

T H E B E L L S Y S T E M

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XLII

MAY 1963

NUMBER 3

Design of a 150-megacycle Pocket Receiver for the BELLBOY
Personal Signaling Service A. E. KERWIEN AND L. H. STEIFF 527

A "Thermodynamic" Theory of Traffic in Connecting Networks
V. E. BENEŠ 567

Flexural Vibrations of a Propped Cantilever R. L. PEEK, JR. 609

Numerical Computation of Phase from Amplitude at Optical
Frequencies D. E. THOMAS 637

Bounds on Communication D. SLEPIAN 681

Analysis of Delay in Mathematical Switching Models for Data
Systems D. G. HAENSCHKE 709

Contributors to this Issue

737

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

- H. I. ROMNES, *President, Western Electric Company*
J. B. FISK, *President, Bell Telephone Laboratories*
J. E. DINGMAN, *Executive Vice President,
American Telephone and Telegraph Company*

EDITORIAL COMMITTEE

- A. C. DICKIESON, *Chairman*
F. T. ANDREWS, JR. L. E. DYER
A. J. BUSCH R. W. EHRLICH
L. R. COOK K. E. GOULD
R. P. CROSS W. C. HITTINGER
R. L. DIETZOLD J. R. PIERCE

EDITORIAL STAFF

- G. E. SCHINDLER, JR., *Editor*
L. M. COLE, JR., *Assistant Editor*
A. ROSENTHAL, *Production Editor*
J. T. MYSAK, *Technical Illustrations*
T. N. POPE, *Circulation Manager*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y., E. J. McNeely, President; Allen G. Barry, Vice President and Secretary; J. J. Scanlon, Vice President and Treasurer. Subscriptions are accepted at \$5.00 per year, single copies \$1.25 each. Foreign postage is \$1.08 per year or 18 cents per copy. Printed in U.S.A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XLII

MAY 1963

NUMBER 3

Copyright 1963, American Telephone and Telegraph Company

Design of a 150-megacycle Pocket Receiver for the BELLBOY Personal Signaling System

By A. E. KERWIEN and L. H. STEIFF

(Manuscript received January 28, 1963)

A highly sensitive pocket receiver with a code-responsive signaling device has been designed for the 150-megacycle BELLBOY personal signaling system. Ten transistors in a single-IF, superheterodyne receiver circuit convert the FM signaling wave to produce excitation of a sound generator through a tuned reed selector circuit. Printed wiring and novel packaging techniques are employed to produce a receiver of acceptable size and weight. A rechargeable battery with provision for home charging or a primary battery is used for power supply.

A discussion of design problems and an analysis of circuit performance is included. Sensitivity sufficient to signal in a 20-microvolt per meter field is achieved.

I. INTRODUCTION

This paper will describe the electrical and physical features of the 55A radio receiver and certain associated apparatus, which were designed for use in BELLBOY personal signaling systems operating in the 150-megacycle common carrier band of frequencies. The receiver is pocket-sized and provides, in effect, an extension of the telephone bell to the customer's pocket so that he may be alerted while away from his office, home, or base location. An incoming signal, bearing the customer's specific code, triggers the receiver. The receiver then emits a continuous tone, in response to which the customer calls his base telephone to receive his message, or responds in some other prearranged manner.

The requirements and applications of this service have been covered

in a previous paper.¹ In the present paper, discussion is concerned with the requirements placed on size, cost, and performance of the receiver, and the various circuit and equipment features which were used to achieve these ends. Major problems encountered in both the electrical and mechanical design are discussed in some detail, and an analysis of the circuit performance is included.

The 55A radio receiver is a ten-transistor superheterodyne circuit packaged in a molded plastic case. It has a self-contained antenna and is powered by a battery mounted in a detachable case. When a rechargeable battery is used, a charger is provided, which will accept the battery case and permit the battery to be charged by placing the assembly in any convenient 117-volt ac outlet.

II. OBJECTIVES

To meet system objectives for personal signaling service, the receiver must be able to respond to a calling signal when hand-held or pocket-borne, when located within or outside of buildings, whether some distance from or close to the transmitter site, and in many and varied environments. The receiver must therefore be very sensitive. It must also possess good stability against temperature variations, against shock and vibration of transportation or handling, and against normal drift in voltage of the battery source.

In addition to the above general requirements, the following performance objectives were set on the receiver:

(a) The receiver should be designed to permit reception on any assigned channel in the frequency range from 152.51 to 152.81 megacycles.*

(b) The receiver should respond to a carrier frequency which is frequency-modulated with three audio-frequency tones, simultaneously applied, each at a deviation of 1.3 kilocycles. Each receiver must respond to only one combination of tones sent out from the base transmitter. (In a fully loaded system, 3200 useful codes¹ are derived from the combinations of the three out of thirty-two available tone frequencies in the range from 500 to 1000 cycles.)

(c) The local oscillator of each receiver must possess a frequency stability of ± 0.0005 per cent, or better. (The base transmitter frequency is maintained to ± 0.0001 per cent, or one part per million.)

(d) The receiver must respond to this specified wave for any environmental field strength between 26 and 100 db above one microvolt per meter.

*The actual receiver is capable of being tuned to frequencies somewhat beyond this range, but performance in such circumstances would be subject to restrictions, especially regarding occupancy of the image frequency band.

(e) With 30-kc channel spacing, the receiver must have a selectivity of at least 80 db against an adjacent channel carrier.

(f) All requirements should be met in an ambient temperature range from 50 to 110 degrees F.

(g) Radiation from the receiver must meet requirements of Part 15 of the FCC rules governing restricted radiation devices. For the frequencies of interest in this receiver, these requirements are: the field due to the 75-mc local oscillator must not exceed $50 \mu\text{v}/\text{m}$ at a distance of 100 feet. The field due to the 150-mc conversion frequency must not exceed $100 \mu\text{v}/\text{m}$ at a distance of 100 feet.

(h) In addition to the FCC requirements, radiation from the receiver must not be strong enough to cause the sensitivity of a second similar receiver at 5 feet distance to be reduced by more than 6 db.

(i) The receiver should operate from a self-contained, rechargeable battery with provisions for home charging. The receiver should operate without recharging the battery for at least 10 hours. As an alternate the receiver should operate from a disposable battery, which should provide at least 75 hours of operation before replacement.

(j) The signaling sound output of the receiver, when it is carried in an inside pocket, should be clearly audible in a reasonably strong noise environment.

The needs of the customer, as well as economic considerations, affected the design of the receiver package. To suit the customer, who must carry the receiver, it needed to be as small and light as possible, and attractive as well. It was required to be completely self-contained, with no appended antenna or battery box. However, the cost and ease of manufacture, as well as reliability in operation and ease of repair, are factors which tended to place a limit on the smallness and compactness of so complex a unit. Naturally, the final design represents an economically feasible compromise between these opposing influences.

Such objectives naturally posed very difficult design problems. The premium on small size and weight limited the available power from the battery. Therefore, to obtain a suitable interval of service before recharge, the current drain of the receiver had to be minimized. Also, the size and number of circuit components had to be kept small, which called for utmost simplification of the circuits.

III. ELECTRICAL DESIGN FEATURES

3.1 *General Circuit Description*

The selectivity and sensitivity requirements dictated the choice of a superheterodyne circuit with at least one low intermediate frequency. In

the early work, conventional approaches using either a crystal filter or an electromechanical filter were explored, but the complexity and large number of components involved seemed contradictory in the face of space and cost limitations. Therefore a less conventional solution involving a single, very low intermediate frequency was adopted.

The circuit of the receiver is considered for convenience of discussion as consisting of the following major parts: RF circuits, IF circuits, discriminator, reed circuit and sound oscillator. These sections will be described briefly, referring to the circuit schematics, Figs. 1 to 4.

The RF circuit (Fig. 1) consists of the antenna, two RF amplifier stages, the RF mixer, and the local oscillator. Shielding is indicated by the broken lines. The output of the mixer is the 6-kilocycle IF, which is delivered to the IF amplifier. The first IF stage is included within the shielding of the RF compartment. In the IF circuits (Fig. 2), the input signal is amplified by one transistor amplifier, passed through a low-pass filter which acts as the IF filter, and is then amplified by two more IF amplifiers. The fourth IF stage is operated as a limiter. This stage is an overloaded amplifier which, for medium to strong signals, produces a square-wave-like output. This output is delivered to the discriminator (FM detector), shown in Fig. 3. The audio output from the discriminator is passed on to the reed circuits (Fig. 4). Here the tone content is amplified to a strength sufficient to operate the tuned armatures of the reed selector units.² Only when the signal contains the proper code will all three reeds be simultaneously stimulated. In that case, a circuit through the reed contacts delivers an impulse to the sounder circuit, which triggers it into oscillation. This causes an audio transducer to emit a continuous tone which signals the customer. To stop the audio sound output, the user must operate a miniature pushbutton, which then resets the circuit and places the receiver in readiness for the next call.

3.2 *IF Plan*

While the use of a single conversion and the low (6-kc) intermediate frequency in this receiver was a practical solution to the space and cost problem, it brought with it an interesting set of associated problems, some advantages and some disadvantages. One advantage was that the IF filter became a simple low-pass structure, inexpensive when compared to an electromechanical or crystal bandpass filter which would be required for a higher IF. Also, the IF amplifier could be designed to use relatively inexpensive alloy junction transistors. From an interference standpoint, the single, low IF remains an advantage only as long as its lone image response falls in an unoccupied space in the spectrum. The

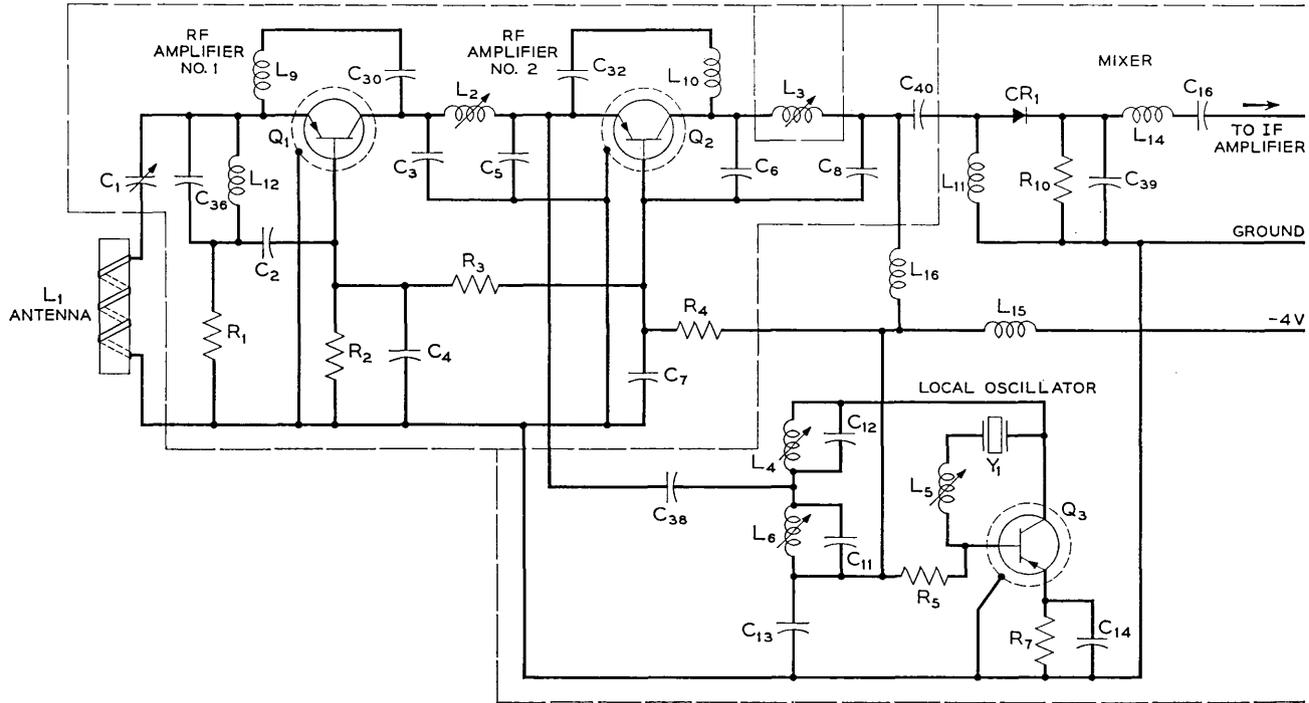


Fig. 1 — RF circuits.

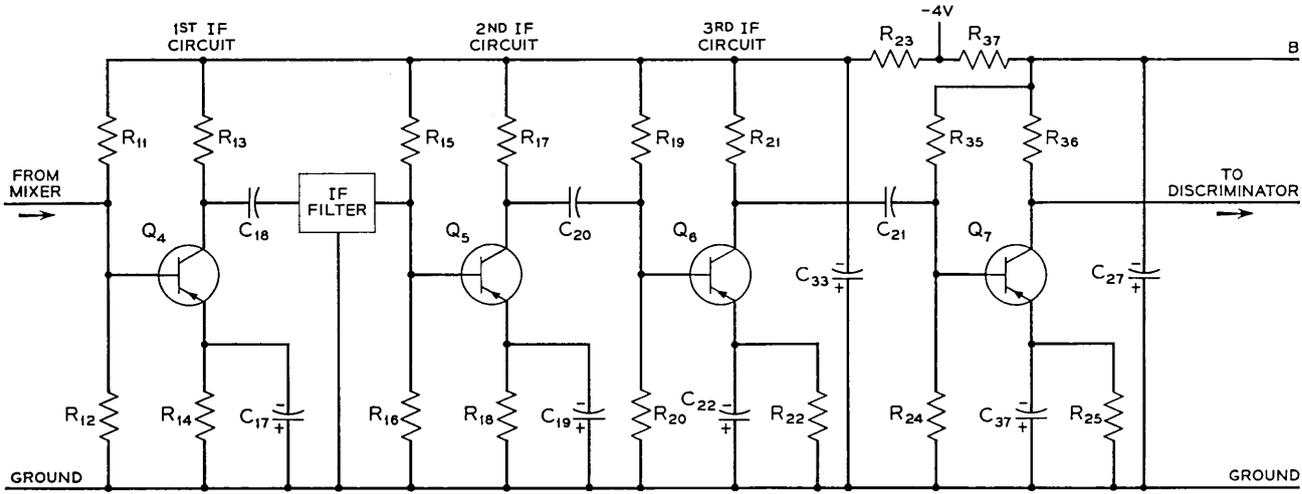


Fig. 2 — IF circuits.

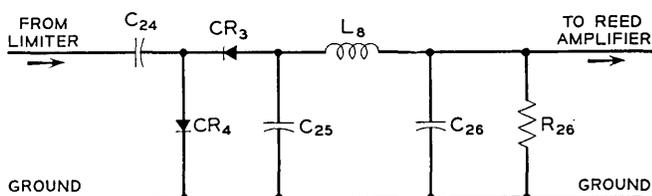


Fig. 3 — Discriminator.

choice of the 6-kc IF frequency was predicated on the assumed use by this service of one of the common-carrier mobile channels. Thus the image frequency, which is spaced 12 kc from the received carrier, falls within a channel width of ± 15 kc (see Fig. 5).

Conventional superheterodyne receivers, with intermediate frequencies considerably higher than that used herein, obtain image rejection by means of RF selectivity. Such selectivity eliminates thermal noise contributions in the vicinity of the image frequency. Since no rejection of the image frequency exists in this design, a 3-db penalty in the noise figure must be accepted.

Another disadvantage of the low IF comes about from $1/f$ noise³ modulation in the high-frequency beating oscillator. This noise modulation, characterized by sidebands which are strongest in the vicinity of the oscillator frequency, is detected by the mixer and appears as extraneous noise energy in the IF amplifier. In receivers employing higher intermediate frequencies, the $1/f$ noise modulation is less significant.

Because the frequency of the local oscillator is so close to the incoming signal frequency, no attenuation of the oscillator frequency is achieved in the RF amplifier tuned circuits. Therefore, the opportunity for spurious outputs of oscillator energy via reverse transmission through the amplifier is greater than would exist if the IF were considerably higher in frequency. This can produce interference in other nearby receivers as discussed in the next section.

To include the necessary sidebands of the intermediate frequency, and at the same time to attenuate the signaling tone frequencies, the IF amplifier was designed to cut off frequencies below 2 kc (Fig. 6). The discriminator output filter was designed to attenuate the IF residue above 2 kc. In the crossover region near 2 kc, the tandem gain of the IF amplifier and the reed amplifier remained sufficient to require very careful control of these characteristics to avoid instability due to inadvertent over-all feedback.

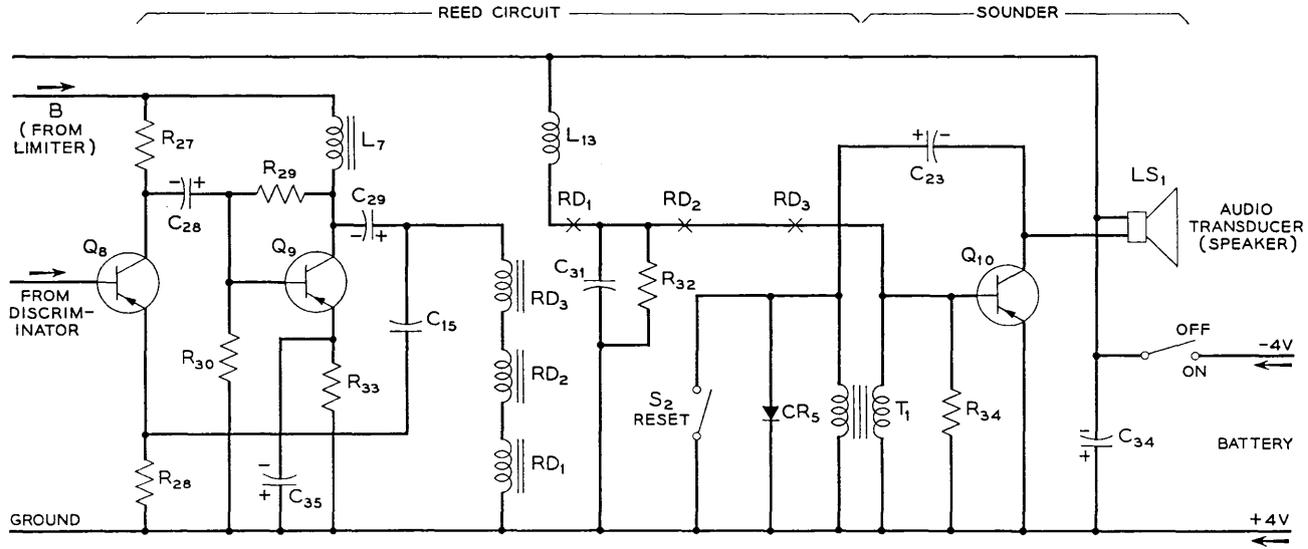


Fig. 4 — Reed circuit and sound oscillator.

