second and frequency shift ± 300 cycles per second. Part of the line bandwidth was reserved for decision signals sent in the backward direction over two normal voice-frequency telegraph channels. Each block of the test message consisted of mixed 0's and 1's in approximately equal quantities. As with the Harlow tests

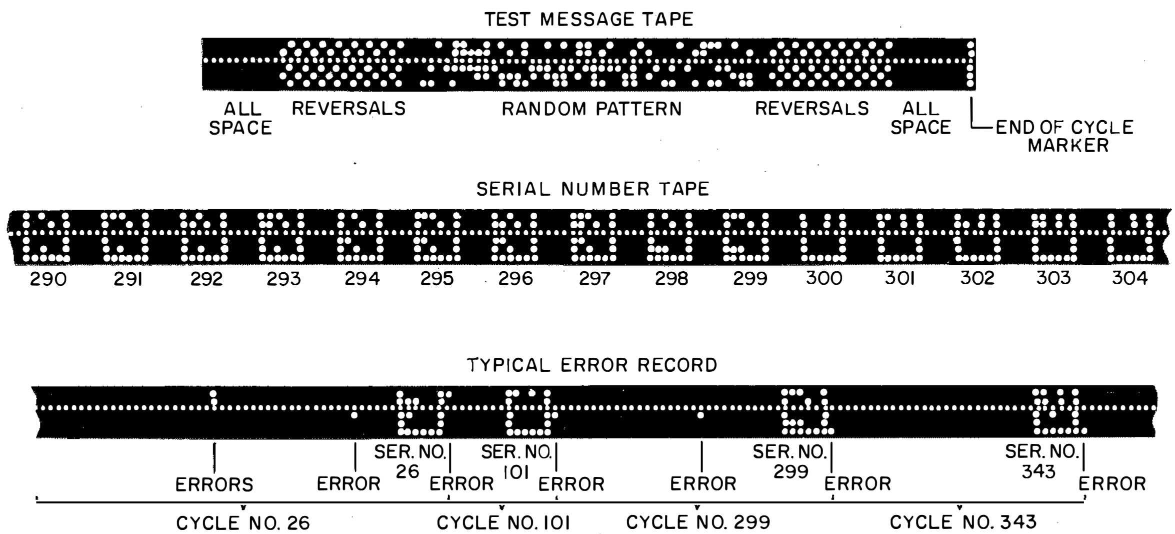


Figure 2—Test-message tape, serial-number tape, and typical error record for the error-measurement programme at Harlow.

measurements were made over loop connections so that transmitting and receiving equipments were at the same location. The maximum transmit level of the system was $-1 \cdot 0$ neper ($-8 \cdot 7$ decibels); the actual level used in the tests was $-2 \cdot 0$ nepers ($-17 \cdot 4$ decibels) giving a level of $-3 \cdot 6$ nepers ($-31 \cdot 3$ decibels) at the final exchange of the tested connections. Over 150 million bits were transmitted and measurements were made of the element error rate and of block error rates for the two cases of 63- and 9-bit blocks.

Recordings taken of the disturbances experienced on these connections were used to make further laboratory measurements over simulated line conditions in which the effect of varying line levels on element and block error rates and on error distribution was studied. Silent periods during which transmission was error-free were eliminated from these recordings so that the laboratory measurements were concentrated tests in which disturbances were present for most of the time. These tests involved the transmission of 458 million bits; because of the concentration of the tape recordings the laboratory tests corresponding to $7 \cdot 8 \times 10^9$ bits transmitted over actual line connections. A part of the recording of line disturbances is illustrated in Figure 3.

4.3 Element Error Rates

The error measurements were made over a large number of different kinds of connection and under a variety of conditions. The circuits tested included switched, leased, local, toll, and international circuits, and were variously derived from the loaded cables, multi-channel carrier systems, et cetera. The tests were made at different speeds, at different times of day, and the transmit levels were varied so as to reproduce in the telephone network signal levels of the range that would be encountered in practice. It is to be expected that there were wide differences also as between one test and another in the errors measured.

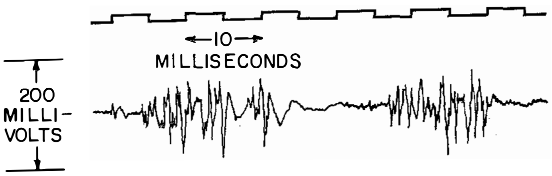


Figure 3—Typical noise on a local telephone connection.

The average element error rate for tests involving the transmission of over 500 million bits over leased connections was less than two errors in 10^5 elements. For toll switched connections the average probability of error was higher, about 7 errors in 10^5 elements. Because the transmit levels had to be reduced on many tests to induce errors, these figures are probably a pessimistic indication of the general quality of the public network for the transmission of data. On numbers of tests, each involving the transmission of perhaps a million or so bits, transmission was completely free of error; on others, disturbances caused hundreds of errors in a single test. Indications of the average performance of a network are, therefore, not too meaningful; design of a data transmission system should be based on the performance to be expected under near-limiting conditions. It would be little compensation to some unfortunately placed subscriber consistently experiencing high error rates on his calls to be told that the average error rate over the network was much lower.

When results of tests made over similar routings and at the same levels are compared, some dependence of error rate on the time of day is indicated; this dependence follows the well-known pattern of daily fluctuations of telephone traffic. The error rate at the busiest and quietest daytime periods may differ by a factor of 3 or 4. The correlation tends to be upset, however, by disturbances during the less-busy periods, which are probably attributable to maintenance operations.

Variations of speed have little effect on error rate, confirming the laboratory white-noise measurements by Standard Radio & Telefon, Sweden. So long as the speed is not excessive in relation to the effective bandwidth of the connection, the larger number of errors caused by disturbances at the higher transmission speeds are compensated by the extra bits transmitted during disturbance-free periods.

The most powerful influence on error rate is the signal level at the receiving end of the connection. Because of system loading and crosstalk considerations, Special Study Group *A* has already recommended limits of -6 dbm at a zero relative point for the signal level on a switched connection and 1 milliwatt at the input into a subscriber's line. On some connections the loss between subscribers will be as high as 30 or 40 decibels. Taking into account the contribution of subscribers' lines to the overall loss of a connection the signal level at the final switch in a switched connection approaching these limiting conditions could reasonably be assumed to be in the range -31 to -34 decibels referred to 1 milliwatt.

Of the tests taken at Harlow a number, involving 64 million bits at 1000 bauds, were over switched toll connections in which the level at the last switch was -34 decibels referred to 1 milliwatt. The average element error rate for these tests was just over $1\frac{1}{2}$ errors in 10^4 bits. The measurements at Stuttgart were confined to switched calls over local connections, which were not set up randomly but deliberately chosen so as to involve exchanges known to be much noisier than average. The element error rate for tests involving 150 million bits at 800 bauds and with a signal level at the final switch of $-31 \cdot 3$ decibels referred to 1 milliwatt was 1 error in 10^3 bits.

The use of pre-emphasis to raise the level of the higher transmitted frequency at the expense of the lower frequency, maintaining the same mean transmitted power, had a beneficial effect on error rate. When the discrimination in favour of the higher frequency amounted to 6 decibels the error rate improved by a factor of about 3. It was found that even better results follow from carrying out part of the equalization, at least, at the receiving end (post-emphasis) where the noise as well as the signal can be influenced.

4.4 Block Error Rates

The measurement programmes were designed to give information not only on overall error probabilities but also on their distribution in time. If errors are well spaced the error rate over short periods of time will differ very little from the overall mean. If they tend to occur in bunches very-high error rates are experienced for short periods of time for the same long-term mean. The distribution of errors is clearly relevant to the design of error-correction equipment. It was studied by dividing transmitted data into blocks of different sizes and examining the incidence of blocks containing errors. The measured block error rates in different series of tests made at Harlow and Stuttgart are given in Table 1 together with the measured element rate and a predicted block error rate calculated from the element error rate and an assumed random distribution of errors.

Comparison of the actual and predicted block error rates shows that errors tend to be more bunched together, and conversely that more blocks are completely error-free, than would be expected from a random distribution. Leased circuits have lower error rates than switched but the errors that do occur are even more tightly bunched than those on switched circuits. Even under limiting transmission conditions a very high proportion of blocks, about 98 per cent, will get through without errors.

The general conclusion that can be drawn from the results of these measurement programmes is that the European switched- and leased-circuit telephone networks are suitable media for the transmission of data. Over switched connections, transmission at speeds of the order of 1000 bauds is practicable, whilst over leased circuits much higher speeds should be attainable. For long periods transmission at these speeds can be expected to be error-free; even under near-limiting conditions of overall loss of

TABLE 1
BLOCK ERROR RATES

Total Bits and (Bits Per Block)	Circuits and (Level at Final Switch in Decibels Referred to 1 Milliwatt)	Mean Element Error Probability	Block Error Rate	
			Predicted for Random Distribution	Actual
191×10^6 (50)	64 Switched Toll Connections (−20 dbm)	$3{\cdot}15 \times 10^{-5}$	$1{\cdot}58 \times 10^{-3}$	$0{\cdot}74 \times 10^{-3}$
71×10^6 (50)	30 Switched Toll Connections (−28 dbm)	$7{\cdot}67 \times 10^{-5}$	$3{\cdot}84 \times 10^{-3}$	$2{\cdot}43 \times 10^{-3}$
64×10^6 (50)	28 Switched Toll Connections (−34 dbm)	$15{\cdot}3 \times 10^{-5}$	$7{\cdot}65 \times 10^{-3}$	$3{\cdot}96 \times 10^{-3}$
491×10^6 (50)	106 Leased Toll Connections	$1{\cdot}74 \times 10^{-5}$	$0{\cdot}87 \times 10^{-3}$	$0{\cdot}24 \times 10^{-3}$
151×10^6 (63)	Local Switched Connections (−31·3 dbm)	120×10^{-5}	$75{\cdot}6 \times 10^{-3}$	$21{\cdot}0 \times 10^{-3}$
$5{\cdot}3 \times 10^6$ (63)	With tape-recorded noise. Sending level adjusted to final-switch level of			
	(−22·6 dbm)	$11{\cdot}0 \times 10^{-5}$	$6{\cdot}93 \times 10^{-3}$	$4{\cdot}80 \times 10^{-3}$
	(−38·2 dbm)	310×10^{-5}	195×10^{-3}	$29{\cdot}0 \times 10^{-3}$

The measurements of the first 4 entries were made at Harlow and those for the final 2 entries were made at Stuttgart.

the connection the long-term mean error rate is unlikely to be more than 1 error in 1000 bits. High concentrations of errors are liable to occur over short periods, however, and these have to be taken into account in any arrangements to reduce the error liability.

5. Error-Correction Methods

In data-transmission systems used for radar and telemetering each item of data is quickly superseded by later information; it is pointless to correct erroneously transmitted data if this involves postponing more up-to-date information. In other applications the data will itself contain redundancy permitting the correction of errors during processing after reception. The natural error rate of the network will often be acceptable in such cases. However, when the received data are not subject to automatic renewal or human inspection, or contain no redundancy, means of correcting most of the errors that are liable to occur during transmission become necessary.

Extremely low error rates after correction could be achieved with suitably elaborate error-correction equipment. The question is what performance the user needs and what he is willing to pay.

An element-error reduction of 10^4 would give, under near-limiting conditions, an overall element error rate of about 1 in 10^7 and a character error rate somewhat better than 1 in 10^6. Bearing in mind the error liabilities in parts of data-handling systems other than the transmission link, and that for many data calls the overall performance will be much better, this seems a reasonable objective.

There are three methods by which errors might be corrected.

(A) By an error-correction code that identifies and corrects those bits received in error.

(B) By an error-detection code that detects the presence of errors without establishing the particular bits in error. The blocks in which the presence of errors has been detected are repeated.

(C) By a "trouble detector" that detects the presence of transmission disturbances likely to cause errors. The blocks received when these disturbances have been present are repeated.

Both the methods involving retransmission of blocks require a backward channel for sending information back to the transmitter. Also, as any disturbance on the line provokes a repetition the flow of information must be discontinuous with these methods; this requires more-complex terminal equipment. Correction by error-correction code, on the other hand, involves forward transmission only (simplex operation), the information flowing continuously.

Error-correction codes require a percentage of check or redundant bits in relation to information bits that depends on the highest concentration of errors to be detected; on the other hand, error-detection codes of relatively much lower redundancy detect large numbers of errors as easily as small. In practice, error-correction codes capable of dealing with the concentration of errors likely to occur in course of transmission over telephone lines require such a high percentage of redundant bits as to reduce very significantly the effective rate of transmission of information bits, even when transmission is substantially error-free. Further, when error bursts occur that are beyond the capacity of the error-correction code, errors get through to the receiving data equipment. However, the effect of such bursts on systems using detection codes is merely to postpone the arrival of correct data. For these reasons, error correction by the detection and repetition of faulty blocks is more suitable for data transmission over telephone circuits.

5.1 Error-Detection Codes[8,9]

Blocks containing errors can be identified at the receiving terminal of a data-transmission system by the inclusion within the block of a number of check bits; these check bits occupy predetermined positions within the block and have values dependent on the information bits included in the same block. If the dependence of the values of the check bits on the information bits does not hold in the received block, then the block is recognized as one containing errors and is repeated. In general, the greater the number of check bits in relation to the information bits, that is, the higher the redun-

Simple Parity Code

0	0	0	0	0	*1*
0	0	0	0	1	*0*
0	0	0	1	0	*0*
0	0	0	1	1	*1*
0	0	1	0	0	*0*
0	0	1	0	1	*1*
0	0	1	1	0	*1*
0	0	1	1	1	*0*
0	1	0	0	0	*0*
0	1	0	0	1	*1*

Constant-Ratio Code

0	0	0	0	1	1	1
0	0	0	1	0	1	1
0	0	0	1	1	0	1
0	0	0	1	1	1	0
0	0	1	0	0	1	1
0	0	1	0	1	0	1
0	0	1	0	1	1	0
0	0	1	1	0	0	1
0	0	1	1	0	1	0
0	0	1	1	1	0	0

Hamming Code

0	0	0	*1*	0	*1*	*1*	*0*
0	0	0	*1*	1	*0*	*0*	*1*
0	0	1	*0*	0	*1*	*0*	*1*
0	0	1	*0*	1	*0*	*1*	*0*
0	1	0	*0*	0	*0*	*1*	*1*
0	1	0	*0*	1	*1*	*0*	*0*
0	1	1	*1*	0	*0*	*0*	*0*
0	1	1	*1*	1	*1*	*1*	*1*
1	0	0	*0*	0	*0*	*0*	*0*
1	0	0	*0*	1	*1*	*1*	*1*

2-Coordinate Code

0	0	0	0	1	1	1
0	0	0	1	0	1	1
0	0	0	1	1	0	1
0	0	0	1	1	1	0
0	0	1	0	0	1	1
0	0	1	0	1	0	1
0	0	1	0	1	1	0
0	0	1	1	0	0	1
0	0	1	1	0	1	0
0	0	1	1	1	0	0
1	*1*	*1*	*1*	*1*	*1*	*1*

Figure 4—Examples of simple error-detection codes. The Hamming code contains four parity words in each row as follows, the bits being numbered from the right:

(1) Bit 1 is a parity bit for the complete row.
(2) Bit 2 is a parity bit for bits 4, 6, and 8.
(3) Bit 3 is a parity bit for bits 4, 7, and 8.
(4) Bit 5 is a parity bit for bits 6, 7, and 8.

The 2-coordinate code is a combination of the simple parity and constant-ratio codes.

These codes are basic arrangements which can be expanded to form more-sophisticated codes; for example, the Hamming code can be modified to provide more security against adjacent errors. Other codes similarly providing protection against bursts of errors use two or more sets of interleaved parity words.

TABLE 2
PERCENTAGE OF ERRORS REMAINING UNDETECTED FOR ROW- AND COLUMN-PARITY DETECTION CODES, MEASURED AND ESTIMATED

Block Size and Percentage Redundancy	Percentage of Undetected Errors			Sample
	Estimated with Equiprobable Distribution within Blocks	Measured		
		Row Parity	Column Parity	
50-Bit Blocks 10-Per-Cent Redundancy	5.03×10^{-2}	15.5×10^{-2}	4.1×10^{-2}	21 271 Errors from 122 Tests (326×10^6 Bits) at 1000 Bauds Over Switched Circuits
100-Bit Blocks 10-Per-Cent Redundancy	9.6×10^{-3}	9.4×10^{-2}	5.15×10^{-3}	8070 Errors from 103 Tests (480×10^6 Bits) at 800 Bauds Over Leased Circuits

dancy, the more surely will erroneous blocks be recognized, but this is at the expense of fewer information bits per block and, therefore, a reduced rate of information flow for the same baud speed.

There are numerous possible ways of arranging the dependence of check bits on information bits; some of the simpler arrangements are given in Figure 4. In the simple row-parity code each check bit is given a value, 0 or 1, such that the number of 1's in each row will be odd, that is, odd parity check. The check bits could as well be in the last row, rather than in the last column, in which case the value of each check bit depends on the information bits in the same column. In the constant-ratio code each line consists of the same numbers of 0's and 1's; a block is identified as containing errors when one or more of the lines in the same received block contain a number of 1's different from that permitted by the code. Each of these codes detects the presence of any single or odd number of errors in a block, but two or larger even numbers of errors in one block may form an undetectable combination; this is the case when two errors occur in the same parity word, for example, a row of the row-parity block. With the constant-ratio code the two errors in a block can form an undetectable combination only when they are in the same row and of opposite kind, that is, one of the errors has changed a 0 into a 1 and another a 1 into a 0.

For any code it can easily be calculated, for each number of errors per block, how many different error patterns can occur and the proportion that are undetectable. If it is assumed that one pattern is as likely to occur as any other and if the distribution of errors per block, that is, the proportions in which errors fall, 1, 2, or 3, et cetera, to a block, are known from line measurements, then the error-rate improvement effected by a code can also be calculated. It is obvious that on this basis row- and column-parity codes give the same calculated improvement. Table 2 compares the calculated with the measured undetected error rate for these codes, taking into account error patterns as they actually occurred in line tests. The measured rates were obtained by computer analysis of error tapes produced in course of the measurement program at Harlow and the "errors-per-block" distribution for the estimated improvement was obtained from the same error record.

The comparison shows that a code may have an error-detection capability either worse or better than that estimated depending on whether the undetectable patterns it admits are more or less likely to occur in practice than when errors are randomly distributed within the block.

The measured error-rate improvements effected by several relatively simple codes applied to different block sizes are given in Figure 5. For an error improvement of the order of 10^4, however, a more complex code is necessary. Such a code under examination is a $3 + 5 + 7$ system of interlaced sets of parity words, all regularly spaced. The first set of three parity words for a 105-bit block is arranged so, the check bits being italicized.

1, 4, 7, 10 . . . 100, *103*
2, 5, 8, 11 . . . 101, *104*
3, 6, 9, 12 . . . 102, *105*

The second and third sets of 5 and 7 parity words respectively are arranged similarly, bits numbers 91 through 105 of the complete block being check bits.

Many codes can be constructed in this way, differing from each other in the number of interlaced sets of parity words, the spacings between bits of the same parity word and, consequently, in the percentage of redundancy introduced into a block, and in error-detection capability. The bits of a parity word may be equally spaced, as in the $3 + 5 + 7$ code; in other codes, as in the so-called "scrambled binary," the bits are spaced irregularly. In addition there are methods under study of using the information bits of a block to improve the security of a detection code.

A failure of a more complex code, such as one involving 3 sets of parity words, to detect errors occurs when some even number of errors, 4, 6, 8, 10, et cetera, form within a block a pattern that defies detection by each set of parity words. If the undetected error rate is to be as low as 10^{-7}, an undetectable combination can be admitted only once or so in 10^5 or more errors, or only once in about every 10^8 transmitted bits. It is clearly impracticable to undertake measurements of line errors on the necessary scale to record undetectable patterns in sufficient numbers to compare the capabilities of different coding arrangements. For this reason the Harlow Laboratories have evolved a method of generating typical error patterns artificially. The error tape records of actual line measurements have been analysed according to the numbers of correctly received bits separating errors so as to give the error separation curve of Figure 6. Using this curve in conjunction with a series of random numbers generated by a computer, error sequences having separations typical of those occurring on telephone connections can be generated. Unwanted information, such as odd numbers of errors per block, is discarded at the time of generation so that a concentrated and sufficiently large sample of a given number of errors per block is obtained for any block size.

Using such a source of errors the error-detection capabilities of different possible codes are being measured and compared on the basis of error distributions characteristic of the disturbances liable to occur on telephone circuits. The outcome should be the choice of a code

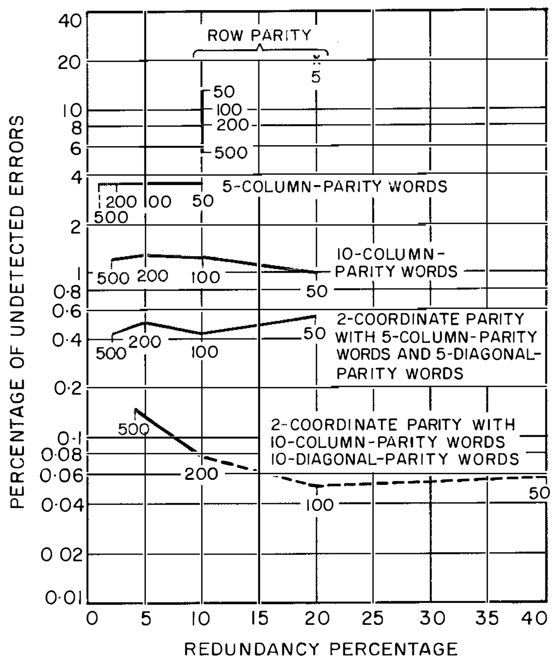


Figure 5—Error-rate reduction factors of selected error-detection codes for 1000-baud transmissions. The numbers 5 through 500 indicate the block sizes.

leading to an error-control system of simple design, adequate error-detection capability, and economical in the number of check bits in relation to information bits.

5.2 Error Control Using a Trouble Detector

Error-detection codes function by virtue of the redundant information contained in each block. As errors are caused by loud noises and circuit breaks, or level variations, an alternative possibility for the correction of blocks containing errors is the detection of disturbances capable of causing errors, blocks received when such

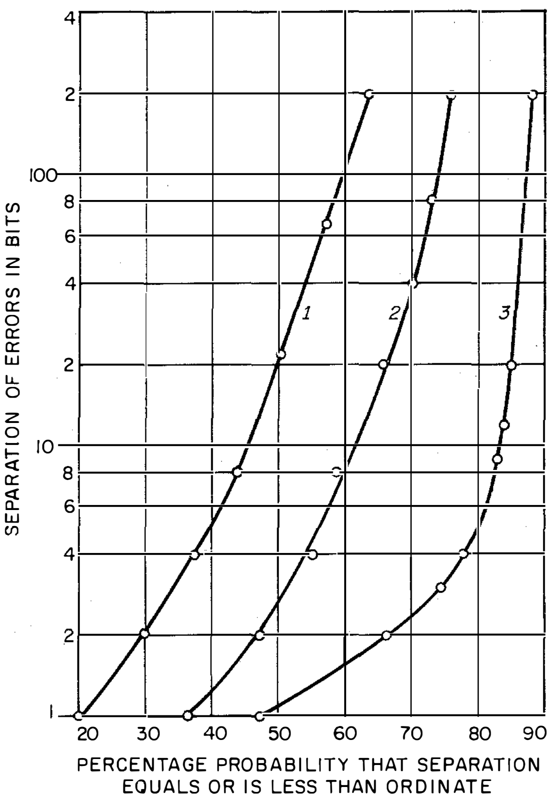


Figure 6—Distribution of error separations. For $F =$ the total number of errors in the sample and $p_e =$ the rate of element errors, curve 1 is for switched connections at 1000 bauds with $F = 15\,897$ and $p_e = 5{\cdot}9 \times 10^{-5}$. Curve 2 is for switched connections at 250 bauds with $F = 5367$ and $p_e = 8{\cdot}25 \times 10^{-5}$. Curve 3 is for leased connections at 1000 bauds with $F = 3512$ and $p_e = 1{\cdot}3 \times 10^{-5}$.

disturbances have occurred being retransmitted. Possible methods of detecting disturbances are the monitoring of a parallel channel, not used for data signals, amplitude detection of a frequency-modulated signal, measurement of the characteristic distortion of the received signal, et cetera.

Studies of the combined use of various types of trouble detectors and an error-detection code have been made by Standard Elektrik Lorenz [10] and by Standard Radio & Telefon. Block repetition is initiated when either the trouble detector operates or the parity check of an error-detection code fails. The advantage of such a combined arrangement is that a relatively high improvement in the error-rate performance can be achieved with error-detection codes of low redundancy.

The penalty incurred for this improvement is a liability to unnecessary repetition of error-free blocks; this is due to the inability of the trouble detector to distinguish disturbances that have caused errors from those that have not. A disturbance capable of causing an error does not necessarily cause one. Both the improvement in error rate attributable to the trouble detector and the percentage of blocks unnecessarily repeated depend on the setting chosen for the detector threshold and on the block size.

The most promising application of the trouble detector is in data-transmission systems using very small block sizes, perhaps 10 bits. Such systems have the advantage of needing little information storage capacity in the terminal equipment but, because of the small block size, require much more highly redundant error-detection codes, for the same error-rate improvement, than do systems using large block sizes. Figure 5 demonstrates this. The introduction of a trouble detector has a correspondingly more beneficial effect on the percentage of information bits per block.

6. Block Size

The choice of block size is important both as regards the effective rate at which information

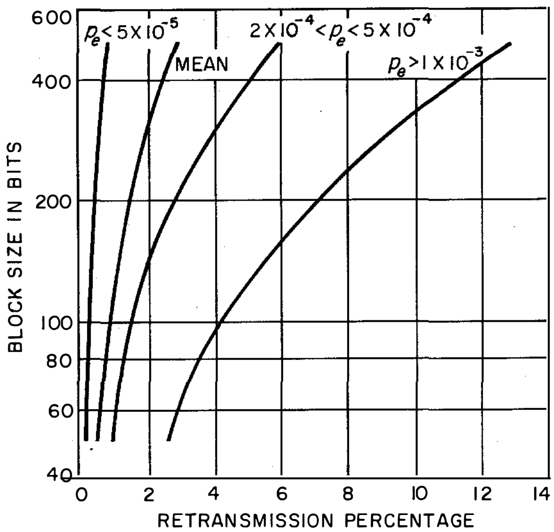


Figure 7—Percentage of blocks requiring retransmission as a function of block size and element error probability based on tests on switched connections at 1000 bauds. Retransmission always involves two blocks, the block in error and the following block.

bits are transmitted and on terminal-equipment costs.

The smaller the block the less is the percentage of blocks requiring repetition. Figure 7 shows how the percentage of blocks retransmitted varies with the block size for different element error probabilities. The curves were obtained from separate analyses, according to the assumed block size, of error tapes obtained from tests over switched connections. The curves allow for the repetition both of the faulty block and that transmitted immediately afterwards, whether it contains errors or not, as would be done in a practical duplex system. The percentage of retransmitted blocks remains small, less than 1 per cent, even for large block sizes when the overall element error rate is low; for element error rates of the order of 1 in 10^{-3} however, it becomes increasingly significant for blocks larger than 100 bits.

Another advantage of small block size also is that it fits more conveniently into different message lengths. A system based on automatic transmission of standard block lengths handles complete blocks only; if there are insufficient data to fill a block, then the block has to be completed with artificial data, usually all 0's or 1's. A large block size, of the order of 1000 bits perhaps, is clearly a disadvantage for a system used primarily to transmit short messages.

An error-control system based on the repetition of faulty blocks involves the storage of each transmitted block at the sending terminal until it is known whether its repetition is required or not. The smaller the block size the less significant, in relation to the total cost of the system, will be the expense of providing such storage facilities. In systems using electronic-type stores, which are not only expensive but increase in cost in proportion to the capacity of the store, the cost of storage will usually be the dominant consideration affecting block size.

However, other considerations tend to favour the larger block sizes. The error-rate improvement brought about by a code of a given percentage redundancy is much less for small than for large blocks. In each of the spaced parity error-detection arrangements of Figure 5, the number of check bits remains the same for each of the different block sizes; as the block size increases the percentage redundancy reduces in proportion, but the effect on error-rate improvement is relatively small.

The design of a data-transmission system using block repetition is simplified if it can be ensured that the need for repeating a block is known at the transmitting end before the transmission of the next block is completed. If, because of propagation delays, information received from the receiving terminal does not necessarily relate to the most-recently completed block but instead to the first, second, or third preceding block possibly, then the terminal equipment must be made capable of determining from which block repetition is to begin.

When the information returned by the receiving terminal is a straight-forward acceptance or rejection of a block, the interval, at the transmitting terminal, between the end of the transmission of a block and the recognition of a request for repetition, will be at least twice the propagation time for the connection, to which are added processing delays at the terminals. Such a system would be one in which the redundant information is transmitted over the forward channel, with the information bits, the parity checks being carried out at the receiving terminal. An alternative method is to transmit only the information bits forward, the receiving terminal calculating redundancy and transmitting these bits over the return channel. In this case the delay is twice the propagation time and processing delays as before, plus the transmission time for the number of redundant bits per block at the speed of the return channel. This transmission time is likely to be quite significant as the major part of the available channel bandwidth will be used for the forward transmission.

The limits at present recommended by the Comité Consultatif International Télégraphique et Téléphonique for group delays on continental telephone connections are 150 milliseconds for the international part of the connection and 50 milliseconds for each national part. However, the present extensive use of high-velocity plant in Europe makes it unlikely that the two-way propagation time of 80 milliseconds is frequently exceeded. Where longer times are encountered, on intercontinental connections for example, other complications such as the likelihood of a time-assigned-speech-interpolation (TASI) type circuit being involved in the connection, will probably require special arrangements for data transmission. A convenient block size would be one ensuring that information sent back to the sending terminal is received before the transmission of the next block is completed for all continental connections. For a 1000-baud system this condition requires a block size of 80 bits at least, even for a system transmitting the redundancy forward.

7. Conclusions

The suitability of the switched and leased telephone networks for data cannot be deduced from the design characteristics of telephone circuits and has had to be determined by means of extensive programmes of practical measurement. These have confirmed that the telephone networks are suitable media for data transmission; even under conditions of overall loss, et cetera, approaching the limits that might be encountered in practice, the error rate and the error distribution are such as to permit, with suitable error-control equipment, the transmission of data with error probabilities in an order acceptable to the users of data-processing equipments.

Much of the data transmitted over telephone circuits, particularly over leased circuits, and probably over switched circuits also, will be between separated locations of the same organizations. It may be arguable, therefore, that compatibility between data handling and transmitting installations at different locations is necessary only within organizations. However, as automation and data-processing methods spread to smaller establishments, the demand for the shared use of centrally located computers and data-processing equipment will make general compatibility a more pressing requirement. This general compatibility will require, in the data-processing field, standardization of machine language, that is, the alphabet and code, and, in the transmission field, of the modulation method, operating frequencies, and speeds.

Another feature requiring standardization would be the method of error control. The effectiveness of different possible methods varies over wide limits according to the burst characteristics of the errors to be corrected. It is appropriate, therefore, that this standardization should be effected by the Comité Consultatif International Télégraphique et Téléphonique on the basis of the line error characteristics determined in the measurement programmes that have been carried out by its members.

This standardization should cover the error-detection code and the block size to be used in an error-control system based on the repetition of blocks containing errors.