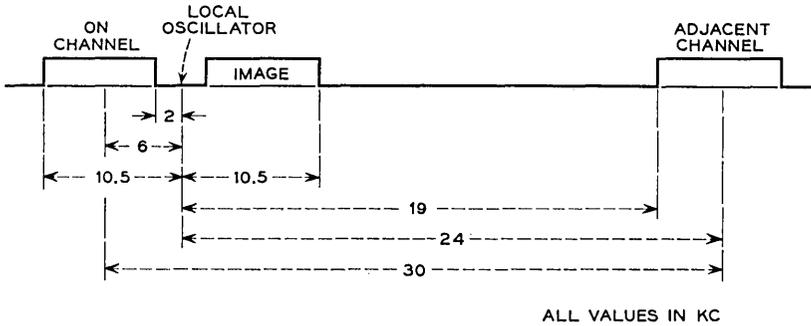


Fig. 5 — RF channel spectrum.

3.3 Frequency Stability

As many as twenty base transmitters may be used in a large metropolitan area to provide the required coverage. To prevent the generation of beat tones in the receiver, which might interfere with signaling, the frequencies of base transmitters are held to an accuracy of one part per million (± 0.0001 per cent).¹

Although oscillator radiation from the receiver has been kept within the stated requirements, there may be instances (as when two customers meet in conversation) in which a beat due to the difference of two local oscillators will occur in the mixer stages of each. If this beat is high enough in frequency it will be transmitted through the IF, causing de-

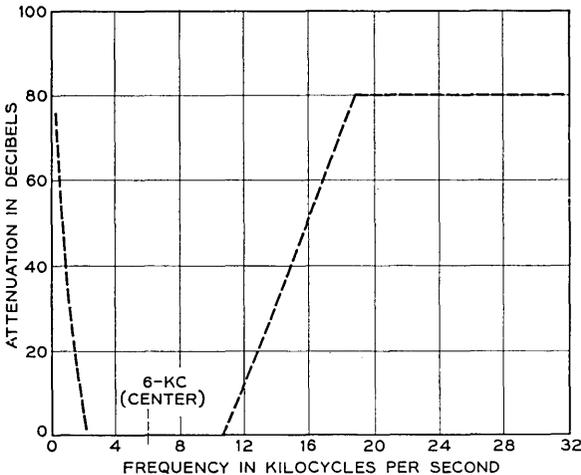


Fig. 6 — IF response requirements.

sensitization. If it is too low to pass through the IF, it may still be carried to the reeds as a tone modulation on the desired signal. In this latter case, if its frequency coincides with that of one of the three reeds, it could conceivably cause false signaling in response to any code containing the other two reed frequencies.

In an attempt to maintain these beat frequencies below the IF transmission band, the oscillators are adjusted within ± 125 cps at the factory. The crystals maintain this frequency within ± 600 cps of the original adjustment throughout the temperature range $+10$ to $+45$ degrees C. If the temperature coefficients are oppositely signed, it would then be possible for two such oscillators to differ by as much as 1450 cps at some temperature within the range. In addition to this, if two crystals did not enter service at the same time, an aging factor could add to or subtract from this difference. Thus, it is evident that desensitization due to direct feed into the IF is possible. However, it is considered improbable that the deviations due to extreme temperature, opposite-signed coefficients, and age difference would all add during a chance meeting of two customers.

In considering the probability of false signaling due to a beat within the reed frequency range (500 to 1000 cps), it is necessary to remember that each of the three reeds responds only to an extremely narrow frequency band (about 1.2 cps). Thus the probability of the beat falling into one of these slots is indeed small. Since such interference also depends on the coincidence of a number of low probability factors, it is not expected to be a serious field problem.

The positioning of the IF signal in the band of transmission defined by the filter and the low frequency cutoffs of the amplifier is affected by all deviations of the oscillator and transmitter combined. Thus the total of all such deviations, including the peak deviation due to modulation, ideally should be contained within the IF transmission bandwidth. This bandwidth is approximately 8.5 kilocycles.

TABLE I

Cause of Deviation	Max. Range
Base Transmitter:	
Oscillator (tolerance $\pm 0.0001\%$)	0.3 kc
Peak modulation (deviation ± 3.9 kc)	7.8 kc
Receiver:	
Temperature (± 4 ppm between 10°C and 45°C)	1.2 kc
Crystal aging (± 5 ppm first year)	1.5 kc
Total	10.8 kc

Table I lists the factors involved in determining the required receiver IF bandwidth.

If we add the inaccuracy of the initial setting of the oscillator frequency, which is held within ± 125 cps, it is obvious that peaks of modulation may often be in danger of spilling outside the IF transmission bandwidth. Fortunately, peak modulation due to the addition of three sinusoidal tones occurs only a small percentage of the time. Thus these peaks may be degraded without serious loss of signaling sensitivity.

Experiments have been performed in which the frequency has been deliberately moved off-center in the IF band. By this means it has been demonstrated that a displacement of ± 1.5 kc produces less than 2-db degradation of signaling sensitivity.

3.4 *Battery Considerations*

The limitation on space and weight was one of the most serious factors in the choice of a suitable battery. A 3-cell, nickel-cadmium battery supplies about 3.7 volts and possesses the advantage that it can be recharged on a routine basis. For this purpose a simple home charger is provided, which may be supplied to the customer by the telephone company. A mercury battery with a nominal voltage of 4 volts is also available. This battery will provide service for about two weeks of average usage, before replacement.

Because of the limited battery capacity, circuits were required which provided the necessary gain with the lowest possible power drain. Special circuit designs were evolved, in some cases, to accomplish these objectives. For example, it was determined early in the development that greater gain in the RF circuits, for a given dc power input, could be obtained by operating the two diffused-base, germanium transistors in series from a dc standpoint, rather than in parallel.

Although the receiver is already in commercial service, development is continuing to improve the characteristics and life of the rechargeable battery. The outcome of this development may necessitate modifications in the battery case and also in the charger.

3.5 *Power Level Diagram*

Fig. 7 shows a block diagram of the receiver and an associated graph giving the power level in dbm of both signal and noise at significant points through the circuit. The noise is shown for the absence of signal. The signal is shown for the just-operate condition of the reeds, and the signal powers given at the reed driver amplifier (RDA) input and output

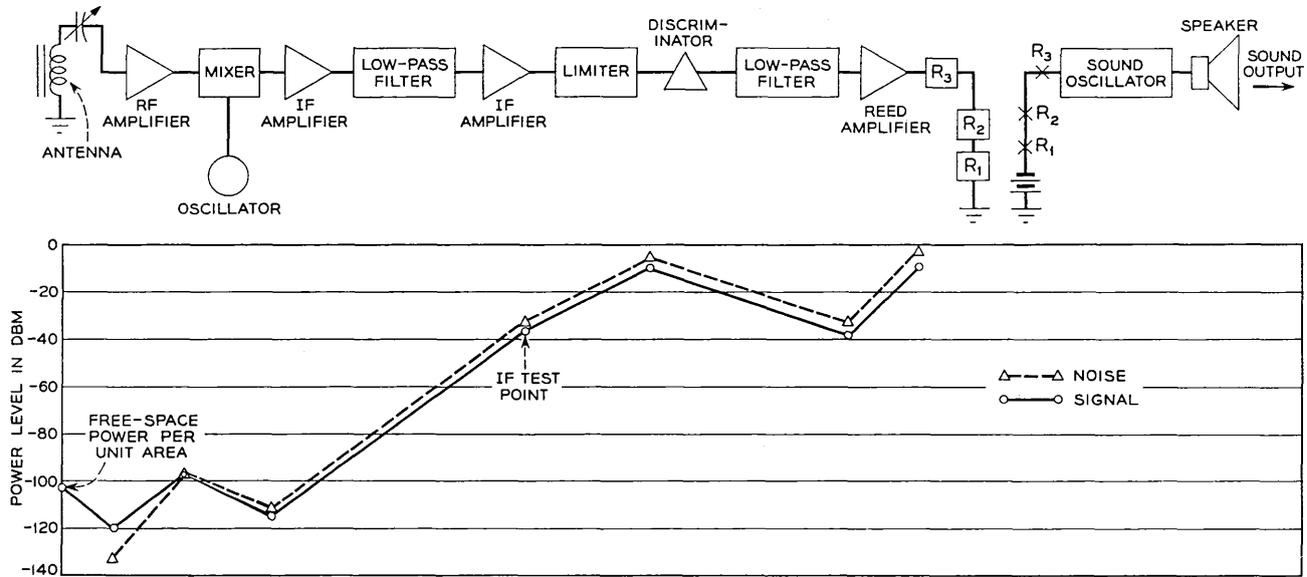


Fig. 7 — Receiver power level diagram.

are those for one of the three signaling tones. The signal values shown prior to the discriminator are for the power in the RF and IF wave regardless of modulation.

A readily measured reference is the level at the IF test point, where a signal-to-noise ratio of about -2 db exists for the just-signaling condition.

The levels in the early sections of the receiver are based on measurements of the IF amplifier gain, on laboratory measurement of the RF gain available in circuit jigs, and on computed values of noise at the RF input. The method of constructing the level diagram fixed the mixer output on the basis of measurements at IF, and the mixer input on signaling sensitivity input data and the measured value of RF gain. The difference between mixer input and output indicates 18-db conversion loss. The difference between signal-to-noise ratios at the first RF input and at the IF test point indicates a noise figure of 15 db, which agrees reasonably well with the result of the noise figure analysis to follow.

A signal level point is provided at the left of the chart showing the power per unit area in space carried by a signal wave whose field strength is at the required value, i.e., 26 db above 1 microvolt per meter. This is derived in the discussion of antenna effectiveness to follow.

3.6 Noise Figure Analysis

Since there is no image rejection, the available RF noise power must be calculated in a bandwidth twice that of the IF. For 8.5-kc IF bandwidth, therefore, the RF noise bandwidth is 17 kc. The available noise power (at 290 degrees Kelvin) is

$$\begin{aligned} P_n &= KTB \\ &= 1.38 \times 10^{-23} \times 290 \times 17 \times 10^3 \times 10^3 \text{ mw} \\ &= 6.8 \times 10^{-14} \text{ mw} \end{aligned}$$

$$10 \log P_n = -131.7 \text{ dbm.}$$

The input impedance of the RF amplifier transistor is determined by measurement from data in Fig. 9, which will be discussed later in connection with the neutralization of the RF stages. This impedance is

$$Z_i = 25 + j38.6.$$

Data on typical receivers indicated that the RF power input to this impedance required to cause 3-db increase of the energy measured at IF stage 3 output was approximately

$$P_s = -117.5 \text{ dbm.}$$

Assuming no further change in the signal-to-noise ratio at points beyond the IF test point, the difference in db between P_n and P_s is the noise figure

$$F_t = -117.5 - (-131.7) = 14.2 \text{ db.}$$

This correlates with the 15-db change in signal-to-noise ratio which appears between the RF input and the IF test point on the level diagram. Measured noise figures on a few sets ranged from 11 to 14 db.

The noise figure results from contributions of excess noise in each of the earlier stages of the receiver. The importance of each contribution is shown by the well-known formula⁴ for over-all noise figure,

$$F_t = F_1 + \frac{F_2 - 1}{A_1} + \frac{F_3 - 1}{A_1 A_2} + \dots$$

in which F_1, F_2, F_3 , etc. are noise figures for the successive individual stages, and A_1, A_2 , etc. are the power gain ratios of the successive individual stages.

Assume $A_1 A_2 = 23$ db, which is the gain of the two RF stages shown on the level diagram, and assume that the individual stage gains, A_1 and A_2 , are equal. Then

$$\begin{aligned} A_1 &= 14.14 \text{ (power ratio)} \\ A_1 A_2 &= 200 \text{ (power ratio)}. \end{aligned}$$

Also assume F_1 and F_2 are each 8 db or a power ratio of 6.3.

Then, for $F_t = 14.2$ db (or 26.3 power ratio), the contribution of the remainder of the set may be calculated:

$$26.3 = 6.3 + \frac{5.3}{14.14} + \frac{F_3 - 1}{200}.$$

Solving, we get

$$F_3 = 35.9 \text{ db.}$$

It is thus apparent that the third term, containing the noise figure of the mixer, is the heaviest contributor to the over-all result. About 18 db of this is due to the conversion loss, as indicated on the level diagram. A considerable amount is attributed to the noise figure of the diode. Another very considerable portion is the result of $1/f$ noise modulation carried by the local oscillator energy.

Although, in the above discussion, no stage beyond the mixer was considered, there is at least a noticeable contribution from the first IF transistor which may be harmful if not controlled.

3.7 Antenna Effectiveness

The antenna effectiveness is largely dependent on which way the receiver, or the person wearing it, is facing relative to the transmitter. As has been noted by Mitchell and Van Wynen,¹ the presence of the human body provides gain in some orientations, while in others it provides shielding, resulting in rather severe loss. Antenna effectiveness, averaged over all orientations, is a useful criterion.

According to Schelkunoff and Friis,⁵ the effective area of a receiving antenna is the ratio of the maximum power received at its terminals from a linearly polarized wave, to the power per unit area in the wave. Thus

$$A = (240\pi P_r/E^2)$$

where

A = Effective area of the antenna in square meters.

E = Field intensity of the wave in microvolts per meter.

P_r = Power received by the load connected to the antenna terminals.

A receiver of average sensitivity will signal satisfactorily in a field of +26 db relative to 1 $\mu\text{v}/\text{m}$ averaged over all orientations (or 20 microvolts per meter).

Thus

$$E = 20 \times 10^{-6} \text{ volt per meter.}$$

As shown on the level diagram, assume the signaling power to be -120 dbm at the antenna output, or

$$P_r = 1.0 \times 10^{-15} \text{ watt.}$$

Thus the effective area of the antenna, when worn on the body and averaged for all orientations, is

$$\begin{aligned} A_{av} &= \frac{240\pi(1.0 \times 10^{-15})}{400 \times 10^{-12}} \\ &= 1.886 \times 10^{-3} \text{ square meter.} \end{aligned}$$

The effective area of a half-wave dipole antenna is

$$A = 0.13\lambda^2.$$

For 150 mc,

$$\lambda = 2 \text{ meters and}$$

$$A = 0.52 \text{ square meter.}$$

Therefore the gain of the receiver antenna averaged over all orientations,

with respect to a half-wave dipole, is

$$G_{\text{av}} = 10 \log (0.001886/0.52) = -24.4 \text{ db.}$$

Some experimental information showed that, on the average, the gain at optimum orientation with respect to the field is about 6.8 db above the average gain. Thus

$$G_{\text{max}} = -24.4 + 6.8 = -17.6 \text{ db (at optimum).}$$

Note that the power per unit area carried by the wave is

$$\begin{aligned} W &= \frac{E^2}{240\pi} = \frac{400 \times 10^{-12}}{240\pi} \\ &= 0.053 \text{ } 10^{-12} \text{ watt} \\ &= -132.8 \text{ dbw} \\ &= -102.8 \text{ dbm.} \end{aligned}$$

This point is plotted as antenna input power on the level diagram, Fig. 7.

A number of other antenna types were tested in the course of the development. The present design is probably not as great in effective area as some other configurations which were tested. It was adopted in preference to types which produced undesirable coupling of the antenna to other circuits of the receiver, resulting in instability, and other types which suffered detuning due to body presence.

3.8 RF Neutralization

Partial neutralization of the RF amplifier transistors was accomplished by providing an inductor between emitter and collector of the common-base amplifier, as shown in Fig. 1.

From a statistical analysis based on a modest sample of transistors in the early stages of production, element values were assigned to an equivalent circuit of a typical transistor. This network is shown in Fig. 8(a). Fig. 8(b) shows Z_F as the neutralizing element applied. This network is resolved in Fig. 8(c) into two parallel networks N_1 and N_2 .