For the important case of serial binary transmission at medium speed over switched telephone circuits frequency modulation is considered to be the most effective, paying due regard to equipment, cost, and complexity.

8. Acknowledgments

The error-measurement programme could not have been undertaken without the willing cooperation of several of the telecommunications administrations of Europe, namely, those of Great Britain, France, Sweden, and Western Germany, in making circuit facilities available for test purposes. This cooperation is gratefully acknowledged.

This paper is a review of a coordinated programme of work undertaken by a number of research engineers in different companies of the International Telephone and Telegraph System; the author did not himself participate in that programme. His indebtedness to colleagues for their help in the preparation of the paper is therefore much more than is ordinarily due from an author; this help also he gratefully acknowledges. The colleagues in question are too numerous to mention individually; they include all the authors mentioned in the list of references.

9. References

1. Report of Meeting of CCITT Working Party 43 (Data Transmission) Geneva, 21–28 March 1960; Comité Consultatif International Télégraphique et Téléphonique Document GT 43, Number 20; 30 March 1960.

2. E. P. G. Wright, "The Cause, Prevention and Correction of Errors Recorded During Data Transmission Tests." Paper presented at the Seventh Annual International Congress on Electronics, Nuclear Energy, Television, Radio, and Cinematography; Rome, June 1960.

3. A. Girinsky and P. Roussel, "High-Speed Transmission of Numerical Data Over Telephone Channels," *Electrical Communication,* volume 36, number 4, pages 242–262, 1960.

4. M. Jeppsson, "Frequency-Shift Modulation of Binary-Coded Signals for Transmission Over Telephone Circuits," *Electrical Communication,* volume 37, number 2, 1961.

5. E. P. G. Wright and A. D. Marr, "Error Distribution Statistics Resulting from Tests in Progress on Switched Telephone Connections Used for Medium Speed Data Transmission," Standard Telecommunications Laboratory Technical Report 112, March 1960.

6. E. P. G. Wright, "Error Rates and Error Distributions on Data Transmitted Over Switched Telephone Connections." Paper presented to the International Symposium on Data Transmission organized by the Benelux Section of the Institute of Radio Engineers, Delft, September 1960.

7. H. Marko, "Error Measurements with a Data Transmission System with Error Detection and Error Correction on Switched Lines," Standard Elektrik Lorenz Technical Report MS/GRÜB/L, April 1961.

8. T. A. Maguire and E. P. G. Wright, "The Examination of Error Distributions for the Evaluation of Error Detection and Error Correction Procedures." Paper presented to the International Symposium on Data Transmission organized by the Benelux Section of The Institute of Radio Engineers, Delft, September 1960.

9. E. P. G. Wright, "Data Collection and Transmission," *Computer Journal,* volume 4, number 2, pages 103–109, July 1961.

10. H. Lange, "Reduction of Error Probability by Combined Application of an Error Detecting Code and a Trouble Detector." Standard Elektrik Lorenz Technical Report MS/GRÜB-61-1, April 1961.

11. H. Marko, "Data Transmission by Telephone Lines," Nachrichtentechnische Fachberichte, Band 19 (1961), Verlag Friedr. Vieweg und Sohn, Braunschweig.

Frequency-Shift Modulation of Binary-Coded Signals for Transmission over Telephone Circuits

M. JEPPSSON
Standard Radio & Telefon AB; Bromma, Sweden

1. Introduction

Automation of business processes has created a demand for rapid and accurate transmission of engineering, scientific, statistical, and accounting information. Although enormous quantities of information are being moved from point to point every day by conventional telegraphy, many present and future data-transmission requirements cannot be met at normal telegraph speeds. To increase the speed, a variety of transmission systems have been proposed. We will be concerned here with a range of systems based on frequency-shift modulation.

A complete data-transmission system consists of a telephone circuit connected at each end to modulating and demodulating equipment. Error-detection and -correction facilities may also be included as well as input and output devices, such as tape readers and punches. However, the present paper is concerned only with the modulator, demodulator, and line—the modem link.

A variety of important questions arises in choosing an appropriate type of modulation. Some views will be given concerning different modulation methods with special attention drawn to frequency-shift modulation. The influence of nonideal behavior of the transmission medium will be discussed, results of measurements will be analysed, and finally some proposals as to standardization of modem equipment will be made.

2. Theoretical Analysis

2.1 Modulation Methods

Frequency-shift modulation will be compared to phase-shift modulation. The latter may be divided into a variety of types, the most important being two-phase, four-phase, and vestigial-sideband modulations. It is interesting to note that two-phase modulation is identical to on-off amplitude modulation with suppressed carrier.[1] Amplitude modulation with carrier will not be discussed here as it is generally agreed to be more sensitive to noise and level variations than frequency- or phase-shift modulations, on which latter two there are differing opinions.

2.2 Characteristics of Transferred Signals

There are various characteristics of the signals that must be distinguished at the receiving end of a data-transmission system. The most obvious one is the condition of modulation, which has to be examined by the receiver at certain instants. For this purpose, a timing signal must be developed. Furthermore, the start of characters and words has to be recognized to reproduce the message accurately. In phase-shift systems, another quantity that must be produced is an appropriate phase reference. This additional requirement is an inherent disadvantage to all types of phase-shift-modulation systems, as is well known.

The timing signal can be developed in the receiver either on a start-stop or on a synchronous basis. For continuous transmission of data, synchronous operation is preferred, as it is more immune to various types of disturbance. The following discussion is based on the synchronous case.

Synchronous timing can be obtained in various ways, such as by using the transitions between different conditions of modulation of a synchronizing signal transmitted at regular intervals. It is preferred, however, to use all transitions of the received stream of data signals because element synchronization can then be performed with a minimum of restrictions on the code of the data signals. The only requirement is that a minimum number of transitions be transmitted in a specified interval of time to maintain synchronism.

It is preferable that the mark-space decisions based on the pulsed output of the detector should take place within the modem-link terminal. This would permit the received signal characteristics at the modem output to be simply defined in terms of speed, error probability, error distribution, et cetera, rather than in terms of time distortion of various kinds, for example, bias distortion, characteristic distortion, fortuitous distortion, conventional degree of distortion, et cetera. In applications where the error-rate requirement is not critical and no error correction is required, the output of the demodulator provides the only possible junction with the receiving data-processing equipment, so it is particularly desirable that the signal characteristics should be clearly and simply defined at this point. It follows that the receiving timing signal be produced within the modem equipment. It may be possible to substantially simplify the circuits of the equipment that is to be connected on the receive local side of the modem link by extending this timing signal to it.

Another advantage of performing this timing-signal generation and of making the mark-space decisions in the modem equipment is that the correction of bias distortion by means of a feedback loop to the voice-frequency signal detector is facilitated. When the signal has bias distortion, the examination instants have to be very accurately phased with the incoming signals; even when this is done successfully, bias distortion reduces the margin against errors. It is extremely desirable therefore that any bias should be corrected and it seems impossible to do this except by a feedback loop as described.

At the send end, the modulation rate must be arranged to be compatible with the timing circuits of the receiver. For this purpose, a send timing signal must be generated by a clock signal generator. It is desirable that this generator should likewise form part of the modem link so that it, the frequency shift, and possibly the channel mid-frequency and filtering may be conveniently adjusted in those cases where speed adjustments are made to accommodate the system to the line characteristics.

Character, word, and block synchronization can be achieved either by an auxiliary modulation or by transmitting a coded start signal. The first method could be attained in several ways, for example, through the use of an additional frequency or phase modulation or by an additional amplitude level. Neither way is recommended for general use, because of excessively increased complexity of equipment. Some other disadvantages are also inherent in this method. If a coded start signal is used, the identification of character, word, and block is not a problem that concerns the modem equipment.

2.3 Definition of Problem

It is assumed that a continuous stream of signal elements is received and the distribution of both marks and spaces is random. The probability of a mark or a space being transmitted is the same. Time-distortion analysis is not of the same importance as in the voice-frequency telegraph case because the modem link as conceived in 2.2 above provides a regenerated output. Where this is not the case, it is presumed that regenerative repeaters would be used if modem links are connected in tandem.

Generally a complete setup for data transmission will include error-correcting equipment, but it is of major importance that the modem link has as low an error probability as possible. Before considering how the various parameters involved should be chosen to minimize error probability, it is advisable to investigate all the pertinent variables. For a specified communication medium and modulation rate, the error probability can be expressed as a function.

$$p_e\,(a_T, a_L, a_{R1}, a_{R2}, b_T, b_L, b_{R1}, b_{R2}, s_L, d_L, n_L, \Delta f)$$

where a_i is the modulus and b_i is the argument of a complex transfer function $a_i \exp [jb_i]$, where index i stands for transmitter T, line L, predetection receiver $R1$, or postdetection receiver $R2$. The voice-frequency signal level received from the line is represented by s_L, the intermodulation-products level by d_L, and the noise level by n_L. These three levels received from line are here defined as absolute levels relative to 1 milliwatt of power. The last variable Δf is the shift in a frequency-shift-modulated system.

Most of the above-listed variables could be chosen optionally. However, some variables have to be accepted as being uncontrollable, the noise level n_L being one, while d_L can be controlled only indirectly by varying the signal level s_L. The transfer function of the line (a_L, b_L) is controllable to a great extent by varying the midfrequency of a channel and in special cases by equalization. As variables in time, neither a_L nor b_L can be controlled, though there are possibilities for compensation in the receiver. However, when a_L drops to zero as in line breaks, no means for recovering the lost information are possible within the modem link.

Most of the variables are functions of frequency and/or time. Pure parameters are s_L and Δf. The signal level s_L is defined as the received level at a nominal value of line attenuation a_L. The noise level n_L is mainly a time-dependent quantity, constituted by the following components: impulse noise, white noise, and/or sinusoidal noise.

In the following analysis, we will be primarily concerned with how the variables n_L and b_L contribute to errors and how some of the other variables should be chosen to produce minimum error probability.

2.4 White-Noise Performance with Ideal Line

The optimum error probability for ideal systems has been calculated by several theoreticians as a function of the signal-to-noise ratio[1,2,3,5]. The transfer function of the line has generally been assumed to be ideal, which implies a_L being a constant and $b_L(\omega)$ being a straight line. In the ideal case with phase reference not perturbed by noise, the characteristics of phase-shift-modulation systems can be written as

$$p_e = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{-(2R_{mr})^{1/2}} \exp\left[\left(-\tfrac{1}{2}t^2\right)\right]dt \quad (1)$$

where R_{mr} is the normalized signal-to-noise ratio at the input of the receiver in terms of average signal power and average noise power in a flat band equal in width to the rate of modulation. An equivalent definition sometimes used is $R_{mr} = E/N_o$, where E is the energy per received element and N_o is the noise power density per unit bandwidth. The relationship $p_e(R_{mr})$ according to (1) has been represented in Figure 7. The curve obtained in this manner is of special interest since there is no question of its validity. Thus this curve may be used as a reference to which other theoretical and measured curves may be compared.

The optimum characteristic obtained from (1) cannot be achieved in practice due to unavoidable disturbances of the phase reference. The error probability of differential coherent phase-shift-modulation systems, also represented in Figure 7, can be written[2] as

$$p_e = \tfrac{1}{2} \exp\left[-R_{mr}\right]. \quad (2)$$

In this case, the phase of the received voice-frequency signals during each unit interval is compared with the phase during the preceding unit interval.

As can be seen from Figure 7, the error probability in this latter case is less advantageous, the characteristic curve is moved 1 to 3 decibels toward higher values of R_{mr}.

Due to mathematical impediments, it is very difficult to calculate the optimum error-probability curve for frequency-shift modulation, because it is a nonlinear modulation method. There are not many theoreticians that have tried an extensive analysis of this problem.[1] However, many[2,4,6] have investigated the 2-tone case under the heading of frequency-shift keying, which is a modulation method based on 2 on-off amplitude-modulated channels, the tone being sent alternately on the two. This modulation method exhibits certain advantages only when the transmission medium is subject to selective fading and should not be confused with ordinary frequency-shift modulation employing a limiter, discriminator, and postdetection low-pass filter.

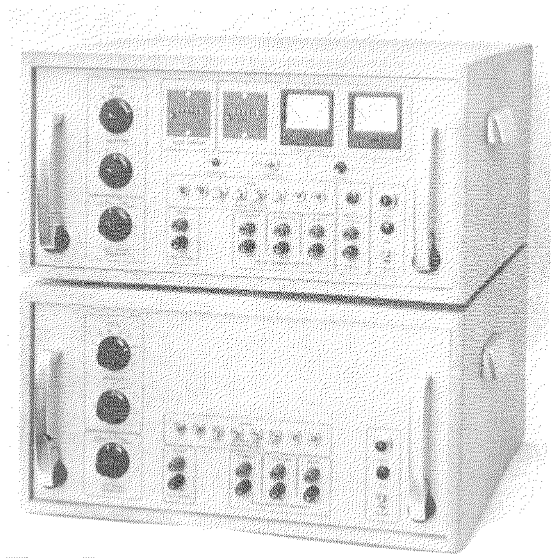


Figure 1—Error-rate analyzer with the receiving set above the sending set.

3. Measuring Methods

3.1 General Remarks

Some results obtained from an extensive measuring program will be given in Sections 4 and 5. Section 4 is concerned with measurements made during the development of the modem part of a frequency-shift-modulated data-transmission system. The main problem has been to arrive at an optimized design regarding noise sensitivity. Some interesting results of measurements made on a model operated over various telephone circuits are discussed in Section 5. The aim has been to analyze to what extent the nonideal behavior of the transmission medium impairs system performance and whether special precautions are required.

Data-transmission quality can be expressed in terms of element-error probability p_e. However, it is also important to know about error-burst distribution in choosing suitable error-correcting schemes. This knowledge can be obtained by measurements on blocks of elements, counting the number of blocks in error. An error-distribution analyzer capable of measuring block-error rate has been developed for this purpose. With this instrument,

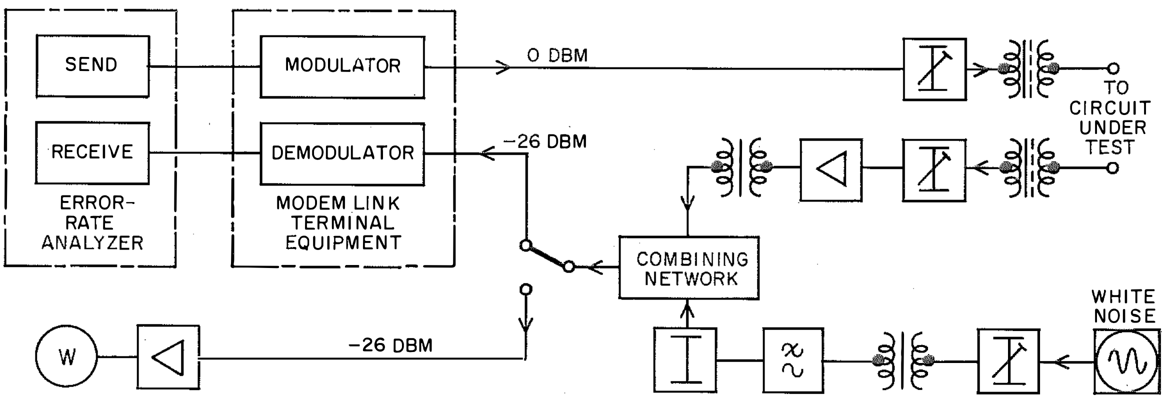


Figure 2—Data transmission measuring setup with white-noise injection. The wattmeter is a thermocouple instrument for measuring 1 milliwatt.

counting is performed simultaneously on blocks of 1, 8, 16, 32, 64, 128, 256, and 512 elements.

However, the result of these measurements will be the subject of a separate paper. The measurements to be discussed below have all been obtained on an element-error rate basis. For this purpose, an error-rate analyzer, shown in Figure 1, has been used.

In the telephone network, the main sources of errors are impulse-noise disturbances and short interruptions. For normal working conditions, attenuation and phase distortion do not cause errors, but might appreciably limit the resistance against other types of disturbances. Throughout the measuring program, white-noise injection was used to optimize various parameters and to relate error-rate impairment due to nonlinear distortion with that due to white noise. With regard to Section 4, the assumption has been made that a system optimized in terms of white-noise disturbance will be a good system for use on telephone lines. The results, as reported in Section 5, also provide a suitable guide when designing the practical system.

3.2 Measuring Setup

To investigate the susceptibility to error and the relative performance for a variety of working conditions, white noise, which is easily measured, was injected into the system. The measuring setup is shown in Figure 2. The error-rate analyzer is connected to the modem link terminal equipment and to some auxiliary apparatus. In Figure 3, the complete modem link terminal is shown.

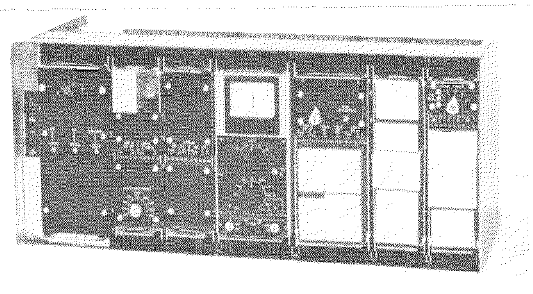


Figure 3—Modem link terminal for 4-wire duplex or 2-wire simplex operation. Designed for bay mounting, the front cover, which exposes only the indicating instrument and the controls beneath it, has been removed.

The send circuits of the error-rate analyzer generate binary-coded direct-current test signals, comprising a repeated cycle of 256 different 8-element characters. Hence the total number of elements within the cycle amounts to 2048. All possible combinations of marks and spaces within the character are being produced during each cycle ($2^8 = 256$). Less-complicated test signals could also be produced but the above-mentioned cycle was used unless otherwise stated.

The test signals go to the demodulator input of the modem link terminal and the voice-frequency output signals pass through an adjustable attenuator and a line transformer to the circuit under test. At the receive end of the loop, the signals are passed via another line transformer and an adjustable attenuator through an amplifier to a hybrid, where the white noise is injected. The disturbed signals are binary demodulated and sent as direct-current signals to the receive circuits of the analyzer. After regeneration, the signals are compared to delayed replicas of the outgoing signals within this instrument.

The output from the hybrid can be switched to a power-level measuring device, comprising an amplifier and thermocouple instrument. The incoming voice-frequency signals and the noise can be measured separately by attenuating the noise and the signals alternately. The measured noise power level refers to a certain equivalent-noise bandwidth, which is controlled by the attenuation characteristic of the low-pass filter through which the output from the wide-band noise generator is passed. Two low-pass filters with different cutoff frequencies have been used, the equivalent-noise bandwidths are 2850 and 4200 cycles per second. The signal-to-noise ratios thus

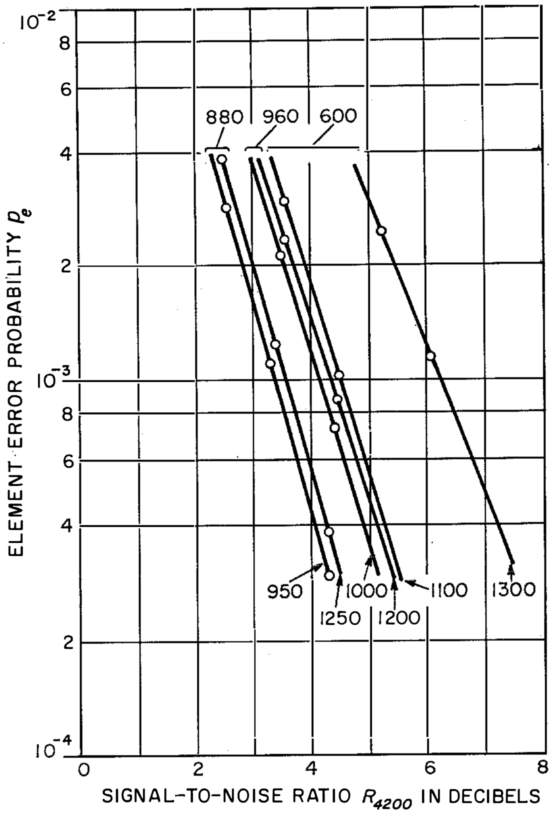


Figure 4—Element-error probability versus signal-to-noise ratio in decibels for an equivalent-noise bandwidth of 4200 cycles per second with disturbing white noise. The pairs of curves bracketed at the top are for the frequency shifts indicated in cycles per second. The designations at the bottom of the curves are the modulation rates in bauds.

worked out from the measurements of power level can be symbolized by R_{2850} and R_{4200}, respectively. For various purposes, it is advantageous to recalculate these ratios and normalize them to either the commonly used noise bandwidth of 3000 cycles per second, or to a bandwidth equal to the modulation rate, which is suitable for some of the studies. The signal-to-noise ratio in these two cases will thus be symbolized by R_{3000} and R_{mr}, respectively.

4. Measurements on Frequency-Shift-Modulation Modem Link Terminals Connected in a Local Loop

4.1 Optimum Relation Between Shift and Channel Bandwidth

At an early stage, a bread-board model of the modem link terminal was constructed and equipped with exchangeable filters having different bandwidths. A series of measurements were undertaken to explore what shifts should be chosen. In Figure 4, some results of the basic measurements have been plotted to illustrate the principles. Using one particular set of filters, the element-error probability has been measured at two or three different signal-to-noise ratios for each combination of the two parameters, frequency shift and modulation rate. Through each set of the indicated measured points, a curve has been drawn and at $p_e = 10^{-3}$ the signal-to-noise ratio R_{4200} has been determined. After normalization to an equivalent noise bandwidth of 3000 cycles per second, the slightly modified ratio R_{3000} has been plotted as a function of the modulation rate in Figure 5. The required signal-to-noise ratio for an element-error probability of 10^{-3} is found to be almost independent of the modulation rate below 1100 bauds. The optimum frequency shift is 880 (± 440) cycles per second.

The filters have been designed like those used in multichannel voice-frequency telegraph systems; the bandwidths were increased by certain factors. In the case of a standard channel spacing of 120 cycles per second, the result reported above would correspond to a modulation rate of 75 bauds and an optimum shift of 56.5 cycles per second. It is interesting to note that the optimum shift of 56.5 cycles per second is only a few percent below the shift of 60 (± 30) cycles per second that has been recommended by the Comité Consultatif International Télégraphique et Téléphonique for frequency-shift-modulated voice-frequency telegraph channels of this particular bandwidth. Hence, these standard channels provide a very efficient medium for synchronous transmission.

4.2 Optimum Modulation Rate and Bandwith

As shown in Figure 5, any modulation rate below a certain figure requires about the same signal-to-noise ratio to keep the error probability below a specified level. However, at low modulation rates the bandwidth of the particular receiver used is unnecessarily large; a narrower bandwidth would reduce the noise reaching the detector. The following question then arises: What is the optimum channel bandwidth for a particular modulation rate? The signal-to-noise ratios of Figure 4 were normalized to a noise bandwidth equal to the modulation rate. By graphic representation, it has been found that the system exhibits a pronounced minimum signal-to-noise ratio R_{mr} for a certain relation between modulation rate and channel bandwidth. The curves in Figure 6 illustrate how this ratio varies. The modulation rate over channel bandwidth is used as a variable on the abscissa and is expressed as a percentage of the optimum. For the particular bandwidth related to above this optimum has been obtained at 1260 bauds, which in the case of standard channels spaced 120 cycles per second corresponds to 81 bauds.

The results reported above have been obtained from a system designed without a send filter. However, in multichannel systems, send filters have to be used and in such cases the optimum signal-to-noise ratio may be obtained at a slightly lower modulation rate.

4.3 Optimum White-Noise Performance

The optimum value of the signal-to-noise ratio R_{mr} obtained in Figure 6 is of special interest. Maintaining the frequency shift and modulation rate thus found to give optimum performance, the error probability has been measured against the ratio R_{mr}. For comparison, the result is shown in Figure 7 with two curves representing theoretical optimum

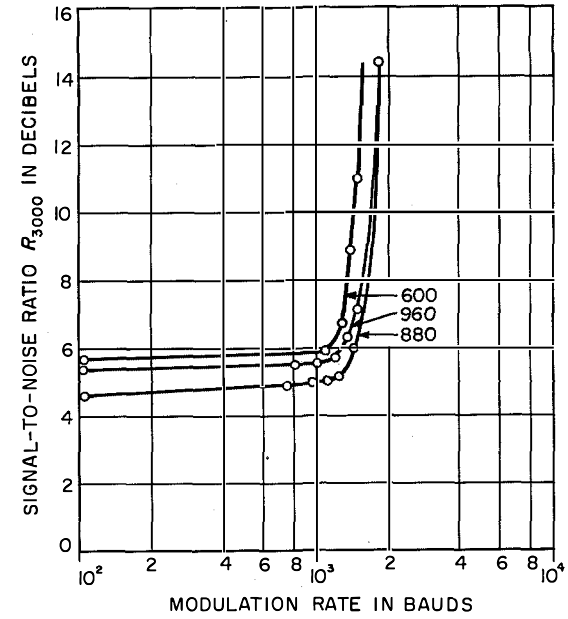


Figure 5—Signal-to-noise ratio in decibels for a bandwidth of 3000 cycles per second and an error probability of 10^{-3} versus modulation rates in bauds for the indicated frequency shifts with disturbing white noise.

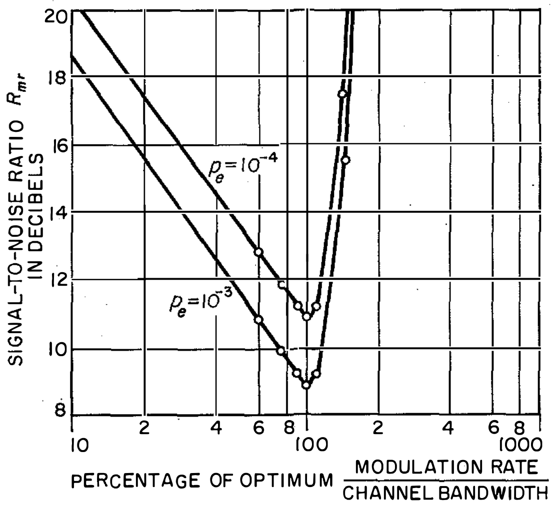


Figure 6—Signal-to-noise ratio for a bandwidth equal to the modulation rate plotted against the percentage of the optimum modulation rate to bandwidth.

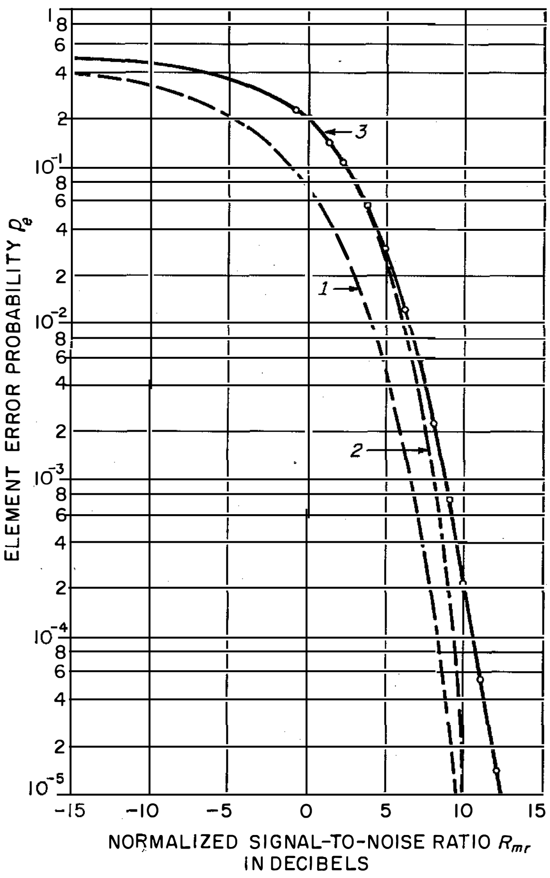
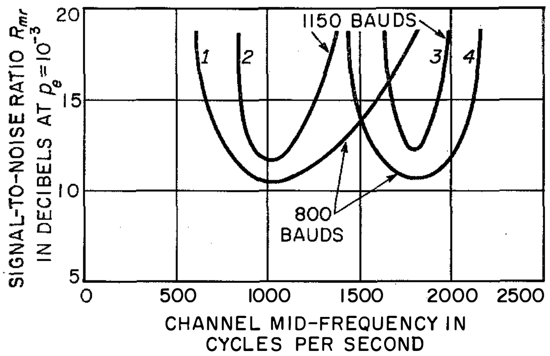


Figure 7—Element-error probability plotted against the normalized signal-to-noise ratio. Curve 1 is the theoretical optimum for phase-shift modulation with ideal phase reference. Curve 2 is for the theoretical optimum case of phase-shift modulation with differential coherent detection. Curve 3 gives the measured values for frequency-shift modulation.



characteristics for phase-shift modulation. Curve 1 relates to phase-shift modulation in the hypothetical case when an ideal phase reference is provided, and curve 2 relates to phase-shift modulation when differential coherent detection is used. A comparison shows that the measured frequency-shift-modulation characteristic, curve 3, happens to follow the theoretical differential coherent phase-shift-modulation curve fairly well at high values of error probability, and that it is only 2 to 3 decibels inferior to the theoretical ideal phase reference phase-shift-modulation curve. The frequency-shift-modulation characteristic has been obtained by measurements in a system where the filters have been designed in a straightforward manner, but employing a special type of limiter. By insertion of a pre-emphasizing type of send filter and by modification of the post-detection low-pass filter[1], it will no doubt be possible to move the frequency-shift-modulation curve a few decibels to the left, and a measurement would probably prove that there are no theoretical or practical advantages from using phase-shift modulation, so far as white-noise disturbances are concerned.

5. Measurements on Frequency-Shift-Modulation Modem Links with Various Loop-Connected Lines

5.1 White-Noise Performance at Various Modulation Rates and Channel Mid-Frequencies

The measuring setup described in Section 3 has been used for measurements on a variety of connections. White-noise performance at various modulation rates and channel mid-frequencies has been measured, with and

Figure 8—Normalized white-noise performance versus channel mid-frequency. Curves 1 and 2 were measurements on a phantom circuit of 93 kilometers (58 miles) of loaded multiple-twin cable. Curves 3 and 4 were measured on 4 cascade-connected carrier circuits.

without equalization. The result will be the subject of a separate report, which will also include error-rate measurements without white-noise injection.

However, some typical results of measurements of this kind will be given here to illustrate the difference in characteristics obtained on loaded-cable circuits and carrier circuits. From Figure 8, it can be seen that the optimum channel mid-frequency of a particular loaded circuit is about 1000 cycles per second while a typical carrier circuit gives best performance at 1800 cycles per second. It should be noticed that there is no common channel mid-frequency usable for both these circuits except at the lower speed.

5.2 Impairment of White-Noise Performance Due to Nonideal Phase Characteristic

A large number of measured characteristics have been used to find a relationship between the white-noise performance and the phase nonlinearity of the transmission medium. Graphic illustrations in Figure 9 define two angles, $\Delta b_L'$ and $\Delta b_L''$, whose sum has been taken as an approximate measure of the phase distortion. The main graph gives the relationship between the impairment of white-noise performance and the defined phase distortion. More than forty measured points have been worked out and a mean curve drawn. The tendency of the curve indicates that the impairment will go to infinity as the phase distortion approaches π radians.

The same basic measurements have also been treated in a different manner in an attempt to obtain a relationship between noise performance and envelope delay distortion. In this case the spread appeared to be greater and it was not possible to find any significant mean curve. It is believed that envelope delay does not correspond to a physical reality in this case where differences in delay between various parts of the same spectrum, comprising a single channel, are considered.