The y -parameters of these two networks were calculated and a well-known theorem of matrix algebra was applied. This states that each of the y -parameters of the combined network is equal to the sum of the corresponding parameters of the two component networks. Applying this, the feedback parameter (Y_{12}) for the combined network was computed in terms of the neutralizing element (Z_f) and equated to zero. Solving,

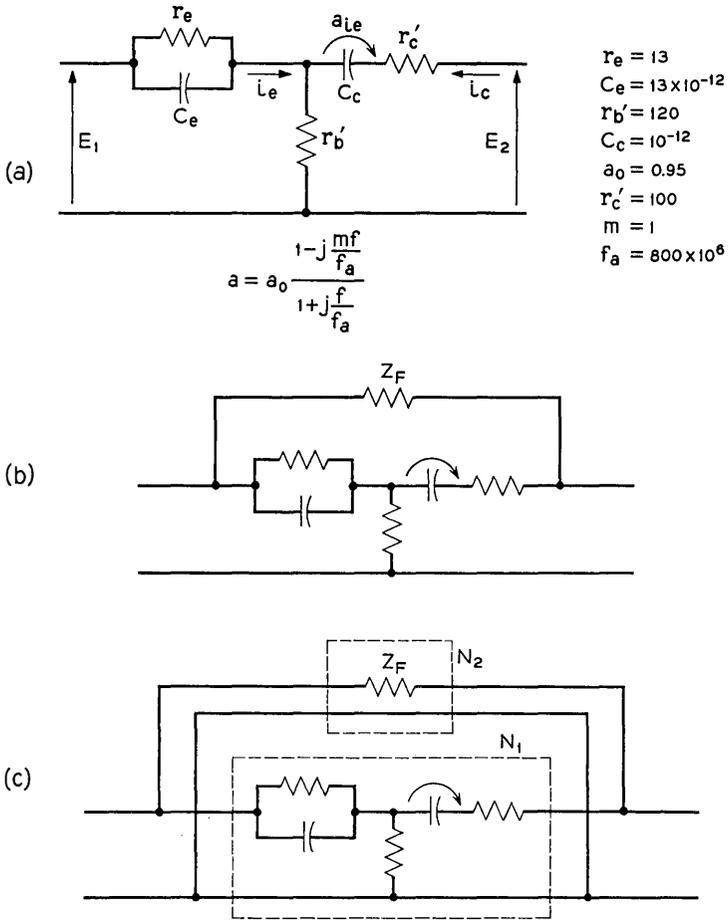


Fig. 8 — (a) Assumed equivalent circuit for RF amplifier transistor; (b) assumed equivalent circuit with neutralizing impedance added; (c) component networks used in analysis of neutralization.

the admittance ($1/Z_f$) required for perfect neutralization was found to be

$$(1/Z_f) = -0.002155 - j0.000873.$$

This represents a coil of reactance $+j1142$ ohms in parallel with a resistor of -463 ohms connected between emitter and collector. Since a negative resistance is impractical in this circuit design, the coil alone was used to give only partial neutralization.

By taking into account the capacitance of the inductor and its mounting, very good agreement was found between the computed value of inductance and the value which was found to be most effective by experiment.

Results of impedance measurements of the input and output of an amplifier stage which employed a transistor of median characteristics, according to the above-mentioned analysis, are shown in Figs. 9 and 10. The test circuits are shown on the figures. These show that in each case the use of the neutralizing coil ($L_N = 0.68 \mu\text{h}$) has little effect on the measured value of reactance, but the variation of resistance is considerably improved. It is to be noticed that the measuring terminals, in each

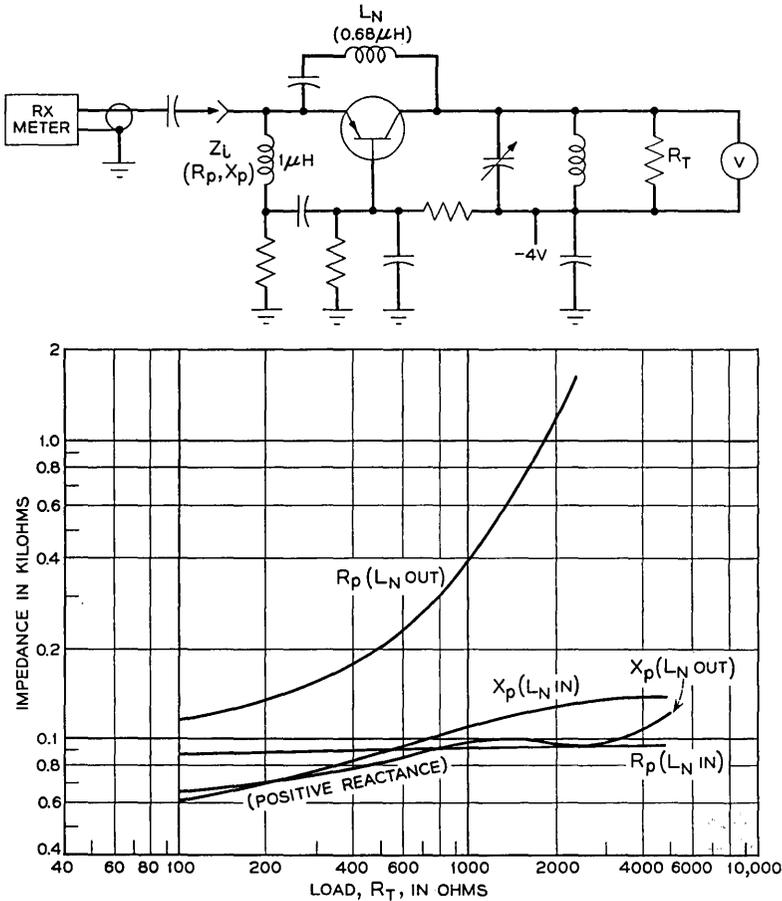


Fig. 9 — RF amplifier, neutralization effect on input impedance.

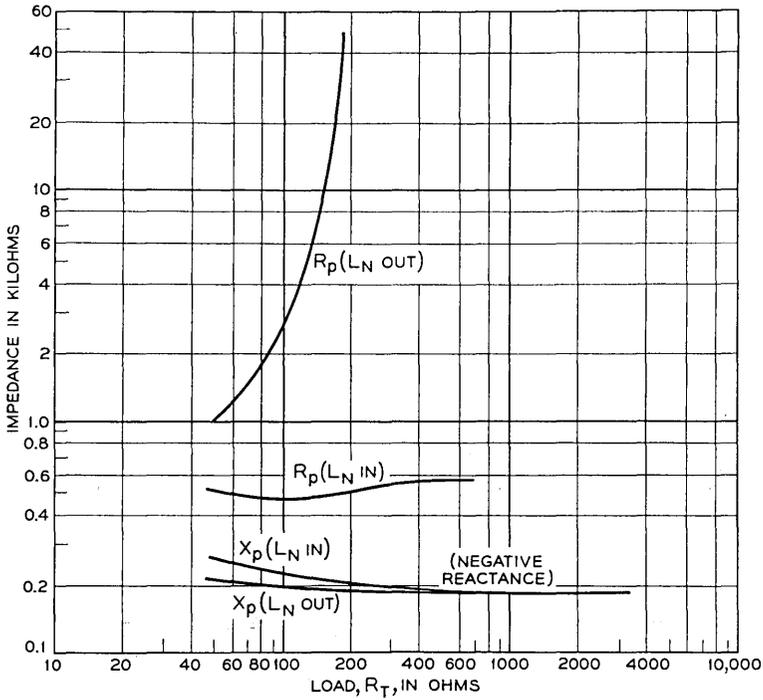
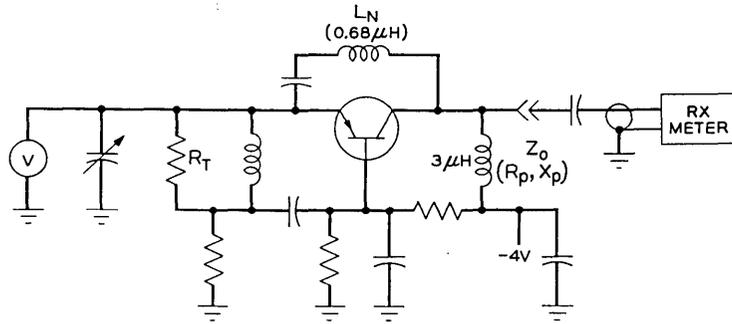


Fig. 10 — RF amplifier, neutralization effect on output impedance.

case, are shunted by a choke, whose reactance is included in the measured values together with circuit strays.

It is interesting to compare the measured values of input impedance with those calculated for the “median” transistor from its equivalent circuit. The calculated impedance is

$$Z_i = 25 + j38.6$$

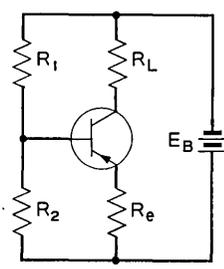
which corresponds to 85 ohms resistance in parallel with $+j55$ ohms reactance. From the curves, assuming 500 ohms load, the values are 89 ohms in parallel with $+j82$ ohms reactance.

3.9 Bias Stabilization

For uniformity of performance with variations due to ambient temperature, battery voltage, and the normal tolerances on components, some stabilization of the transistor bias is necessary. The design must accommodate the difference in battery voltage between the mercury battery (4 volts) and the standard nickel-cadmium rechargeable battery (3.7 volts). Since neither battery affords a surplus of energy to be consumed in biasing circuits, only a modest degree of stabilization was possible.

The general principles of bias stabilization are treated in many texts on transistor circuit design: for example, in Chapter 6 of Shea.⁶ The basic bias circuit used widely in the 55A receiver is shown in Fig. 11. The effect on the collector current (I_c) of the variations of the saturation current (I_{co}) due to temperature, is designated as a factor S , which it is desired to minimize. The best stability is thus achieved when the emitter resistor (R_e) is made as large as feasible and the parallel combination of R_1 and R_2 is made as small as feasible.

Fig. 12 shows the bias circuit used in the RF stages, where the transistor currents are connected in series. An emitter resistor (R_1) stabilizes the current of the first transistor, while the base bias voltages of both transistors are fixed by the resistor chain $R_2, R_3,$ and R_4 across the battery. The factor S for the first transistor is estimated to be about 4.5. The first collector current which is stabilized to this degree is auto-



$$S = \frac{dI_c}{dI_{co}}$$

$$S = \frac{k}{k-\alpha}, \text{ WHERE } k = 1 + \frac{R_e}{R_2} + \frac{R_e}{R_1}$$

$$k - \alpha = 1 - \alpha + \frac{R_e}{R_2} + \frac{R_e}{R_1}$$

FOR GOOD VALUES OF α , $1 - \alpha \approx 0$

$$S \approx \frac{1 + \frac{R_e}{R_2} + \frac{R_e}{R_1}}{\frac{R_e}{R_2} + \frac{R_e}{R_1}} = 1 + \frac{1}{\frac{R_e}{R_2} + \frac{R_e}{R_1}}$$

$$\therefore S \approx 1 + \frac{R_x}{R_e}$$

IN WHICH,
 $R_x = R_1$ IN PARALLEL WITH R_2

Fig. 11 — 55A receiver — bias stabilization, single stage.

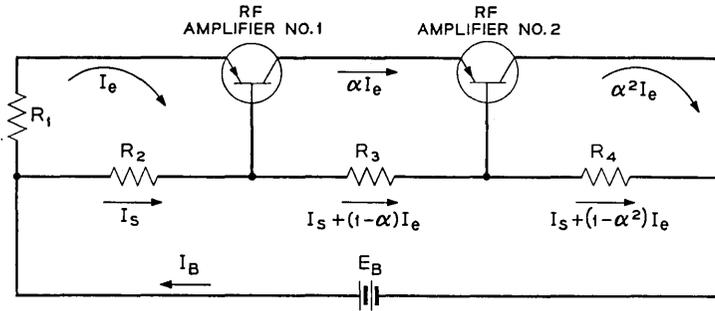


Fig. 12 — 55A receiver — RF amplifiers transistor bias circuit.

matically the emitter current of the second transistor. Thus, we might say that the emitter current of the second RF transistor is stabilized by the output resistance of the first transistor, which is relatively high.

On the basis of Fig. 11, the factor S for the IF amplifier stages is about 9.5. This poorer stabilization of the IF stages is in part compensated for by the fact that the IF transistor characteristics are carefully controlled, particularly with respect to I_{co} . Temperature variation tests showed that satisfactory stability has been attained.

3.10 Semiconductor Devices

In the design of this receiver, advantage was taken of the best available semiconductor devices, and in fact, the demand created through this application has had considerable influence on the characterization of the devices used. To meet the requirement of small size of the receiver, it was necessary to miniaturize the encapsulation of the transistors and diodes.

Gold-bonded germanium diodes are used both in the mixer and in the discriminator.

A set of seven germanium alloy junction transistors, coded as the 28A transistor, is used in the IF amplifier, limiter, reed amplifier and sound oscillator stages. These are mechanically and electrically the same as the WE 17A transistor but individually identified as to the range of the parameter h_{fe} (beta). Although each set uses the full range of beta found in normal 17A production, by the identification it is possible to install the low-beta and high-beta transistors in the stages which can benefit most from these characteristics.

The RF transistors (WE 26A), when connected in the common-base amplifier circuit, are capable of about 12 db per stage gain. These transis-

tors are designed to maintain this characteristic at the low values of collector-to-base voltage which result from operating the two RF amplifiers in series.

IV. CIRCUIT DESCRIPTION

The following sections will describe the circuit in some detail, making use of the Figs. 1 to 4, which together form a complete schematic of the receiver.

4.1 *RF Section*

Fig. 1 shows the radio frequency circuits of the antenna, the two RF amplifiers, the mixer, and the local oscillator.

The self-contained antenna is a "loop-stick" type, consisting of about $2\frac{1}{2}$ turns of copper ribbon wound in a helix on a high-Q ferrite core 2 inches long and $\frac{3}{8}$ inch in diameter.

The inductance of the antenna winding is tuned to series resonance by a variable capacitor (C_1), connecting to the input of the first RF amplifier stage. Since the input of the transistor is shunted by a capacitor (C_{36}), the transistor is tapped across a portion of the capacitive branch of the resonant antenna mesh.

The RF amplifier stages consist of two diffused-base pnp transistors (WE 26A) in a common-base configuration. The coupling networks between the two transistors and between the second transistor and the mixer provide the necessary impedance transformations. The inductors are self-supporting coils which are tuned by manually stretching and compressing their length by use of an insulated pick inserted through openings provided in the shield around the RF circuits. The ratio of the two capacitance values in the pi networks determines the transformation ratio.

As discussed earlier, partial neutralization is accomplished by a fixed inductor in series with a dc blocking capacitor, connected from emitter to collector of each transistor. Neutralization is required to minimize interaction of tuning of the antenna, interstage, and mixer circuits. It also minimizes leakage of local oscillator energy to the antenna.

The amplified RF signal is combined with the second harmonic of the 75-megacycle, crystal-controlled, local oscillator* in the mixer diode (CR_1) to produce the intermediate frequency as the difference between these two frequencies. The second harmonic energy from the local oscil-

* At the start of this development, a 150-megacycle crystal with the required stability was not considered to be feasible.

lator is capacitively fed to the emitter of the second RF amplifier transistor. It is then amplified simultaneously with the signal before being impressed on the mixer diode. The shunt capacitor (C_{39}) which follows the mixer diode provides RF ground, which causes the full RF voltage to be developed on the diode. Series inductor L_{14} passes to the IF amplifier the difference frequency that is developed on the mixer load resistor (R_{10}).

The method used here, which derives the second harmonic directly from the oscillator and amplifies it to a value suitable for mixing, possesses advantages over the direct injection of the 75-megacycle oscillator frequency into the mixer. The latter method would require third-order mixing, with inherently greater conversion loss than is achieved by the present method, which involves second-order mixing. Another advantage of the circuit used is that the 75-megacycle energy radiated from the receiver is more effectively attenuated.

A quartz crystal, oscillating on the fifth overtone in the 75-megacycle range, is used with a WE 26A diffused-base transistor to form the local oscillator circuit. The circuit may be thought of as a common-emitter amplifier in which the crystal provides a feedback path from collector to base. A slug-tuned coil (L_5) in series with the crystal is used to set the frequency.

The collector-to-emitter impedance of the oscillator consists of two resonant meshes. One (L_4 and C_{12}) is resonant near 75 megacycles while the other (L_6 and C_{11}) is resonant near 150 megacycles. The voltage developed across the latter is fed to the injection point on the second RF amplifier.

This type of oscillator possesses inherent $1/f$ noise modulation³ which is apparently a function of the individual transistor, and is particularly troublesome because of the low value of the intermediate frequency. The resistor (R_7) between emitter and ground provides feedback which reduces the noise modulation to within tolerable limits. This resistor is bypassed at RF.

4.2 IF Section

Fig. 2 shows the schematic circuit for the three stages of IF amplification and the limiter stage. All four stages are similar, using alloy junction transistors in the common-emitter configuration with bias stabilization as shown in Fig. 11.

Filtering in the IF amplifier is controlled primarily by the package filter, which appears on the schematic as a block between the first and

second IF stages. This is a low-pass filter with the ground terminal common to input and output. The filter cuts off at approximately 10 kilocycles.

Although the filter is a low-pass network, the over-all transmission of the IF amplifier exhibits a bandpass characteristic (Fig. 6). The low-frequency cutoff of about 2 kc is caused by the blocking capacitors between transistors.

The limiter is operated as an overdriven amplifier. Because of the high gain developed in the preceding stages, even the no-signal noise experiences a small degree of amplitude limiting in this stage. Thus any signal which rises out of the ambient noise is limited in this stage. Stronger signals are limited in progressively earlier stages.

4.3 Discriminator

Fig. 3 shows the discriminator circuit, which converts the FM IF signal into the original tone frequencies of the coded signal. It is seen to be a form of rectifier circuit using diodes with the load connected through a low-pass output filter. The circuit configuration resembles that of a voltage doubler rectifier. This circuit also bears a strong resemblance to that of a "storage counter" described in the literature.⁷

The low-pass output filter provides a cutoff of about 2 kc, in order to prevent the passage of the IF to the following stages. The filter has essentially zero loss to the recovered signal tones whose frequencies lie between 500 and 1000 cycles per second.

It is not necessary for the applied signal to be limited for this discriminator to function, but it may simplify understanding if the applied IF signal is considered to be a square wave whose frequency varies according to the modulating wave form.

The output capacitor (C_{25}) is continually charged by the rectification of the IF wave and discharged by current flow through the load resistor (R_{26}). The output filter separates C_{25} and C_{26} at the IF, but effectively connects them in parallel at the signal frequency. Therefore the charge is shared at the slower rate, and the rate of discharge is in effect determined by a time constant, $(C_{25} + C_{26})R_{26}$.

Referring to the input of the discriminator circuit, it is seen that a positive pulse will cause the shunt diode (CR_4) to conduct while the series diode is non-conducting. Thus, the series input capacitor (C_{24}) is charged at a rate dependent on its capacitance value multiplied by the effective resistance of the diode (CR_4). This rate is made high by choice of a small value of capacitance, so that approximately full charge is reached during the positive half-cycle of the IF wave.

Upon the reversal of the input wave, the shunt diode becomes non-conducting and the series diode (CR_3) conducts. Because of the voltage reversal, the stored charge on C_{24} adds to the drive voltage, and assists in charging the output capacitor (C_{25}). The measure of charge delivered to the output capacitor is determined by the change in the quantity stored on C_{24} during the cycle. Since the charge on C_{24} becomes completely reversed from its initial value, to a nearly equal but oppositely poled value, the net charge delivered to the load capacitor is nearly twice the maximum charge stored on the input capacitor.

The time constants of input capacitor charge and output capacitor discharge are chosen so that, for the unmodulated IF carrier, the output voltage on the load resistor (R_{26}) stabilizes at about half its maximum possible value.

During modulation, the intermediate frequency varies at the signal rate from its unmodulated value (6 kc) to a maximum value (near 10 kc) and then to a minimum value (near 2 kc). When the frequency increases, the increments of charge are delivered to the load capacitor (C_{25}) more rapidly and the output voltage therefore rises. Similarly, when the IF frequency decreases, the increments of charge arrive less frequently, and the output voltage falls because of the drain to a lower value. Time constants are chosen to allow these variations to follow the signal wave frequency.

4.4 Reed Circuit and Sound Oscillator

Fig. 4 shows the circuits of the reed amplifier stages, the reed selectors and the sound oscillator. The output from the discriminator is applied to a two-stage transistor amplifier using the transistors Q_8 and Q_9 in common-emitter configuration. The first of these transistor amplifiers gets its base bias from the rectification of the discriminator, combined with the base current flowing in the load resistor of the discriminator. A resistor (R_{28}) biases the emitter of this amplifier and provides bias stabilization, and at the same time furnishes the impedance across which feedback is introduced from the output of the second amplifier (Q_9). This second amplifier is biased on both base and emitter in the same manner as the IF amplifiers. The three windings of the reed selectors (RD_1 , RD_2 , and RD_3) forming the load are coupled by a blocking capacitor (C_{29}) which keeps the direct current from saturating the cores of these selectors. The capacitor C_{15} provides negative feedback, effective at the IF frequencies. This provides stability against IF regeneration and reduces noise without reducing the reed frequency gain.

The sound oscillator circuit uses a transistor oscillator whose positive

feedback is accomplished through a transformer (T_1). This couples the collector back to the base of the transistor (Q_{10}). When the reed selectors are quiescent, the oscillator is also at rest. When a signal is received which causes all three reeds to respond, the simultaneous operation of their contacts causes an impulse to be applied to the base of the oscillator transistor. This impulse is amplified and returned through the transformer to the base in proper phase to start a buildup of oscillation. Once started, the oscillation continues regardless of the excitation of the reeds. The hearing-aid type transducer which is the load of the oscillator gives forth an audible tone which alerts the customer. The customer may then stop the oscillation and reset the circuit for further signaling by simply closing the reset switch (a miniature pushbutton-type). Capacitor C_{31} and resistor R_{32} are connected to ground from a point between the contacts of RD_1 and RD_2 . These furnish a reservoir of charge whenever RD_1 is energized, so that if RD_2 and RD_3 become simultaneously energized there will be adequate pulse energy to set off the sounder oscillator. A diode (CR_5) across the primary of the feedback transformer (T_1) is polarized to absorb impulses caused by mechanical shock and thus diminish the probability of false signaling due to this cause. However, the receiver is automatically triggered when the receiver is first energized. This serves as an indication of the condition of the battery, since the oscillator will not function with a discharged battery.

V. TEST METHODS AND TECHNIQUES

5.1 *The Testing Problem*

In most other FM receivers such as, for example, those used in mobile telephone service, the receiver is tested as a unit without connection to its antenna. Test requirements are based on magnitudes or frequencies of energy applied to the input terminals of the receiver. Similarly, in such applications the efficiency of the antenna is determined by its energy yield into a standard terminating impedance, when the antenna is immersed in a standard strength of radio field.

In the development stages of the 55A pocket receiver, these same approaches were followed. Field measurements were made in which antennas of various types were compared with a half-wave dipole and with each other by connecting them through a transforming device to the input of a field strength measuring set. Also, the sensitivity, selectivity, and noise figure of the receiver were tested by connecting the appropri-

ate test generator to the emitter input of the first RF amplifier through a suitable coupling transformer.

While these methods were useful in giving relative results, their absolute significance was always in doubt. The antenna, when connected to a cable leading to a field strength set, experiences a field which is distorted by the coupling to that cable. Also, it is then difficult to evaluate it in its true relation to the human body. Furthermore, the tests on the receiver were always in doubt because of the difference between the input coupling used and that which exists in the normal connection to the antenna in the assembled set.

To overcome these uncertainties, a method was devised by which the assembled receiver could be bench tested as a complete unit in a suitable test jig. These results were then correlated with the field performance of the receiver when carried normally by a person.

5.2 RF Circuit Tuning

The RF amplifiers are tuned with the local oscillator disabled by operation of a switch on the test jig. This switch places an RF ground on the collector of the oscillator transistor. As a result of a signal coupled to the antenna, rectified current flows in the mixer diode, and is measured across the diode load resistor (R_{10}). The antenna capacitor (C_1) and the interstage tuning coils (L_2 and L_3) are then adjusted to maximize this current.

5.3 Local Oscillator Adjustments

With no signal input to the antenna, and with the oscillator operating, the rectified current of the mixer diode is a measure of the injection of local oscillator energy. This is brought to final adjustment by varying the slug position in the coil (L_6) of the oscillator circuit. A coarse adjustment of the injection is available in the initial alignment by selecting the value of the oscillator emitter resistance (R_7).

The local oscillator may be adjusted to the correct frequency by the slug of the coil (L_5) which is in series with the crystal.

5.4 Calibration of Test Jig

Considerable discussion of signaling sensitivity and the field tests which were made to measure it is given in the paper by Mitchell and Van Wynen.¹ Signaling sensitivity is defined as the field in db above one microvolt per meter required to just trigger the receiver. A number of

receivers whose sensitivity had been tested under free field conditions were used to calibrate the test jig. Thus the voltage from the modulated signal generator to the input of the jig, which is needed to produce triggering, could be interpreted in terms of the free field strength. In this way, meaningful measurements of signaling sensitivity are made in the jig setup for aligning the receivers in the laboratory, or in production.

5.5 *Noise Measurement*

By means of the jig, an rms type voltmeter may be connected to the IF test point at the collector of the third IF transistor. This measures the no-signal noise voltage at the test point. When a signal is supplied to the antenna coupling coil, the energy required to cause a 3-db increase in the voltage at the test point is a measure of sensitivity, which is related to noise figure as discussed elsewhere. Oscilloscopic observations at the IF test point show qualitatively the fact that neither the input to double the energy nor the input of modulated signal required for triggering is visible as a change from the random noise pattern. This shows qualitatively the fact that triggering occurs even for signal levels commensurate with the average noise in the IF band, or lower, as is indeed shown on the level diagram. Thus signals strong enough to produce limiting and FM quieting are not essential for the operation of the receiver. This is to a large extent an advantage derived from the exclusion of much of the noise energy by the very sharp frequency response of the reed selectors.

It is interesting to note that even after FM demodulation and band limiting by the 2-kc low-pass filter, the noise lies in a band nearly 2000 times greater than the bandwidth of a reed (about 1.2 cycles).

VI. MECHANICAL DESIGN

6.1 *General Features*

The mechanical design of the 55A receiver was influenced strongly by the inclusion of several required features. The receiver was designed to enable the customer to recharge and change batteries easily and to enable the telephone company to insert the reed selectors without the use of tools. It was necessary to include a changeable number card which could be exposed for viewing, but which would automatically remain hidden from view during normal operation of the receiver. The on-off switch and the audio transducer were placed in the top end of the receiver for optimum accessibility and audibility, respectively, when the receiver is pocket-borne. Moreover, it was necessary to reconcile such seemingly

incompatible objectives as small size and reliability, light weight and ruggedness, and low cost and high performance. Obviously, at the outset of the mechanical design, it was not possible to set absolute values on all these objectives. The design became a problem in optimizing, and the realistic approach of making the receiver as small, light, and inexpensive as possible, consistent with high performance, reliability, and ruggedness, was taken. It was necessary to refrain from "gilding the lily," performance-wise, even if this penalized size or weight only slightly. For in the final analysis, if the receiver were incapable of being carried in a pocket, its market would diminish.

The receiver contains an antenna, thirty-eight capacitors, thirty-three resistors, four diodes, fifteen inductors, ten transistors, one transformer, one crystal unit, three reed selectors, one audio transducer, two switches, and one filter (which itself contains three inductors and five capacitors), all mounted on a 4.85×2.24 -inch printed wiring board. The printed wiring substrate is $\frac{1}{16}$ -inch thick epoxy glass. This material was selected, rather than the less expensive and more commonly used XXXP phenolic substrate, because of its superior mechanical and electrical properties.

6.2 *Circuit Layout and Shielding*

The circuitry was laid out in a smooth, logical pattern, so that the mechanical flow from top to bottom is in the same sequence as the electrical flow. As shown in Fig. 13, the antenna is at the top of the receiver. In sequence toward the bottom, the antenna is followed by the RF amplifier, RF mixer, local oscillator, IF amplifier and limiter, discriminator, reed circuit, and battery. The audio transducer is placed against the top of the receiver case where it will be most easily heard by the customer. The reed selectors are placed near the bottom of the circuitry, adjacent to the battery. Removal of the battery case uncovers a number card, which, upon its removal, in turn uncovers an access port through which the reed selectors may be inserted or extracted.

The RF amplifiers, oscillator, and first IF amplifier are contained in a three-compartment copper enclosure so that portions of these circuits are shielded from each other and from the remainder of the circuitry. The can cover, not shown in Fig. 13, is soldered into place, and the can is soldered to a "ground plane" on the printed wiring board. Ground planes run on both sides of the printed wiring board, covering as much area as possible. This minimizes the ground circuit impedance, reduces the coupling between ground paths, and contributes to the stability of the circuit.

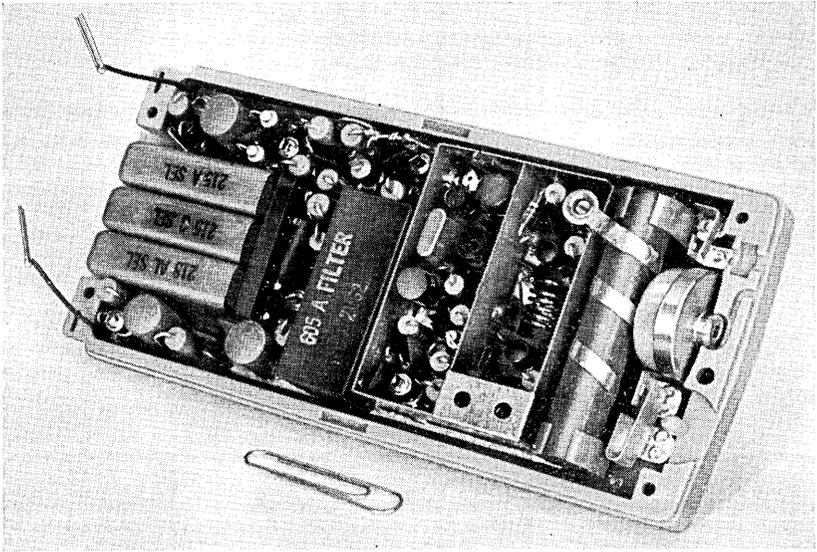


Fig. 13 — Layout.

6.3 *Space Conservation*

Space was conserved on the printed wiring board by using miniaturized components and by mounting all axial-lead components perpendicular to the printed wiring board. In some instances, where the tops of several components were electrically common, the lead of one component was bent into a common bus, which was connected to the tops of all the components in the group, thus eliminating land areas and further conserving board space. This arrangement is shown in Fig. 14. Although

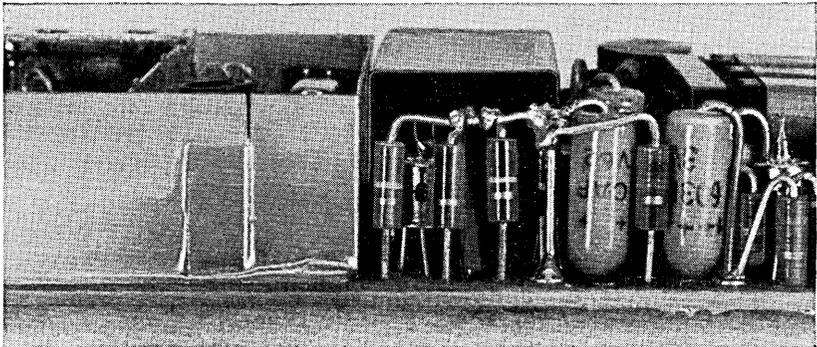


Fig. 14 — Component wiring for space conservation.

conductor spacings and widths were made as large as possible, for maximum board reliability, space considerations necessitated basing the printed wiring pattern on minimum path width, spacing, and land area diameter of 0.030, 0.040, and 0.075 inch, respectively.

6.4 Case Design

The operational requirements had a direct bearing on the design of the receiver case. The case consists of three main parts — two dish-shaped covers, which enclose the circuitry, and a battery case. These are shown in Fig. 15. The battery case is designed to accommodate a nickel-cadmium rechargeable battery. The battery case is a plug-in unit, and enables a discharged battery to be unplugged from the receiver, battery case and all, and inserted into a battery charger which, in turn, plugs directly into a 117-volt ac wall outlet. If necessary, a fully charged battery, in another battery case, can be plugged into the receiver for uninterrupted service. For special circumstances, a battery case designed to accommodate a nonrechargeable mercury battery is available.

The battery case is equipped with nickel-silver prongs which mate with contacts in the receiver case. The battery is equipped with slotted nickel tabs as shown in Fig. 16. Connection between the battery and the nickel-silver prongs is effected by means of screws which fasten the battery tabs to an extension of the nickel-silver prongs. The battery tabs and the connecting screws are located off-center with respect to the battery. This allows the battery to be inserted into the battery case only if it is properly oriented. Furthermore, the receiver case is designed so that the

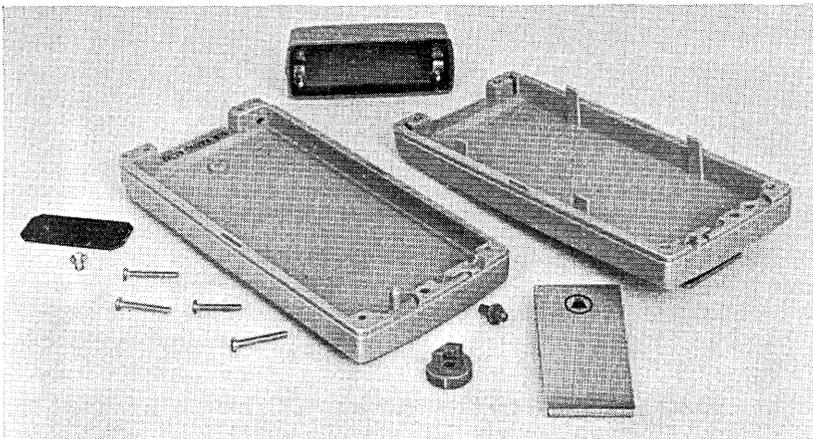


Fig. 15 — Case parts.

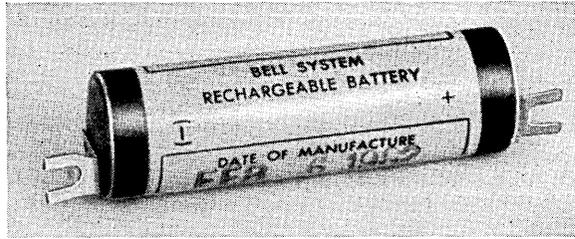


Fig. 16 — Battery.

prongs of the battery case can make contact with the circuitry only if the battery case is properly oriented. This series of orienting devices render it virtually impossible to inadvertently damage a receiver with an improperly oriented battery.

6.5 Contact Design

The mating contacts in the receiver were designed for minimum space consumption. They are made of extra hard spring-tempered nickel-silver wire, and are in the shape of the letter U. They are housed in cavities in the wall of the receiver case in a manner which permits the battery contacts to be inserted, through holes in the receiver case, into the mouth of the U. The cavities in which they are housed are large enough to permit the contacts to float into proper alignment with the pins of the battery holder. Connection is made to the circuitry by means of flexible jumper wires. The U-shaped contacts are shown in Fig. 17.

To be assured of reliable electrical contact between the prongs of the battery case and the U-shaped contacts in the receiver, the contact force and the working stress in the contacts were calculated. Calculations revealed a contact force of 0.460 lb. The contacts produce a wiping action upon mating and exert this force at two points. These features tend to increase the reliability of the electrical contact, and it is expected to provide trouble-free service. Calculations indicated a maximum working stress of 67,000 psi in the spring. Inasmuch as this is less than the safe working stress, 80,000 psi, for the nickel-silver alloy of the spring contact, it can be assumed that the spring will not lose its properties.

6.6 Materials and Special Features

The receiver case is molded of an acetal resin. This is a tough thermoplastic material which exhibits a high resistance to abrasion and cold flow. The battery case is molded of nylon, also a thermoplastic material

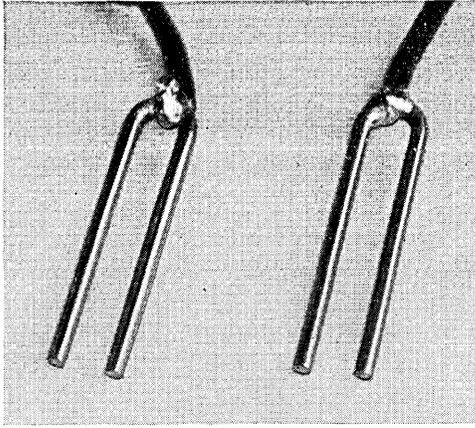


Fig. 17 — Contacts.

with excellent bearing properties. These properties were used to advantage in the design of the fastening device which permits the battery case to be snapped onto or off from the receiver case. The snap mechanism is molded as an integral part of the battery case and the receiver case, and obviates the need for hardware of any kind. A cross-sectional view of the snap is shown in Fig. 18.