It seems very difficult to find a theoretical relationship between white-noise performance and phase distortion in the generalized case. However, some preliminary considerations have led to the following suggestion, which has been found to be in considerable agreement with the experimental result.

$$\Delta R = -20 \log_{10} \left(\frac{\cos \Delta b_L' + \cos \Delta b_L''}{2} \right), \quad \text{decibels.}$$

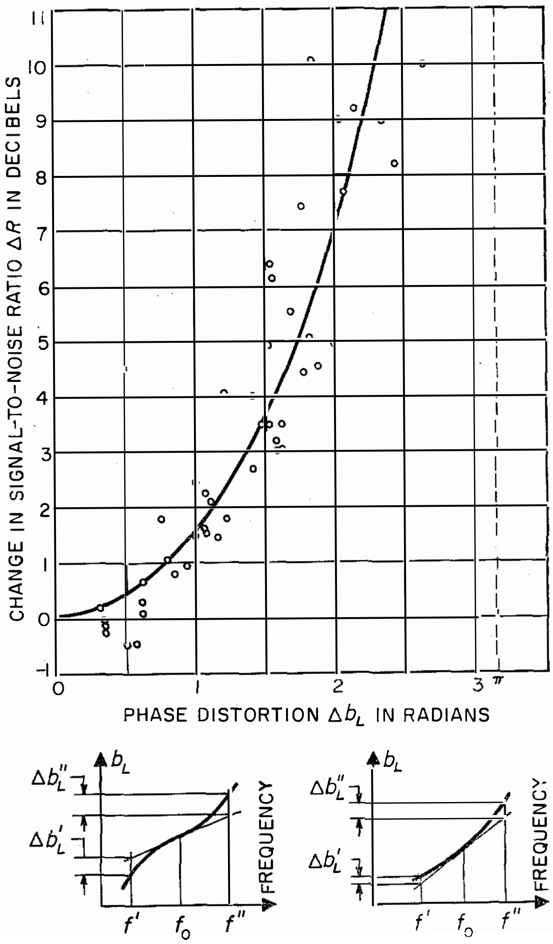


Figure 9—Impairment of white-noise performance ΔR versus phase distortion. As shown in the small illustrations, the phase distortion Δb_L for a typical telephone channel $= \Delta b_L' + \Delta b_L''$. f_0 = channel mid-frequency and $f_0 - f' = f'' - f_0$ = modulation rate/2.

6. Standardization of Modulation Rates

6.1 General Remarks

To establish a suitable standard for modem equipment, it is necessary to know the particular modulation rates that will be encountered in practice. It would be advantageous if various clock-signal frequencies could be derived from one common stable device. Then all modulation-rate figures would be related to each other.

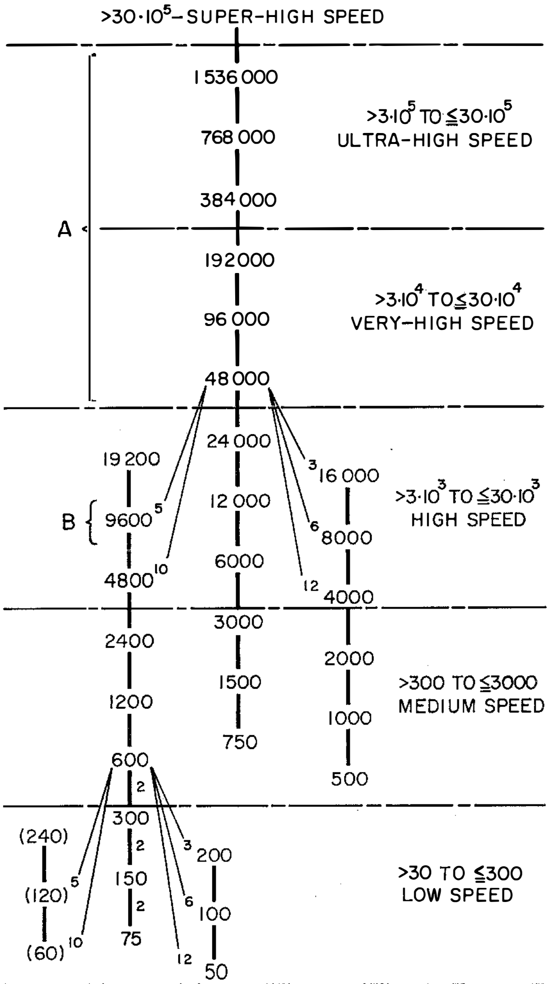


Figure 10—Proposed standard modulation rates in bauds. Group *A* are suitable modulation rates for time-division multiplex systems including pulse-code modulation and delta modulation for telephony. Group *B* are suitable for pulse-amplitude modulation audio-frequency sampling rates.

The above requirements are fulfilled by certain types of geometric progressions for modulation rates. If the ratio between contiguous figures is 2, the number of modulation rates would be too limited. If each such interval is divided in three parts, a more suitable progression would result, where the ratio between any two successive figures is equal to $2^{1/3} = 1.26$. However, to simplify these relationships, it would be preferable to introduce the ratios $4/3 = 1.33$, $5/4 = 1.25$, and $6/5 = 1.20$, which would be approximately the same as the aforementioned one. If these three ratios are used to build up a progression and the modulation rates are taken in units of 3, it will be found that each unit is exactly double the previous one. It is thereby possible to choose a progression that includes most of the modulation rates encountered in practice, as can be seen from Figure 10. From a baud figure sufficiently high in the column that starts at 750, any of the proposed modulation rates can be derived if repeated division is performed by the prime numbers 2, 3, and 5. Hence, clock signals for any of the proposed modulation rates can easily be developed from a crystal oscillator with a frequency equal to one of the figures in the upper part of this particular column.

As shown in Figure 10, modulation rates within the range of 30 to 300 bauds should be derived from the 75-baud column instead of the 750-baud column. There are two reasons for this; first of all, modulation rates that are generally used with the 50-baud column are related to the 75-baud column with the same low prime numbers that have been encountered throughout. A second reason will be given in Section 7.2.

6.2 Proposals

(A) The modulation rates shown in Figure 10 are all proposed for standardization but baud figures up to 3000 are the ones of current interest concerning transmission over telephone circuits.

(B) It is proposed to divide the total progression of modulation rates into decadal ranges and to introduce designations for each range as shown in Figure 10. The proposed low-speed range covers the modulation rates encountered in multichannel and other parallel transmitting systems, while the medium-speed range covers the modulation rates generally used in a telephone channel, that is, those utilized for serial transmission.

7. Standardization of Modem Link Terminals

7.1 Modulation Method

It is proposed that frequency-shift modulation be used if two-condition modulation is applicable. Section 4.3 shows that it is possible to achieve excellent noise performance with frequency-shift modulation and that no other modulation method now used is likely to perform so well.

By using multicondition modulation, it is possible to realize a higher modulation rate by trading good noise performance for a less-stringent bandwidth requirement. Both 4-frequency as well as 4-phase systems have been proposed. However, it is believed that systems of this kind should be used only in special cases, and no recommendations for them will be made here.

7.2 Bandwidth, Shift, and Modulation Rates

It would be an advantage if a variety of systems could be built from a few standardized types of modem channels. Thus, for telephone circuit applications, modem-link bandwidths and shifts should be chosen from among 6 sets of standard values for different modem channels. The white-noise characteristics of these 6 channels are given in Figure 11,

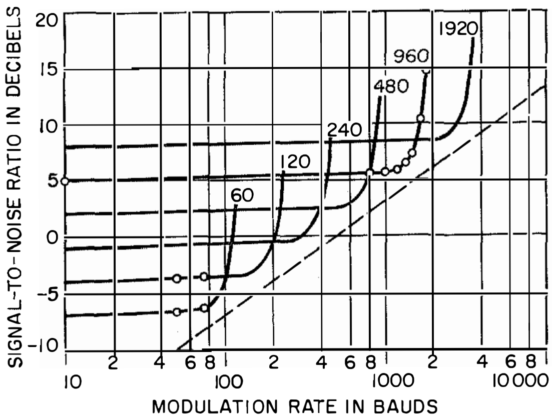


Figure 11—Signal-to-noise ratio in decibels for a bandwidth of 3000 cycles per second and an error probability of 10^{-3} versus modulation rate in bauds for the indicated frequency shifts with disturbing white noise. The dashed line is the theoretical optimum for differential coherent phase-shift modulation.

which also provides information on suitable modulation rates for each type of channel.

The channel proposed for the lowest group of modulation rates is supposed to be identical with the channel already recommended by the Comité Consultatif International Télégraphique et Téléphonique for 120-cycle-per-second spaced multichannel systems using a frequency shift of 60 cycles per second. The other channels have been derived from this low-speed channel by doubling the channel shift and bandwidth 5 successive times. These different types of modem channels will be referred to and identified by a frequency shift, for example, FS 240, which stands for a frequency shift of 240 (± 120) cycles per second. For low-speed applications there are three channels to choose from, namely, FS 60, FS 120, and FS 240. Basically these channels are intended for multichannel systems with spacing of 120, 240, and 480 cycles per second, respectively. For single-channel applications, the remaining three are of prime interest, namely, FS 480, FS 960, and FS 1920.

The continuous curves shown in Figure 11 are calculated from the measured performance of an FS 960 channel without a send filter. When the conventional type of send filters are used, which is normal in multichannel systems, the maximum speed for each channel is slightly reduced. However, measurements performed on standard voice-frequency telegraph systems indicate that at speeds below the maximum the white-noise performance is the same. The results of such measurements with send filters performed at 50 and 75 bauds are shown in Figure 11.

Each modem channel may be used at modulation rates from zero to an upper limit that is proportional to the bandwidth. However, as can be seen from Figure 11, each particular channel offers the most favorable white-noise performance within a limited range, covering 3 of the modulation rates proposed as a standard.

It should be observed that each of the 9 modulation rates recommended within the low-speed range corresponds to a figure that is 10 times as high within the medium-speed range, while the shifts and bandwidths of channels suitable at medium speeds are 8 times the corresponding ones at the low speeds. This implies that the relationship between speed and shift in a low-speed channel corresponds to 80 percent of the relationship within a medium-speed channel. This decrease in recommended modulation rates within the low-speed range has been found to be suitable due to the almost obligatory use of send filters within this range. This is one reason why, as has been discussed in Section 6.1, the low-speed baud figures have been derived from the 75-baud column, as illustrated in Figure 10.

Also for channels within the medium-speed range there are applications where filtering has to be performed on the send side, for example, in modem links comprising forward and backward channels. In such cases, the performance at the upper speed limit may be slightly reduced.

The modem channels are listed in Table 1.

TABLE 1

Modem Channel Type	Frequency Shift in Cycles per Second	Range of Modulation Rates for Optimum Performance in Bauds
FS 60	60	50–75
FS 120	120	100–150
FS 240	240	200–300
FS 480	480	500–750
FS 960	960	1000–1500
FS 1920	1920	2000–3000*

* Systems incorporating the FS 1920 modem channel should be used only on phase-equalized high-quality private circuits, for example, the normal 300-to-3400-cycle-per-second carrier type.

7.3. Channel Mid-Frequencies

In 24-channel systems, which can be comprised of channels of the FS 60 type, the following channel mid-frequencies have been recommended by the Comité Consultatif International Télégraphique et Téléphonique: 420, 540, 660 . . . 3180 cycles per second. By use of FS 120 and FS 240 channels, it is possible to design 12- and 6-channel systems, respectively, extending over the same frequency band that is used in the 24-channel system. Hence, in a 12-channel system with 240-cycle-per-second spacing, it is proposed that the frequency allocation of each FS 120 channel should be the same as two FS 60 channels in a 24-channel system. Moreover, in 6-channel systems with 480-cycle-per-second spacing, each FS 240 channel should correspond to four FS 60 channels. If intermodulation due to nonlinear distortion is considered, the suitability of the frequency allocations dealt with above can be criticized. This also applies to the standardized 24-channel system. However, it is believed that under all circumstances, a low channel level must be chosen to avoid intermodulation disturbances. It should be noted that distortion due to intermodulation products of the third order cannot be reduced by any other means, no matter which frequency allocation is used.

In the case of medium-speed channels, it is necessary to choose the mid-frequency with due regard to the characteristics of the telephone circuit. As has been shown in Section 5.1, this is of utmost importance when the modulation rate is increased. On private circuits, there should be no difficulties in setting the channel mid-frequency to give optimum performance. On the other hand, switched-circuit applications require that a suitable standard be adopted. Even in this latter case, it seems advisable to be able to choose between a few different frequencies, according to the composition of the telephone network being used.

7.4 Proposed Range of Modem Link Terminals

To meet the variety of data-communication needs, the system types given in Table 2 are proposed.

8. References

1. E. D. Sunde, "Ideal Binary Pulse Transmission by AM and FM," *Bell System Technical Journal*, volume 38, pages 1357–1426; November 1959.

2. J. G. Lawton, "Comparison of Binary Data Transmission Systems," *Proceedings of the Second National Conference on Military Electronics*, pages 54–61; 1958.

3. V. A. Kotelnikov, "The Theory of Optimum Noise Immunity," McGraw-Hill Book Company, New York, New York, 1959.

4. S. Reiger, "Error Probabilities of Binary Data Transmission Systems in the Presence of Random Noise, *Institute of Radio Engineers Convention Record*, Part 8, pages 72–79; 1953.

TABLE 2

System Type	Modem Channel Type	Maximum Number of Channels in Each Direction	Channel Mid-Frequencies in Cycles per Second
Voice-frequency telegraph system for 4-wire or 2-wire duplex application. Three versions are proposed.	FS 60	24	420, 540···3180
	FS 120	12	480, 720···3120
	FS 240	6	600, 1080···3000
Data-transmission system for 4-wire duplex or 2-wire simplex application. Three versions are proposed.	FS 480	1	840, 960···1920*
	FS 960	1	1080, 1200···1920*
	FS 1920	1	1800, 1920*
Data-transmission system for 4-wire duplex or 2-wire simplex application. Backward channel for transmission of error-control signals added. Two versions are proposed.	FS 480 FS 120	1 Forward 1 Backward	1200, 1320···1920* 480
	FS 960 FS 120	1 Forward 1 Backward	1800, 1920* 480
Data-transmission system for 2-wire duplex application.	FS 240	1	1080, 1800

* Alternative mid-frequencies in steps of 120 cycles per second.

5. G. F. Montgomery, "A Comparison of Amplitude and Angle Modulation for Narrow-Band Communication of Binary-Coded Messages in Fluctuation Noise," *Proceedings of the IRE*, volume 42, pages 447–454; February 1954.

6. H. J. Held, "Fehlersicherheit binärer Übertragungen bei verschiedenen Modulationsarten," *Nachrichten Technische Zeitschrift*, heft 6, pages 286–292; 1958.

9. Acknowledgements

The author is grateful to Mr. Stig Janson and Mr. Kurt Silversjö of the Swedish Telecommunications Administration for many helpful discussions and review of the manuscript. He expresses his gratitude for assistance offered by the staff of the Royal Swedish Air Force, which sponsored much of the basic development work and all the line measurement work. He is also indebted to the staff of the Transmission Division of Standard Radio & Telefon AB.

Physikalische Grössen und Einheiten, Einleitenlexikon (Physical Quantities and Units, Encyclopaedia of Units)

The author of this book, recently published in German, is Dr. A. Sacklowski of Standard Elektrik Lorenz AG in Stuttgart, Germany.

There are two main divisions of the book; one of 52 pages presents the principles for formulating and rules for applying quantities, units, and their systems. The larger division of 136 pages is a catalog of units, abbreviations, prefixes, and supplemental information useful for an understanding of them.

All the terms are listed alphabetically in dictionary style but the treatment is much more extensive than would be found in a dictionary.

The book concludes with a list of references and standards and several tables for converting units among the various systems that are in use in different countries.

The book is 12½ by 18 centimeters (5 by 7 inches). It is published by Deva-Fachverlag, Post Office Box 209, Stuttgart, Germany, and is priced at DM 12,80.

Wide-Aperture Direction-Finding Systems

In 1942, H. G. Busignies took out a patent that showed how aerial motion could be simulated by commutating signals to or from a number of fixed aerials, thus forming the first basis of the so-called quasi-Doppler direction-finding systems.

Following some experimental work in London, patents were filed in Great Britain by Standard Telephones and Cables in 1944 covering the principles involved in the design of automatic direction-finders and beacon systems of the quasi-Doppler and differential-phase type. An initial study was made with a view to developing a direction-finder for the very-high-frequency aircraft band, but the main effort was later diverted to improving circuit techniques, and constructing a high-frequency equipment that was installed in West Africa in 1951. Results were promising enough to warrant further development of a whole range of equipment covering high-, very-high-, and ultra-high-frequency bands with a maximum of common equipment, based on the principle of measuring phase differences between consecutive aerials in a ring and using electronic switching.

While Standard Telephones and Cables was engaged in work in the high-frequency band, which was later to lend itself so readily to developing equipment at higher frequencies, Standard Elektrik Lorenz had begun development of equipment for the very- and ultra-high-frequency aircraft bands based on the principle of a smooth transference of signal from consecutive aerials using capacitive coupling and a frequency discriminator to demodulate the generated Doppler frequency.

Each type of system is described in one of the following two papers. It is clear that both techniques have peculiar advantages. Now that the two types of equipment are well established in production, a realistic and practical assessment of them can be made for incorporation in future equipments that will utilize the best features of both systems.

Practical Evolution of the Commutated-Aerial Direction-Finding System *

C. W. EARP
D. L. COOPER–JONES
Standard Telephones and Cables Limited; London, England

The paper sets out to show how the more fundamental problems encountered in the development of the commutated-aerial direction-finding system have been solved during a long-term development plan.

Reasons are given for the choice of a circular array of aerials, which are commutated sequentially to the receiving equipment. It is shown how such an arrangement permits great economy in the number of aerials required for a given system performance, maximum adjacent aerial spacing being much larger than might at first appear permissible, and the method for computation of the signal bearing is described in some detail.

Significant features of the development are described, particularly where problems are peculiar to the system, such as realization of adequate selectivity despite aerial commutation and the avoidance of the effect of aerial interaction.

Final practical circuits are described for signal processing and demodulation, and for the display of bearing, and it is shown that the solutions adopted have permitted realization of the performance predicted some 12 years ago.

Considerable practical experience in the field has now been obtained, and measurements have been made to substantiate most of the theoretical predictions with regard to performance.

1. Introduction

A considerable amount of experimental work was carried out on the commutated-aerial direction-finding system between 1945 and 1947, and a paper[1] published in 1947 described the state of development at that time. The system already appeared to offer a revolutionary advance in the accuracy of radio direction-finding, but it did not appear to be possible, immediately, to set about the final development of practical equipment for normal commercial operation. A number of theoretical aspects still needed to be studied, and sufficiently reliable circuits were not available for signal processing, computation of bearing, and satisfactory display of the bearing. Moreover, for one or two aspects of a practical development it appeared that new circuit elements would need to be developed.

The present paper sets out to describe significant advances made during a long-term development plan. In practice, by the end of 1954, both theoretical and practical ideas had crystallized to the extent that serious practical development could start with reasonable assurance that the final product could give results truly representative of the early theoretical promise.

From 1954 effort was used to produce prototype models of direction-finding equipment for the ultra-, very-, and high-frequency bands, the equipment using a maximum of common apparatus for the three projects. Installation of production very- and ultra-high-frequency systems commenced early in 1957, and at the present date a total of 120 ground-based systems are installed in 10 countries, and naval shipborne installations account for another 12. A number of these systems provide simultaneous very- and ultra-high frequency direction-finding channels.

One ultra-high-frequency version is perhaps worthy of particular mention. Primarily designed for naval use, the aerial system is comprised of 18 dipoles threaded round a 'Tacan' mast. With access to such a good site for direction-finding purposes, performance is equal to that of a ground-based system on a good site.

2. Basic Principles

Before continuing to a description of the development of the commutated-aerial direction-finder, it will be opportune to define exactly

* The present paper is a revision of one previously published in *The Proceedings of the Institution of Electrical Engineers,* volume 105, Part B, Supplement No. 9, 1958.

what is considered to constitute the system. This will obviate confusion with suggestions that have been made and which, though involving the commutation of aerials, were not considered worthy of consideration for the final development. For the purpose of the paper the commutated-aerial direction-finder comprises systems in which the phase of the signal field is sequentially sampled at a number of points, by a single receiver, and in which the signal bearing is computed from the measurements of relative phases thus made. In no case is there a standard or reference phase by which measurements of phase of a single sample could be meaningful.

It is significant that the system thus defined can be realized without having to set up phase or amplitude balance between two independent tunable amplifiers. Measurement of relative phase of two simultaneously derived samples could not be made without the use of such balanced amplifiers. Although, in practice, an auxiliary receiver may be used in a commutated-aerial direction-finder system for other reasons, it is never used to measure the phase difference between two simultaneous samples of the signal field.

3. Choice of Aerial Configuration

When a given aerial space is available, the most accurate evaluation of signal bearing is evidently possible if the signal phase pattern is examined over the whole of that space. For obvious reasons, a random configuration of sampling positions, or any complex pattern, would make bearing computation too difficult. The most likely practical configurations appeared to be a pair of orthogonal linear arrays and the simple, uniformly spaced, circular array of aerials.

The pair of orthogonal linear arrays would at first appear to have the advantage that measurements made on the two arrays could be associated with the respective orthogonal deflecting elements of an indicator such as a cathode-ray tube. This configuration was finally rejected, however, for two reasons, namely:

A. The effect of interaction between unused aerials and the aerial being sampled (one of the most serious problems in commutated-aerial direction-finders and dealt with in a later section) would be to produce an uncontrollable and frequency-sensitive octantal error.

B. Simple computation of bearing would involve the use of adjacent aerial spacing of less than half of a wavelength. Hence—for some applications, at least—the number of aerials required would be uneconomic.

The use of a circular array considerably reduces the problem of errors due to aerial interaction, for lower-order cyclic errors such as the octantal error cycle of 'orthogonal' or '2-phase' systems are completely balanced out. When using a ring of 18 aerials, for example, the lowest-order error has 18 complete cycles per azimuth, so that, for a given amount of interaction, the maximum error is reduced by a large factor. Furthermore, as will be shown later, if the number of aerials in a ring exceeds 6, it is possible to use a process of computation permitting unambiguous determination of bearing despite wide spacing between adjacent aerials. Theoretically, bearing could be computed unambiguously when the ring contains only 5 aerials, but this case is unlikely to have practical value. The practical spacing of aerials is limited only by the amount the signal field is disturbed by site imperfections or propagational complexities.

3.1 Circular Array

Uniform successive sampling of the phase of the signal at points uniformly spaced round a circle has been compared with the gyration of a single aerial round a circular track, and commutated-aerial direction-finding has been considered by some as a Doppler system. The authors are conscious of so much confusion in some minds that the reader is warned that the concept is dangerous unless he can easily discard the term Doppler frequency, which is strictly a misnomer. In communication technique, frequency modulation of an oscillator is quite distinct from phase modulation of a stable frequency source: phase modulation is path-length modulation, and the source of oscillation is unmodified. Similarly, Doppler modulation is path-length modulation and the source of a signal is unmodified: hence Doppler modulation originates as phase modulation.

The phase modulation imposed on a signal by smooth gyration of an aerial is sinusoidal, the deviation being $\pi d/\lambda$, where d is the diameter of gyration, and being sinusoidal, can, of course, be translated to an equivalent frequency modulation. The commutated-aerial direction-finding signal also bears phase modulation of deviation $\pi d/\lambda$, but the waveform of the modulation contains transients that cannot reasonably be considered as frequency modulation.

In practice, the commutated-aerial direction-finder receiver first attempts to measure the differential phase envelope of the signal, the time-constant of differentiation usually being equal to the time of connection to a single aerial, that is, the time interval of a single horizontal step of the phase envelope. (If this differential phase envelope can be measured, the original phase envelope can be constructed by integration; but integration is unnecessary for bearing computation, since the differential envelope has a fixed and known phase relation with respect to the envelope itself.)

Now, when adjacent aerial spacing exceeds $\lambda/2$, differential phase steps may lie outside the range ± 180 degrees, so that the differential phase envelope cannot be directly measured

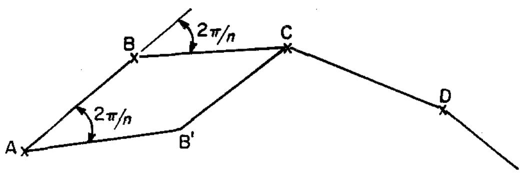


Figure 1
Analysis of phase steps for other operating modes.

without ambiguity. Leaving this case for the moment, let us assume that phase steps do not exceed 180 degrees. In this particular case, which represents the most simple operating regime of a commutated-aerial direction-finder receiver, bearing is, in fact, measured by establishment of the differential phase envelope and comparison of this in cyclic repetition phase with the phase of the cyclic aerial commutator.

Practical experience has shown that, when the bearing is derived by measurement of the first differential of the phase envelope, it is undesirable to have adjacent aerial spacing greater than $\lambda/3$. A maximum phase step of 120 degrees gives an adequate margin of safety against breakdown of a linear discriminator, which can deal with phase steps approaching ± 180 degrees. The margin of 60 degrees is required for phase errors due to maladjustment of equipment, noise (particularly at full range), site imperfections, and multi-path propagational effects. Now the diameter of a commutated-aerial direction-finder array is substantially determined by the performance required for the lowest frequency of operation, while the number of aerials required is inversely proportional to the maximum phase step that can be handled by the bearing computer. For frequency-band ratios much exceeding 2, a very large number of aerials would be required if the maximum phase step were limited to 120 degrees. In the following section, it is shown that this limit can be raised to 360 degrees without unreasonable complication, so allowing a threefold reduction in the number of aerials, with a substantial saving in cost.

3.2 Use of Adjacent Aerial Spacing Greater than $\lambda/2$

In Figure 1, A, B, C, and D represent typical sampling points on a circular array. The signal wave-train derived from such a system has a phase envelope defined by successive step values ϕ_A, ϕ_B, ϕ_C, ϕ_D, et cetera, of the signal phases at A, B, C, D, et cetera, with reference to some unspecified constant signal phase.

In the final equipment, by circuits that will be described later, a new wave-train is developed that bears a phase envelope defined by successive phase values $\phi_A - \phi_B$, $\phi_B - \phi_C$, $\phi_C - \phi_D$, et cetera. (Note that these values may be ambiguous when aerial spacing is large, but only to the extent that multiples of 2π radians could be added or subtracted without effect.)

This wave-train is now subjected to a delay equal to one aerial connection period, when input and output phases of this delay line may be subtracted to yield a succession of values, $(\phi_A - \phi_B) - (\phi_B - \phi_C)$, $(\phi_B - \phi_C) - (\phi_C - \phi_D)$, et cetera, which constitutes the second differential with regard to time of the original signal envelope. Again in Figure 1, B′ is determined by completion of the parallelogram ABCB′.

Now for any direction of a single signal ray,

$$\begin{aligned}&(\phi_A - \phi_B) - (\phi_B - \phi_C)\\ &= (\phi_A - \phi_B) - (\phi_A - \phi_{B'})\\ &= \phi_{B'} - \phi_B\end{aligned}$$

$$\begin{aligned}(\phi_{B'} - \phi_B)_{max} &= \mathrm{B'B} \times \frac{2\pi}{\lambda}\ \mathrm{rad}\\ &= 2\mathrm{AB}\sin\frac{\pi}{n} \times \frac{2\pi}{\lambda}\end{aligned}$$

where n is the number of aerials in the circle.

If $n = 18$, $\mathrm{B'B} \simeq \frac{1}{3}\mathrm{AB}$; furthermore, if $n > 6$, $\mathrm{B'B} < \mathrm{AB}$. Thus, when the number of aerials in the ring is greater than 6, phase ordinates have smaller maxima than the first differential values $\phi_A - \phi_B$, et cetera, and when using 18 aerials, the maximum values cannot exceed about one-third of a maximum first-differential step. Hence, in an 18-aerial system in which adjacent aerials are spaced at λ, the second-differential ordinates cannot exceed 120 degrees.

The phase of the second differential of the phase envelope is, of course, fixed with respect to the phase envelope itself, so that the signal bearing is measured by low-frequency phase-comparison of the second-differential envelope with the cyclic aerial commutating waveform.

It may be noted that when using 18 aerials and permitting a maximum measured phase step of 120 degrees, the diameter of the array may be as much as 6λ.

4. Problem of Aerial Interaction

One of the outstanding problems in the development of commutated-aerial direction-finding systems has been the design of efficient individual aerials that, when not in use, could be disconnected in such manner that they become ineffective as signal reflectors. Aerials not in use behave very much as unwanted site obstacles, producing a distortion of the signal field, and hence equipment errors somewhat similar to site errors. In the circular array, however, some interaction is permissible, for errors due to different aerials tend to cancel out except for the regular cyclic error, which is repeated a number of times per azimuth equal to the number of aerials.

As already described by Bain,[2] when the number of aerials is large, one practical solution is provided by terminating the aerials not in use with a suitable dummy load. This method was adopted for a high-frequency (3-to-30-megacycle-per-second) system using 24 aerials in a ring having a diameter of 100 metres.

Later experience showed that a ring of 18 aerials is completely adequate for first-class performance of very- and ultra-high frequency equipments, even on rather inferior sites, but also that, when using this number of aerials, the terminating technique does not remove interaction by an adequate factor.

Thus it became necessary to effect some form of electronic de-sensitisation of aerials not in use.

One satisfactory solution for such a case is the use of thin unipoles rather less than $\lambda/4$ long at the highest working frequency, each top-loaded with a large circular disc. Each aerial is inactivated by the use of a germanium diode that disconnects it at about one third of its length from the ground-plane. A small choke returns diode current from aerial to ground, and a critical choice of inductance for this component, though permitting sharp resonances in disconnected aerials both below and above the working frequency band, renders interaction exceedingly low over the desired signal-frequency range.

The same principle has been used for construction of dipoles, and other methods are now available.

5. Aerial Commutation

Commutation of aerials involves the generation of switching pulses for each aerial involved, and the application of these pulses to produce efficient connection of the aerials in succession to the receiver, at the same time arranging to disconnect or terminate aerials not in use to avoid aerial interaction.

An early version of the commutated-aerial direction-finding system employed 9 aerials, but subsequent models use 12, 18, or 24 elements depending on the application. The switching rate is controlled by a reasonably stable free-running oscillator at a frequency of approximately 1 kilocycle per second. Shaping circuits after the oscillator result in a train of sharp trigger pulses at a repetition rate of approximately 1 kilocycle per second for application to a ring counter to produce the switching pulses. Early attempts at using a hard-valve counter were abandoned in favour of using gas tubes, which were known to be extremely reliable. A basic ring of 6 was produced using conventional techniques, and it was arranged that the blocks could be connected in series to produce rings of 12, 18, or 24 as required. Outputs in the form of rectangular pulses of approximately 1 millisecond duration and 80 volts amplitude were taken from resistors on the cathode side of the gas tubes. To isolate the counter from the actual aerial switch, cathode-followers were interposed.

Since the semi-conductors used to switch the aerials are essentially low-voltage 'current' devices, it was considered that transistors could readily be used instead of gas tubes. A basic ring-of-6 counter has been constructed using junction-type transistors as a replacement for the gas-tube counter, resulting in a considerable saving in space and power supplies, with—it is hoped—even greater reliability.

Efficient commutation of the aerial elements for 100-to-400-megacycle-per-second signals presented serious difficulties, which have now been alleviated by the development of semi-conductor devices having low capacitance, low insertion loss, and high switching ratios. At lower frequencies (2 to 15 megacycles per second) it was found possible to use hard valves as switching elements, a cathode-follower being placed at the base of each aerial and an inverted amplifier at the inner ends of the lines feeding a common load impedance. Such devices, however, were prone to cross-modulation, and their capacitance precluded their use at higher frequencies.

Early attempts (about 1946) at using semi-conductors were abandoned because of the lack of mechanical robustness and wide tolerances on the electrical performance. Since then the art has progressed rapidly, and germanium impurity diodes are now available having capacitances of less than 1 picofarad, forward resistances of less than 5 ohms for 15 milliamperes and back-to-front ratios of at least 40 decibels. Their dimensions are such that they can conveniently be mounted in the transmission lines between the aerial elements and the receiver. Welded junctions result in good mechanical properties, and the operating range of temperature is adequate.

Each aerial gate consists of three diodes arranged as a T network, one form being shown in Figure 2. It will be seen that two diodes, $X1$ and $X2$, are in series with the line connecting the aerial to the common point B, while $X3$ is connected to earth. The switching pulse is applied to the junction of the diodes A via a buffer resistor $R1$. When the gate is shut, point A is negative with respect to earth, $X3$ is conducting, and $X1$ and $X2$ are held off. When a positive switching pulse is applied, point A becomes positive, $X3$ is non-conducting, and $X1$ and $X2$

become conducting. The aerial is thus connected to the common point B and thence to the receiver. $R2$ is 100 ohms and serves to terminate the aerial and feeder when the gate is shut for reasons given earlier. Its use decreases the switching ratio obtainable, but since this is at least 30 decibels in the circuit shown, this is not important. It should be noted that the rectifier $X2$ is used, not only to provide the signal switching ratio, but also to disconnect the unused line from the common point B. There still remains a considerable capacitance between B and earth, owing to the gating rectifiers, and this tends to modify the impedance match from the aerial to the receiver. Care must be taken in the design of the switch box containing the gates to minimize this mismatch, which, using the latest diodes available, is still tolerable for 18 aerials at 400 megacycles per second.

The switching system just described was used in a 24-aerial 3-to-30-megacycle-per-second system with success. In present 18-aerial very- and ultra-high-frequency systems, $R2$ of Figure 2 is eliminated and a diode is introduced at the aerial end of the feeder, permitting more complete suppression of aerial interaction in the smaller geometric frequency band.

6. Selectivity of the Receiver

To use a maximum of common apparatus for commutated-aerial direction-finder equipments for the ultra- very-, and high-frequency bands,

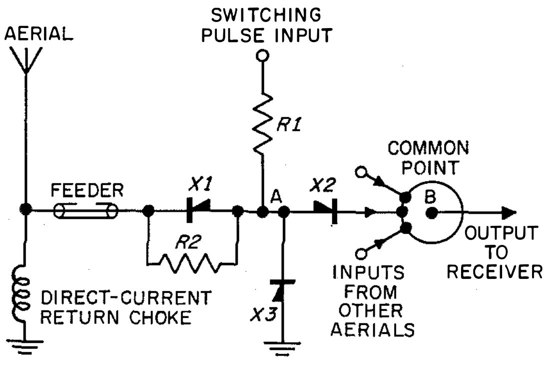


Figure 2—Typical aerial gate circuit.

an aerial switching rate of 1 millisecond per aerial was chosen. Such a rate gives a cyclic commutation frequency of about 55 cycles per second (for an 18-aerial system) and permits an accurate bearing to be taken on a signal of $\frac{1}{5}$ second duration, when using the integrated type of display. Such a speed was considered adequate for all normal purposes.

Aerial pulses of 1 millisecond duration were considered the shortest permissible for the high-frequency band, when the receiver bandwidth would be about 3 kilocycles per second. At this switching rate, however, a very serious problem arose with regard to selectivity. It was impossible to switch on aerials smoothly without great inefficiency, and, incidentally, giving unequal treatment of the various aerials. Hence, each aerial must be switched on in a very short time, and after 1 millisecond, off again in a very short time. The effect of such switching is to generate higher-order switching sidebands of a number of off-tune signals. A simple Fourier analysis of the repetitive square switching pulses at each aerial shows at once that reasonably satisfactory overall selectivity could not be obtained without resorting to less-steep pulses and durations of the order of 10 milliseconds, which must be considered far too slow for satisfactory speed of bearing indication. Some signal pulse shaping for minimization of the frequency spread could be achieved after the signals from different aerials are combined in a single channel, but control of pulse shape still appeared too difficult to achieve the result considered necessary.

After noting experimentally that off-tune signals tended to produce switching-transient interference only near the leading and trailing edges of wanted signal pulses, it was decided to gate out such interference after frequency selection of the wanted signal. Thus, after each transient operation of the aerial switch, switching one aerial off and at the same instant switching the next aerial on, the output from the receiver was allowed to reach its steady state of response before opening a gate to accept only such steady-state response. In a typical case of high-frequency reception, the first half of each millisecond pulse was rejected as containing possible transient interference, the second half-pulse being used for bearing determination. Hence, steady-state selectivity was obtained, despite bandwidth requirements of 3 kilocycles per second at 3 decibels down and about 12 kilocycles per second at 80 decibels down.

Theoretical limitations of this method have not been studied, but it should be mentioned that performance corresponding to the above figures could not be achieved without a rather unusual design of the selective circuits. Different types of receiver were found to have vastly different transient response characteristics. For example, a receiver using an efficient mechanical filter at intermediate frequency is likely to be totally unsatisfactory. Although it may theoretically be possible to design an efficient low-loss band-pass filter comprising a few lumped reactances to have a suitable phase characteristic, it is probable that the use of such low-loss components would involve an impractically high precision of adjustment. In commutated-aerial direction-finder development, it was assumed that, in general, a low-loss lumped filter would be the most likely to show the effect of multiple internal reflections, and hence a maximum duration of transient response. Hence, the intermediate-frequency system of the high-frequency commutated-aerial direction-finder receiver was made up of coupled pairs of tuned circuits, pairs being carefully buffered by either resistive attenuating pads or valves.

On equipment designed for the very- and ultra-high-frequency bands, it has not been found necessary to use the transient gating technique, owing to the much larger bandwidths permitted because of the necessary frequency tolerance.

7. Basic 2-Receiver System

It has been shown that the simplest method of processing the information of a commutated-aerial direction-finding system is by comparison of the phases of the signal picked up in adjacent aerials in the array. This operating mode is satisfactory until the maximum phase step exceeds 180 degrees and causes breakdown in the phase detector. A basic circuit for achieving this mode is shown in Figure 3. It will be seen to involve two receivers with a common beating oscillator, one connected to the direction-finder aerial array and the second to a single element. This technique was adopted for three reasons. First, it enabled the bearing information to be transferred to a frequency having crystal-oscillator stability before attempting the demodulation process. As will be seen later, the demodulation process involves the use of a delay network through which the signal must pass, the phase distortion at signal carrier frequency being fairly accurately defined. To set up the phase relationship between the input and output of this 1-millisecond delay network accurate to 10 degrees, signal frequency must be accurate to about 25 cycles per second. Secondly, the auxiliary receiver can be used to provide an audio monitoring output of good quality, whereas the output from the normal direction-finder receiver has switching tone superimposed. Thirdly, any phase or frequency modulation on the received signal does not appear in the beat frequency after combination of the two receiver outputs. The principle of operation is as follows.

Since the receivers have a common oscillator, the outputs are at identical frequencies, say f_1, that from the direction-finder receiver bearing the phase modulation imposed by the aerial array commutation. The output from the auxiliary receiver is applied to a frequency changer, together with the output from a crystal oscillator at a frequency f_2. The resulting beat at $f_1 - f_2$ is selected in a filter, the output from which is applied to a further frequency changer, together with the output at f_1 from the direction-finder receiver; the resulting beat at f_2 is selected, and now has the stability of the original oscillator at f_2 and bears all the modulation imposed by the aerial array. If the auxiliary aerial is within a few wavelengths of the direction-finder array, any frequency or phase modulation of the incoming signal will affect both direction-finder and auxiliary receiver virtually simultaneously and will not appear on the beat output at f_2. Typical values of f_1 and f_2 would be 2 megacycles per second and 130 kilocycles per second respectively.

To compare the phase of the signal from successive aerials, it is necessary to hold up the information for a time equal to the time of connection of an aerial, that is, 1 millisecond. This is achieved by applying the signal at f_2 to a delay line, which may simply consist of a number of band-pass circuits in cascade. The delayed and undelayed signal pulse trains are applied to the phase detector, which provides an output amplitude proportional to the phase difference between the two inputs up to $\pm$ 180 degrees electrical. The output from the phase detector is thus a stepped waveform at the commutation frequency, having a phase related to the direction of arrival of the signal at the array. This waveform is applied to a phase-meter device, either mechanical or electronic, together with fixed-phase orthogonal reference waveforms. Display of the bearing information can be either by meter or cathode-ray tube. The reference waveforms can be readily obtained from the aerial commutation circuits, say, by tuning a single pulse to the commutation frequency.

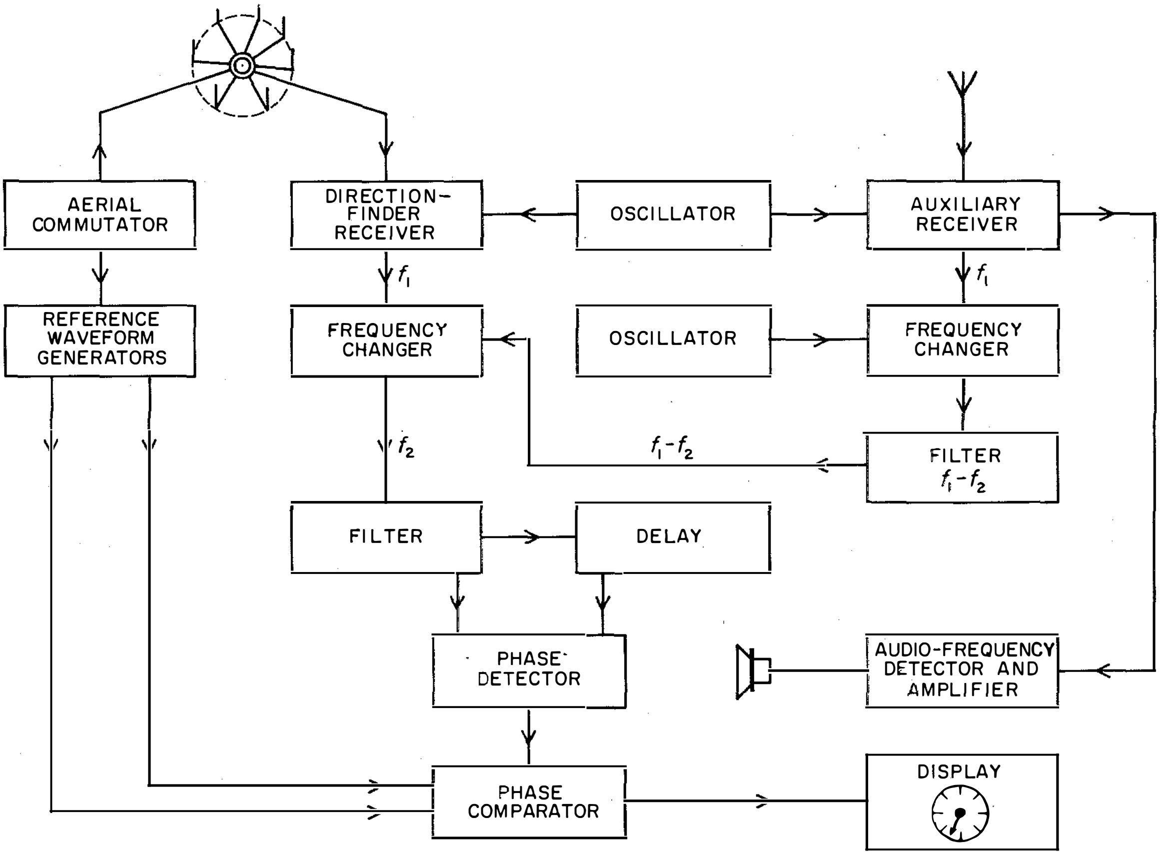


Figure 3—Basic commutated-aerial direction-finding system.

8. Practical Signal-Processing Circuits

A practical form of the signal-processing circuits is shown in Figure 4. This illustrates operation in the basic mode of the previous section and also a mode employing the second differential of the original signal modulation.

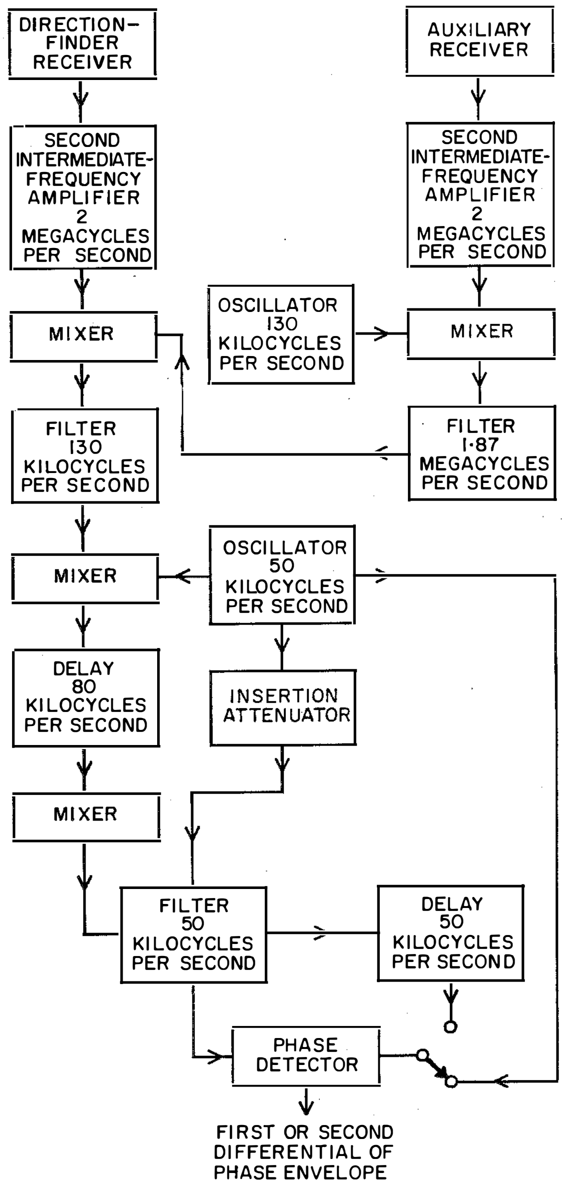


Figure 4—Practical signal-processing circuits.

The principal difference between the basic circuit and the practical form is the addition of a further frequency changer. This has several advantages, quite apart from enabling the second mode of operation to be obtained. It enables the special phase detector described in the next section to be worked at a more-convenient frequency, and also permits a small constant signal of suitable phase to be applied to the signal path to maintain operation of the phase detector in the absence of a signal.

Considering the operation in detail, it will be seen that the outputs from the direction-finder and auxiliary receivers are at a nominal 2 megacycles per second, that from the auxiliary receiver being changed to 1·87 megacycles per second using a crystal oscillator at 130 kilocycles per second. This is an appropriate amount of offset to enable the beat (1·87 megacycles per second) to be readily separated despite receiver bandwidths of 50 kilocycles per second, but it is rather high for maintenance of a precise phase relationship between the input and output terminals of a delay network, and it is also too high a frequency at which to operate the special phase detector.

The 130-kilocycle-per-second beat is selected and applied to a frequency changer with the output from a 50-kilocycle-per-second crystal oscillator. The beat at 80 kilocycles per second is selected and applied to a delay line having a delay of one pulse width, that is, approximately 1 millisecond. This delayed signal train and the undelayed train at 130 kilocycles per second are applied to a frequency changer and the beat at 50 kilocycles per second is selected. This 50-kilocycle-per-second signal is now in the form of a train of wave packets bearing the differential phase between signals picked up in adjacent aerials in the commutation cycle; that is, one stage of differentiation, and hence effective compression of deviation, has been achieved. In the simplest, or 'basic,' regime of operation, this wave train is applied as one input, and the fixed 50-kilocycle-per-second oscillation is applied as the other, to the phase detector or demodulator.

In addition, a small amount of 50-kilocycle-per-second oscillator signal is injected into the signal path in correct phase to maintain the phase demodulator in the event of signal failure from the receiver. The significance of this cross-injection will be better realized after considering the trigger type of phase detector finally adopted, and described in a later section. Under certain conditions, the injected signal can have particular advantage: for example, when badly fading signals are being received, the error of display is minimized, since the tendency to display a large error corresponding to a false signal sample is reduced. A frequency of 50 kilocycles per second was selected as a reasonable value to operate the trigger-type detector, and also to provide easy filtering of bearing components after detection.

The second mode of operation is readily obtained by applying the 50-kilocycle-per-second signal train to another delay network having a delay of a further millisecond. If the output from this delay line is applied as the second input to the phase detector in place of the crystal oscillator signal, the detector will give an output proportional to the second differential of the original phase envelope.

9. Signal Demodulation and Bearing-Display Circuits

Having processed the signal in the manner discussed above, the next step is to demodulate the signal by phase detection to yield the first or second differential of the original phase envelope according to the mode of operation. In early work, this phase detector took the form of the well-known sum-and-difference detector, which yields an output proportional to the sine of the input phase difference. Later, it was realized that linear phase discrimination would have certain advantages, but the full significance was not appreciated until a satisfactory solution was evolved.

Considerable attention has been paid to methods of detecting the phase modulation envelope and subsequent combination with the reference waveforms to obtain the bearing information. For most purposes time-constants of the order of at least $0 \cdot 5$ second are inserted before presentation of the bearing, the information normally being in the form of orthogonal quasi-direct-current components applied to the coils of a meter-type display or to the plates of a cathode-ray tube. However, it is often desirable to have a faster indication, particularly in the high-frequency band, where two signals may be present simultaneously. If one of the signals is in the form of an on-off keyed carrier, it is often possible to read the bearing of the second station during the keying intervals of the first, provided that the display response is fast enough.

Consider an output from the phase detector of the form $V \sin(2\pi ft + \theta)$, where f is the frequency of commutation and θ the bearing of the incoming signal relative to an arbitrary reference waveform $V \sin 2\pi ft$. Direct multiplication of the two waveforms results in the outputs $\frac{1}{2}V \cos \theta$ and $\frac{1}{2}V \cos(4\pi ft + \theta)$. Similarly, multiplication of the signal waveform by a reference waveform $V \cos 2\pi ft$ results in components of the form $\frac{1}{2}V \sin \theta$ and $\frac{1}{2}V \sin(4\pi ft + \theta)$. There are thus direct-current components

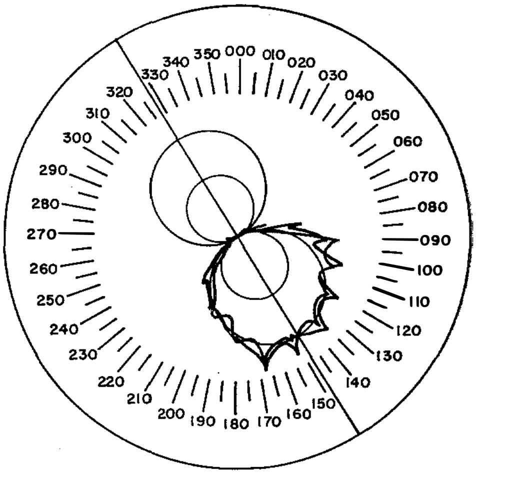


Figure 5—Typical offset circle display.

of the bearing and terms at double the commutation frequency. Direct application of these outputs to the plates of a cathode-ray tube will result in a bearing display ideally in the form of two circles superimposed on each other and offset from the centre of the tube at the bearing angle by the direct-current components (Figure 5). Using the type of cursor shown it has been found possible to read the bearing with fair accuracy. Departure from ideal circles is due to disturbance in the ideal phase pattern by site imperfections and aerial interaction. Such a display is capable of a fast response, since no filters need be inserted other than to remove intermediate-frequency components after the phase detector, which will be of the order of 50 kilocycles per second. If this rapid display is not required, filters having time-constants of about 0·5 second may be inserted, resulting in removal of the double-frequency term and leaving the direct-current components only, which will paint a single spot on the face of the tube. This can readily be converted to a radial line pointer using established techniques.

Direct multiplication of the demodulated signal envelope with a reference waveform presents a number of practical difficulties, and a technique has been devised[3] to integrate the process with the actual circuit performing the phase detection.

The phase detector takes the form of a pair of triodes connected as a bistable flip-flop or trigger circuit. The two signal inputs are squared and differentiated to provide trains of very sharp pulses to operate as inputs to the trigger grids of the flip-flop. This results in anti-phase square waveforms at the anodes of the trigger valves, having a variable on-off ratio in accordance with the differential phase modulation of the two inputs. It is arranged that with no modulation the pulse trains are phased 180 degrees apart, hence giving unity on-off ratio of the square output waves. These square waveforms are then combined with the reference waveforms, as shown in Figure 6, which indicates the basic circuit for combining the signal with, say, the sine component of the reference waveforms. A similar circuit is needed for the cosine component. For push-pull operation, a total of four circuits is required to incorporate, in addition, the negative sine and cosine components. To simplify the description it will be assumed that the phase modulation and the reference waveforms are sinusoidal, as for the pure Doppler case, the stepped waveforms merely involving harmonic terms at the commutation frequency.

The sine component of the reference waveform is applied to the primary of the transformer $T1$, providing a push-pull output relative to the earthed centre-tap on the secondary. $C1$ and $C2$ are capacitors which ensure that the reference side of the device presents a very low impedance to the 50-kilocycle-per-second detector square waves. It is essential that the reference voltages at F and G are of smaller amplitude than the square wave inputs at A and B, from the detector.

Consider an instant when a positive input is applied at A, the input at B then being negative. The potential at C will rise positively until it reaches the same potential as F has at that moment, where it will be effectively clamped by the conduction of rectifier $X1$. Because the potential at G must always be less negative than D, rectifier $X2$ will not be conducting; point D will be at the same potential as B. During the next half-cycle of output from the discriminator, A is negative and B is now positive. $X1$ ceases to conduct and $X2$ will allow point D to assume the potential of G. Since the reference frequency is very low compared with the frequency from the discriminator, the potential at G can be assumed to be equal in amplitude to that at F but opposite in sign.

$X3$ and $X4$ enable point E to take up the potential of C or D, whichever is the more positive at any instant. It thus alternates between the potentials at F and G at the frequency of the discriminator output, producing a waveform resembling that shown in Figure 7. In a practical case the frequency of the discriminator output is 50 kilocycles per second and that of the reference waveforms might be 55·5 kilocycles per second, so there will be many more cycles of discriminator frequency within the reference envelope than are shown in the simplified illustration.

If the discriminator output has unity on/off ratio, the waveform at E will have equal positive and negative areas and the net output will be zero after integration in resistor $R4$ and capacitor $C3$. If the output has an on/off ratio either larger or smaller than unity there will be a net positive or negative area per half-cycle of the reference waveform, resulting in a sinusoidal output at the reference frequency after integration. By reference to Figure 8 this can be shown mathematically as follows.

Consider one cycle only of the discriminator output and let f be the frequency of the reference waveform.

Then the amplitude of each half-cycle at any instant t will be $\sin 2\pi ft$.

If the discriminator output is misphased from the unity on/off condition by an amount ϕ, the width a, in radians, will be $(\pi - \phi)$ and the width z will be $(\pi + \phi)$, $a + b$ being 2π.

The average area of the waveform thus becomes

$$\frac{(\pi + \phi) - (\pi - \phi)}{2\pi} \sin 2\pi ft$$

that is

$$\frac{\phi}{\pi} \sin 2\pi ft.$$

Thus, by controlling the phasing of the discriminator an output can be obtained of any desired amplitude up to the peak values of the reference waveform.

If now the phasing (ϕ) is varying according to the bearing modulation, that is, $\phi = \sin(2\pi ft + \theta)$, where θ is the desired bearing angle, the output is

$$\frac{1}{\pi} \sin(2\pi ft + \theta) \sin 2\pi ft$$

that is

$$\frac{1}{2\pi} \{\cos(4\pi ft + \theta) - \cos\theta\}.$$

This is a direct-current component of the bearing information $(1/2\pi) \cos\theta$ together with a double-frequency term which is removed by further filtering.

By repeating the circuit three more times and applying reference inputs of $-\sin 2\pi ft$, $\cos 2\pi ft$, and $-\cos 2\pi ft$, push-pull sine and cosine components of the bearing information are produced.

In practice, $X3$ and $X4$ are replaced by a double triode acting as a long-tail pair producing the output across the cathode load at a low impedance. It will be noticed that there are no tuned circuits involved, and provided that the input from the discriminator exceeds the reference input, the circuit is virtually error-free.

10. Advantages of the Linear Detector and Multiplier Technique

Linear detection of the phase envelope of the signal translates all the bearing information to the fundamental commutation frequency, hence

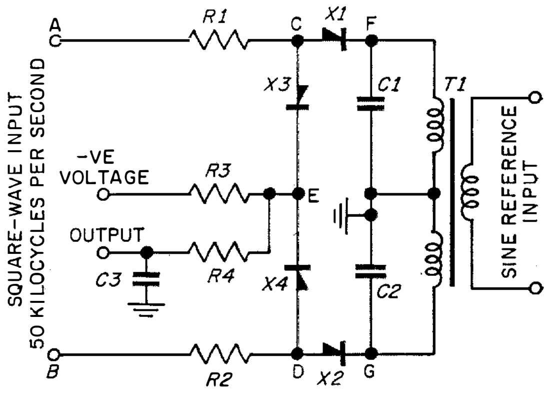


Figure 6—Simplified circuit of phase resolver.

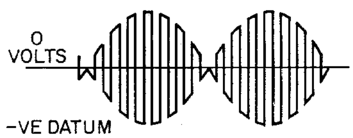


Figure 7

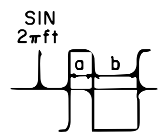


Figure 8

Figure 7—Waveform from resolver before integration.
Figure 8—Analysis of single cycle of resolver output.

providing maximum accuracy with simplicity of circuit. Sinusoidal phase discrimination gives a reduced output at the fundamental frequency, as compared with error components, plus odd harmonics. The best computation of bearing would involve the use of the harmonics as well as the fundamental.

Linear phase detection involves the use of a limiter, which completely avoids an amplitude variance component of error. Bain[4] has suggested that systems employing a signal-amplitude weighting factor may give greater accuracy than the commutated-aerial direction-finding system, but it is not clear how this weighting factor should be introduced. It is the authors' belief that an amplitude weighting required to reduce error due to small site imperfections would have to be more intelligently applied than by the simple positive weighting of phase steps according to signal amplitude. The effect of a single obstacle is to impose an alternating ripple on the 'correct' phase envelope of the signal, and the amplitude ripple due to the same obstacle is in quadrature with the phase ripple, hence increasing the error if weighting is used. On the other hand, it may well be that with very large amplitude variations due to multipath propagation it is useful to suppress the display of phase value that corresponds to near-cancellation of signal. In this latter case, as is explained in Section 8, the desired amplitude weighting is provided by a small artificial signal injected into the demodulator.

The linear detector described permits a practical circuit for multiplication of the phase envelope with the reference waves. This permits a fast display of the goniometer type (see Figure 5), and also assists in a serious problem of filtering.

When the signal phase envelope (or its differential) is produced, harmonic terms are also present, owing to site imperfections, aerial interaction, et cetera. Now, most methods of phase measurement against a reference phase would yield errors from the harmonics, so that careful filtering of the fundamental from all other components becomes necessary. The multiplier technique translates only the fundamental component to direct current, so that the necessary filtering is reduced to separation of direct current from multiples of the commutation frequency. This is an important factor, even for the slow integrated type of display, since low-frequency filtering is liable to give variable phase distortion with temperature.

A still-further advantage of linear phase detection is that it permits a theoretically perfect capture effect by the stronger of two signals, so that the bearing displayed is completely unaffected by interference (see Section 11.4).

11. Performance of the System

11.1 Site-Error Suppression

Hopkins and Horner[5] have made a theoretical estimate of site-error performance for the gyrated aerial or true Doppler direction-finder, but with the commutated-aerial direction-finding one must consider whether anything is lost by the use of a finite number of signal samples.

For the gyrated aerial, a single signal ray may be considered to be phase-modulated sinusoidally by $\pi d/\lambda$ radians, where d is the diameter of the array and λ is the wavelength. Fourier analysis of such a modulated wave shows that, when the deviation is considerable, sidebands of significant amplitude will extend up to $\pi d/\lambda$ times the gyration frequency. An interfering ray from an obstacle produces a similar pattern of sidebands, but of varying relative phase with respect to the wanted signal sidebands. The resultant total wave now contains phase variations corresponding to significant harmonics up to $\pi d/\lambda$ times the gyration frequency. To reproduce this phase envelope with accuracy, it is necessary to sample the phase at more than $2\pi d/\lambda$ times the gyration frequency. Hence the number of aerials must be greater than $2\pi d/\lambda$.

But the length of the circumference is πd, so that the adjacent aerial spacing must be less than

$$\frac{\pi d}{2\pi d/\lambda} = \frac{\lambda}{2}.$$

Consideration must now be given to the important practical case where adjacent aerial spacing is somewhat greater than $\lambda/2$. When the diameter of a commutated-aerial direction-finder array is large enough to provide a considerable improvement of site-error performance as compared with an Adcock system, maximum error for a single interfering ray is obtained when the angle between the wanted and the spurious rays is rather small. In this particular case, the phase error in the phase envelope of the signal alternates very slowly in sense as compared with cases corresponding to large angular separation between rays. Hence the sampling rate may be reduced without detriment to performance.

The reduced sampling rate, that is, the wider spacing between aerials, will, of course, cause deterioration of performance where angular separation between rays is large, but since the error is already very small, it seems reasonable to suppose that overall performance (as determined by the root-mean-square error for all angles) will suffer very little.

It now seems reasonable to conclude from this qualitative analysis that, when using a given number of aerials, overall site-error suppression will continue to improve as adjacent aerial spacing is increased from $\lambda/2$ to λ. In other words, the necessary down-grading of the true Doppler performance is very small for aerial spacing between $\lambda/2$ and λ. In any case—as will be seen from a typical practical example—the exact site-error suppression at the highest working frequency of a commutated-aerial direction-finding system is of little significance, the performance being such as to make all reasonable site imperfections quite undetectable.

Consider a typical commutated-aerial direction-finding system using 18 aerials and covering the range from 100 to 400 megacycles per second. If the adjacent aerial spacing is λ at 400 megacycles per second, true Doppler performance can be expected between 100 and 200 megacycles per second, this corresponding to aerial diameters between $1 \cdot 5\lambda$ and 3λ. Reference to the Hopkins and Horner paper[5] suggests a mean error suppression as compared with the Adcock system varying between about 7 and 18 times. Above 200 megacycles per second the suppression factor will continue to rise, to a maximum value somewhat less than the 48 indicated by the Hopkins and Horner[5] graph at a diameter of 6λ, or 400 megacycles per second.

11.2 Suppression of Errors due to Multi-Path Propagation

Bain[2,6] has made a detailed study of the effects of multi-path propagation on commutated-aerial direction-finders. His theoretical conclusion—which has now been verified qualitatively in practical equipment—is that significant improvement over the Adcock system is obtained when the aerial array has a diameter of 3λ to 4λ and becomes considerable at a diameter of 10λ. It is interesting to note that Bain concludes that aerial spacing up to at least λ is permissible without deterioration in performance. This, of course, would be expected from the argument in the last section, since the unwanted signal rays have small angular separations from the wanted or strongest ray.