The battery case was subjected to 20,000 snap-on-snap-off cycles. The pull-off force dropped from an initial value of seventeen pounds to a final value of nine pounds. This drop in pull-off force is not judged to be serious, since even the lower value is considered adequate to hold the battery case securely in place. It is estimated that the average customer will recharge his battery once a day. At this rate, 20,000 on-off cycles

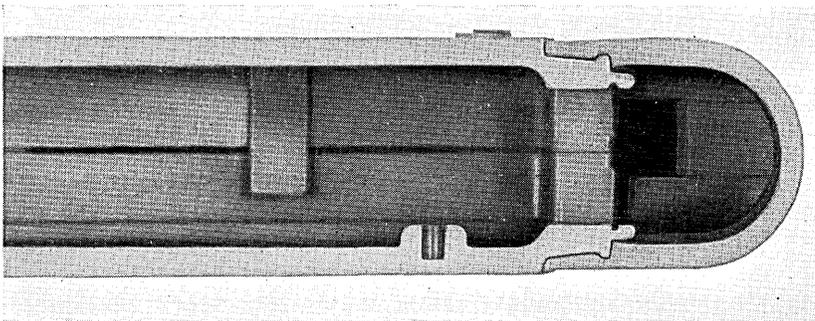


Fig. 18 — Snap fastener of case.

will not be achieved for over fifty years. For customers who are unable to exert the necessary force to unsnap the battery holder, a coin slot is provided, permitting the customer to pry it off with minimum exertion.

When carried in a pocket or on a belt, the receiver is held in place by means of a die-cast zinc clip. It has a brushed satin finish with the Bell System emblem depressed and colored dark blue. The clip is shown in Fig. 19.

A special switch, which is reliable, durable, compatible with printed wiring, and capable of blending harmoniously with the physical appearance of the receiver case, was designed. The switch contacts are a gold-silver-platinum alloy. They are welded to phosphor bronze flat springs, which are in turn mounted on the printed wiring board. The switch is actuated by a thumb wheel, which rotates on a molded axle protruding from the edge of the case. The thumb wheel is also molded of nylon and is designed with a protrusion which hits against the inner surface of the receiver case when the thumb wheel reaches either the "on" or "off" position. This device limits the rotation of the wheel and imparts a comfortable feel and a pleasant click to the switch. A model of this switch has been on life test in the laboratories for several months. As of the time of this writing, it has undergone over five million on-off cycles without any discernible degradation in performance. The switch is shown in Fig. 20.

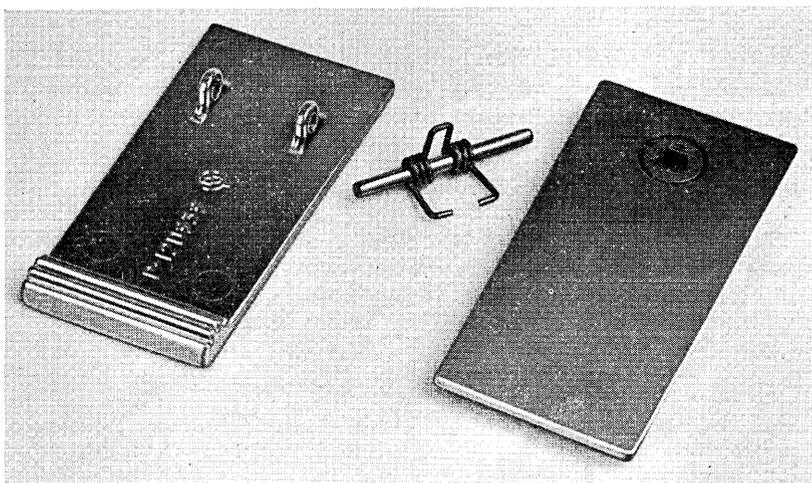


Fig. 19 — Pocket clip.

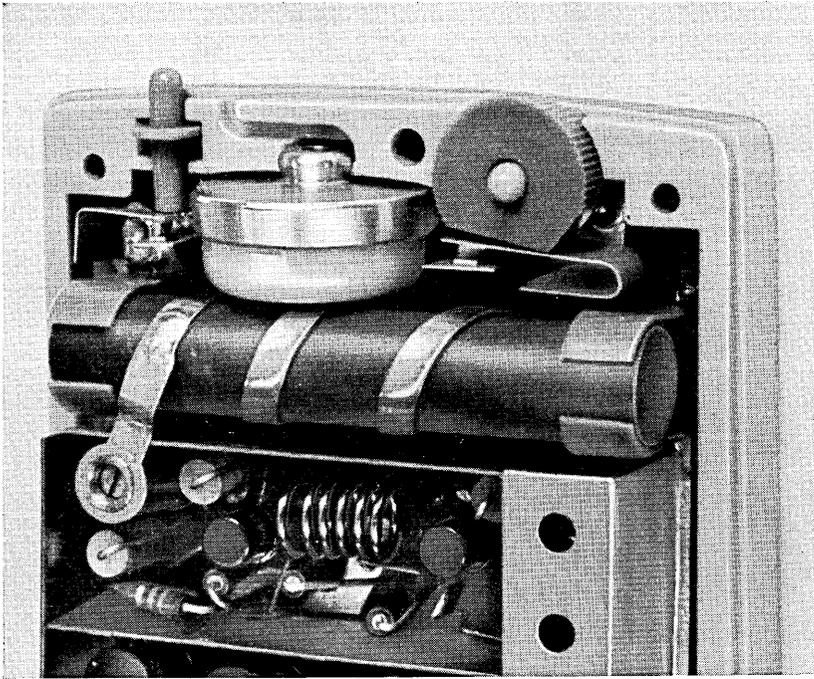


Fig. 20 — Switch.

6.7 Battery Charger

For normal use each customer is supplied with a cordless plug-in battery charger. The charger circuit consists of a diode rectifier, two resistors, and a neon indicator lamp. The circuitry is housed in a molded nylon case, which serves the dual purposes of case and structural support for the components. The charger plugs directly into a wall outlet — a feature which imposed two important restrictions on the design. The charger had to be made of nonflammable materials, and the “fall-out” torque — the product of the weight of the charger and the distance from the wall outlet cover to the center of gravity of the charger — had to be held to a low value. A fall-out torque not exceeding six inch-ounces was set as the design target, for it was judged that below this value the probability of a charger inadvertently falling out of a wall outlet would be negligible.

Nonflammability was achieved by specifying nylon for the housing. The inside of the housing is used as the structural support for the cir-

cuitry, eliminating the need for additional supporting media, such as printed wiring boards or brackets. In this manner, it was possible to produce a charger which weighed only 3.1 ounces, including the battery and the battery case, and which exerted a fall-out torque of only 1.45 inch ounces.

The battery case, containing the discharged battery, is disengaged from the receiver and plugged into the charger. The prongs in the battery case make contact with U-shaped contact springs, which are identical to those used in the receiver. The same system of orienting devices used to prevent insertion of an improperly oriented battery in the receiver is similarly used in the charger. The charger is equipped with a neon indicator lamp, which glows only when charging current is flowing. The charger is shown in Figs. 21 and 22.

The use of molded plastics contributed significantly to the realization of the objectives. The plastic parts are attractive, rugged, light in weight, intricately shaped and inexpensive. The case halves, in addition to serving as a closure, were designed with built-in functional refinements which eliminated the need for attached hardware in such places as the battery case snap, the switch axle, the speaker support, the pocket clip axle bearings, and the printed wiring board support. After tooling costs, these features are obtained virtually free of charge.

6.8 *Subjective Qualities*

The BELLBOY personal signaling receiver is a consumer product, to be worn on the person of the customer. Outwardly, the receiver was given a tailored appearance to satisfy the needs of the well-dressed customer. The appearance of boxiness was averted by adding barely discernible compound curves to the surface. These curves actually add slightly to

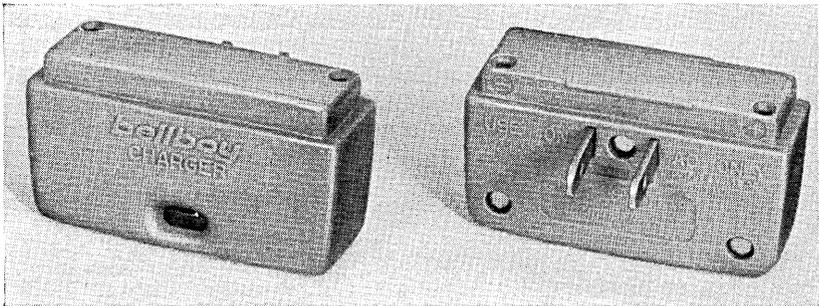


Fig. 21 — Charger.



Fig. 22 — Battery being charged.

the over-all dimensions of the receiver, but impart an appearance of elegance and compactness. The finished product, shown in Fig. 23, is $5\frac{7}{8}$ inches long, $2\frac{1}{2}$ inches wide, and $\frac{1}{8}$ inch thick, and weighs 11 ounces, including the battery.

The receiver case and battery case are different shades of gray, the battery case being the darker of the two. This not only adds to the appearance of the receiver, but avoids what would otherwise have been a troublesome color-matching problem which would have been manifest upon supplying replacement battery cases.

VII. CONCLUSION

In both the electrical and mechanical design of this receiver, emphasis has been placed on reliability in fulfilling service objectives and on convenience to the customer. In connection with the latter, the esthetic qualities essential to consumer acceptance have not been overlooked. The reactions of the consumer have been sampled by means of field trials



Fig. 23 — Finished product.

using development models. In the first commercial installation, which went into service at Seattle in April, 1962, the performance of the receiver has been satisfactory, and it is expected that additional installations will follow.

VIII. ACKNOWLEDGEMENTS

The receiver described in this paper is the result of the combined efforts of a number of people in several departments of Bell Telephone Laboratories, with the cooperation of members of the Western Electric Company at the Merrimack Valley Works. The authors, in presenting these results, acknowledge these contributions without attempting to give individual credit.

REFERENCES

1. Mitchell, D., and Van Wynen, K. G., A 150-mc Personal Radio Signaling System, *B.S.T.J.*, **40**, September, 1961, p. 1239.
2. Bostwick, L. G., A Miniature Tuned Reed Selector of High Sensitivity and Stability, *B.S.T.J.*, **41**, March, 1962, p. 411.
3. Bennett, W. R., *Electrical Noise*, New York, McGraw-Hill, 1960, p. 101.
4. Friis, H. T., Noise Figures in Radio Receivers, *Proc. I.R.E.*, **32**, July, 1944, p. 419.
5. Schelkunoff, S. A., and Friis, H. T., *Antennas: Theory and Practice*, New York, John Wiley, 1952.
6. Shea, R. F., *Principles of Transistor Circuits*, New York, John Wiley, 1953.
7. Millman, J., and Taub, H., *Pulse and Digital Circuits*, New York, McGraw-Hill, 1956, p. 346.

A "Thermodynamic" Theory of Traffic in Connecting Networks

By V. E. BENEŠ

(Manuscript received October 22, 1962)

Two new theoretical models for representing random traffic in connecting networks are presented. The first is called the "thermodynamic" model and is studied in detail. The second model is formulated in an effort to take methods of routing into account and to meet certain drawbacks of the "thermodynamic" model in describing customer behavior; since it is more realistic than the first, it leads to results that are vastly more complicated and must be described in another paper.

The "thermodynamic" model is worth considering for four reasons:

(1) *It is faithful to the structure of real connecting systems. Indeed it is an improvement over many previous models in that it only considers physically accessible states of the connecting network, while the latter suffer the drawback that a large fraction of the network states on which calculation is based are physically meaningless, being unreachable under normal operation.*

(2) *It gives rise to a relatively simple theory in which explicit calculations are possible.*

(3) *The "thermodynamic" model provides a good simple description of traffic in the interior of a large communications network.*

(4) *It has an analogy to statistical mechanics which permits us to be guided by the latter theory as we try to use the model to understand the properties of large-scale connecting systems.*

The two models to be described differ in only one respect. In the first (the "thermodynamic") model, the system moves from a state x to a state y that has one more call in progress, at a rate λ ; the effective calling-rate per idle inlet-outlet pair is thus proportional to the number of paths usable in x from that inlet to that outlet. In the second model, the calling-rate per idle inlet-outlet pair is set at λ , and is then spread over the paths usable in x from that inlet to that outlet in accordance with some routing rule. This provides a mathematical description of routing, and avoids the unwelcome feature that a single customer's calling-rate depends on the state of the network.

The "thermodynamic" model is based on the single postulate that the "equilibrium" probabilities of the states of the connecting network maximize the entropy functional for a fixed value of the traffic carried. These probabilities have the same geometric or exponential form as the canonical Maxwell-Boltzmann distribution of statistical mechanics. The theory developed applies to any connecting network regardless of its structure or configuration. The number of calls in progress is analogous to the energy of a physical system. As in statistical mechanics, important averages can be expressed as logarithmic derivatives of a generating function analogous to the partition function of physics.

It is possible to give an interpretation of the maximum entropy postulate in terms of random behavior at the inlets and outlets of the connecting network; this interpretation leads to a stochastic process z_t of the familiar Markov type, for which the canonical distribution is invariant. The transition rate matrix of z_t is self-adjoint in a suitable inner product space, so that the approach of z_t to equilibrium is easily studied, with resulting applications to traffic measurement.

I. INTRODUCTION

Like the physicist, the traffic engineer is faced with the study of an extremely complex system which is best described in statistical terms. The great success of the theoretical methods of statistical physics has given rise to a fervent hope, sometimes voiced among traffic theorists, that similar methods exist and can be found for the study of congestion. Indeed, the problems are much the same: one desires a small amount of "macroscopic" information about averages, based in a rational way on vast complexities of detail. A. K. Erlang was probably influenced by statistical mechanics when he introduced his method of "statistical equilibrium" into traffic theory. This method has had great success in dealing with problems of the birth-and-death type, like trunking and queueing, but as applied to more complex cases it has led mostly to algebraic and combinatorial difficulties. Nothing as elegant or powerful as statistical mechanics has resulted so far.

We shall present two traffic models in this paper. The first is the outcome of a deliberate attempt to ape the methods of physicists in statistical mechanics, and thus to realize, at least in part, the hope mentioned above. It is called the "thermodynamic" model, and it is treated in detail. The second model is introduced later in the paper in an attempt to avoid certain drawbacks that appear in the interpretation of the "thermodynamic" model. Since it has independent interest and leads to involved, more realistic results, it is studied in detail in a later paper.

The approach taken in the "thermodynamic" model bears a close analogy to the methods of statistical mechanics, and is based on the single postulate that the "equilibrium" probabilities of the states of the connecting network maximize the entropy functional for a fixed value of the traffic carried. We develop a theory, briefly summarized in the next paragraph, by deducing the consequences and interpreting the meaning of the one basic assumption.

The state probabilities that maximize the entropy for a given carried load form a distribution function over the set of states that has the same geometric or exponential form as the *canonical* (or Maxwell-Boltzmann) distribution of statistical mechanics. The theory applies to any connecting network, regardless of its structure or configuration. The number of calls in progress is analogous to the energy of a physical system. As in statistical mechanics, important averages can be expressed as logarithmic derivatives of a generating function analogous to the partition function of physics. It is possible to give an interpretation of the maximum entropy postulate in terms of random behavior at the inlets and outlets of the connecting network. This interpretation leads to a stochastic process z_t of the familiar Markov type, and is such that any stochastic process based on it satisfies the maximum entropy postulate. The transition rate matrix A of z_t is self-adjoint in a suitable inner-product space; its characteristic values are real and non-positive, and can be studied by classical variational methods. In terms of these characteristic values the approach of z_t to equilibrium can be studied, with resulting applications to traffic measurements. It turns out that the covariance of any function of z_t is strictly positive. The paper ends with a time-dependent or non-stationary generalization of the maximum entropy postulate that has close analogies with the statistical "derivation" of thermodynamics.

II. PRELIMINARIES

In order to give an adequate summary and discussion of our theory in Section III, it is necessary to present first its concepts, terminology, and notation. Virtually all the notions about to be discussed have appeared in earlier papers by the author^{1,2} so only a brief résumé is given here.

Let S be the set of possible (or permissible, or both) states of a connecting network, and let x, y, \dots be variables ranging over S . The elements of S are partially ordered by inclusion \leq , where $x \leq y$ means that x can be obtained from y by removing zero or more calls. Furthermore, the states $x \in S$ can be arranged in an intuitive manner in the *state-diagram*, the Hasse figure for the partial ordering \leq . This figure is constructed by partitioning the states in rows according to the number of

calls in progress. The unique zero state (in which no calls are in progress) is placed at the bottom of the figure; above it comes the row of states with one call in progress; and so on. The figure is completed by drawing a graph with the states as nodes, and with lines between states (in adjacent rows) that differ in exactly one call. In an earlier work² we made the assumption that in a given state at most one call could be in progress between a given inlet and outlet; it is convenient to discard this assumption here.

If the connecting network under study is in a state x , it can move only to states which are *neighbors* of x , i.e., are obtainable from x by adding a new call or terminating a call in progress. It is useful to divide the neighbors of x into two sets A_x and B_x where

A_x = set of states immediately *above* x , i.e., accessible from x by adding a new call,

B_x = set of states immediately *below* x , i.e., accessible from x by a hangup.

For any set X , the notation $|X|$ is used to denote the number of elements of X . The states $x \in S$ can be defined² as sets of chains on a graph, one chain for each call in progress. Hence it is natural to use $|x|$ to mean the number of calls in progress in x . The k th level L_k is the set of all states with k calls in progress, i.e.,

$$L_k = \{x \in S : |x| = k\}.$$

III. SUMMARY AND DISCUSSION

We start, in Section IV, with a brief informal discussion of what is meant, heuristically as well as precisely, by "equilibrium."