11.3 Suppression of Polarization Error

Polarization errors in the Adcock direction-finder arise from aerial response to horizontal polarization, and exist, theoretically, only through lack of aerial balance. However, it is difficult in a practical case to set up a balance that suppresses horizontal pick-up by more than about 30 decibels as compared with response from vertical polarization. Furthermore, it is not normally useful to set up a better apparatus balance, for irregularities of a practical site will, in effect, produce an unbalance of this magnitude. Although concerned with sites that would be considered rather poor for direction-finding purposes, Saxton and Harden[7] have made measurements showing an average polarization discrimination of less than 20 decibels.

Now, it can easily be shown that, for equal suppression of horizontal polarization in an Adcock system and a commutated-aerial direction-finding system of radius 1 electrical radian, performance of the two would be equal, and that error suppression of the commutated-aerial direction-finder increases linearly with the diameter of the aerial system. It is likely, particularly for aerial arrays of small diameter, that ease of achievement of a particular factor of aerial balance is not the same for the two systems, but for normal sizes of commutated-aerial direction-finding array it is certainly possible to suppress horizontal pick-up by the practical factor of 30 decibels. Hence it is a reasonable approximation to say that the commutated-aerial direction-finding system gives an improvement factor proportional to diameter, and equal to π times the diameter of array in wavelengths. Thus, in the typical practical case of an 18-aerial commutated-aerial direction-finding system for 100 to 400 megacycles per second, polarization performance exceeds that of the Adcock system by a factor that varies between about 4½ and 18. Bain[2] has considered a high-frequency system of 24 unipoles, and also concludes that polarization errors are much reduced as compared with the Adcock system.

11.4 Suppression of Errors due to Interference

In most forms of automatic direction-finder, bearing is determined by measurements on the amplitude envelope imposed on the signal by the aerial. When two signals of slightly different frequency pass through the selective circuits of the receiver, the stronger of the two exercises a strong capture effect in that the demodulated output is representative almost entirely of the stronger signal. However, this capture can never be complete, since the root-mean-square value of the wanted signal is effectively increased by a variable amount according to the instantaneous power level of the weaker signal. Hence, when two signals are present, the automatic presentation of bearing represents something intermediate between the bearings of the two signals, considerably weighted in favour of the stronger signal.

Now, in the commutated-aerial direction-finding system, an interfering signal of weaker amplitude than the wanted signal can only modulate the phase at a particular aerial symmetrically about the correct mean. Hence, neglecting complete breakdown of the phase discriminator when presented with phase values of more than ± 180 degrees, the stronger signal exercises a complete capture and the correct bearing of the stronger signal is displayed.

Since no amplitude modulation is imposed on the commutated-aerial direction-finding signal by the aerial, it is more likely that the wanted signal will be stronger than the unwanted one over the whole modulation cycle than in systems dependent on amplitude modulation.

This effect of complete capture proves to be of very special value in an automatic 'fixer' system comprising, say, about 8 direction-finding centres. In such a system it is most unlikely that very strong intereference could be experienced simultaneously at all the sites, and in cases of serious jamming an automatic fixer system should tend to display both the wanted signal position and the source of the interference, each with accuracy, at the same time.

11.5 Overhead Performance

Direction-finders operating in the very-high-frequency region, and more recently the ultra-high-frequency bands, have commonly been used for 'homing' aircraft, and the erratic indications when the aircraft passes overhead have often been used to indicate the overhead condition. In an average case the Adcock direction-finder starts to give 'wild' indications when the elevation of the aircraft is greater than about 45 degrees.

It has been found necessary to exercise extreme care in the use of wildly fluctuating bearings, since the characteristics of many sites are such that an occasional 'false overhead' has been indicated at very low, even grazing, angles of elevation. Such false indications arise from a number of different causes. Most often they have been traced to the effect of slightly 'saucered' terrain, which can magnify the intensity of ground reflection until cancellation with the direct ray is produced. In others, the characteristics of the aircraft transmitting aerial have been such as to cause radiation of an excess of horizontal polarization in some directions. In another observed case it seemed probable that false operation of a short-basc-line equipment was produced by multi-path effects when the signal passed via irregular banks of trees on a distant horizon. It seems likely that a very thorough exploration of any site could yield false operation of an Adcock type of equipment at some particular azimuth and elevation.

In 1956, extended simultaneous field trials were carried out with an ultra-high-frequency commutated-aerial direction-finding system using 12 aerials on a circle of about λ diameter and a typical short-base system. In these trials, maximum signal elevation angle for dependable operation on the commutated-aerial direction-finding system was about 75 degrees. With the short-base equipment, operation was usually 'safe' up to 45 degrees of elevation, but occasional 'false overheads' were recorded for small angles of elevation, including consistent false operation with an aircraft flying at 10 000 feet at a distance of 100 miles in one particular azimuth. The performance of the commutated-aerial direction-finding equipment was such that no false overheads were ever recorded, and the steadiness of display was such that using trace length (which was proportional to phase deviation of the signal) as indicative of the cosine of angle of elevation, accurate estimation of any elevation between 30 and 60 degrees could be made.

The present trend is to use rather larger aerial dimensions, and furthermore, the preferred technique is to use a set of unipoles on a large horizontal circular counterpoise. The effect of the counterpoise is to fill in high-angle slots in the vertical polar diagram of aerials, and therefore to provide nearly perfect overhead performance. The only noticeable effect of high-angle signals is the shortening of display trace according to the cosine of the angle of elevation.

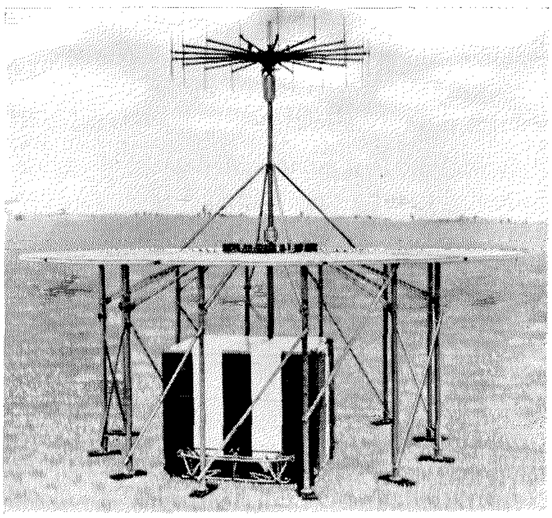


Figure 9—A transportable very- and ultra-high-frequency commutated-aerial direction-finder.

The first practical example of nearly perfect overhead performance occurred during October 1957, when a commutated-aerial direction-finding equipment was used to track the first Russian satellite with considerable success. Using the 20·005-megacycle-per-second transmission from the satellite, it was found possible to plot azimuth with good accuracy up to the maximum elevation obtained, despite the use of only the vertically polarized component of the signal. The particular equipment used unipoles mounted at ground level, hence completely avoiding high-angle slots, and furthermore, the aerial diameter was approximately 7λ.

12. Conclusions

The paper has described how the more-fundamental problems in the development of the commutated-aerial direction-finding system have been solved, and shows that a performance consistent with early predictions has now been obtained.

It is regretted that space does not permit the inclusion of a representative record of practical results, but it can now be stated that the comparative immunity of the system from site and polarization errors in the very- and ultra-high-frequency bands is already demonstrating a considerable improvement in the accuracy of aircraft fixer systems.

A typical modern commutated-aerial direction-finding system is shown in Figure 9. The system illustrated is a transportable very- and ultra-high-frequency station providing simultaneous operation of two direction-finding channels in each of these bands.

13. References

1. C. W. Earp and R. M. Godfrey, "Radio Direction-Finding by the Cyclical Differential Measurement of Phase," *Electrical Communication,* volume 26, pages 52–75; March 1949: and *Journal of the Institution of Electrical Engineers,* Part IIIA, volume 94, pages 705–721; 1947.

2. W. C. Bain, "Possible Errors of a Particular Wide-Aperture Direction Finder," *Proceedings of the Institution of Electrical Engineers,* Part C, volume 103, pages 313–324; March 1956: Monograph 170R.

3. C. W. Earp and D. L. Cooper-Jones, British Patent Application 9095; 1957.

4. W. C. Bain, "Theoretical Design of Direction-Finding Systems for High Frequencies," *Proceedings of the Institution of Electrical Engineers,* Part B, volume 103, pages 113–119; January 1956: Paper 1960R.

5. H. G. Hopkins and F. Horner, "Direction-Finding Site Errors at Very High Frequencies," *Proceedings of the Institution of Electrical Engineers,* Part III, volume 96, pages 321–332; July 1949: Paper 773R.

6. W. C. Bain, "Calculation of Wave-Interference Errors on a Direction-Finder Employing Cyclical Differential Measurement of Phase," *Proceedings of the Institution of Electrical Engineers,* Part III, volume 100, pages 253–261; September 1953: Paper 1545R.

7. J. A. Saxton and B. N. Harden, "Polarization Discrimination in V.H.F. Reception," *Proceedings of the Institution of Electrical Engineers,* Part B, volume 103, pages 757–760; November 1956: Paper 2184R.

Wide-Base Doppler Very-High-Frequency Direction Finder*

F. STEINER

Standard Elektrik Lorenz AG; Stuttgart, Germany

Although it is quite well known that bearing errors occurring with multipath propagation can be reduced by using wide-base antenna arrays, only systems operating with phase or frequency modulation have come into practical use.

A method of progressively scanning a circular antenna array is described. This scanning, which produces a frequency modulation in which the frequency deviation is a function of the bearing information, is accomplished by a rotating capacitive switch, and no vacuum tubes or crystal diodes are employed. Moreover, a special light indicator replaces the usual cathode-ray tube. Designed primarily for the very- and ultra-high frequencies, the equipment is quite simple and of high reliability.

1. Introduction

Bearing errors resulting from multipath propagation have been discussed in various papers.[1,2] These errors, which amount in the most unfavorable cases to several times the angle between the direct and the reflected waves, discredited direction finders especially in the very- and ultra-high-frequency bands. Since the direction finder in the absence of such errors is a very useful aid in navigation, suggestions have been made[3,4] to reduce these errors by wide-base antenna arrays. The mathematical treatment of the problem appears complicated; however, it can be simplified by comparing maximum bearing errors of a wide-base antenna array with those of a narrow-base array.[5,6] Figure 1 shows the improvement coefficient obtained against D/λ, where D is the diameter of the antenna array and λ is the operating wavelength. It will be seen that this coefficient increases proportionally with D/λ when $D/\lambda > 1$. It is interesting to note that this characteristic is essentially independent of the method employed; it is approximately valid for the wide-base Adcock array as well as for methods operating with phase or frequency modulation. However, in the former, the ambiguity increases with the base and cannot be eliminated except with complicated methods; in the latter, unambiguous bearings are obtained regardless of the base size.

These methods can be applied practically by the cyclical differential measurement of phase[7] and by the utilization of the Doppler effect so that the bearing information is contained in a frequency modulation resulting from an actual or simulated antenna motion.[8,9]

In the differential measurement of phase, a considerable number of antennas are uniformly distributed along the circumference of a circle, the phase of the voltage received by each antenna being cyclically compared with that from the next one. A phase discriminator in the output circuit of a receiver converts these phase differences into direct-voltage values proportional to the phase differences. These direct-voltage levels give a sequence of alternations or an alternating voltage, containing the bearing information. This method requires special means to obtain the frequency stability that is

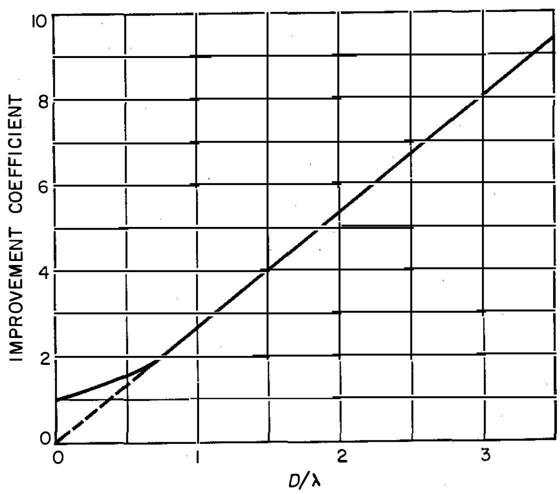


Figure 1—Ratio of maximum bearing errors of an Adcock to a wide-base direction finder plotted against the width of the antenna array in operating wavelengths.

*Reprinted from Institute of Radio Engineers *Transactions on Aeronautical and Navigational Electronics,* volume ANE-7, pages 98–105; September 1960.

necessary for the measurement of phase independent of a varying transmission frequency.

In the systems using the Doppler principle, the antenna rotates or such rotation is simulated by the sequential scanning or switching of individual antennas. This results in phase modulation that appears as a frequency modulation of the received voltage and as an alternating voltage in the output of a frequency discriminator, the latter supplying the bearing information. Here the difficulties encountered are concerned with conflicting conditions that affect the selection of the rotational frequency.

Basically, there is no difference between these two methods. The frequency modulation of the Doppler system results from phase modulation obtained through differentiating while the other system produces similar modulation by direct differential measurement of phase.

This paper will show how the difficulties inherent in the Doppler direction finder have been avoided and how a particularly simple unit has been designed on this principle.

2. Conditions for the Doppler Direction Finder

To ensure simple evaluation of the frequency modulation generated by the rotation of an antenna, the frequency deviation should have a reasonable relation to the bandwidth of the receiver used. Although the evaluation of the frequency containing the bearing information is a selective process, interfering waves, such as undesired frequency modulation of the transmitter or noise, may produce undesired effects if nonlinearities exist in the amplifiers preceding the resolver and if the amplitude of the interfering wave is several times that of the signal.

This could only be prevented by overdimensioning the amplifiers and the phase-sensitive demodulators. In a very-high-frequency receiver, for example, the bandwidth can hardly be narrowed to less than 50 kilocycles per second; this is primarily due to the limited frequency stability of transmitters and receivers. Hence, it would be reasonable to operate with frequency deviations Δf of the order of ± 10 kilocycles per second.

It can be shown that the rotation of an antenna results in a frequency deviation Δf that can be determined by the modulation index m. The latter can be calculated from the diameter of the antenna-array circle D and is $m = \Delta f/fn = \pi D/\lambda$. Various experiments have shown that a $D/\lambda = 3$ (corresponding to an improvement coefficient of 8 in Figure 1) will yield satisfactory results in most cases; hence, an $m = 10$ should be taken as a basis. This would mean a rotational frequency $fn = \Delta f/m$ of about 1000 cycles per second. Considering the required diameter of the array, this cannot be achieved with mechanical rotation of a dipole, but it can well be approximated by cyclical scanning of antennas arrayed in a circle. The approximation is almost perfect if means are provided for a continuous transition from antenna to antenna and if the spacing between adjacent antennas does not essentially exceed $\lambda/3$. This close spacing for the 30 antennas in the system under discussion required special dimensioning of the antenna elements to reduce the mutual impedances between them. The desired frequency deviation of $\pm$ 10 kilocycles per second can thus be achieved by a sufficiently high equivalent rotational frequency. It will be shown, however, that the transient response does not readily tolerate this high rotational frequency.

The transient response of a receiver is determined by the sums of the derivatives of the phase angles depending on the angular frequency $d\phi/d\omega$ of the individual selective circuits. It produces a phase shift of the modulating frequency that increases with frequency and with decreasing bandwidth of the receiver. In the example of the very-high-frequency receiver, it is of the order of 10 degrees per kilocycle per second. This value is neither constant nor subject to calibration, but varies with such factors as limiting, automatic control, and frequency drift. To be on the safe side, at least half of the maximum value should be taken into account, which is 5 degrees per kilocycle per second. Allowing for a maximum inherent bearing error of 0.5 degree, a tolerable rotational frequency for the antenna of not more than 100 cycles per second will be obtained. This is in sharp contradiction to the requirement stated above. The fulfilment of this requirement is possible without the said errors, however, if the problem is solved by the method described below. Figure 2 shows the relations between the desirable and the permissible rotational frequencies depending on the received frequency.

3. Solution of the Problem

It is possible to obtain a sufficiently large frequency deviation resulting in no bearing errors if the simulated motion of the antenna is not along a circle but is to-and-fro on a straight line. This can be provided by sequential switching of a number of antennas mounted along such a straight line. If this line of antennas is made to rotate, as in Figure 3A, the frequency modulation will disappear when the direction of incidence is perpendicular to the antenna line. The bearing information can be determined by just the magnitude of the frequency deviation, and the site criterion by the sign, not using the phase; hence, it is possible to operate with scanning frequencies of several kilocycles per second without bearing errors.

If enough time (several seconds) is available to take a bearing, the line of antennas can be rotated to the minimum automatically by means of a servomechanism. If it were mechanically possible, such a line of antennas could be rotated at a speed corresponding to 100 cycles per second and no noticeable bearing errors would be generated since errors due to the transient response would not be produced by the high scanning frequency but only by the low rotational frequency.

The combination of the two motions, that is, the to-and-fro motion and the slow rotation, can be obtained by a simple arrangement shown in

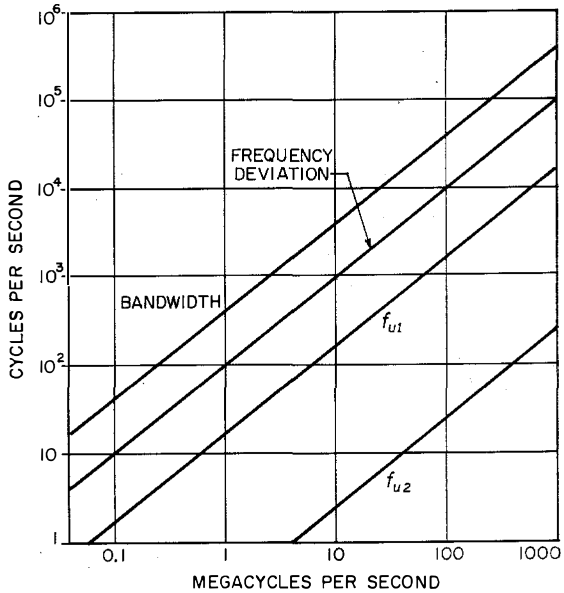


Figure 2—Permissible rotational frequencies as a function of operating frequency. Curve f_{U1} is the rotational frequency for 20 antennas in a circular array having a base of 2 wavelengths giving an improvement ratio of 5.5 and a group-delay error of ±8 degrees. Curve f_{U2} is for 320 antennas on a 32-wavelength base giving an improvement coefficient of 88 and a group-delay error of ±0.5 degree.

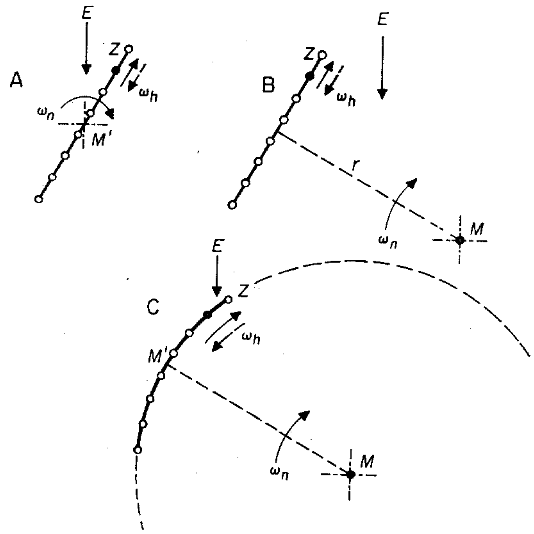


Figure 3—Scanning of antennas along a line that is an arc of a circle.

Figure 3. The antenna line Z is scanned at a high frequency ω_h. However, it does not rotate around its center M' as in Figure 3A, but on the circumference of a circle having the radius r and the center M as in Figure 3B. This arrangement provides, in practice, the same results as if the antenna line rotated around M', and as if it were no longer straight, but of a curvature corresponding to this circle of the radius r (Figure 3C).

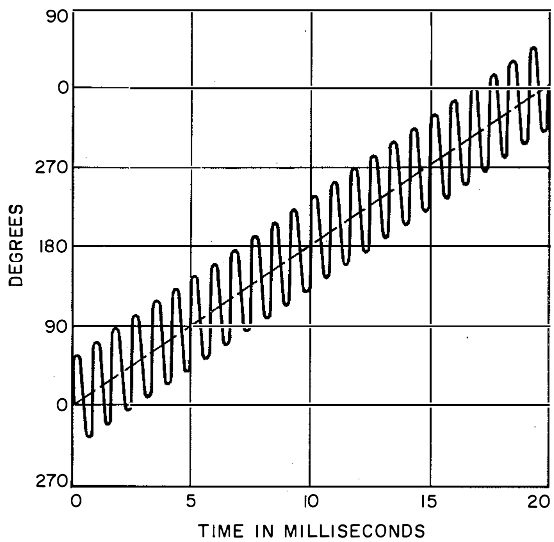


Figure 4—Virtual position of the receiving antenna during one cycle of scanning.

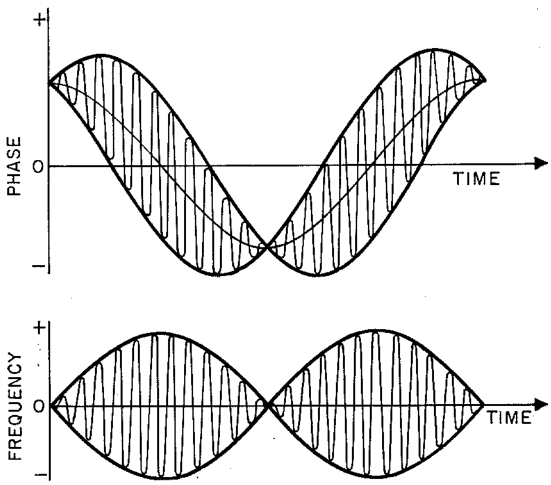


Figure 5—Changes in phase and frequency as a result of scanning. The fine pattern corresponds to the to-and-fro scanning and the envelope to the slower rotational frequency.

The desired combination of the two motions can be simulated by a scanning switch connecting sequentially to that number of antennas of a circular array that would be effective if there were physical motion. Based on these considerations, an antenna array was constructed. The calculations of Section 2 had suggested 30 antennas as the optimum number. Assuming an antenna line of one sixth the length of the circumference, the deviation of this line from a straight line may be neglected.

A progressive scanning program was employed to provide the two motions. After scanning 5 antennas in a backward direction, 6 antennas are scanned in the forward direction. In other words, 5 backward steps are followed by 6 forward steps. After this total of 11 steps in which the antenna line has been scanned once, the scanning program has advanced along its circular path by 1/30th of the circumference. It follows that the frequency deviation obtained by rapid scanning is 11 times as high as with only the slow rotation, provided each dipole is connected to the receiver for the same time interval. In this example, the ratio of the scanning frequency to the rotational frequency is 30:1. The to-and-fro motion can also be made approximately sinusoidal by suitable choice of switching times for the individual dipoles.

For this case, Figure 4 shows the virtual position of the receiving antenna as a function of time; Figure 5 indicates the resulting phase modulation and the corresponding frequency modulation. At a rotational frequency of 50 cycles per second, the output voltage of a frequency-demodulating receiver will be an audio frequency voltage of 1500 cycles per second modulated by an envelope of 50 cycles per second; the bearing information is contained in this 50-cycle-per-second envelope.

The antenna scanning at the desired sequence could be accomplished, for instance, by switching diodes controlled by pulse generators, which are of a somewhat complicated design. However, this statement is true only for sudden switching from one antenna to the next. Continuous scanning, more closely resembling the actual motion, can be achieved with a capacitive commutator of a suitable design.

In the installation described above, the capacitive commutator had 30 segments in each of two stators. The segments of one stator are connected individually to the 30 antennas, and the 30 segments of the other stator are connected to one another and to the receiver. Conducting bars on the rotor are arranged to overlap the segments on the stators and provide capacitive coupling from one antenna at a time to the receiver. These rotor bars are positioned to produce the desired progressive scanning. Figure 6 shows the relative positions of the stator and rotor, which are of cylindrical form. In the position shown, antenna 2 is connected to the receiver.

To simplify the description, Figures 6 and 7 are based on the assumption that 3 forward steps are followed by 2 backward steps. Figure 7 shows how the uniform motion of the rotor results in successive connections to the receiver of the antennas 1, 2, 3, 4, 3, 2, 3, 4, 5, 4, 3, 4, 5, 6, et cetera. Figure 8 is a photograph of a commutator. The switching capacitances are compensated by series inductances, and the output parallel capacitances by parallel inductances. This arrangement provides good wideband matching. For a frequency range of about 1:2, the attenuation is in the order of 3 decibels. No nonlinear elements are employed in this circuit; hence, all possibility of cross modulation is precluded.

The cost of this type of commutator is not critically dependent on the number of antennas to be connected to it.

The antennas constituting the circular array should have very low mutual impedances to each other. They should also be connected to coaxial transmission lines in such a manner that the reflection coefficient over a substantial frequency range is low. The first requirement is necessary to avoid distorting the field produced by the signal that actuates the antennas. The

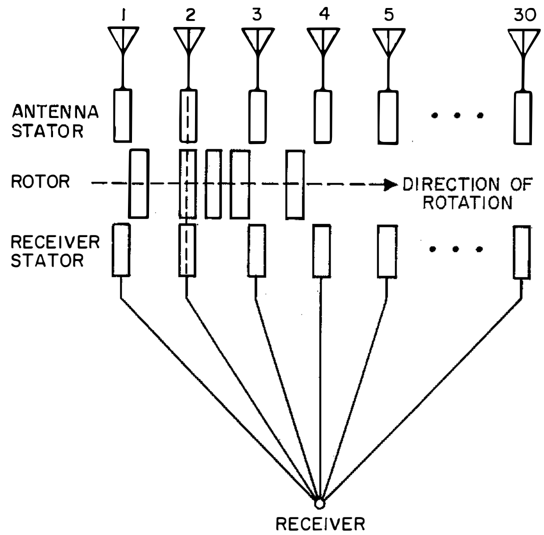


Figure 6—Capacitive commutating arrangement. The spacing of the rotor elements produces 3 forward steps followed by 2 backward steps.

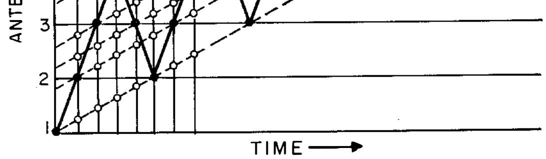


Figure 7—Scanning pattern. The five circles in vertical alignment correspond to the rotor bars, which move upward across the antenna segments on the stator. The black dots indicate when each rotor bar is closely coupled to an antenna segment.

Figure 8
The rotor and stator of the capacitive commutator.

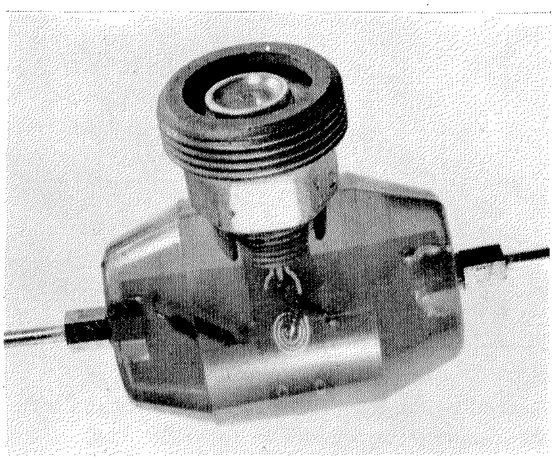


Figure 9—Prototype antenna transformer cast in transparent resin.

second requirement permits the antennas to be at a substantial distance from the commutator, without unduly limiting the operating frequency range.

Although the design of high-frequency antennas is well known, development work had to be done on very- and ultra-high-frequency antennas. Naturally, the requirements could not be met except at the expense of some sensitivity. The dipoles were shortened somewhat so that their radiation resistance could be regarded, to a first approximation, as a resistance connected in series with a capacitive reactance. The desired wideband property and decoupling is achieved by an additional series resistance.

Each antenna is connected to its transmission line through a balance-unbalance transformer that presents at the antenna side a damped series-resonant circuit due to its leakage inductance, the antenna impedance, and an additional damping resistance. At the transmission-line side, its reactance is compensated by a parallel resonant circuit formed by the winding inductance and a capacitor.

These measures ensure not only the wideband frequency characteristics, but also serve to bring about the desired decoupling of the antennas from each other. The resulting loss in sensitivity is kept within tolerable limits. Figure 9 shows a prototype transformer using a transparent casting resin, and Figure 10 shows the actual transformers employed in this installation. On the basis of this design, the individual antenna is so simple that there can be no objection to a large number of antennas.

4. Display of Bearing Information

The use of a rotating capacitor for antenna scanning permits the simultaneous generation of the reference voltages required for the utilization of the output signals from the receiver.

Referring to Figure 5, it will be remembered that the receiver output voltage carries the bearing information in the envelope of the phase variations. If two identical reference voltages are generated so that their envelopes differ by a phase angle of 90 degrees, the products of any one of them and of the receiver output voltage will be in the form of direct-current components corresponding to the sine or the cosine, respectively, of the angle of incidence. Figure 11 is a diagram showing how the rotor of the antenna-commutating assembly is driven by an asynchronous motor that also drives the reference-voltage generator and the search coil of a goniometer that produces reference sine and cosine functions in its two field coils. These react individually with the receiver output in the light indicator to position the light spot on the indicator scale.

The driving motor is of the outer-rotor type; that is, the short-circuited armature rotates around the periphery of a fixed field winding. The armature is mechanically coupled to a 1500-cycle-per-second tone generator. The voltage of this generator is applied to the rotating primary winding of a low-frequency goniometer. The fixed secondary winding of the latter is closed on itself and is tapped at 4 points spaced 90 degrees apart. Each pair of taps that are 180 degrees apart provide one of the two reference voltages. This design avoids the use of slip rings and brushes.

Although the tendency would be to use a cathode-ray oscilloscope for an indicator, there are some desirable features to the use of a different arrangement.

An ideal product-forming setup is represented by a wattmeter in which the torque acting on the shaft is strictly proportional to the currents through stator and rotor and to the cosine of the phase shift. Since the bearing information is contained in a bandwidth not exceeding 2 cycles per second, a transient time of 0.5 second is sufficient. This can be achieved easily with such devices. Hence, the mechanical properties of the device provide the function of the low-pass filters otherwise indispensable in the output circuit of a phase-sensitive demodulator.

The combining of the torques of two such devices, corresponding to the sine and the cosine

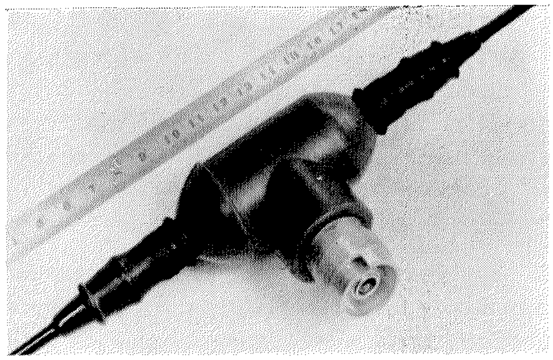


Figure 10—Final design of transformers used in described installation.