In Section V we formulate and discuss the *maximum entropy postulate*, according to which a suitable "equilibrium" distribution $\{q_x, x \in S\}$ of probability over S is obtained by choosing the probability vector q so as to maximize the entropy functional

$$H(q) = - \sum_{x \in S} q_x \log q_x$$

for a given value of the average number of calls in progress, i.e., for

$$\sum_{x \in S} |x| q_x = m.$$

Various heuristic arguments are adduced to support the *prima facie* reasonableness of this principle. In Section VI it is shown that the maximizing probability vector q is given by

$$q_x = \frac{\lambda^{|x|}}{\Phi(\lambda)}, \quad x \in S$$

$$\Phi(\xi) = \sum_{x \in S} \xi^{|x|}$$

and λ is a constant determined uniquely by the equation

$$m = \lambda(d/d\lambda) \log \Phi(\lambda).$$

Because of their close similarity to corresponding notions from statistical mechanics, the vector q and the function $\Phi(\cdot)$ are henceforth called the *canonical distribution* and the *partition function*, respectively.

In Section VII we have collected together various properties of the partition function, most of them based on the partial ordering \leq of S . Among these are expressions for $\Phi(\cdot)$ in terms of the Möbius function for \leq , and in terms of several sets of "characteristic polynomials" associated with \leq and S .

The canonical distribution q is placed in a dynamic context in Section VIII. This is done by defining a Markov process z_t (taking values on S) for which q forms a *stationary* distribution. The transition rate matrix A (infinitesimal generator) of this process allows one to give interpretations of this dynamic context in terms of calling rates and mean holding-times. An informal description of the process z_t is this: if it is in state x , it moves to a state $y \in A_x$ at a rate λ , and to a state $y \in B_x$ at a rate set at unity by convention.

A full discussion of the analogy between the "thermodynamic" theory of traffic and statistical mechanics is given in Section IX. For purposes of illustration, we mention that the number of calls in progress corresponds to the energy of a statistical mechanical system, and that the constant λ is related to the calling rate and corresponds to the temperature (up to a monotone transformation).

The reasonableness of z_t as a description of an operating connecting network is discussed and criticized in detail in Section X. Two possible interpretations of the inlets and outlets are considered: in one, the inlets and outlets are the ultimate terminals of the system, beyond which there is no more switching equipment; in the other, the inlets and outlets are switching centers such as PBX's, frames, or individual crossbar switches, acting as sources of traffic for a network under study. In the first interpretation, there can be at most one call in progress on an inlet or an outlet; in the second, there may be several.

Regardless of which interpretation of the inlets and outlets is adopted, the transition rate matrix A for z_t must be interpreted as saying that

the calling rate between an inlet and an outlet in a given state x is proportional to the number of free paths that x provides between that inlet and outlet. This assumption is unacceptable for the interpretation of inlets and outlets as ultimate terminals; it is not entirely unreasonable if the inlets and outlets are local switching centers.

Section XI is devoted to describing, as an alternative to z_t , a Markov stochastic process on S based on the assumption that the calling rate between an idle inlet terminal and an idle outlet terminal is a constant λ . This calling rate is then spread over the possible ways of realizing the call in question in the current state of the network in accordance with some method of routing. A mathematical description of such a method of choosing routes for calls is given. This description leads directly to a transition rate matrix Q for a process x_t in which every idle inlet-outlet terminal pair has a calling rate λ in every state. The possibility that z_t may be a useful perturbation of x_t is considered.

In Section XII it is observed that the rate matrix A for z_t is a self-adjoint operator in a suitable finite-dimensional inner product space. This implies that the characteristic values of A are real and nonpositive, and leads to bounds on the rate of approach of z_t to equilibrium. These bounds can be applied to estimate the covariance of z_t , and the sampling error incurred in measuring carried loads by averaging z_t , or discrete observations of z_t . In particular it is shown that the dominant (i.e., that of smallest nonzero magnitude) characteristic value r_1 of A satisfies

$$-(m/\sigma^2) \leq r_1 < 0,$$

where m and σ are (respectively) the mean and standard deviation of the load associated with the equilibrium probability vector q for z_t , so that

$$m = \sum_{x \in S} |x| q_x,$$

$$\sigma^2 = \sum_{x \in S} (|x| - m)^2 q_x.$$

In Section XIII we give a formula for the covariance of any process f_t defined by applying a function $f(\cdot)$ to z_t , i.e.,

$$f_t = f(z_t).$$

This covariance is *always* positive. Applications of this formula to traffic averages are described briefly in Section XIV. Finally, Section XV considers a time-dependent generalization of the variational principle on which the "thermodynamic" theory of traffic is based.

We conclude this section with an appraisal of the "thermodynamic"

theory presented herein. This will take the form of a list of comments, first *pro*, then *con*, and then a defense.

1. There exist theories^{3,4} for connecting networks in which it is assumed that the links of the system are busy or idle with a given probability, all *independently* of one another. It can be verified that an overwhelming fraction of the states of the system so considered are in fact *not* physically meaningful states that the system can reach under normal operation. The theory presented here is based only on permitted, physically meaningful states, and so is not open to this serious objection.

2. The theory provides a *uniform* method of treating any connecting network in that the calculation of equilibrium probabilities *always* reduces to that of the partition function. In most other treatments the nature of the algebraic process of calculating probabilities depends heavily on a detailed account of the network configuration; in our theory it depends on the network only via the numbers $|L_1|$, $|L_2|$, \dots .

3. The maximum entropy principle can be given a certain informal, *a priori* justification. It provides a "conservative, worst possible case" approach to problems and processes of fantastic complexity. This is because it can be interpreted as enjoining that an "equilibrium" distribution of probability for given carried traffic correspond to a condition of maximum ignorance of the actual state of the connecting network.

4. The canonical distribution q that results from the maximum entropy postulate can be embedded in a dynamic model of traffic by defining a Markov process z_t for which q is the invariant or stationary distribution. This dynamic model is described by a transition rate matrix which is a self-adjoint operator, a fact which makes it possible to study the time-dependent behavior of z_t in a simple approximate way, with applications to traffic measurement, for instance.

5. A very serious drawback of the "thermodynamic" theory is that its natural interpretation in terms of calling rates appears to be unreasonable in most practical cases. For this reason it will probably remain an amusing curiosity, rather than become an engineering tool.

6. The problem of calculating the partition function $\Phi(\cdot)$ is, as in statistical mechanics, very difficult except in cases of unrealistic simplicity. Thus, even if its assumptions are granted, our "thermodynamic" theory does not afford much progress in calculating quantities of interest.

7. The theory can take into account only one of the many different possible methods of routing calls in operating networks. Thus it cannot help the designer choose among alternative methods.

By way of defense against the objections just raised, these points can be made:

(i) Comment 5, that the interpretation of the "thermodynamic" theory in terms of calling rates is unreasonable, depends on a natural, but not necessarily valid or compelling, assignment of causes for new calls.

(ii) Although the calculation of $\Phi(\cdot)$ is hard, it is at least a definite combinatorial problem solvable in principle by counting; thus at least part of the problem of obtaining state probabilities is disposed of.

(iii) It is doubtful whether routing methods make as much as an order of magnitude of difference in carried loads in large systems; hence it is reasonable to ignore them in a relatively crude theory such as the present one. (See however, Beneš, Ref. 11.)

The theory presented in this paper should be judged by its success in practice as well as by its agreement with our preconceptions. I believe that in spite of the major failings mentioned, the theory musters interest enough to warrant its presentation to traffic engineers, if only because its concepts and results may prove useful in more realistic approaches.

IV. EQUILIBRIUM

Quantities that are of interest in the design of a connecting network, such as the average load carried, the variance of the load, or even the probability of blocking, can often be calculated from a knowledge of some "equilibrium" or "stationary" state probabilities $\{q_x, x \in S\}$ for the network of interest. These probabilities are usually assumed or proven to be of "equilibrium" type in the sense that they have some physically reasonable invariance property.

Since the concepts of stationarity and equilibrium can assume many precise forms of varying strength, it is important to consider briefly some of these senses. The strongest notion, of course, is that of strict stationarity of a stochastic process, defined by the condition that all the finite-dimensional distributions be independent of time, i.e., be translation-invariant. A whole class of weaker notions can be obtained by requiring only that the distributions of dimension not greater than n be invariant. The notion of wide-sense stationarity, defined by the condition that the covariance depend only on the difference of its arguments, is still another concept of stationarity, formulated for a moment rather than a distribution. Again, Markov processes are described as homogeneous or stationary if their transition probability operators are time-invariant.

"Equilibrium" is a word that usually connotes a stable, quasi-static random behavior which is perhaps a condition of attraction for a process, in the sense that a process not in equilibrium tends toward it. Ergodic

Markov processes with denumerable state spaces are typical examples. It is to be remarked, though, that use of the word "equilibrium" usually implies a nondegenerate limiting behavior for a process y_t under study as $t \rightarrow \infty$. Thus a time-homogeneous Markov process may not have a genuine "equilibrium" distribution because it in some sense "blows up," e.g., the process may take values on the integers and the probability mass may move out toward $+\infty$, even though the transition probabilities are time-independent. In such a case, clearly, no first-order distribution can be assigned which is time-invariant.

The analytical expression of "equilibrium" often takes the form of a statement to the effect that an operator has zero as a characteristic value. Perhaps the most familiar example of such a statement arises in the case of a Markov process in continuous time with a transition rate matrix A ; the equilibrium equation is $Aq = 0$, for a probability vector q .^{*} This equation, together with its connections to semigroups, to Markov processes, and to the notion of statistical equilibrium used in traffic theory, is discussed immediately below.

A traditional analytical method in telephone traffic theory is that of "statistical equilibrium," due to Erlang.⁵ This method may be described heuristically as follows: A notion of equilibrium is defined by the property that the rate of flow of probability into (or onto) a state equals that out of (or from) the state; this equilibrium is expressed in a set of equations among the state probabilities, the so-called statistical equilibrium equations; the "equilibrium" state probabilities are then taken to be (or defined by) the solution of these equations.

The method of statistical equilibrium can be interpreted in the mathematically rigorous context of semigroups of positive operators, here the matrices of transition probabilities $\{Q(t), t \text{ real}\}$ for a Markov process x_t taking values in S , with

$$Q(t) = (q_{xy}(t))$$

$$q_{xy}(t) = \Pr \{\text{state of system is } y \text{ at } t \text{ if it was } x \text{ at } 0\}.$$

The *generator* A of the semigroup is the matrix of transition *rates* or the derivative

$$A = \lim_{t \rightarrow 0} \frac{A(t) - I}{t} = Q'(0).$$

The matrix A expresses the relative probabilities of the various changes

^{*} We are using the convention $(Aq)_x = \sum_y a_{yx}q_y$, rather than the more usual

$$(Aq)_x = \sum_y a_{xy}q_y.$$

that can take place in a small amount of time, and indeed

$$Q(t) = I + At + o(t) \quad \text{as } t \rightarrow 0.$$

In terms of the generator A the equation of statistical equilibrium can be written as $Aq = 0$, which expresses the fact that the vector q of state probabilities is an eigenvector of A corresponding to a zero eigenvalue of A . From the semigroup property

$$Q(t + s) = Q(t)Q(s)$$

it follows that

$$Q(t) = \exp At$$

$$\sum_y q_y q_{yx}(t) = q_x, \quad x \in S, \quad t \text{ real}$$

the last equation expressing the *invariance* of q under the transition probability matrices $Q(\cdot)$.

V. THE MAXIMUM ENTROPY PRINCIPLE

In the method of statistical equilibrium, the state probabilities are calculated *a posteriori* from a prior equation expressing an equilibrium or invariance principle. This equation is either postulated or is derived from assumptions that lead to a Markov stochastic process as a model for the operating network.

In the present work we use a variational principle rather than an equilibrium principle as a basis for calculating "equilibrium" state probabilities. In drawing this distinction we refer only to the immediate form of the assumptions and derivations, and imply no absolute distinction, since an "equilibrium" principle can almost always be given a "variational" form. For example, if A is a transition rate matrix for an ergodic Markov process, and A is self-adjoint with respect to an inner product (\cdot, \cdot) , then the "equilibrium" probability vector q , i.e., the solution of $Aq = 0$ is equally well described as the vector which maximizes the Rayleigh quotient

$$\frac{(Aq, q)}{(q, q)}.$$

It will turn out that the probabilities $\{q_x, x \in S\}$ derived from our variational principle also have an invariance property expressible, as in the example given, in terms of the self-adjoint generator A of a Markov semigroup by the equation $Aq = 0$. This equation can be interpreted

as a "statistical equilibrium" equation, and the elements of A related to calling-rates and hangup rates, in the various states $x \in S$.

However, instead of starting with a suitable matrix A to represent the infinitesimal dynamic behavior, and solving $Aq = 0$ in order to obtain an equilibrium distribution $\{q_x, x \in S\}$ over the states of the system, we shall directly choose a certain q , to be used as an "equilibrium" distribution for calculating quantities of interest, according to this criterion: The entropy functional

$$H(q) = - \sum_{x \in S} q_x \log q_x$$

is to be as large as possible subject to the conditions

$$\begin{aligned} q_x &\geq 0, & x \in S \\ \sum_{x \in S} q_x &= 1, \\ \sum_{x \in S} |x| q_x &= m, \end{aligned}$$

where m is a given number, the average load carried. The first two conditions ensure that only *bona fide* probability distributions are considered, while the third enjoins that q give rise to m as the mean number of calls in progress in equilibrium. This criterion or method for choosing a probability distribution over S we call the *maximum entropy principle*; it is exactly analogous to that used in statistical mechanics, provided that the number of calls in progress is interpreted as the energy of the mechanical system. We have already stated that this principle leads to a unique q which is exactly the same as would be obtained by a particular choice of A , given later, and solving $Aq = 0$; this matrix A has a definite interpretation in terms of system behavior during small periods of time.

A measure of justification for using the maximum entropy principle can be obtained from five arguments:

(1) Insofar as a high value of the entropy functional is an indication of a low degree of information, so far can use of the principle be interpreted as postulating that an equilibrium distribution $\{q_x, x \in S\}$ corresponds to a condition of maximum ignorance subject to a given average number of calls in progress. The principle may thus be said to represent a "safe" or "worst case possible" approach to the problem. Exactly the same principle is used in statistical mechanics to obtain the canonical distribution. In both cases it is a reasonable and systematic way of throwing up our hands.

(2) The principle is appealing for the obvious reasons of unity, uniformity, simplicity, and elegance.

(3) It leads to a theoretical structure similar to that of statistical mechanics. As in the physical theory, statistical quantities of interest are calculated from a *partition function*, characteristic of the network under study, that depends on purely combinatorial properties of that network.

(4) The principle of maximum entropy leads to a unified theory applicable to all connecting networks. That is, the resulting "equilibrium" distribution depends algebraically on the structure of the network in a way that in a sense is uniform for all networks.

(5) The principle can be given a dynamic context in terms of Markov processes. This context permits the study of the approach to equilibrium in time, with important applications to sampling error.

VI. THE CANONICAL DISTRIBUTION

In the next few sections we develop some of the principal consequences of the maximum entropy principle, and examine their similarity to statistical thermodynamics. In the present section we determine the distribution $\{q_x, x \in S\}$ which maximizes $H(q)$ for a given average load carried. The following lemma is no doubt well-known, especially to physicists; since its application in traffic theory is relatively new, its detailed proof is included for completeness.

Lemma 1: Let $f(\cdot)$ be a nonnegative function defined on S , and let

$$\Psi(\xi) = \sum_{x \in S} \xi^{f(x)}.$$

The maximum of

$$H(q) = - \sum_{x \in S} q_x \log q_x$$

subject to the conditions

$$q_x \geq 0, \quad x \in S$$

$$\sum_{x \in S} q_x = 1 \tag{1}$$

$$\sum_{x \in S} q_x f(x) = m_f, \quad (m_f \text{ a given positive number in the closed convex hull of the range of } f(\cdot),) \tag{2}$$

is

$$H(q) = \log \Psi(\omega) - m_f \log \omega,$$

where ω is the unique positive solution of

$$\omega(d/d\omega) \log \Psi(\omega) = m_f.$$

The maximum is achieved by the choice

$$\begin{aligned} q_x &= [\omega^{f(x)} / \Psi(\omega)], \\ &= \exp \{-a_1 f(x) - a_2 - 1\}, \quad x \in S \end{aligned} \quad (3)$$

where a_1, a_2 are (the values of Lagrange's multipliers) determined by any two of the relations

$$\begin{aligned} a_2 &= \log \Psi(e^{-a_1}) - 1, \\ m_f &= \sum_{x \in S} f(x) \exp(-a_1 f(x) - a_2 - 1), \\ \omega &= \exp\{-a_1\}. \end{aligned}$$

Proof: With a_1 and a_2 as Lagrange's multipliers, we form the expression

$$h = - \sum_{x \in S} q_x \log q_x - a_1 \sum_{x \in S} f(x) q_x - a_2 \sum_{x \in S} q_x,$$

differentiate with respect to each q_x , and set the resulting derivatives equal to zero. This gives the equations

$$\partial h / \partial q_x = -(\log q_x + 1 + a_1 f(x) + a_2) = 0, \quad x \in S \quad (4)$$

whose solution is (3). The multipliers a_1 and a_2 are to be determined from the conditions $\sum_{x \in S} q_x = 1$ and $\sum_{x \in S} f(x) q_x = m_f$. The first gives

$$1 = e^{-a_2-1} \sum_{x \in S} \exp[-a_1 f(x)],$$

$$a_2 = \log \Psi(e^{-a_1}) - 1,$$

while the second yields

$$\begin{aligned} m_f &= e^{-a_2-1} \sum_{x \in S} f(x) \exp[-a_1 f(x)] \\ &= \frac{\sum_{x \in S} f(x) \exp[-a_1 f(x)]}{\sum_{x \in S} \exp[-a_1 f(x)]}. \end{aligned}$$

Setting $\omega = \exp\{-a_1\}$, it is found that ω should be a solution of the equation

$$\omega(d/d\omega) \log \Psi(\omega) = m_f > 0. \quad (5)$$

From the fact that

$$\begin{aligned} \frac{d^2}{da^2} \log \Psi(e^{-a}) &= \frac{1}{\Psi(e^{-a})} \sum_{x \in S} \left(f(x) + \frac{d}{da} \log \Psi(e^{-a}) \right)^2 e^{-af(x)} \\ &> 0 \end{aligned}$$

it is easily shown (Khinchin,⁶ p. 77) that there is exactly one solution ω of (5), and that ω is positive.