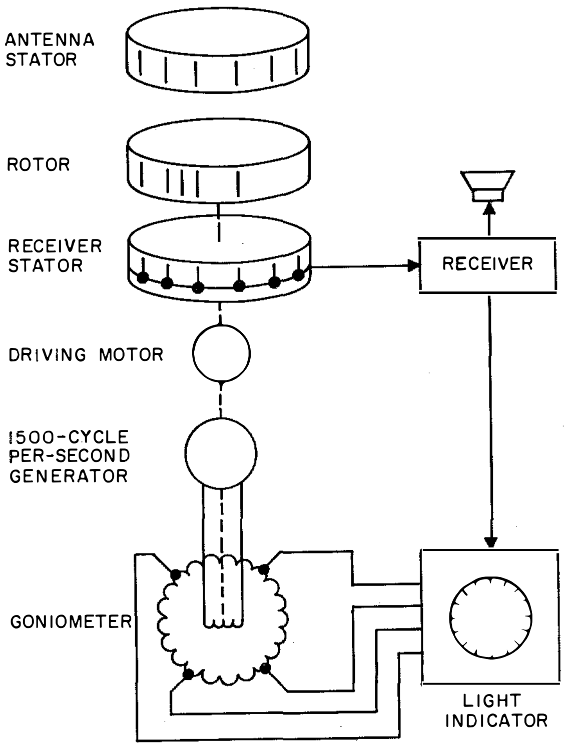


Figure 11—The elements above the driving motor connect the receiver to each antenna in prescribed order. The alternating current from the generator below the motor flows through the rotating coil of the goniometer. The voltage induced in the field coils of the goniometer are used as references for the watt-meter-type light-spot indicator. The position of the light spot corresponds to the angle of the incoming wave to the direction-finder antenna.

of the angle of incidence, is accomplished optically by mounting a small mirror on each device as shown in Figure 12. The two mirrors deflect a spot of light in the two axes of a system of coordinates. Several such devices can be combined for simultaneous projection of markers differing in size and/or color.

If supplemented by a simple electromagnetic equipment, this light indicator can also be used to store the displayed bearing information until a new signal is obtained from the receiver. Arrangements can be provided by which the displayed information is retained until a new signal is obtained from the receiver. Thus, the bearing information obtained from a brief transmission can be observed for any desired length of time. If cathode-ray oscilloscopes are used for the same purpose, much-more-complicated equipment is required.

The final choice of an indicator depends on the information content of the received wave and the use to be made of it. If only direction is desired, any display on a 360-degree indicator will suffice. If the quality of the operating signal is to be checked, a milliammeter measuring the current in the automatic-gain-control circuit will indicate the strength of the signal. The same results could be obtained with a cathode-ray tube but only at much greater expense. Nevertheless, any phase-measuring system can be adapted as an indicator for this direction finder.

5. General View

Figure 13 is a block diagram of a direction finder consisting of the units described, and Figure 14 shows the outside of a station with the antenna array supported from the roof, the capacitive commutator being mounted in the turret carrying the antennas. Figure 15 is a sectional view of the installation. The rack is equipped with two ordinary amplitude-modulated-wave receivers, a panel with the frequency-

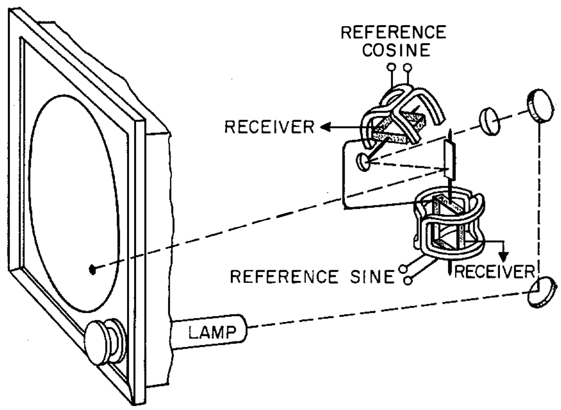


Figure 12—Light-spot type of indicator. Each wattmeter type of moving-coil system positions a small mirror in one coordinate in response to the signal from the receiver and that from one phase of the reference generator. These two mirror positions determine the position of the light spot on the scale. The entire system is mounted within a lighttight box.

modulated-wave output circuits for both receivers, and a panel with remote control switches, et cetera. Two display units are used as slave indicators for monitoring purposes in this installation. The number of slave indicators can be arbitrarily increased and installed, for instance, in the control tower of an airport. The latter application is shown in Figure 16.

Vacuum tubes are employed only in the receivers and there are 18 tubes per receiver. The reliability is, accordingly, very high. The direction finder operates unattended and can be switched from a remote operating position that uses a slave indicator.

The slave indicators for operation over several kilometers of telephone lines are equipped with transistor-type amplifiers having negative feedback. All of them permit individual selection of the desired channel and of the bearing reference direction, irrespective of the operation of the other slave indicators.

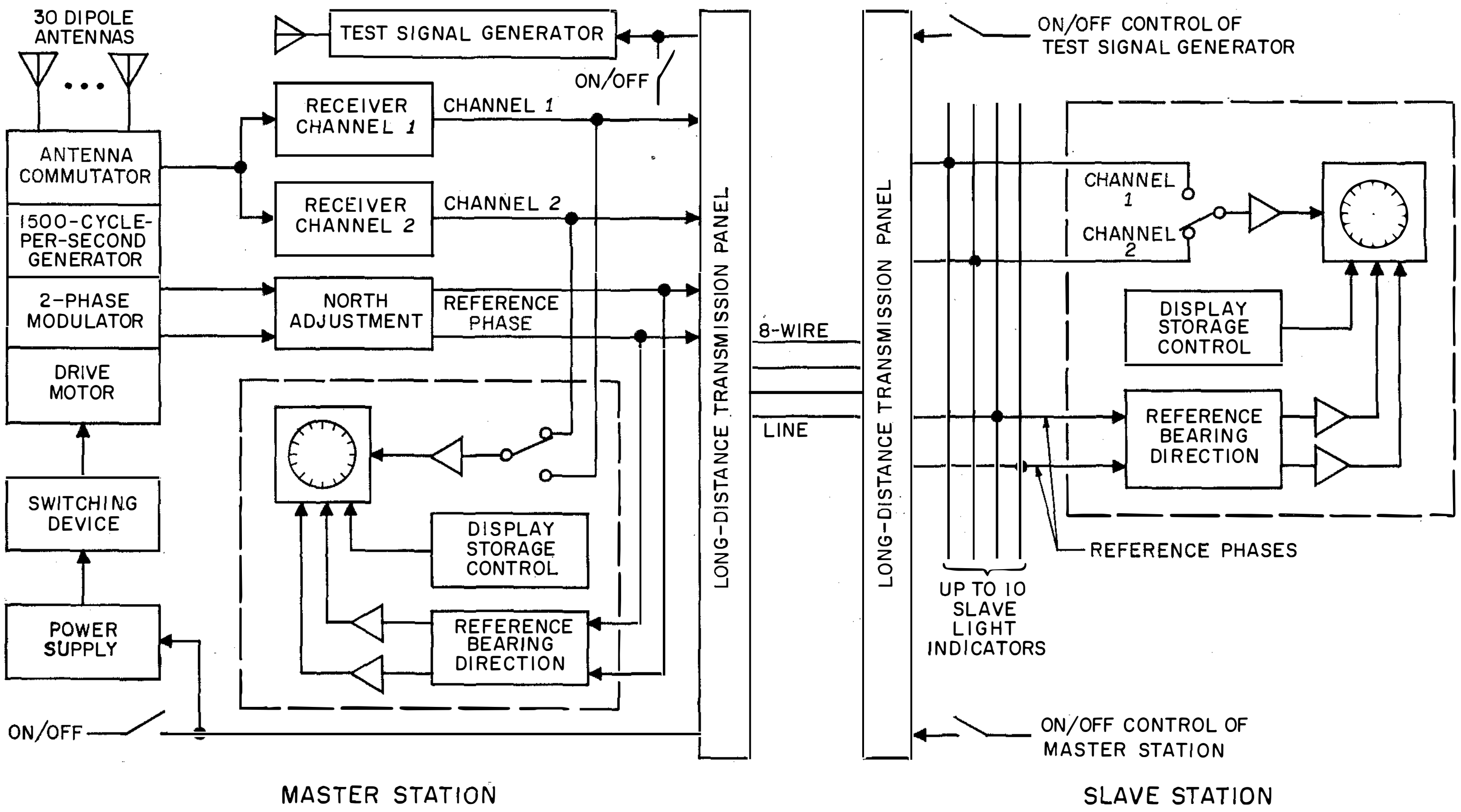


Figure 13—Block diagram of master and slave stations connected by telephone lines, which can also handle a speech circuit simultaneously.

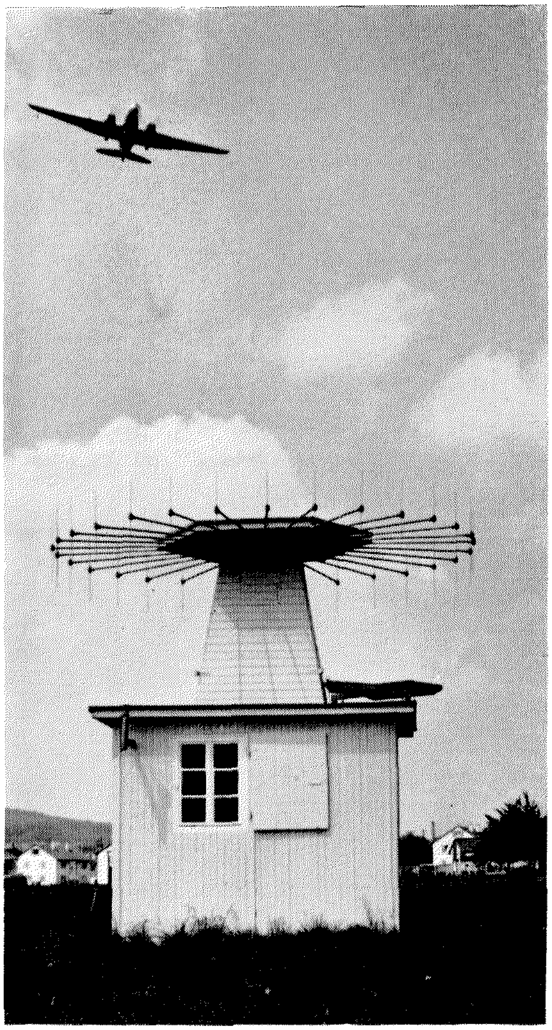


Figure 14
Photograph of Doppler direction-finding station.

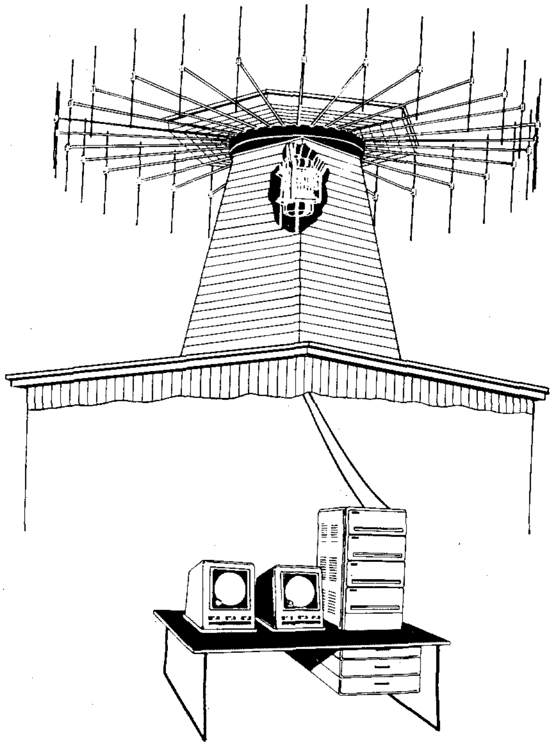


Figure 15—Disposition of elements in a station.

6. Conclusion

The conditions for the selection of a suitable rotational frequency of a direction finder based on the Doppler principle are contradictory: The frequency should be high to reduce interference and low to prevent errors depending on tuning adjustments. It has been shown that this contradiction was circumvented by progressive scanning of the antennas. The rotating capacitive antenna switch used for this purpose eliminates nonlinear components in the antenna circuits and permits simple generation of the reference voltages required for the bearing designation. Indication is accomplished by an arrangement of wattmeters. This design permitted the construction of an automatic wide-base circular-array direction finder using vacuum tubes only in the receiver.

The large number of antennas of simple construction and the symmetry achieved with only mechanical means permit the inherent direction-finding errors to be kept small; they amount to about ± 0.5 degree.

The bearing fluctuations caused by unfavorable site conditions are only about a tenth of those encountered with an ordinary Adcock direction finder.

Figure 17 shows results of flight tests made by the German Aviation Administration with such

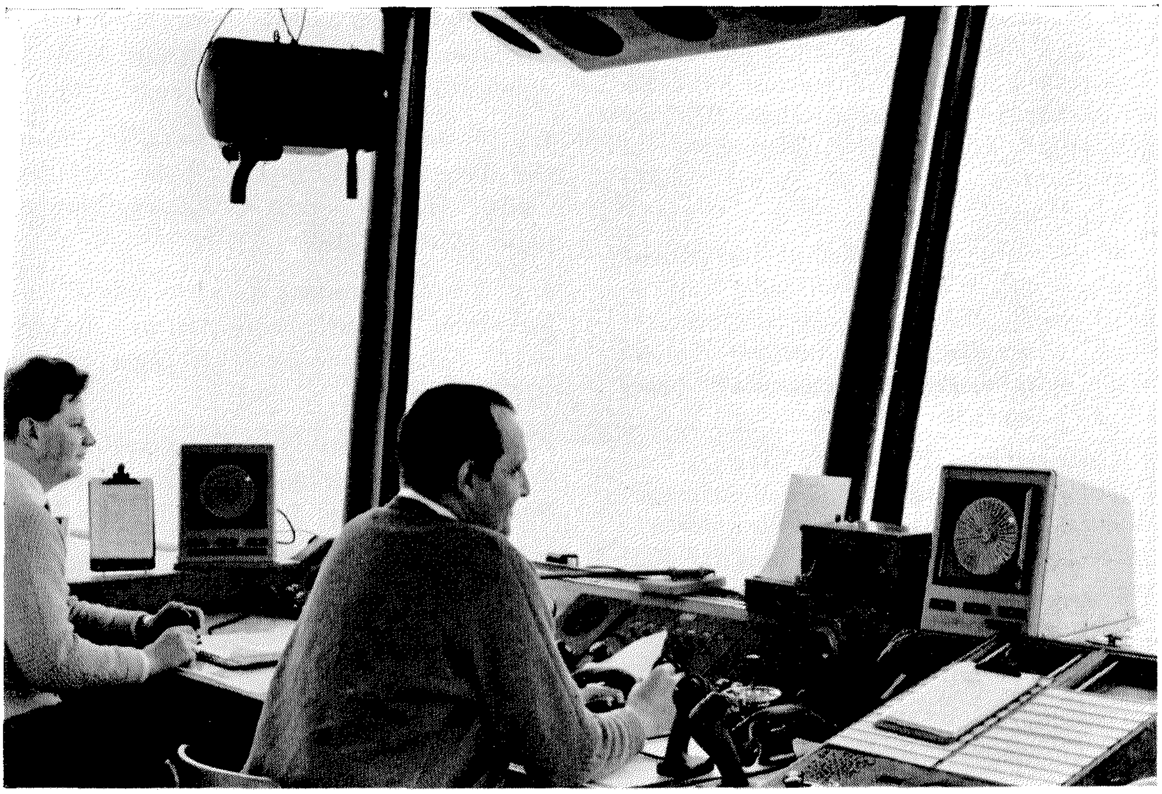


Figure 16—Installation of two slave light indicators in the control tower at an airport.

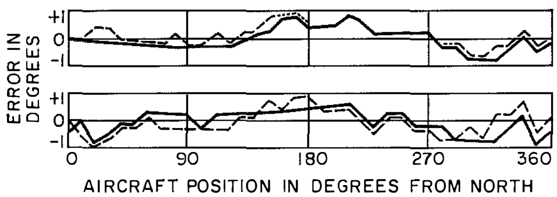


Figure 17—Error curves made on four test flights at Stuttgart airport. The circular flights were at a radius of 4 miles from the station and the broken and solid lines are for the two directions of circling. The flight altitudes were 3000 feet (upper curve) and 4000 feet (lower curve). Frequency was 122.3 megacycles per second.

equipment at the Stuttgart airport and with unfavorable site conditions.

The application possibilities of the direction finder, for instance in air-traffic control, have already been discussed.[10,11] This equipment combines the advantages of high direction-finding accuracy and, as a result of its simple design, exceptional reliability.

7. References

1. T. Heiligtag, "Über die Gründe der Missweisung beim Richtungs empfang," *Jahrbuch der drahtlosen Telegraphie,* volume 21, pages 77–100; 1923.

2. J. Grosskopf, "Über die Vorgänge im Interferenzfeld von Raumwellen," *Hochfrequenztechnik und Elektroakustik,* volume 62, pages 173–183; 1943.

3. H. Rindfleisch, "Die Grossbasis-Peilanlage 'Wullenwever,'" *Nachrichtentechnische Zeitschrift,* volume 9, pages 119–123; March, 1956.

4. C. W. Earp and R. M. Godfrey, "Radio Direction Finding by the Cyclical Differential Measurement of Phase," *Journal of the Institute of Electrical Engineers,* volume 94, Part IIIA, pages 705–721; 1947. Also, *Electrical Communication,* volume 26, pages 52–75; March, 1949.

5. W. C. Bain, "Calculation of Wave-Interference Errors on a Direction-Finder Employing Cyclical Differential Measurement of Phase," *Proceedings of the Institute of Electrical Engineers,* Part III, volume 100, pages 253–267; September, 1953.

6. F. Steiner and H. Stittgen, Über die Verringerung von Ortungsfehlern bei Mehrfacheinfall durch Grossbasissystem," *Nachrichtentechnische Zeitschrift,* volume 11, pages 417–423; August, 1958.

7. C. W. Earp and D. L. Cooper-Jones, "Practical Evolution of the Commutated-Aerial Direction-Finding System," *Proceedings of the Institution of Electrical Engineers,* Part B, volume 105, pages 317–325; March, 1958.

8. J. A. Fantoni and R. C. Benoit, Jr., "Doppler Type High Frequency Radio Direction Finder," *IRE Convention Record,* Part 8, pages 165–171; 1956.

9. F. Steiner, "Grossbasispeiler nach dem Dopplerprinzip," *Nachrichtentechnische Fachberichte Zeitschrift,* volume 11, pages 85–90; 1958.

10. I. F. Bond, Type PV 1-B Automatic VHF Cathode-Ray Direction Finder, *Buecherei der Funkortung,* volume 5, number 3, pages 109–118; 1955.

11. H. Diekmann, "Der Dopplerpeiler in der Flugsicherung," *Buecherei der Funkortung,* volume 7, number 9, pages 93–106; 1959.

Printed Circuits

J. M. C. Dukes has published a book on printed circuits that is based on his broad experience with Standard Telephones and Cables and its associated companies in England. It is divided into 10 chapters, 2 through 5 being on manufacture and 6 through 10 devoted to design and application. The chapters and 2 appendixes are titled as follows.

1. Introduction
2. The Direct Deposition of a Metallic Circuit on Bare Insulating Surfaces
3. The Controlled Removal of Metal from a Previously Metallized Surface
4. Associated Circuit and Panel Fabrication
5. Component Assembly, Dip Soldering, and Final Protection
6. Preparation of the Circuit Information
7. Materials for Printed Circuits
8. Printed Components for Low and Medium Frequencies
9. Strip Transmission Lines
10. Printed Microwave Systems

Appendix 1. Testing Procedures and Specifications

Appendix 2. Some Early Printed Patents

The book is available from Macdonald & Co. (Publishers), Ltd., 16 Maddox Street, London, W. 1, England, for 40 shillings.

Logarithmic Navigation for Precise Guidance of Space Vehicles*

W. G. GREEN

ITT Federal Laboratories, Fort Wayne, Indiana

1. Symbols

g = Nominal gravitational acceleration of earth = 32.2 feet per second per second

h = Altitude above the surface (assumed spherical) of the astronomical body

k, α_1, α_2, et cetera = Logarithmic guidance proportionality factors

m, n = Integers

t = Time in seconds

x = Lateral distance normal to local vertical of space vehicle

ζ = Damping ratio in quadratic frequency-response operator

η = Angle from local vertical of space vehicle to velocity vector

ω = Natural frequency in quadratic frequency-response operator (in radians per second)

r = Distance from vehicle to landing point measured along line of sight between the two

B = Instrumentation-created bias error

$\dot{\sigma}$ = Rate of change of line of sight in inertial space

$\ddot{\lambda}$ = Acceleration component normal to line of sight and velocity vector

$\ddot{Z}$ = Acceleration component normal to local vertical and velocity vector

R = Radius of astronomical body

β = Angle subtended by local vertical of vehicle and local vertical through impact point

ϕ = Angle subtended by line of sight and local vertical of impact point

ψ = Angle subtended by line of sight and reference local horizontal of impact point

Q = Variable controlling time of arrival

Subscript o = Initial conditions

Subscript f = Impact conditions

Sign Conventions: Positive range direction assumed away from planet

$$\text{Operator } (\dot{\ }) = \frac{d}{dt}(\ )$$

$$\text{Operator } (\ )^* = \frac{(\ )}{\frac{1}{\omega^2}\frac{d^2}{dt^2} + \frac{2\zeta}{\omega}\frac{d}{dt} + 1}.$$

2. Introduction

Various aspects of space-flight guidance have been discussed in the literature wherein the optimizing criteria have been minimum fuel, restraint of atmospheric heating or decelerations, minimum engine thrust, et cetera, with little consideration being given to minimum sensitivity to system parameter errors and variations. This last factor often becomes of paramount importance near the terminus of control where slight deviations from minimum fuel, minimum heating, et cetera can be tolerated for precise matching of the vehicle's path to the desired trajectory. An example of this is the orbital rendezvous of two vehicles, which requires as predominant factors precise control of the range between the vehicles and zero relative velocity as they meet rather than a few less pounds of fuel. This paper suggests and analyzes certain guidance relationships that offer broad tolerance to instrumentation error and system-parameter variation, plus simplicity and quasi-independence of time. Logarithmic guidance appears to be a powerful tool for the solution of the vernier-guidance problem.

The first section of the paper derives certain guidance relationships for controlling the vehicle's acceleration as a function of the error in desired vehicle position and velocity. An analysis of the effects of instrumentation

* Presented at Institute of Radio Engineers National Convention in New York, New York, on 22 March 1961. Reprinted from *IRE Transactions on Aerospace and Navigational Electronics*, volume ANE-8, number 2, pages 59–71; June 1961.

The research reported in this paper was partially sponsored by the Navigation and Guidance Laboratory, Wright Air Development Division, Air Research and Development Command, under contract AF33(616)-6032 with ITT Federal Laboratories.

errors on logarithmic guidance is presented in the next section. The third section of the paper discusses application of the guidance laws to various space-flight guidance problems.

3. Analysis of Various Control Functions

The effort to resolve guidance relationships with minimum system-error sensitivity was directed toward functions that control the deceleration of the space vehicle as a function of the instantaneous range between the actual and the desired vehicle position, and the relative instantaneous velocity between the two positions. Under these conditions, system performance is not dependent on exact initial conditions and precisely predicted component performance; further, the system possesses a capability of coping with unpredictable events that may arise during the trajectory. Another important criterion for the control function is that near zero range, the relative velocity vector must approach the orientation of the range vector, and the magnitude of velocity must decrease to zero for desirable terminal conditions.

3.1 Control in a Single Dimension

The general class of control polynomials being considered is composed of various sums of products, each product having a function of range and a function of velocity as its only factors. Due to the practical difficulty of instrumenting several different functions and maintaining required accuracy during the summation process, single product cases are considered of primary interest. This immediate discussion is limited to one-dimensional effects.

This group of control functions of interest can be further categorized in terms of terminal behavior (that is, the relationship existing between range rate and range as they both approach zero). The most desirable terminal condition is zero range rate at impact; two other possibilities exist, one, a finite range rate at impact and the other, zero range rate at a finite range. The general relationship, $\ddot{r} = k[F(\dot{r})]/[G(r)]$, was investigated to determine its terminal behavior for various possible combinations of range rate and range functions. Selections of $F(\dot{r}) = (-\dot{r})^{m+1}$ and $G(r) = (r)^n$, together with integration of the resulting expressions, provides relationships between range and range rate as functions of k, m, n, and the initial conditions. The terminal behavior of these relationships is indicated in Table 1 along with comments concerning the operating time t_f and dependence on the constant of proportionality k.

Examination of Table 1 shows that only three combinations of m and n values permit the desired terminal behavior, and two of these

TABLE 1

Terminal Behavior of $\ddot{r} = k[(-\dot{r})^{m+1}]/r^n$

	$m > 1$	$m = 1$	$m < 1$
$n > 1$	*$t_f \to \infty$ as r and $\dot{r} \to 0$ *requires special k value.* Sluggish (never get there).	$t \to \infty$ as r and $\dot{r} \to 0$ independent of k. Sluggish (never get there).	$r > 0$ when $\dot{r} = 0$ independent of k. Stop before contact.
$n = 1$	$t \to \infty$ as r and $\dot{r} \to 0$ independent of k. Sluggish (never get there).	r and $\dot{r} \to 0$ together for finite t. Independent of k.	$r > 0$ when $\dot{r} = 0$ independent of k. Stop before contact.
$n < 1$	$\dot{r} > 0$ at impact independent of k. Bump.	$\dot{r} > 0$ at impact independent of k. Bump.	r and $\dot{r} \to 0$ together for finite t, *requires special k value.*

* For $m - n \geq 0$, r and $\dot{r} \to 0$ in finite time for special value of k.

combinations are acceptable only if an exact relationship or balance exists between k and the initial conditions. Furthermore, systems requiring a special k balance are very sensitive to instrumentation and sensing errors, and would require excessively severe tolerances in the instrumentation to provide desirable terminal behavior. Thus, the only polynomial function in the table with acceptable characteristics is that having $m = 1$ and $n = 1$ or

$$\ddot{r} = \frac{k\dot{r}^2}{r}. \tag{1}$$

This control function, when integrated, provides the following relationship between range, range rate, the constant k, and initial conditions.

$$\frac{d(\dot{r})}{\dot{r}} = \frac{kdr}{r}$$

$$\log\left(\frac{\dot{r}}{\dot{r}_o}\right) = k \log\left(\frac{r}{r_o}\right)$$

and thus

$$\frac{\dot{r}}{\dot{r}_o} = \left[\frac{r}{r_o}\right]^k. \tag{2}$$

The logarithmic character of (2) as the primary solution of (1) is essential since this type of relationship is one of the simplest that provides acceptable terminal behavior completely independent of k and the initial conditions. Hereafter (1) will be termed the logarithmic control function, providing "logarithmic" guidance.

Investigation of elementary nonpolynomial functions as possibilities for the guidance function reveals that two trigonometric functions would also provide acceptable terminal behavior. However, the instrumentation complexity to obtain these functions is very much greater than that required for the simple polynomial form of logarithmic guidance.

There are certain restrictions on the logarithmic guidance proportionality constant k, namely, that $0.5 < k < 1$ must hold to obtain acceptable terminal behavior. For $k < 0.5$, infinite deceleration at impact is required; $k = 0.5$ calls for constant nonzero deceleration over the entire operation period and is an unstable operating point with respect to impact conditions; $k \geq 1$ calls for infinite operating time.

The logarithmic guidance law can be written as a function only of range and its derivatives. Thus, it is evident that range is the fundamental independent variable for logarithmic guidance rather than time, the latter being little more than an auxiliary function of range. Deceleration, range rate, and time-to-impact are single-valued, being maximum initially and zero at impact. Deceleration and range rate are small during the latter phases of the flight and are less sensitive to errors. An added advantage of this hovering approach is the increased time available at short ranges permitting the system instrumentation to provide increased accuracy and resolution.

3.2 Multidimensional Control

Logarithmic guidance has been shown to be very well suited for terminal control in one dimension; however, the system may require control in two or more dimensions. Logarithmic guidance can be generalized to include multidimensions by proper choice of coordinate systems. If the multidimensional range can be written in terms of a vector, that is, a magnitude and one or more angular or phase relationships, then the magnitude of this range vector may be taken as the principal axis of control and is controlled by (1). Equation (1) will drive the magnitude of the range vector and its first two time derivatives to zero; however, the traverse components of velocity also require some type of control. The control requirements for these traverse components are that their terminal behavior be acceptable: namely, be zero when the magnitude of the range is zero and approach zero without interference with the principal velocity orientation being along the principal axis. This requirement on terminal behavior forces the traverse-guidance equations to be functions of the magnitude of the range vector and have the characteristics of logarithmic guidance. A simple polynomial that has these features is

$$\dot{L} = k\alpha L\left(\frac{\dot{r}}{r}\right) \tag{3}$$

with

$$\frac{L}{L_o} = \left(\frac{r}{r_o}\right)^{k\alpha}$$

where L is some function of the traverse components of the velocity vector.

Multidimensional logarithmic guidance can be accomplished by the use of (1) to control the path along the principal axis and (3) to control each traverse component.

4. Effects of Instrumentation Errors

The effects of instrumentation errors on logarithmic guidance have been determined through a generalized error analysis. In addition, specific investigations of the effects of bias and dynamic errors have been made and are presented in the Appendixes (Sections 8 and 9).

The general instrumented form of logarithmic guidance could be expressed as

$$\ddot{r}_m = k\frac{\dot{r}_m^{\,2}}{r_m} \tag{4}$$

where the subscript m denotes the measured functions and if the relationships between the measured values and the actual values are expressed as

$$r_m = r + \epsilon_r \tag{5}$$

where ϵ denotes the error in the measurement.

Then (4) can be written

$$\ddot{r} = \frac{k\dot{r}^2}{r}\,\xi^* \tag{4A}$$

where

$$\xi^* = \frac{\left(1 + \frac{\epsilon_{\dot{r}}}{\dot{r}}\right)^2}{\left(1 + \frac{\epsilon_r}{r}\right)\left(1 + \frac{\epsilon_{\ddot{r}}}{\ddot{r}}\right)}.$$

If ξ^* is restricted to small variations about a constant and the extremes of these variations in ξ^* are ξ_1 as an upper limit and ξ_2 as a lower limit, then the solutions of (4A) using ξ_1 and ξ_2 as constants will bound $\dot{r}$.

$$\left[\frac{r}{r_o}\right]^{k\xi_2} \geq \frac{\dot{r}}{\dot{r}_o} \geq \left[\frac{r}{r_o}\right]^{k\xi_1}. \tag{6}$$

Time of flight can also be determined as

$$\left[\frac{1}{1 - k\xi_1}\right]\left(\frac{r_o}{\dot{r}_o}\right) \geq t_f \geq \left(\frac{r_o}{\dot{r}_o}\right)\left[\frac{1}{1 - k\xi_2}\right] \tag{7}$$

and the required accelerations as

$$\frac{k\dot{r}_o^{\,2}}{r_o}\,\xi_1\left[\frac{r}{r_o}\right]^{2k\xi_1 - 1} \leq \ddot{r} \leq \frac{k\dot{r}_o^{\,2}}{r_o}\,\xi_2\left[\frac{r}{r_o}\right]^{2k\xi_2 - 1}. \tag{8}$$

The following restrictions on ξ_1 and ξ_2 are evident.