A relative extremum of $H(q)$ in $q \geq 0$ subject to (1) and (2) must satisfy equations (4). Since these have only one solution there is only one such extremum. To show that it is a maximum it is enough to show that the matrix of second derivatives of $H(q)$ with respect to the components q_x of q is negative definite. However, this is straightforward, since

$$\frac{\partial^2 H}{\partial q_x \partial q_y} = \begin{cases} 0 & \text{if } x \neq y \\ -\frac{1}{q_x} & \text{if } x = y \end{cases}.$$

In Lemma 1 we let

$$f(x) = |x| \\ = \text{number of calls in progress in state } x$$

and we obtain

Theorem 1: Let $m > 0$; let

$$\Phi(\xi) = \sum_{x \in S} \xi^{|x|};$$

and let λ be the unique (positive) root of

$$m = \lambda(d/d\lambda) \log \Phi(\lambda).$$

The maximum of $H(q) = -\sum_{x \in S} q_x \log q_x$, subject to the conditions that q be a probability vector over S and that $\sum_{x \in S} |x| q_x = m$, is

$$H_{\max} = \log \Phi(\lambda) - m \log \lambda$$

and is achieved by the vector q with components

$$q_x = \frac{\lambda^{|x|}}{\Phi(\lambda)}, \quad x \in S.$$

This is the distribution of probability over S that is determined uniquely by the maximum entropy principle; as noted before, it is the *canonical* distribution. The function $\Phi(\cdot)$ is called the *partition function* of the connecting network whose states form the set S . Since m determines λ uniquely and *vice versa*, we can use λ as the parameter that determines the average traffic level instead of m . Indeed, m is a monotone increasing function of $\lambda \geq 0$. Also it can be seen that moments of the distribution of the number of calls in progress (other than the mean) can be calculated from $\Phi(\cdot)$ by logarithmic differentiation.

VII. PROPERTIES OF THE PARTITION FUNCTION

In this section we exhibit various identities and relationships that are typical of the partition function $\Phi(\cdot)$. This function is the generating function of the number of states in a given level; that is

$$\Phi(\xi) = \sum_{k=0}^w \xi^k |L_k|, \quad w = \max_{x \in S} |x|.$$

Thus the problem of calculating λ , $\Phi(\cdot)$, and q in our model reduces to the calculation of the sequence

$$|L_0|, |L_1|, \dots$$

and *vice versa*.

Remark 1:

$$|L_k| = \frac{1}{k} \sum_{y \in L_{k-1}} |A_y|$$

$$\sum_{x \in S} |A_x| = \sum_{x \in S} |x| = \Phi'(1)$$

The first part of this result was proven as Theorem 1 in Ref. 2, and it implies the second part.

The Möbius function $\mu(\cdot)$ of the partially ordered system (S, \leq) is defined recursively by

$$\mu(0) = 1, \quad \mu(x) = -\sum_{y < x} \mu(y) \quad \text{if } x > 0, \quad x \in S.$$

We have remarked in previous work (Ref. 2, Section VII) that if S is a class of network states, then $\mu(\cdot)$ takes on the simple form

$$\mu(x) = (-1)^{|x|} |x|!$$

We define the generating function $M(\cdot)$ by

$$M(\xi) = \sum_{x \in S} \xi^{|x|} \mu(x).$$

Since

$$\Phi(\xi) = \sum_{x \in S} \xi^{|x|} = \sum_{x \in S} (-\xi)^{|x|} \frac{\mu(x)}{|x|!}$$

it can be seen that (except for a change of sign in the generating variable) $\Phi(\cdot)$ is the exponential generating function associated with $M(\cdot)$. Thus we have

Remark 2:

$$M(\xi) = \int_0^{\infty} e^{-u} \Phi(-\xi u) du$$

Proof:

$$\begin{aligned} M(\xi) &= \sum_{k=0}^w \xi^k |L_k| \cdot (-1)^k k! \\ &= \sum_{k=0}^w (-\xi)^k |L_k| \int_0^{\infty} e^{-u} u^k du \\ &= \int_0^{\infty} e^{-u} \Phi(-\xi u) du. \end{aligned}$$

In analogy with Birkhoff,⁷ p. 15, (12), we define for each $x \in S$ a characteristic polynomial by the recursion formula*

$$p_x(\xi) = \xi^{|x|} - \sum_{y < x} p_y(\xi).$$

This is related to the Möbius function $\mu(\cdot)$ by the fact that if $\mu_y(\cdot)$ denotes the Möbius function for the set $\{x: x \geq y\}$, then

$$p_x(\xi) = \sum_{y \leq x} \xi^{|y|} \mu_y(x).$$

However, the partial ordering of the cone $\{x: x \geq y\}$ is again of the same form as that of S ; i.e., there are exactly $(|x - y|)!$ ascending chains between y and x , all of length $|x - y|$. Hence, by ref. 7, p. 15, (11),

$$\mu_y(x) = \frac{(-1)^{|x-y|}}{(|x - y|)!} = \mu(x - y)$$

and

$$\begin{aligned} p_x(\xi) &= \sum_{y \leq x} \xi^{|y|} \mu(x - y) \\ &= \xi^{|x|} + \sum_{y < x} \xi^{|y|} \mu(x - y). \end{aligned}$$

Let now

$$\begin{aligned} q_x(\xi) &= \sum_{y < x} \xi^{|y|} \\ P_x(\xi) &= \sum_{y < x} p_x(\xi). \end{aligned}$$

* Actually, Birkhoff's polynomial $p_x[\xi]$ equals $\xi p_x(\xi)$. The definition we use is more convenient for our purposes.

The Möbius inversion formula gives

$$\xi^{|x|} = \sum_{y < x} \mu(x - y) q_y(\xi)$$

$$p_x(\xi) = \sum_{y < x} \mu(x - y) P_y(\xi).$$

To calculate $q_x(\cdot)$ explicitly, we note that if $0 \leq k \leq |x|$, then, using the cup \cap for set intersection,

$$|L_k \cap \{y: y \leq x\}| = \binom{|x|}{k},$$

i.e., there are exactly $\binom{|x|}{k}$ states with k calls up below any state x .

Hence

$$q_x(\xi) = \sum_{y < x} \xi^{|y|} = (1 + \xi)^{|x|} - \xi^{|x|}.$$

Let us write

$$\xi^{|x|} = \sum_{y < x} r_y(\xi),$$

where $r_y(\cdot)$ are functions to be determined. Using the Möbius inversion formula once more, we find that one choice of the r_x 's is

$$r_x(\xi) = \sum_{y < x} \mu(x - y) \xi^{|y|},$$

$$= p_x(\xi) - \xi^{|x|},$$

so that

$$\xi^{|x|} = \sum_{y < x} (p_y(\xi) - \xi^{|y|})$$

$$= P_x(\xi) - q_x(\xi)$$

and

$$P_x(\xi) = (1 + \xi)^{|x|}$$

$$= \sum_{k=0}^{|x|} \binom{|x|}{k} \xi^k$$

$$= \sum_{k=0}^{|x|} \xi^k \cdot \text{number of elements of } L_k \text{ less than or equal to } x.$$

It is apparent that

$$\Phi(1 + \xi) = \sum_{x \in S} (1 + \xi)^{|x|} = \sum_{x \in S} \sum_{k=0}^{|x|} \binom{|x|}{k} \xi^k.$$

Since for $0 \leq k \leq |x|$ there are precisely

$$\binom{|x|}{k}$$

elements in L_k that are below x , we have

$$\Phi(1 + \xi) = \sum_{x \in S} P_x(\xi).$$

The preceding results yield the following identities for $\Phi(\cdot)$:

$$\begin{aligned} \Phi(\lambda) &= \sum_{x \in S} \lambda^{|x|} = \sum_{x \in S} \sum_{y < x} \mu(x - y) q_y(\lambda) \\ &= \sum_{x \in S} \sum_{y < x} r_y(\lambda) \\ &= \sum_{x \in S} \sum_{y < x} \{p_y(\lambda) - \lambda^{|y|}\} \\ &= \sum_{x \in S} P_x(\lambda) - \sum_{x \in S} q_x(\lambda) \\ &= \Phi(1 + \lambda) - \sum_{x \in S} \sum_{y < x} \lambda^{|y|}. \end{aligned}$$

VIII. A REVERSIBLE MARKOV PROCESS FOR WHICH THE CANONICAL DISTRIBUTION IS INVARIANT

We shall describe an ergodic reversible Markov process z_t , taking values in the set S of states, and having the property that its stationary distribution over S is *precisely* the canonical distribution derived from the maximum entropy postulate. This Markov stochastic process can be used to place the canonical distribution into a dynamic context by exhibiting it as invariant under a semigroup of positive operators, viz., the transition matrices of the Markov process in question. The transition rate matrix A of this process, i.e., the generator of the semigroup, then provides several interpretations (cf. Section X) of this dynamic context in terms of behavior at the terminals of the networks, i.e., in terms of calling rates and mean holding-times.

Let $x \in S$ be a possible state of the network. In Section II we have introduced the sets of states A_x and B_x with

A_x = set of states immediately *above* x , i.e., accessible from x by adding a new call,

B_x = set of states immediately *below* x , i.e., accessible from x by a hangup.

The process z_t to be considered can be described heuristically by saying that if $z_t = x$ then z_{t+} is moving to each $y \in A_x$ at a rate $\lambda > 0$,

to each $y \in B_x$ at a rate unity, and to any other state at a rate zero. Its transition rate matrix $A = (a_{xy})$ is given by

$$a_{xy} = \begin{cases} -|x| - \lambda |A_x| & \text{if } y = x \\ 1 & \text{if } y \in B_x \\ \lambda & \text{if } y \in A_x \\ 0 & \text{if } y \notin A_x \cup B_x \text{ and } y \neq x. \end{cases}$$

With this matrix we can define a Markov process z_t in the usual way. (Cf. Doob.⁸) A discussion and critique of possible physical interpretations of the rate matrix A is given in Section X.

The probabilistic interpretation of the rate matrix A is that if $z_t = x$ there is a conditional probability $\lambda h + o(h)$ that $z_{t+h} = y$, for $y \in A_x$; there is a conditional probability $h + o(h)$ that $z_{t+h} = y$, for $y \in B_x$; there is a conditional probability $1 - \lambda h |A_x| - h |x| - o(h)$ that $z_{t+h} = x$; all other events have a conditional probability $o(h)$, as $h \rightarrow 0$. The constant λ is the calling rate per idle path.

An alternative informal description of the Markov process z_t is as follows: the length of time spent in any state x is a random variable independent of all other lengths of time spent in a state, having a negative exponential distribution with a mean

$$\frac{1}{|x| + \lambda |A_x|}.$$

At the end of a stay in x , a new state is chosen (independently of everything except x) from $A_x \cup B_x$ according to the probabilities

$$\frac{\lambda}{|x| + \lambda |A_x|} \quad \text{for elements of } A_x$$

$$\frac{1}{|x| + \lambda |A_x|} \quad \text{for elements of } B_x.$$

The equation $Aq = 0$ is the matrix-vector form of the statistical equilibrium equations for the process z_t . These equations can be written out and solved explicitly, as follows: $Aq = 0$ is equivalent to

$$\{|x| + \lambda |A_x|\} q_x = \sum_{y \in A_x} q_y + \lambda \sum_{y \in B_x} q_y, \quad x \in S. \quad (6)$$

We find by substitution that q_x taken proportional to $\lambda^{|x|}$ gives a solution. Hence the unique normalized (to be a *probability vector*) solution is

$$q_x = \frac{\lambda^{|x|}}{\sum_{x \in S} \lambda^{|x|}} = \frac{\lambda^{|x|}}{\Phi(\lambda)}.$$

This is precisely the canonical distribution of probability over S which was obtained earlier from the maximum entropy principle. Thus, one sense in which the canonical distribution is an *equilibrium* distribution is that it is invariant under the transition probability matrices of z_t .

It will be noticed that the vector q has components which satisfy not only the statistical equilibrium equation (6) for z_t , but also the much stronger condition

$$q_x a_{xy} = q_y a_{yx} \quad x, y \in S,$$

which is an analog of the principle of detailed balance. In the language of probability, this condition is that of *reversibility*; that is, it is equivalent to the condition that the process z_t look the same whether seen forward or backward in the sense that for any two times t and s

$$\Pr \{z_t = x \text{ and } z_s = y\} = \Pr \{z_t = y \text{ and } z_s = x\}.$$

The reversibility of z_t has important statistical consequences, explored in Sections XII–XIV. However, an immediate consequence is the following form of the Boltzmann H -theorem for z_t :

Remark: Let

$$H_x(t) = - \sum_{y \in S} q_{xy}(t) \log q_{xy}(t),$$

where

$$q_{xy}(t) = \Pr \{z_t = y \mid z_0 = x\}.$$

Then

$$(d/dt)H_x(t) \geq 0.$$

The proof of this is well-known, being just Pauli's proof of the quantum-mechanical H -theorem from the principle of detailed balance. (See Tolman,⁹ p. 464.)

IX. ANALOGY WITH STATISTICAL MECHANICS

As its name suggests, the canonical distribution of probability over S , implied by the maximum entropy principle, resembles the canonical ensemble of statistical mechanics and thermodynamics. This analogy extends to several other concepts arising either in traffic theory or in statistical mechanics, and will now be described. It is assumed that the reader is familiar with the rudiments of statistical mechanics; a lucid account can be found in Khinchin.⁶

Let us consider a conservative mechanical system embedded in a heat bath, and assume that it is described by a canonical ensemble. It can

exchange energy with its surroundings; its energy is a randomly varying quantity. The basic identification we make is of the *number of calls in progress* in a connecting network with the *energy* of this mechanical system. In other words, new calls in the operating network are analogous to increments of energy in the mechanical system, while hangups represent decrements of energy. The *average energy* is identified with the *average load* carried by the network.

The surfaces of constant energy in the phase-space of the mechanical system are analogous to the levels L_k , i.e., the sets consisting of the various states with k calls in progress for $k = 0, 1, 2, \dots$. The number $|L_k|$ of ways of putting up k calls, on which our theory rests, is the analog of the area of a surface of constant energy. Just as the canonical density function is constant over the surfaces of constant energy and maximizes the entropy for a given average energy, so is the canonical probability vector q constant over each L_k and maximizes $H(q)$.

The partition function of statistical mechanics is defined (cf. Ref. 6, p. 79) by

$$Z(\alpha) = \int_{\Gamma} e^{-\alpha H(x)} dV(x),$$

where Γ is the phase-space, $x \in \Gamma$ is a typical state, $H(x)$ is the total energy of state x (here given by the Hamiltonian function), and dV is the volume element of phase-space. In a similar way, the partition function $\Phi(\cdot)$ is the generating function of the numbers $|L_k|$, $k = 0, 1, 2, \dots$. The set S of states corresponds to the phase-space Γ , $H(x)$ is analogous to $|x|$, the volume measure on Γ is analogous to the counting measure on S , and $e^{-\alpha}$ replaces ξ .

In Khinchin's development⁶ of statistical mechanics the temperature is defined as inversely proportional to the unique root θ of the equation

$$(d/d\theta) \log Z(\theta) = \text{average energy.}$$

Specifically, the absolute temperature T is given by

$$\theta = (kT)^{-1},$$

where k is Boltzmann's constant. In our model for a connecting network the analog (modulo a logarithmic transformation) of θ is the solution λ of

$$(d/d\lambda) \log \Phi(\lambda) = \text{average load carried.}$$

Thus it is tempting to identify $(\log \lambda)^{-1}$ as proportional to the "temperature" of the traffic system.

The matrix A , introduced in Section VIII as the "transition rate"

matrix for the process z_t , provides a sense in which the canonical distribution q is of "equilibrium" type. The *reversibility* of z_t is analogous to the detailed balance property of transition matrices in statistical mechanics. (Cf. Tolman,⁹ pp. 165 and 521.) This property also implies that a form of the Boltzmann H -theorem is valid for z_t , as we saw in Section VIII.

The analogies between our thermodynamic model of traffic and statistical mechanics can be collected in the following tabulation:

STATISTICAL MECHANICS	TRAFFIC THEORY
Energy	Calls in progress
Partition function	Generating function $\Phi(\cdot)$ of number of ways of putting up k calls, $0 \leq k \leq w$
Entropy	$-\sum_{x \in S} q_x \log q_x$
Temperature	$\{\log (\text{calling rate per idle path})\}^{-1}$
Area of surface of given energy	$ L_k =$ number of ways of putting up k calls
Transition rate matrix	A
Detailed balance	Reversibility of z_t
Equilibrium	$Aq = 0$
Heat bath	Idle customers' needs
Phase space Γ	Set S of possible states
Volume measure on Γ	Counting measure on S

X. DISCUSSION AND CRITIQUE

It is now reasonable to consider possible physical interpretations of the stochastic process z_t and of the transition rate matrix A in terms of items describing behavior at the inlets and outlets of the connecting network, such as calling rates, holding-times, and routing rules. Obviously, transitions of z_t from a state x into B_x represent hangups, while transitions from x into A_x represent new calls; the entries of A indicate the "rates" at which these events occur in the different states. However, the reasonableness, and so the acceptability, of z_t as a model for traffic depends on the interpretations of z_t and A in physical terms. Hence we must inquire whether (and how) the rates entered in A can be viewed as realistically describing the terminations of calls in progress, the occurrence of new calls between inlets and outlets of the network, and their routing or disposition.