(A) If $\dot{r}/\dot{r}_o$ is to become zero when $r = 0$, then (6) requires that

$$k\xi_1 > 0 \quad \text{and} \quad k\xi_2 > 0.$$

(B) If the required acceleration is to remain finite, then (8) requires that

$$2k\xi_2 - 1 > 0 \quad \text{or} \quad \xi_2 > \frac{0.5}{k}.$$

(C) If the time of flight is to be real and finite, then (7) requires that

$$\xi_1 < \frac{1}{k}.$$

These restrictions can be summarized by

$$\frac{0.5}{k} < \xi_2 < \xi_1 < \frac{1}{k} \quad \text{or} \quad \frac{\xi_1}{\xi_2} < 2.0.$$

Thus, the logarithmic guidance system can cope with errors that change the desired acceleration by nearly a factor of 2 and attain the desired terminal conditions. The types of error that are represented by ξ^* are: scale-factor errors, noise, limiting or saturation errors, and most other types of scalar errors.

The effects of bias errors (constant offset errors) on the system were investigated and the following effects noted in Appendix (Section 8).

(A) Bias decelerations that add to the desired deceleration cause the range rate to go to zero before the range does, while those bias decelerations that subtract allow zero range rate at impact. If the velocity goes to zero before impact, the guidance law will cause undesirable behavior during the remaining portion of the flight. This possibly suggests that a system requirement exists for rocket-motor cutoff whenever the velocity becomes zero. Figure 2 in Appendix (Section 8) shows this situation quantitatively for the case of additive deceleration bias, giving a plot of the stopping range values versus deceleration bias. Three curves are plotted to illustrate the effects of initial conditions and system parameters on stopping-range magnitude. It might be expected that the subtractive deceleration bias would cause the guidance law to demand large decelerations near impact to overcome the bias effect, but it can be shown that no such danger arises.

(B) Bias errors in the determination of the range rate, which add to the magnitude of the actual range rate, cause zero rate to occur at some nonzero range (assuming no negative range rates are permitted); bias errors that subtract cause impact at nonzero range rates. Table 4 in Appendix (Section 8) shows the magnitudes of the stopping ranges versus bias error for various initial conditions and k values.

(C) Subtractive bias errors in range determination produce zero velocity at nonzero ranges, while additive errors produce nonzero impact range rates. Table 5 in Appendix (Section 8) indicates the relative magnitudes of these effects.

The effects of dynamic-response errors on logarithmic guidance were demonstrated through analog-computer studies. A complex quadratic response function was inserted into the system equations and the natural frequency was varied to simulate a family of such responses. The primary objective was the approximate determination of the minimum natural frequency for which the range rate does not go to zero at impact. The conclusions of these investigations (discussed in more detail in Appendix (Section 9) are:

(A) Dynamic time lags between desired acceleration and actual acceleration of less than 2 percent of total time of flight will not affect the terminal behavior of the system.

(B) Time lags of nearly 5 percent of ideal time of flight can be tolerated between the measured range rate and the actual range rate without interference with the desired terminal conditions. However, these dynamic errors in range-rate measurement cause a considerable increase in the actual flight time of the system as shown in Figure 5. This increased flight time would reduce the efficiency of the system and force a compromise between fuel economy and the narrow bandwidths desirable for sensor instrumentation. It is expected that this compromise would be for bandwidths much wider than those indicated for the limited-response conditions.

(C) Dynamic errors in range determination are much more critical than the other dynamic errors studied. Any dynamic error in range produces deviation from the ideal terminal conditions; finite range rates exist at impact. However, for time lags of less than 2 percent of the time of flight, the impact range rates are less than 0.1 percent of the initial range rate.

5. Mission Applications of Logarithmic Guidance

There appears to be some 7 space guidance missions for which the extremely precise control of the terminus of the trajectory by logarithmic guidance is worthwhile. These missions are:

(A) Steering a space vehicle such that it becomes coincidental in space, time, and velocity with an orbiting satellite. This is the so-called rendezvous problem between two space vehicles that wish to transfer cargo or become mechanically integrated.

(B) Controlling the velocity vector of a space vehicle to provide a soft landing on a planet, a soft landing being defined as vehicle touchdown with impact velocities of a few feet per second. Logarithmic guidance would be applicable when atmospheric entry problems of heating and high aerodynamic drag were neglectable.

(C) Providing a soft landing with provision for avoiding certain undesirable terrain features, such as mountains, craters, chasms, cliffs, et cetera. This problem principally differs from problem (B) in that certain restrictions are required on the flight profile.

(D) Providing a soft landing at a particular geographical location on the surface of a planet. Here, again, logarithmic guidance would not be applied until after severe atmospheric entry conditions had subsided. It would also be necessary that the vehicle be initially in the vicinity of the desired landing area. This problem might exist, if it becomes necessary to supply logistics to a lunar or interplanetary outpost.

(E) Guidance of the space vehicle to a soft landing at a particular location with provision for making the approach from a given geographical direction. This might involve landing upon a runway or avoidance of terrain obstacles near the prepared landing area.

(F) Further sophistication of problem (E) may be considered if several landing vehicles are considered and the flight time becomes important.

(G) Matching a desired trajectory precisely in position, velocity, acceleration, and time. This is the most sophisticated of the missions and actually represents the sum of the requirements of the other missions.

5.1 Velocity Control for Satellite Rendezvous

The terminal phase of the rendezvous operation is characterized by the requirements to reduce the distance r along the line of sight between the vehicle and the satellite to zero in a finite time interval, while forcing the vehicle-satellite velocity vector to become coincident with the line of sight and reduce its magnitude to zero.

Equation (9), properly instrumented, will force the range between the vehicles and the component of velocity along the line of sight to zero in the manner discussed previously. Since there are no preferred directions of approach, the traverse control problem can be effectively reduced to a planar case, that is, the plane defined by the relative velocity vector and the range vector, by maintaining accelerations $\ddot{\lambda}$ normal to this plane at zero. Application of (10) to some function of the traverse velocity will then provide the desired terminal conditions.

Justification for this planar approach follows from Reference 1 in Section 7 wherein the effects of the earth's gravitational field on both vehicles are shown to be nearly equal (for satellites from 100 to 4000 miles (161 to 6436 kilometers) in altitude, terminal ranges of less than 100 miles (161 kilometers), and rocket thrusting of 5 to 10 g's, the differential effects of gravity may be neglected). Another justification exists for short times to impact, namely, that the flight path of the satellite closely resembles a straight line.

One possibility for controlling the traverse component of velocity that seems promising is to control the angular rate of change of the line of sight $\dot{\sigma}$ in inertial space by the use of (11). Reduction of this angular rate to zero will place the vehicle on a collision course with the satellite in a manner analogous to proportional navigation of missiles. The chief difference will be that the vehicle will not attain this course until range and range rate are zero. The parameters of the guidance laws must be selected such that the traverse components approach zero more rapidly than does the range rate. The system equations are given below.

Range control $\ddot{r} = \frac{k\dot{r}^2}{r}$, $(0.5 < k < 1.0)$ (9)

Lateral control $\ddot{\sigma} = \frac{k\alpha\dot{r}\dot{\sigma}}{r}$, $\left(\alpha > \frac{1}{k} - 1\right)$ (10)

where α is an arbitrary constant.

$$\text{Planar control } \ddot{\lambda} = 0. \quad (11)$$

From a design viewpoint, it is desirable to select k and α as small as convergence limits permit, since

$$\text{Time of flight} = \frac{r_o}{\dot{r}_o}\left(\frac{1}{1-k}\right). \quad (12)$$

Motor acceleration

$$\approx \left[\left(\frac{k\dot{r}^2}{r} - r\dot{\sigma}^2\right)^2 + (k\alpha\dot{r}\dot{\sigma} + 2\dot{\sigma}\dot{r})^2\right]^{1/2} \quad (13)$$

and longer times of flight require increased motor fuel and higher accelerations require larger motors. Selection of initial values by the designer must also be undertaken judiciously since they also affect time of flight and the magnitude of required motor accelerations.

5.2 Velocity Control for Planetary Soft Landings

The problem of landing on a planet with zero touchdown velocity is very similar to the rendezvous problem discussed above. Altitude instead of range becomes the primary control variable, and the requirement that the velocity approach the orientation of the primary variable (in this case, the vehicle's local vertical) is more stringent due to terrain avoidance and landing gear considerations. This increased emphasis on approach angle, and the fact that the desired landing area is not a point but may be anywhere on the planetary surface indicates that the traverse control variable should be the angle η between the velocity vector and the vehicle local vertical. The system can again be considered planar if accelerations out of the plane formed by the velocity vector and the local vertical are kept at zero. This system has been analyzed extensively in Reference 2 of Section 7. The system equations are

Altitude control $\ddot{h} = \frac{k\dot{h}^2}{h}$, $(0.5 < k < 1.0)$ (14)

Lateral control $\ddot{\eta} = \frac{k\alpha\dot{h}\dot{\eta}}{h}$, $\left(\alpha > \frac{1}{k} - 1\right)$ (15)

Planar control $\ddot{Z} = 0.$ (16)

The motor used to supply the vertical accelerations will have to supply enough acceleration to offset planet gravitational forces as well as the vertical accelerations called for by (14). This "constant" acceleration will provide an acceleration "floor" for the system and reduce the requirements for variations in motor thrust.

5.3 Soft Landing with Terrain Avoidance Capability

The majority of the space flights to the moon or the planets in the next decade will be exploratory missions and will necessitate soft-landing capabilities on uncharted and unprepared planet surfaces. While the gross landing sector can be selected prior to launch from mission considerations, the actual landing site will have to be chosen during the terminus of the flight from in-flight measurements of the planetary terrain characteristics. In practice, the actual landing site will probably not be selected as such; instead, restraints will be placed on the flight profile to prevent landing at undesirable sites and to avoid terrain obstacles.

The flight program may consist of two phases, a surveillance phase and a landing phase. The surveillance phase could be accomplished by applying accelerations in the plane normal to the local vertical to produce a spiralling flight path over the surface. The spiral path would allow examination of the gross sector while proceeding with the deceleration process.

The second phase or the actual landing phase would be initiated after the examination was complete. The following landing scheme would offer promise.

(A) The spiralling process used during the surveillance phase would be continued until the predicted impact point intercepted a satisfactory landing area. The vehicle would then be controlled to a soft landing by use of the equations listed below.

$$\text{Altitude control } \ddot{h} = \frac{k\dot{h}^2}{h}, \quad (0.5 < k < 1.0) \quad (17)$$

$$\text{Lateral control } \ddot{\eta} = \frac{k\alpha\dot{h}\dot{\eta}}{h}, \quad \left(\frac{\alpha > 1 - 1}{k}\right) \quad (18)$$

$$\text{Planar control } \ddot{Z} = 0. \quad (19)$$

These equations are defined similarly to those of Section 5.3.

(B) The impact point prediction doesn't need to be precise and can be predicted by

$$(h + R)\beta = \dot{h} \tan \eta, \quad R\frac{d\beta}{dh} \approx \eta \quad (20)$$

$$\Delta(R\beta) \approx \frac{\eta h}{1 + k\alpha} \quad (21)$$

where R = planet radius (see Figure 1)

β = angle subtended by the vehicle local vertical and the local vertical through the impact point

$\Delta(R\beta)$ = crudely the ground range to the impact point.

5.4 Soft Landing at Geographical Location

Landing at a given geographical location on the planet's surface will require another dimension of control over the systems discussed previously. This added dimension of control is necessary since a vertical soft landing at a preflight selected point is required. Thus, the guidance system must force the range r between vehicle and landing point to zero, the velocity magnitude to zero, and the angle between the velocity vector and the local vertical must be forced to near zero. The equations listed below will provide these terminal conditions.

$$\text{Range control } \ddot{r} = \frac{k\dot{r}^2}{r}, \quad (0.5 < k < 1.0) \quad (22)$$

Vertical approach control

$$\ddot{\phi} = \frac{k\alpha_1\dot{\phi}\dot{r}}{r} + \frac{k\alpha_2\dot{\phi}^2}{\phi}, \quad (23)$$

$$\left(\alpha_1 > \frac{1}{k} - 1\right), \quad \frac{1}{k} > \alpha_2 > \frac{1}{2k}$$

$$\text{Planar rotation control } \ddot{\psi} = \frac{k\alpha_3\dot{\psi}\dot{r}}{r}, \quad (24)$$

$$\left(\alpha_3 > \frac{1}{k} - 1\right)$$

where

ϕ = angle subtended by the vehicle-impact point line of sight and the local vertical of the impact point

ψ = angle subtended by the line of sight and one of the local horizontals of the impact point

These angles are illustrated in Figure 1.

Equations (22) and (24) are the familiar control functions discussed previously. Equation (23) is a combination of previous control functions and its consequences can be noted by integrating (23) to obtain

$$\frac{\dot{\phi}}{\dot{\phi}_o} = \left(\frac{r}{r_o}\right)^{k\alpha_1} \left(\frac{\phi}{\phi_o}\right)^{k\alpha_2} \quad (25)$$

and integrating (25) to obtain

$$\frac{\phi}{\phi_o} = \left\{\tau \left[\frac{1 - k\alpha_2}{1 - k + k\alpha_1}\right] \times \left[\left(\frac{r}{r_o}\right)^{1-k+k\alpha_1} - 1\right] + 1\right\}^{1/1-k\alpha_2} \quad (26)$$

where

$$\tau = \left(\frac{r_o}{\dot{r}_o}\right)\left(\frac{\dot{\phi}_o}{\phi_o}\right).$$

When the range r becomes zero

$$\frac{\phi_f}{\phi_o} = \left\{1 - \tau \left[\frac{1 - k\alpha_2}{1 + k(\alpha_1 - 1)}\right]\right\}^{1/1-k\alpha_2}$$

and ϕ_f is not zero except for special values of the constants. However, ϕ_o is probably always less than 30 degrees and if $\frac{\phi_f}{\phi_o} \leq \frac{1}{10}$, a near-vertical approach will be made. Table 2 shows the permissible variations of $\frac{\phi_f}{\phi_o}$ and τ for various values of $k\alpha_2$ and $k(\alpha_1 - 1)$.

Thus, Table 2 shows that a near-vertical approach can be made by selecting proper system parameters for all initial conditions that cause τ to be greater than $\sim$2.0. It is evident that for the majority of cases of interest, this restriction on τ will not impose any limitations other than that a large component of traverse velocity is needed.

5.5 Soft Landing at Geographical Location with Fixed Approach

This mission is a variation of mission (D) in Section 5 and involves no innovations. Its importance stems primarily from a need to avoid obstacles, both natural and man made. The equations are listed below.

$$\text{Range control } \ddot{r} = \frac{k\dot{r}^2}{r}, \quad (0.5 < k < 1.0) \quad (27)$$

Vertical approach control

$$\ddot{\phi} = \frac{k\alpha_1\dot{\phi}\dot{r}}{r} + \frac{k\alpha_2\dot{\phi}^2}{\phi}, \quad (28)$$

$$\left(\alpha_1 > \frac{1}{k} - 1\right), \quad \left(\frac{1}{k} > \alpha_2 > \frac{1}{2k}\right)$$

Planar rotation control

$$\ddot{\psi} = \frac{k\alpha_3\dot{\psi}\dot{r}}{r} + \frac{k\alpha_4\dot{\psi}^2}{\psi}, \quad (29)$$

$$\left(\alpha_3 > \frac{1}{k} - 1\right), \quad \left(\frac{1}{k} > \alpha_4 > \frac{1}{2k}\right)$$

5.6 Soft Landings with Controlled Time of Arrival

The time of arrival of a space vehicle at the landing port may become important in the future, though admittedly in the present era the possibility of congested space traffic seems remote. The time of arrival for the space vehicle can be controlled by the equations listed below, along with landing velocity, landing point, approach, et cetera.

$$\text{Range control } \ddot{r} = \frac{k\dot{r}^2}{r} + \dot{r}Q, \quad (30)$$

$$(0.5 < k < 1.0)$$

where

$$Q = \frac{2}{(\Delta t)^2}\left[\frac{r}{\dot{r}(1-k)} + \Delta t\right]$$

TABLE 2

$k\alpha_2$	$\frac{\phi_f}{\phi_o}$	$\left(\frac{\phi_f}{\phi_o}\right)^{1/1-k\alpha_2}$	$\frac{\tau}{1+k(\alpha_1-1)}$	τ*	τ**
0.9	0.1	0.8	20	20	2000
0.9	10^{-2}	0.62	38	38	3800
0.9	10^{-3}	0.5	50	50	5000
0.9	10^{-10}	0.1	100	100	10^4
0.75	0.1	0.56	1.8	1.8	180
0.75	10^{-2}	0.3	2.8	2.8	280
0.75	10^{-3}	0.18	3.3	3.3	330
0.75	10^{-4}	0.1	3.6	3.6	360
0.6	0.1	0.4	1.5	1.5	150
0.6	10^{-2}	0.16	2.1	2.1	2.10
0.6	3×10^{-3}	0.1	2.3	2.3	230

* For $k(\alpha_1 - 1) = 0$.
** For $k(\alpha_1 - 1) = 100$.

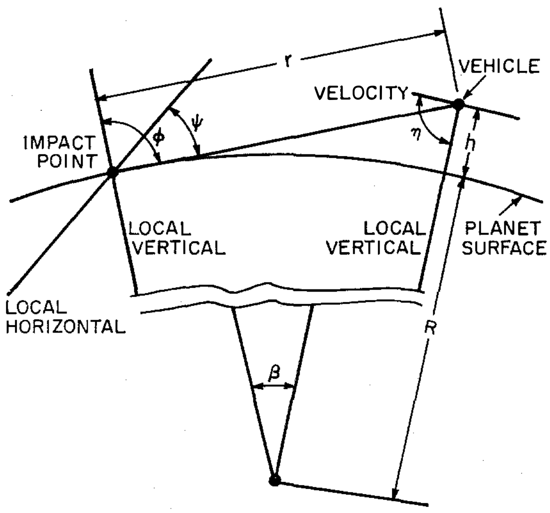


Figure 1—Guidance geometry.

and Δt is the time in seconds between the present and the desired landing instant.

Vertical approach control

$$\ddot{\phi} = \frac{k\alpha_1\dot{\phi}\dot{r}}{r} + \frac{k\alpha_2\dot{\phi}^2}{\phi}, \quad (31)$$

$$\left(\alpha_1 > \frac{1}{k} - 1\right), \quad \left(\frac{1}{k} > \alpha_2 > \frac{1}{2k}\right)$$

Planar rotation control

$$\ddot{\psi} = \frac{k\alpha_3\dot{\phi}\dot{r}}{r} + \frac{k\alpha_4\dot{\psi}^2}{\psi}, \quad (32)$$

$$\left(\alpha_3 > \frac{1}{k} - 1\right), \quad \left(\frac{1}{k} > \alpha_4 > \frac{1}{2k}\right)$$

The first equation controls the range and range rate along with the time of arrival. Since Q is a constant in the ideal system, the relationship between range and range rate is

$$\frac{\dot{r}}{\dot{r}_o} = \left(\frac{r}{r_o}\right)^k (Qt - 1) \quad (33)$$

$$\frac{r_o}{\dot{r}_o(1-k)}\left(\frac{r}{r_o}\right)^{1-k} = \frac{Q}{2}t^2 - t - \frac{r_o}{\dot{r}_o(1-k)}. \quad (34)$$

Thus, Q can be calculated in the computer continuously to reflect the errors in instrumentation that will occur.

5.7 Matching Desired Trajectory in Velocity, Space, and Time

Most of the satellite flights, both lunar and interplanetary, require positioning a space vehicle on a precalculated orbit that must be precisely matched in positions, velocities, and calendar time for maximum efficiency. The principles of logarithmic guidance offer extreme control for this type of mission as well. The equations listed below will perform the desired function.

Range control $\ddot{r} = \dfrac{k\dot{r}^2}{r} + \dot{r}Q, \quad (35)$

$$(0.5 < k < 1.0)$$

where

$$Q = \frac{2}{(\Delta t)^2}\left[\frac{r}{\dot{r}(1-k)} + \Delta t\right].$$

Traverse control

$$\ddot{\phi} = \frac{k\alpha_1\dot{\phi}\dot{r}}{r}, \quad \left(\alpha_1 > \frac{1}{k} - 1\right) \quad (36)$$

Planar control

$$\ddot{\psi} = \frac{k\alpha_2\dot{\psi}\dot{r}}{r}, \quad \left(\alpha_2 > \frac{1}{k} - 1\right) \quad (37)$$

The guidance laws, parameter limitations, and comments concerning the application of logarithmic-guidance principles to the 7 mission types are summarized in Table 3.

6. Summary

The broad tolerance of logarithmic guidance to instrumentation error and system parameter variation, plus its simplicity and quasi-independence of time, show it to be a powerful tool for the solution of those terminal guidance problems wherein extreme accuracy of vehicle kinematics is the supreme objective. The general functionality of the logarithmic-guidance principles is indicated by the 7 basic space flight missions to which it can be applied profitably.

The design of a system using logarithmic-guidance principles should allow provision for certain factors peculiar to this type of control. These factors are:

(A) System time of flight is proportional to the initial values of range and range rate, and care must be used to prevent either extremely brief or excessively long flight times. Most missions require earlier guidance stages (that is, midcourse), and this factor can be controlled by appropriate terminal requirements for the earlier stages.

(B) The velocity vector must initially tend to reduce the range. Provision for this factor can be made in the same manner as for time of flight.

(C) The relative accelerations between vehicle and desired position may be difficult to ascertain. This can generally be circumvented because either the accelerations are known

TABLE 3
SUMMARY OF DATA FOR MISSIONS LISTED IN SECTION 5

Mission	Guidance Laws	Parameter Limitations	Comments
(A) Satellite Rendezvous	$\ddot{r}=\frac{k\dot{r}^2}{r}$ (9) $\ddot{\sigma}=\frac{k\alpha\dot{r}\dot{\sigma}}{r}$ (10) $\ddot{\lambda}=0$ (11)	$0.5<k<1.0$ $\alpha>\frac{1}{k}-1$	k and α should be small as possible for system-weight minimization. Initial conditions (including accelerations) should be designed to minimize gravitational and centrifugal force differences between the vehicles.
(B) Soft Landing Velocity Control	$\ddot{h}=\frac{k\dot{h}^2}{h}$ (14) $\dot{\eta}=\frac{k\alpha\dot{h}\eta}{h}$ (15) $\ddot{Z}=0$ (16)	$0.5<k<1.0$ $\alpha>\frac{1}{k}-1$	Chief differences from (A) is gravitational acceleration offset, necessity for vertical-landing approach, and that any point on the surface may be the landing site.
(C) Soft Landing with Terrain Avoidance	$\ddot{h}=\frac{k\dot{h}^2}{h}$ (17) $\dot{\eta}=\frac{k\alpha\dot{h}\eta}{h}$ (18) $\ddot{Z}=0$ (19) $\Delta(R\,\beta)\approx\frac{\eta h}{1+k\alpha}$ (21)	$0.5<k<1.0$ $\alpha>\frac{1}{k}-1$	Landing accomplished in two phases: one, a surface surveillance spiralling phase; and a phase that is initiated when the prediction equation (21) indicates an acceptable landing area.
(D) Soft Landing at Geographical Location	$\ddot{r}=\frac{k\dot{r}^2}{r}$ (22) $\ddot{\phi}=\frac{k\alpha_1\dot{\phi}\dot{r}}{r}+\frac{k\alpha_2\dot{\phi}^2}{\phi}$ (23) $\psi=\frac{k\alpha_3\dot{\psi}\dot{r}}{r}$ (24)	$0.5<k<1.0$ $\alpha_1>\frac{1}{k}-1$ $\frac{1}{2k}<\alpha_2<\frac{1}{k}$ $\alpha_3>\frac{1}{k}-1$	A computer is necessary to select α_1 and α_2 as a function of initial values of r, $\dot{r}$, ϕ, and $\dot{\phi}$. Design provision should be made to force $r_o\dot{\phi}_o/\dot{r}_o\phi_o$ to be greater than 2.0. This will probably require an initial approach nearer the horizontal than the vertical.
(E) Fixed-Approach Soft Landing at Geographical Location	$\ddot{r}=\frac{k\dot{r}^2}{r}$ (22) $\ddot{\phi}=\frac{k\alpha_1\dot{\phi}\dot{r}}{r}+\frac{k\alpha_2\dot{\phi}^2}{\phi}$ (23) $\ddot{\psi}=\frac{k\alpha_3\dot{\psi}\dot{r}}{r}+\frac{k\alpha_4\dot{\psi}^2}{\psi}$ (24)	$0.5<k<1.0$ $\alpha_1>\frac{1}{k}-1$ $\frac{1}{2k}<\alpha_2<\frac{1}{k}$ $\alpha_3>\frac{1}{k}-1$ $\frac{1}{2k}<\alpha_4<\frac{1}{k}$	Same comments apply here as for mission (D).

TABLE 3—*Continued*

Mission	Guidance Laws	Parameter Limitations	Comments
(F) Controlled Time of Arrival with Soft Landings	$\ddot{r}=\frac{k\dot{r}^2}{r}+\dot{r}Q$ (30)	$0.5<k<1.0$	Same comments apply here as for mission (D). In addition, on-board computer will have to compute Q continuously.
	$\ddot{\phi}=\frac{k\alpha_1\dot{\phi}\dot{r}}{r}+\frac{k\alpha_2\dot{\phi}^2}{\phi}$ (31)	$\alpha_1>\frac{1}{k}-1$	
		$\frac{1}{2k}<\alpha_2<\frac{1}{k}$	
	$\ddot{\psi}=\frac{k\alpha_3\dot{\psi}\dot{r}}{r}+\frac{k\alpha_4\dot{\psi}^2}{\psi}$ (32)	$\alpha_3>\frac{1}{k}-1$	
		$\frac{1}{2k}<\alpha_4<\frac{1}{k}$	
	$Q=\frac{2}{(\Delta t)^2}\left[\frac{r}{\dot{r}(1-k)}+\Delta t\right]$		
(G) Trajectory Matching	$\ddot{r}=\frac{k\dot{r}^2}{r}+\dot{r}Q$ (35)	$0.5<k<1.0$	On-board computer will have to compute equation Q continuously.
	$\ddot{\phi}=\frac{k\alpha_1\dot{\phi}\dot{r}}{r}$ (36)	$\alpha_1>\frac{1}{k}-1$	
	$\ddot{\psi}=\frac{k\alpha_2\dot{\psi}\dot{r}}{r}$ (37)	$\alpha_2>\frac{1}{k}-1$	
	where		
	$Q=\frac{2}{(\Delta t)^2}\left[\frac{r}{\dot{r}(1-k)}+\Delta t\right]$		

rather accurately as a function of altitude (if the target is a planet) or can be neglected for the case of orbital or ballistic trajectories.

(D) The system must be capable of delivering a wide range of accelerations. The range of accelerations required can be reduced by applying the results of the deceleration-bias-error study, that is, since offset bias doesn't interfere with terminal conditions, an acceleration "floor" can be attained by deliberately using a subtractive-bias acceleration. The level of this bias should be carefully selected to minimize the effects of engine cutoff.

7. References

1. R. A. Hord, "Relative Motion in the Terminal Phase of Interception of a Satellite or a Ballistic Missile," NACA TN 4399; September, 1958.
2. F. H. Brady and W. G. Green, "Deceleration Control of Space Vehicles," presented 30 June 1959 at Institute of Radio Engineers National Convention on Military Electronics, Washington, District of Columbia (Confidential).

8. Appendix—Effects of Bias Errors

The effects of bias errors (constant offset errors in measurements) on the system were investigated by analytical means. Bias errors in deceleration modify logarithmic guidance to give

$$\ddot{r} = \frac{k(\dot{r})^2}{r} + B_{\ddot{r}} \qquad (38)$$

TABLE 4

γ/e

$\frac{r_s}{r_o} \diagdown k$	0.5	0.6	0.75	0.9	1.0
10^{-2}	10^{-1}	6×10^{-2}	3×10^{-2}	1.7×10^{-2}	10^{-2}
10^{-3}	3×10^{-2}	1.7×10^{-2}	6×10^{-3}	2×10^{-3}	10^{-3}
10^{-4}	10^{-2}	4×10^{-3}	10^{-3}	2.5×10^{-4}	10^{-4}
10^{-5}	3×10^{-3}	10^{-3}	2×10^{-4}	3×10^{-5}	10^{-5}
10^{-6}	10^{-3}	2.5×10^{-4}	3×10^{-5}	4×10^{-6}	10^{-6}

$$\frac{\dot{r}}{\dot{r}_o}=\left\{\left(\frac{r}{r_o}\right)^{2k} - \frac{2B_{\ddot{r}}r_o}{(2k-1)(\dot{r}_o)^2}\left(\frac{r}{r_o}\right)\left[1-\left(\frac{r}{r_o}\right)^{2k-1}\right]\right\}^{1/2} \quad (39)$$

where $B_{\ddot{r}}$ is the deceleration bias.

For bias decelerations that add to the desired deceleration, the velocity goes to zero before impact. Thus, the stopping range is

$$\frac{r_s}{r_o} = \left(\frac{2\lambda}{2k-1+2\lambda}\right)^{1/2k-1}$$

where

$$\lambda = \frac{B_{\ddot{r}}}{(\dot{r}_o^2/r_o)}.$$

For bias decelerations that subtract from the desired deceleration, the velocity goes to zero at impact. Three curves are plotted in Figure 2 to illustrate the magnitude of the final range error as a function of initial conditions and system parameters.

Bias errors in the determination of range rate produce changes in logarithmic guidance as shown by (40).

$$\ddot{r} = \frac{k(|\dot{r}| + B_{\dot{r}})^2}{r} \quad (40)$$

where $B_{\dot{r}}$ is the range rate bias

$$\frac{|\dot{r}|+B_{\dot{r}}}{|\dot{r}_o|+B_{\dot{r}}}=\left[\frac{r}{r_o}\right]^k \times\exp\left[-B_{\dot{r}}\left(\frac{1}{|\dot{r}|+B_{\dot{r}}}-\frac{1}{|\dot{r}_o|+B_{\dot{r}}}\right)\right]. \quad (41)$$

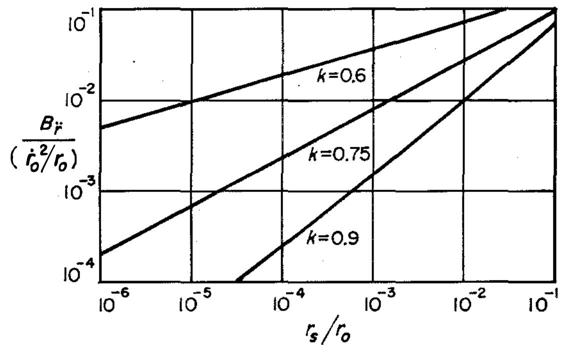


Figure 2—Normalized stopping range versus acceleration bias. $\ddot{r} = (k\dot{r}^2/r) + B_{\ddot{r}}$.

Bias errors that subtract from the magnitude of the actual range rate cause impact at some non-zero range rate. This impact velocity is

$$\frac{\dot{r}_I}{\dot{r}_o} = \frac{B_{\dot{r}}}{\dot{r}_o}.$$

Bias errors that add to the range rate cause the range rate to become zero before impact. This stopping range is

$$\frac{r_s}{r_o} = \left\{\frac{\gamma}{1+\gamma}\exp\left[-\frac{1}{1+\gamma}\right]\right\}^{1/k}$$

where $\gamma = B_{\dot{r}}/|\dot{r}_o|$.

Table 4 shows the magnitude of the effects of these range-rate-bias errors.

Equation (42) is the modified logarithmic-guidance law for bias errors in range determination.

$$\ddot{r} = \frac{k_r^2}{(r + B_{\dot{r}})} \quad (42)$$

$$\frac{\dot{r}}{\dot{r}_o} = \left(\frac{r + B_r}{r_o + B_r}\right)^k. \quad (43)$$

Here subtractive-bias errors produce stoppage at non-zero range and additive errors produce impact range rates. These ranges and range rates are

$$\frac{r_s}{r_o} = \frac{B_r}{r_o} \quad \text{and} \quad \frac{\dot{r}_I}{\dot{r}_o} = \left(\frac{\Delta}{1+\Delta}\right)^k$$

where $\Delta = B_r/r_o$.

TABLE 5

Δ

$\frac{\dot r_I}{\dot r_o}$ \ k	0.5	0.6	0.75	0.9	1.0
10^{-2}	10^{-4}	5×10^{-4}	2×10^{-3}	6×10^{-3}	10^{-2}
10^{-3}	10^{-6}	10^{-5}	10^{-4}	5×10^{-4}	10^{-3}
10^{-4}	10^{-8}	2×10^{-7}	5×10^{-6}	4×10^{-5}	10^{-4}
10^{-5}	10^{-10}	4×10^{-8}	2.5×10^{-7}	3×10^{-6}	10^{-5}

Table 5 illustrates these effects for various system parameters and initial conditions.

The lateral velocity control equation possibilities, $\dot\eta = \frac{k\alpha_1\eta\dot r}{r}$ (angle control) and $\ddot x = \frac{k\alpha_2\dot x\dot r}{r}$ (translational control) are also subject to bias errors with respect to r, $\dot r$, and $\ddot r$ as well as η, $\dot\eta$, $\dot x$ and $\ddot x$ bias errors. Let these bias errors be termed B_r, $B_{\dot r}$, et cetera. Then, in general,

$$\dot\eta = k\alpha_1(\eta + B_\eta)\left[\frac{(\dot r + B_{\dot r})}{\dot r + B_r}\right] + B_{\dot\eta} \quad (44)$$

$$\ddot x = k\alpha_2(\dot x + B_{\dot x})\left[\frac{(\dot r + B_{\dot r})}{r + B_r}\right] + B_{\ddot x} \quad (45)$$

$$\ddot r = k\frac{(\dot r + B_{\dot r})^2}{(r + B_r)} + B_{\ddot r}. \quad (46)$$

Examining the effects of these bias errors singly on the system of equations

(A) For all bias errors except $B_{\ddot r}$ equal to zero

$$\eta_{\text{terminal}} = \begin{cases} 0 \text{ (for } r=0), \quad B_{\ddot r}<0 \\ \eta_o\left[\dfrac{2B_{\ddot r}r_o}{2B_{\ddot r}r_o+(2k-1)(\dot r_o)^2}\right]^{k\alpha_1/(2k-1)}, \\ \qquad\qquad B_r>0 \end{cases} \quad (47)$$

$$\dot x_{\text{terminal}} = \begin{cases} 0 \text{ (for } \dot r=0), \quad B_r<0 \\ \dot x_o\left[\dfrac{2B_{\ddot r}r_o}{2B_{\ddot r}r_o+(2k-1)(\dot r_o)^2}\right]^{k\alpha_2/(2k-1)}, \\ \qquad\qquad B_{\ddot r}>0. \end{cases} \quad (48)$$

η will be zero at impact in any event for angle control since when $r = 0$, the magnitude of velocity is zero. However, additive bias will cause a traverse velocity at impact for translational control.

(B) For all bias errors except $B_{\dot r}$ equal to zero

$$\eta = \eta_o\left(\frac{r}{r_o}\right)^{k\alpha_1} \times\exp\left[-\alpha_1B_{\dot r}\left(\frac{1}{|\dot r|+B_{\dot r}}-\frac{1}{|\dot r_o|+B_{\dot r}}\right)\right] \quad (49)$$

$$\dot x = \dot x_o\left(\frac{r}{r_o}\right)^{k\alpha_2} \times\exp\left[-\alpha_2B_{\dot r}\left(\frac{1}{|\dot r|+B_{\dot r}}-\frac{1}{|\dot r_o|+B_{\dot r}}\right)\right]. \quad (50)$$

By the same argument as cited in (A), angle control will force η to be zero at impact in any event, and translational control will produce traverse velocities at impact when there is additive bias. Subtractive bias does not interfere with traverse terminal conditions.

(C) For all bias errors except B_r equal to zero

$$\eta_{\text{terminal}} = \begin{cases} 0 \text{ for } B_r < 0 \\ \eta_o\left(\dfrac{B_r}{r_o + B_r}\right)^{k\alpha_1}, \quad B_r > 0. \end{cases} \quad (51)$$

$$\dot x_{\text{terminal}} = \begin{cases} 0 \text{ for } B_r < 0 \\ \dot x_o\left(\dfrac{B_r}{r_o + B_r}\right)^{k\alpha_2}, \quad B_r > 0. \end{cases} \quad (52)$$

The non-zero values of η_{terminal} and x_{terminal} occur for additive bias and have the same basic form of independence on B_r as do $\dot r_{\text{terminal}}$ values on $B_{\dot r}$.

(D) For all bias errors except B_η and B_x equal to zero

$$\eta = (\eta_o + B_\eta)\left[\frac{r}{r_o}\right]^{k\alpha_1} - B_\eta \quad (53)$$

$$\dot x = (\dot x_o + B_{\dot x})\left[\frac{r}{r_o}\right]^{k\alpha_2} - B_{\dot x}. \quad (54)$$

Here $\eta_{\text{terminal}} = -B_\eta$ and $\dot x_{\text{terminal}} = -B_{\dot x}$ in general. However, $\dot r_{\text{terminal}} = 0$ and, consequently, translational control will cause parallel velocity orientation at impact, a very undesirable condition.

(E) For all bias errors except $B_{\ddot{\eta}}$ and $B_{\ddot{x}}$ equal to zero

$$\eta = \eta_o\left(\frac{r}{r_o}\right)^{k\alpha_1} + \frac{B_{\ddot{\eta}} r_o}{|\dot{r}_o|\,(k\alpha_1 + k - 1)}\left[\left(\frac{r}{r_o}\right)^{1-k} - \left(\frac{r}{r_o}\right)^{k\alpha_1}\right] \quad (55)$$

$$\dot{x} = \dot{x}_o\left(\frac{r}{r_o}\right)^{k\alpha_2} + \frac{B_{\ddot{x}} r_o}{|\dot{r}_o|\,(k\alpha_2 + k - 1)}\left[\left(\frac{r}{r_o}\right)^{1-k} - \left(\frac{r}{r_o}\right)^{k\alpha_2}\right]. \quad (56)$$

Since $k < 1$ and $k\alpha > 1$, both $\dot{x}$ and η are definitely zero at impact. It should be recalled that $k\alpha$ must be greater than unity for $\dot{\eta}$ to go to zero at impact when $B_{\ddot{\eta}} = 0$. It would appear that angle control has decided advantages over translational control with respect to insensitivity to bias errors.

9. Appendix—Effects of Dynamic Errors

The effects of dynamic-response errors on logarithmic guidance were demonstrated through analog-computer studies. A complex quadratic response function was inserted into the system equations and the natural frequency was varied to simulate a family of such responses. The resulting equations are

$$\ddot{r} = k\left[\frac{(\dot{r})^2}{r}\right]^* \quad (57)$$

$$\ddot{r} = k\,\frac{[(\dot{r})^*]^2}{r} \quad (58)$$

$$r = k\,\frac{(\dot{r})^2}{r^*} \quad (59)$$

where

$$\frac{1}{\omega^2}\frac{d^2 f^*}{dt^2} + \frac{2\zeta}{\omega}\frac{df^*}{dt} + f^* = f.$$

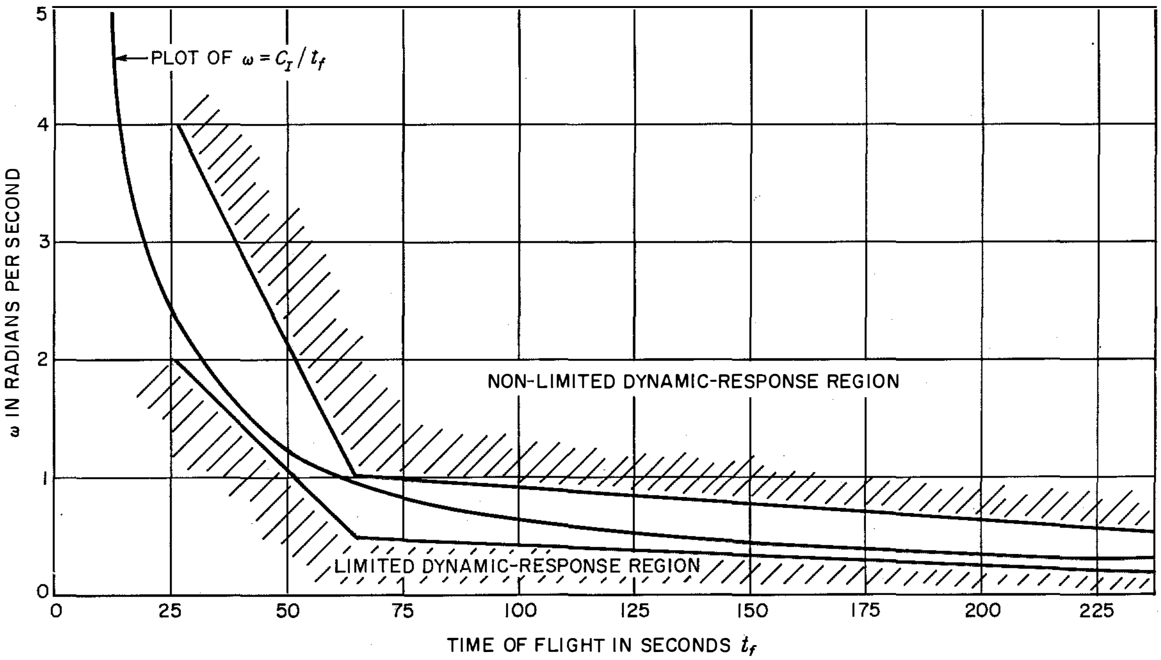


Figure 3—Dynamic-response limitation in the determination of deceleration. $t_f = r_o/[\dot{r}_o(1-k)]$ and

$$\ddot{r} = k\frac{\dot{r}^2}{r}\left(\frac{1}{1 + (2\zeta D/\omega) + (D^2/\omega^2)}\right),$$

where ω = natural frequency, $k = 0.6$, $\zeta = 0.5$ damping ratio, $D = d/dt$, and $C_I = 61$.

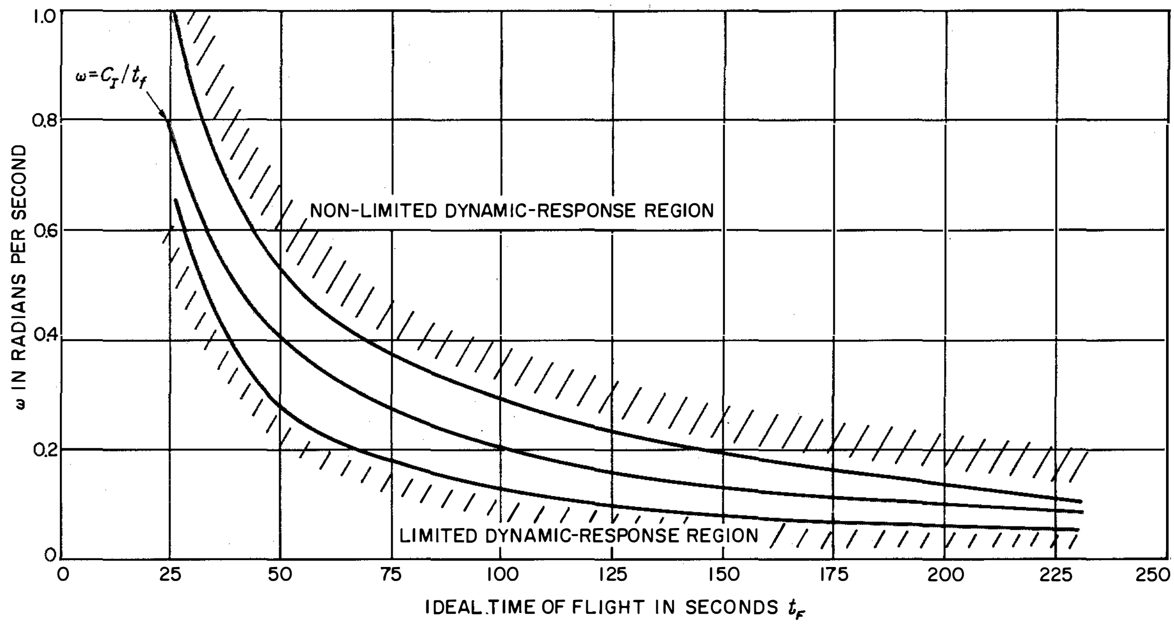


Figure 4—Dynamic-response limitation in the determination of range rate. $t_F = r_o/[f_o(1-k)]$ and $\ddot{r} = \frac{k}{r}\left[\left(\frac{D^2}{\omega^2}+\frac{2\zeta D}{\omega}+1\right)\dot{r}\right]^2$, where ω = natural frequency, $k = 0.6$, $\zeta = 0.5$ damping ratio, $D = \mathrm{d}/\mathrm{d}t$, and $C_I = 20$.

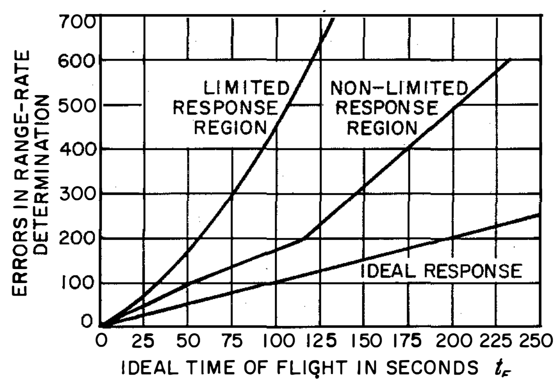


Figure 5—Comparison of ideal and actual times of flight for errors in determination of range rate. $t_F = r_o/[\dot{r}_o(1-k)]$ and $\ddot{r} = \frac{k}{r}\left[\left(\frac{D^2}{\omega^2}+\frac{2\zeta D}{\omega}+1\right)\dot{r}\right]^2$, where ω = natural frequency, $k = 0.6$, $\zeta = 0.5$ damping ratio, and $D = \mathrm{d}/\mathrm{d}t$.

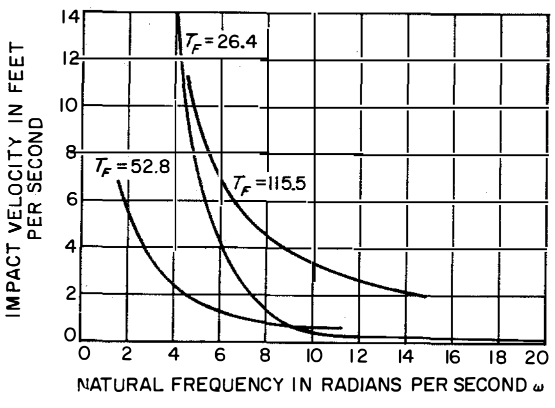


Figure 6—Effects of dynamic errors in range determination. $\ddot{r} = \frac{k\dot{r}^2}{r^*}$, $T_F = \frac{r_o}{\dot{r}_o}\left(\frac{1}{1-k}\right)$, and $r = r^* + \frac{2\zeta}{\omega}\frac{\mathrm{d}r^*}{\mathrm{d}t} + \frac{1}{\omega^2}\frac{\mathrm{d}^2r^*}{\mathrm{d}t^2}$, where $\zeta = 0.5$ damping ratio and $k = 0.6$.

The terminal behavior of the system was of paramount interest in these studies. The primary objective was approximate determination of the minimum natural frequency for which the range rate does not go to zero at impact, to be obtained for a variety of system conditions.

Figure 3 shows the results of studies of (57), giving the effects of dynamic errors in computed decelerations. Since the data presented in Figure 3 were obtained on an analog computer by successive trials of values of the natural frequency, it was not possible to determine the exact value of frequency for which the dynamic response became the limiting factor. Instead, the exact value of the frequency could only be bracketed within some region as indicated by the straight-line curves in the figure. A plot of frequency proportional to the reciprocal of the operating time is also presented for convenience. Operating time was selected as representative of the initial conditions, it being proportional to the ratio of initial range to initial range rate.

Investigation of the effects of dynamic errors in the determination of range rate was performed in the same manner as for the deceleration dynamic-error effects. The regions of limited and non-limited dynamic response are shown in Figure 4. These limited responses cause the range rate to go to zero at a finite range, whereas the limiting response effect for the studies of (57) was that of a finite range rate at impact. Another difference between the two effects is the change in operating time caused by the errors. The deceleration errors produce only a small increase in the operating time; however, range rate errors produce considerable increase in the operating time as is shown in Figure 5. This figure also shows operating time for the limited response case versus the ideal operating time (no dynamic error). This increased operating time would reduce the fuel efficiency of the system and force a compromise between fuel economy and the narrow bandwidths desirable for the sensor instrumentation. It is expected that this compromise would be for bandwidths much wider than those indicated for the limited-response conditions.

Dynamic errors in range determination are much-more critical than the other dynamic errors studied. Any dynamic error in range produces deviation from the ideal terminal conditions. In general, dynamic errors cause finite range rates at impact. Figure 6 shows these impact range rates versus the natural frequency for various initial conditions.

United States Patents Issued to International Telephone and Telegraph System; November 1960 – January 1961

Between November 1, 1960 and January 31, 1961, the United States Patent Office issued 43 patents to the International System. The names of the inventors, company affiliations, subjects, and patent numbers are listed below.

H. H. Adelaar, Bell Telephone Manufacturing Company (Antwerp), Electrical Pulse Distributors, 2 960 623.

A. Beadle, Standard Telephones and Cables (London), Telephone Transmitters, 2 960 579.

N. A. Blake and K. N. Fromm, ITT Laboratories, Flat-Copy Scanner, 2 967 906.

F. H. Bray, R. G. Wright, and G. C. Hartley, Standard Telephones and Cables (London), Automatic Telecommunication Systems, 2 960 575.

A. E. Brewster and P. E. Graham, Standard Telephones and Cables (London), Transistor Oscillator with Impedance Transformation in Feedback Circuit, 2 960 666.

N. J. Cafarelli, Jr., ITT Laboratories, Doppler Frequency Position Fixing Method, 2 968 034.

B. H. Claussen, Standard Telephones and Cables (London), Manufacture of Semi-Conductor Devices, 2 958 633.

P. P. Danesi, Royal Electric Corporation, Socket for Bulb or the like, 2 965 875.

W. Dietrich, Mix and Genest Werke (Stuttgart), Circuit Arrangement for the Automatic Voltage Regulation of Television Picture Tubes, 2 966 624.

R. Ehni, Standard Elektrik Lorenz (Stuttgart), Circuit Arrangement Supervising Tracks of Railroads with Electric Traction, 2 966 581.

H. F. Engelmann, ITT Laboratories, Antenna Array, 2 962 716.

S. J. Erst and G. E. Bowden, Farnsworth Electronics Company, Quadrant Homing System, 2 969 018.

A. Evans, Standard Telephones and Cables (London), Travelling Wave Tube, 2 960 622.

R. E. Gray and G. Q. McColl, ITT Laboratories, Telegraph Communication Systems with Carrier Monitoring, 2 967 908.

W. Hatton, L. B. Haigh, G. F. McCarthy, and A. N. Gulnick, ITT Laboratories, Coordinate Switching System, 2 964 591.

J. A. Henderson, ITT Laboratories, Memory Circuit, 2 965 884.

W. Hinz and H. Ruckert, Mix and Genest Werke, Arrangement for Detecting Characteristic Marking on Articles, 2 966 594.

S. L. Hjertstrand, Standard Radio & Telefon (Bromma), Signalling Circuit for High Capacity Exchange System, 2 962 556.

K. G. Hodgson and D. L. Thomas, Standard Telephones and Cables (London), Electric Carrier Current Communication Systems, 2 960 573.

R. E. Hufnagel, ITT Laboratories, Instantaneous Type Time Compressors and Expanders for Pulse Time Modulation Transmission Systems, 2 959 641.

L. G. Johnson and E. McVey, ITT Federal Laboratories, Frequency Modulated Oscillator System, 2 968 769.

H. A. Kalina and C. B. Goss, ITT Federal Laboratories, Line Voltage Regulator, 2 966 626.

E. Labin and S. Frankel, ITT Laboratories, Frequency Controlling System, 2 958 767.

A. M. Levine, ITT Laboratories, Automatic Machine Control, 2 958 247.

J. Luongo, R. H. Rymer, and F. P. Turvey, ITT Laboratories, Analog to Digital Translators, 2 958 861.

M. Mandel, ITT Laboratories, Automatic Gain Control Circuit, 2 958 772.

F. P. Mason and L. O. Marden, Creed and Company (London), Facsimile Apparatus, 2 966 389.

G. Q. McColl, All America Cables and Radio, Diplex Telegraph Synchronization System, 2 958 728.

O. S. Meixell, R. L. Whittle, and N. J. Cafarelli, Jr., ITT Laboratories, Radio Signal Separator System, 2 962 714.

H. Moons, Bell Telephone Manufacturing Company (Antwerp), Method of Baring Electric Ribbon Cables, 2 958 615.

W. H. P. Pouliart and F. M. Michiels, Bell Telephone Manufacturing Company (Antwerp), Guiding Means for a Magnetic Tape, 2 962 200.

A. J. Radcliffe, Jr. and A. R. Denz, Kellogg Switchboard and Supply Company, Signal System Including a Diode Limiter, 2 964 650.

A. J. Radcliffe, Jr., M. Ribner, and W. V. Sayner, Kellogg Switchboard and Supply Company, Relayless Line Circuit and Call Distributing System, 2 962 557.

D. S. Ridler and R. Grimmond, Standard Telecommunication Laboratories (London), Intelligence Storage Devices with Compensation for Unwanted Output Current, 2 958 853.

D. S. Ridler and B. F. Armsby, Standard Telephones and Cables (London), Toroidal Winding Machine, 2 962 235.

R. D. Salmon and F. J. L. Turner, Creed and Company (London), Apparatus for Checking Recorded Information, 2 960 164.

F. P. Smith, Farnsworth Electronics Company, Geiger Counter, 2 960 609.

O. Snedkerud, Electronic Plotting Station for Automatic Plotting of One or More Movable Objects, 2 958 725.

F. A. Stearn and L. P. Hopkins, Kellogg Switchboard and Supply Company, Telescoping Binding Posts, 2 960 677.

C. G. Treadwell, Standard Telephones and Cables (London), Electric Pulse Code Modulation Systems, 2 960 574.

M. W. Wallace, ITT Laboratories, Electrode Support Arrangement, 2 958 800.

R. W. Wilmarth, ITT Laboratories, Slow Wave Propagating Structure, 2 959 707.

E. P. G. Wright, J. Rice, and N. F. Fossey, Standard Telecommunication Laboratories (London), Telegraphy Encoding Equipment Comprising Magnetic Storage Means, 2 958 726.

Contributors to This Issue

D. L. Cooper-Jones

Dennis L. Cooper-Jones was born in Guernsey, Channel Islands, in 1924. He received a B.Sc. (Eng.) degree from the University of London in 1944.

From 1944 to 1947, he served with the Royal Navy. He then joined Standard Telephones and Cables and in 1951 became group leader in charge of development of commutated-aerial direction-finding systems. In recent years, he has spent much time on evaluation trials and demonstrations in Europe and North America.

Mr. Cooper-Jones is an Associate Member of the Institution of Electrical Engineers.

C. W. Earp

C. W. Earp was born in Cheltenham, Gloucestershire, England, on 14 July 1905. He received the B.A. degree with First Class Honours in 1927 from Cambridge University.

He joined the engineering staff of International Standard Electric Corporation at New Southgate in 1927, working on high-frequency transatlantic transmission. In 1929, he transferred to Paris on receiver design for ship-shore service including single-sideband techniques.

He returned to England in 1933 becoming chief of the advance development section at Kolster Brandes in Sidcup and completed the circle by returning to New Southgate in 1935, where he became section head for development of radio receivers and direction finders for Standard Telephones and Cables. In 1940, he assumed his present position of head of the newly formed advance development section of the radio division.

Mr. Earp is a Senior Member of the Institute of Radio Engineers and a Member of the Institution of Electrical Engineers.

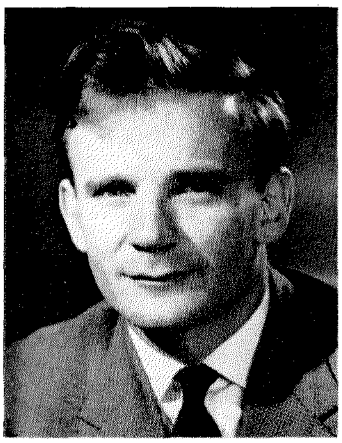
W. G. Green

William G. Green was born in Springfield, Missouri, on 6 October 1928. He received the B.S.E.E. degree from Missouri School of Mines in 1951.

In 1952, he joined McDonnell Aircraft Corporation for analysis and design of aircraft fire-control and guidance systems. From 1954 to 1956, he served Gilfillan Brothers in establishing and managing an analog-computer facility.

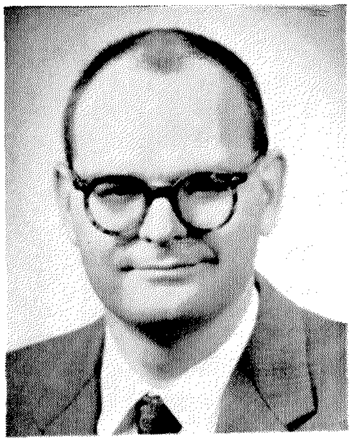
In 1956, he became affiliated with ITT Federal Laboratories in Fort Wayne, Indiana. He has been engaged in studies of guidance and control

D. L. Cooper-Jones



C. W. Earp

W. G. Green



systems of various missiles and has directed an Air Force sponsored research program on terminal space guidance.

Martin Jeppsson

Martin Jeppsson was born on 8 February 1929 in Karlshamn, Sweden. He received the degree of Master of Science in electrical engineering in 1953 from the Royal Institute of Technology in Stockholm.

After three years with L. M. Ericsson, he joined the engineering staff of Standard Radio & Telefon in Stockholm in 1954. He has been engaged in the development of open-wire carrier systems, voice-frequency repeaters, voice-frequency telegraph equipment, and data transmission apparatus. He is now in charge of the development laboratories of the transmission division.

Mr. Jeppsson is a member of the Swedish Association of Engineers and Architects.

W. T. Jones

W. T. Jones was born at Treorchy, Glamorganshire, in 1918. After attending County Grammar School at Porth, he joined the traffic department of the British Post Office Telephones in 1937.

After six years of service with the Indian Army Signals in the Far East and a brief return to the Post Office, he joined the Sudan Government Post and Telegraph Department headquarters in 1949.

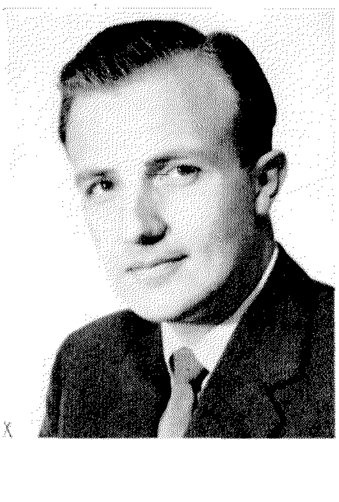
In 1955, he joined Standard Telephones and Cables, where he has been engaged on forward telecommunication system studies, which are part of the participation of the International Telephone and Telegraph Corporation in the work of the Comité Consulatif International Télégraphique et Téléphonique.

F. H. Steiner

Friedrich Hugo Steiner was born on 2 June 1913 in Villach, Austria. He graduated as a certified engineer in 1938 from the Vienna Institute of Technology, from which he also received a Doctor of Engineering degree in 1946.

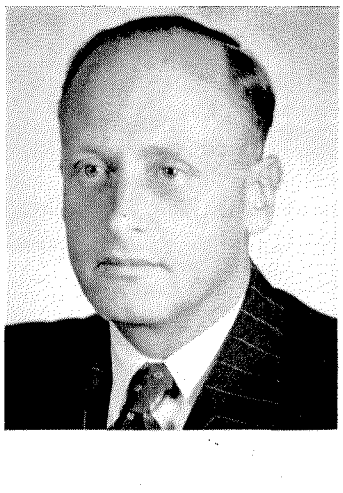
He joined one of the associate companies of the International Telephone and Telegraph Corporation in 1938. With the exception of 1954, he has continued with that system and is now with Standard Elektrik Lorenz in Stuttgart, Germany.

Martin Jeppsson



W. T. Jones

F. H. Steiner



Principal ITT System Products

Telecommunication Equipment and Systems

Automatic telephone and telegraph central office switching systems
Private telephone and telegraph exchanges—PABX, PBX
Carrier systems: telephone, telegraph, power-line
Long-distance dialing and automatic message-recording equipment
Switchboards: manual, central office, toll
Telephones: desk, wall, coin-operated
Automatic answering and recording equipment
Intercommunication, paging, and public-address systems
Microphones and loud speakers
Microwave radio systems: line-of-sight, over-the-horizon
Parametric amplifiers
Data-transmission systems
Teleprinters and facsimile equipment

Military/Space Equipment and Systems

Aircraft weapon systems
Missile fuzing, launching, guidance, tracking, recording, and control systems
Electronic countermeasures
Power systems: ground-support, aircraft, spacecraft, missile
Radar
Simulators: missile, aircraft, radar
Ground and environmental test equipment
Programmers
Infrared detection and guidance equipment
Global and space communication
Nuclear instrumentation
Antisubmarine warfare systems
System management: worldwide, local

Industrial/Commercial Equipment and Systems

Distance-measuring and bearing systems: Tacan, DMET, Vortac, Loran
Instrument Landing Systems (ILS)
Air-traffic control systems
Direction finders
Ground and airborne communication
Data-link systems
Inverters: static, high-power
Power-supply systems
Altimeters
Flight systems (autopilot)
Information-processing and document-handling systems
Electronic computers
Analog-digital converters
Mail-handling systems
Pnuematic tube systems
Broadcast transmitters: AM, FM, TV
Point-to-point radio communication
Mobile communication: air, ground, marine, portable
Closed-circuit television: industrial, aircraft, and nuclear radiation
Instruments: test, measuring
Oscilloscopes: large-screen, bar-graph
Magnetic amplifiers and systems
Alarm and signaling systems
Telemetering

Consumer Products

Television and radio receivers
High-fidelity phonographs and equipment
Refrigerators, freezers
Air conditioners
Hearing aids
Incandescent lamps
Home intercommunication equipment
Electrical housewares

Cable and Wire Products

Multiconductor telephone cable
Telephone wire: bridle, distribution, drop
Switchboard and terminating cable
Telephone cords
Submarine telephone and telegraph cable and systems
Coaxial cable
Aircraft cable
Power cable
Domestic cord sets
Fuses and wiring devices
Wire, general-purpose

Components and Materials

Semiconductors: selenium, germanium, silicon
Power rectifiers, metallic
Transistors
Diodes: tunnel, zener
Capacitors: wet, dry, ceramic
Ferrites
Tubes: power, transmitting, traveling-wave, rectifier, receiving
Picture tubes
Relays and switches: telephone, industrial
Magnetic counters
Resistors
Varistors
Fluorescent starters
Transformers
Quartz crystals
Crystal filters
Printed circuits
Hermetic seals
Magnetic cores

VOLUME 37 • 1961 • NUMBER 2