In general, to construct a Markov process as a model for traffic in a

connecting network whose states form the set S , it is usually sufficient to give, for each state $x \in S$, and each inlet u and outlet v ,

- (i) the hangup rate for the various calls in progress in x ,
- (ii) the calling rate between u and v in state x ,
- (iii) the method for disposing of requests that encounter congestion, receive busy tone, etc.,
- (iv) the method for choosing routes of new calls.

A particular choice of the items (i)–(iv) leads to a transition rate matrix, and so to a Markov process. We shall assess the reasonableness of z_t as a model for traffic in terms of items (i)–(iv) above by exhibiting two choices of (i)–(iv) that both lead to the rate matrix A of z_t .

In the dynamic model z_t described in Section VIII, the role of the inlets and outlets is open to (at least) two different interpretations, each of which induces a corresponding interpretation of the transition rate matrix A .

One possible interpretation of the inlets and outlets is to take them seriously as actual terminals or customers' lines. They are then the outermost portions of the network under study, the original sources for traffic that enters the system, beyond which there is no more connecting or switching equipment. From any inlet, or to any outlet, there can be at most one call in progress. In this case the rate matrix A can be interpreted as saying that in a state x each call in progress is terminating at a unit rate, that the calling rate from an idle inlet u to an idle outlet v is

$$\begin{aligned} \lambda \cdot \text{number of available paths from } u \text{ to } v \text{ in state } x \\ = \lambda \cdot \text{number of states covering } x \text{ which include a } (u,v) \text{ call,} \end{aligned}$$

and that of the possible routes for a new call one is chosen at random (equal probability for all). The reader can verify that this choice of (i)–(iv) does in fact lead to the rate matrix A . Note that this description does not provide for the generation of blocked calls.

The choice of a unit hangup rate per call in progress is tantamount to measuring time in units of mean holding-time, with the convenience that carried and offered loads come out in the standard units of erlangs. This unit hangup rate can be obtained as a consequence of assuming that the holding-times are negative-exponentially distributed with mean unity, mutually independent, and independent of the random process describing new calls. This assumption of "negative exponential holding-times" is a standard one in congestion theory. (See e.g., Syski,¹⁰ p. 9.)

More interesting (and questionable!) is the fact that under this interpretation the calling rate in a state x between an idle inlet u and an idle

outlet v depends on the number of ways in which a call from u to v could be put into the network in state x . This calling rate can therefore change in time as the state changes, even if u and v remain idle. It can be argued that this is an unrealistic feature, and that therefore z_t is not a wholly reasonable model for telephone traffic in a network whose inlets and outlets are interpreted as terminals or customers' lines. For surely the idle calling parties do not know the state of the system, nor the number of paths available for a call between them, and so they cannot (let alone *do not*) adjust their calling rates accordingly.

In a sense, it would be more intuitive and reasonable to assign a calling rate λ to each idle pair (u,v) of terminals (an inlet u and an outlet v) irrespective of the state x of the system. This basic calling rate for each idle pair (u,v) is then distributed over the states that cover x and realize (u,v) [assuming that (u,v) is not blocked, so that there are such states] in accordance with some routing rule. A stochastic process x_t on S based on this idea is described in Section XI, and is studied in detail in a work (Beneš¹¹) to appear later.

From an *a priori* viewpoint, x_t is a more reasonable model for traffic than z_t . The objection (described above) to letting the calling rate for an idle pair depend on the state is severe. Nevertheless it does not necessarily destroy the usefulness of the process z_t for describing traffic. Three comments are relevant here:

(1) If all calls can be put up in at most one way, then x_t and z_t coincide.

(2) If calls can be put up in only a few ways, it may often be possible or useful to regard z_t as a small perturbation of x_t obtained by raising various calling rates. This idea is explored in Section XI.

(3) Even if z_t is not in any precise sense a small perturbation of the *a priori* reasonable model x_t , it deserves to be considered as a model of traffic. It must not be forgotten that the usefulness of a theory rests more on its success in predicting than on its meeting criteria of reasonableness that are adduced *a priori*.

However, it is possible to give the inlets and outlets a second interpretation, different from the one that assigns them the role of "outermost terminals." This interpretation makes z_t a fairly reasonable model of traffic, in the *a priori* sense we are discussing. It consists in letting each inlet or outlet represent a point from which several or many calls can be in progress to other points in the system. Physically, such an inlet or outlet might be a PBX or central office serving a locality. As such, it would itself contain a connecting network which is left out of account in the model. It no longer necessarily makes sense to speak of busy and idle inlets, or outlets.

To give an intuitive rationale for this interpretation and for the assumption about calling rates that corresponds to it, let us pick an inlet-outlet pair (u,v) and think of u and v as (possibly geographically separated) points between which there may be several calls in progress. For example, the network under study might be a toll network, and u and v might be local central offices acting as sources of traffic for the toll system. Or, for a second example, u and v might be distinct switches in a large network inside a central office.

In such situations, it is natural to expect that if in a state there are *many* paths available for a call from u to v , then there is a *larger* probability that a requested call from u to v arise in the next small interval of time h than if there were very few paths between u and v available. In other words, it is reasonable that the calling rate in x for (u,v) calls be a monotone increasing function of the number of paths available in x for such calls.

A particularly simple monotone function is the linear one, and we shall assume that the calling rate for an idle pair (u,v) in x is

$$\lambda \cdot \text{number of paths available in } x \text{ for } (u,v) \text{ call,}$$

and that of the available paths one is chosen at random. Again, no provision is made for the generation of blocked attempts, since these will not affect the state probabilities when blocked calls are refused.

We observe that A_x can be partitioned and written as

$$A_x = \bigcup_{(u,v)} A_x(u,v),$$

where

$$A_x(u,v) = \{y: y \text{ covers } x \text{ and realizes } (u,v)\}$$

with

$$|A_x(u,v)| = \text{number of paths available in } x \text{ for a } (u,v) \text{ call.}$$

Since routes for new calls are chosen at random we find that the transition rate from x to $y \in A_x$ is exactly λ , so that this second interpretation also leads to the rate matrix A .

XI. A MARKOV MODEL BASED ON TERMINAL-PAIR BEHAVIOR

We now revert to interpreting inlets and outlets as the ultimate terminals of the connecting network. In Section X it was suggested that under this interpretation an *a priori* reasonable model (a stochastic process x_t) can be obtained by postulating an effective calling rate $\lambda > 0$ per idle inlet-outlet pair. This can be done by assuming that each

idle inlet calls an arbitrary outlet at a rate λ , and *vice versa*, with attempted calls to busy terminals rejected with no change of state. The total attempt rate in a state x (excluding calls to busy terminals) is

$$\lambda \cdot \left\{ \begin{array}{l} \text{number of idle} \\ \text{inlet-outlet pairs in } x \end{array} \right\}.$$

If I is the set of inlets, and Ω that of outlets, with I and Ω disjoint, this has the quadratic form

$$\lambda(|I| - |x|)(|\Omega| - |x|).$$

As before, we assume a unit hangup rate per call in progress, with blocked calls rejected. The description of x_t can be completed, finally, by specifying a method of routing. This we do by introducing a "routing matrix" $R = (r_{xy})$ with the following properties: Let x be a state, and let Π be the partition of A_x induced by the equivalence relation \sim of "having the same calls up, possibly on different routes"; then

$$\begin{aligned} r_{xy} &\geq 0 \\ r_{xy} &= 0 \quad \text{unless } y \in A_x \\ \sum_{y \in Y} r_{xy} &= 1 \quad \text{for } Y \in \Pi. \end{aligned}$$

We note that $\sum_{y \in S} r_{xy}$ is exactly the number $s(x)$ of attempts which would be "successful" if they arose in state x , and that Π consists of exactly the sets $A_x(u,v)$ for $\{(u,v)\}$ idle and unblocked in x .

The routing matrix R is to have this interpretation: each time the call $\{(u,v)\}$ is to be completed in state x , a state y is chosen independently from $A_x(u,v)$ with probability r_{xy} and the call is routed so as to take the system to state y .

The foregoing assumptions lead to a rate matrix Q for x_t defined by

$$q_{xy} = \begin{cases} 1 & \text{if } y \in B_x \\ \lambda r_{xy} & \text{if } y \in A_x \\ -|x| - \lambda s(x) & \text{if } y = x \\ 0 & \text{if } y \in (A_x \cup B_x)' \text{ and } y \neq x. \end{cases}$$

This matrix is exactly like A except that for $y \in A_x$ the rate from x to y is not λ but (the in general smaller quantity) λr_{xy} , and that the diagonal terms are correspondingly increased so as to keep row sums equal to zero. For each $Y \in \Pi$, r_{xy} for $y \in Y$ represents a distribution of the calling rate of some idle unblocked pair (u,v) over $A_x(u,v) = Y$. Indeed A results from Q if all the r_{xy} are replaced by unity. The process x_t can be defined in terms of its rate matrix Q .

The assumptions leading to the rate matrix Q and to the process x_t have much *a priori* appeal; x_t itself is discussed in detail in a forthcoming paper¹¹ already mentioned. Here we shall merely consider whether z_t may be regarded as a perturbation of x_t . Since each process is determined by its respective rate matrix, and since we are interested mostly in equilibrium behavior, we restrict attention to asking how different are the respective equilibrium distributions over S for x_t and z_t . Thus, if p and q are probability row-vectors satisfying $Qp = 0$ and $Aq = 0$ respectively, how different is p from q ?

To give a precise estimate, we introduce the norms

$$\| M \| = \sum_{x,y} | m_{xy} |$$

$$\| v \| = \sum_x | v_x |$$

for matrices and vectors, respectively. Since $Ap = (A - Q)p$ and $Qq = (Q - A)q$, we find

$$\| p - q \| \leq \frac{2 \| Q - A \|}{1 - \| Q - A \|}.$$

The norm of $Q - A$, in turn, can be seen to be

$$\begin{aligned} \| Q - A \| &= 2\lambda \sum_{x \in S} \sum_{y \in A_x} (1 - r_{xy}) \\ &= 2\lambda \sum_{x \in S} (| A_x | - s(x)) \\ &= 2\lambda \{ \Phi'(1) - \sum_{x \in S} s(x) \} \end{aligned}$$

where

$s(x)$ = number of pairs that are idle and not blocked in x .

Letting

μ = max number of ways a call can be realized

we find $| A_x | \leq \mu s(x)$, and hence

$$\begin{aligned} \| Q - A \| &\leq 2\lambda(\mu - 1) \sum_{x \in S} s(x) \\ &\leq 2\lambda(\mu - 1)\Phi'(1). \end{aligned}$$

Let

$$w = \max_{x \in S} | x |$$

so that $\Phi'(1) \leq w | S |$, and

$$\| Q - A \| \leq 2\lambda(\mu - 1)w | S |.$$

The average contribution (per state) to $\| Q - A \|$ is then

$$\frac{\| Q - A \|}{| S |} \leq 2\lambda(\mu - 1)w.$$

XII. THE APPROACH TO EQUILIBRIUM

It is known from the theory of Markov processes that the matrix $Q(t) = (q_{xy}(t))$ of the transition probabilities

$$q_{xy}(t) = \Pr \{z(t + s) = y \mid z(s) = x\}, \quad t \geq 0$$

of the process z_t satisfies the Kolmogorov equations

$$(d/dt)Q(t) = AQ(t) = Q(t)A, \quad Q(0) = I,$$

and that the study of the *time-dependent* (as opposed to the asymptotic, or equilibrium) behavior of z_t can be carried out in terms of the characteristic values of A . Knowledge of the transition probabilities is essential, for example, in calculating the sampling error incurred in such load averages as

$$\frac{1}{n} \sum_{j=1}^n |z_{n\tau}|, \quad \frac{1}{T} \int_0^T |z_t| dt, \quad (7)$$

where τ is the interval between successive discrete observations of $|z_t|$, and $(0, T)$ is an interval of continuous observation of $|z_t|$. In this section we study the manner in which z_t approaches equilibrium in terms of the two principal characteristic values of A , i.e., that of largest, and that of smallest nonzero, magnitude. Applications to estimating the covariances of functions of z_t , and to studying sampling error for the traffic averages in (7), are described in Sections XIII and XIV, respectively.

Our study of the approach to equilibrium is based on the observation that the matrix A of transition rates for the process z_t is symmetrizable, i.e., is a self-adjoint operator in a suitably chosen inner-product space of finite dimension $| S |$. The probabilities

$$q_x = \frac{\lambda^{|x|}}{\Phi(\lambda)} = \frac{1}{\mu_x}$$

are all strictly nonnegative, and we use their reciprocals μ_x as weights in defining an inner product,

$$(r,s) = \sum_{x \in S} r_x \bar{s}_x \mu_x, \tag{8}$$

and a norm,

$$\|s\| = (s,s)^{\frac{1}{2}}.$$

We now remark that for all states x,y from S ,

$$q_y a_{yx} = q_x a_{xy}$$

or alternatively

$$a_{yx} \mu_x = a_{xy} \mu_y.$$

Indeed, this remark is the basis for the solution q given in Section VIII for the statistical equilibrium equations (6) of the process z_t ; it has the important consequence that A is self-adjoint with respect to the inner product defined by (8), viz.

Lemma 2: $(Ar,s) = (r,As)$, for any $|S|$ -vectors r,s .

Proof: A is a real matrix, so

$$\begin{aligned} \sum_x \sum_y a_{yx} r_y \bar{s}_x \mu_x &= \sum_x \sum_y r_y \bar{s}_x a_{xy} \mu_y \\ &= \sum_y r_y \mu_y \sum_x a_{xy} \bar{s}_x = \sum_y r_y \overline{\sum_x a_{xy} s_x \mu_y} \end{aligned}$$

In a similar way we prove

Lemma 3:

$$(Ar,s) = -\frac{1}{2} \sum_x \sum_y a_{yx} q_y (\overline{s_x \mu_x - x_y \mu_y}) (\mu_x r_x - \mu_y r_y).$$

Proof: Since the matrix whose elements are $a_{yx} \mu_x$ is symmetric, we have

$$\begin{aligned} (Ar,s) &= \sum_x \sum_y a_{yx} \mu_x r_y \bar{s}_x \\ &= \frac{1}{2} \sum_x \sum_y a_{yx} \mu_x (r_y \bar{s}_x + r_x \bar{s}_y). \end{aligned}$$

Now

$$\sum_x \sum_y a_{yx} q_y \mu_x^2 \bar{s}_y r_x = 0$$

because $Aq = 0$, and

$$\begin{aligned} \sum_x \sum_y a_{yx} q_y \mu_y^2 \bar{s}_y r_y &= \sum_y \sum_x a_{yx} \mu_y \bar{s}_y r_y \\ &= 0 \end{aligned}$$

because $\sum_x a_{yx} = 0$. This proves the lemma.

Theorem 2: The characteristic values of A are real and nonpositive. Zero is a simple characteristic value corresponding to the characteristic vector q , normalized to unity.

Proof: The result follows from the known properties of self-adjoint transformations. (See Ref. 12, pp. 153–155.)

The characteristic values of A will all be of the Rayleigh quotient form

$$\frac{(Av, v)}{(v, v)} \leq 0$$

for some vector v ; by Lemma 3 this form is nonpositive. The probability vector solution q of $Aq = 0$ is unique so that zero is a simple characteristic value. Furthermore, if $0 > r_{\max} = r_1 \geq \dots \geq r_{|S|-1} = r_{\min}$ is an arrangement of the characteristic values in decreasing order, the variational description of the characteristic values (Ref. 12, p. 111) implies that with $\|v\|^2 = (v, v)$,

$$r_{\max} = r_1 = \max \{(Av, v) \mid v \perp q, \|v\| = 1\}$$

$$r_{\min} = r_{|S|-1} = \min \{(Av, v) \mid \|v\| = 1\}.$$

The alternative notations r_{\max} and r_{\min} identify the two “dominant” characteristic values, and are introduced for later convenience to enhance the symmetry of the theory.

One can now estimate r_1 from below by substituting suitable trial vectors in the Rayleigh quotient. Choosing a vector v with components

$$v_x = \frac{|x| - m}{\sigma \mu_x}, \quad x \in S,$$

where

$$m = \sum_{x \in S} |x| q_x = \lambda \frac{\partial}{\partial \lambda} \log \Phi(\lambda)$$

$$\sigma^2 = \sum_{x \in S} (|x| - m)^2 q_x = \lambda^2 \left(\frac{\partial^2}{\partial \lambda^2} + \frac{\partial}{\partial \lambda} \right) \log \Phi(\lambda),$$

it is easily seen that $(q, v) = 0$, that $\|v\| = 1$, and that

$$\begin{aligned} (Av, v) &= -\frac{1}{2} \sum_x \sum_y a_{yx} q_y \left(\frac{|y| - m}{\sigma} - \frac{|x| - m}{\sigma} \right)^2 \\ &= -\frac{1}{2\sigma^2} \sum_y q_y (|y| + \lambda |A_y|). \end{aligned}$$

In equilibrium, the average rate of new calls equals the average rate of

hangups, as can be verified from the equilibrium equations $Aq = 0$. That is,

$$\sum_{y \in S} |y| q_y = \lambda \sum_{y \in S} q_y |A_y|,$$

and we find

$$-\frac{m}{\sigma^2} \leq r_1 < 0,$$

a generalization of a result known (Ref. 13, p. 147) for the simple busy signal trunk group (classical Erlang model).

In general, letting $f(\cdot)$ be any function defined on the set S of states, but not identically a constant, we define

$$m_f = \sum_{x \in S} f(x) q_x$$

$$\sigma_f^2 = \sum_{x \in S} (f(x) - m_f)^2 q_x.$$

Choosing now a vector v with components

$$v_x = \frac{f(x) - m_f}{\sigma_f \mu_x}$$

we obtain (by repetition of previous reasoning)

$$-\frac{1}{2\sigma_f^2} \sum_y q_y \left(\sum_{x \in B_x} [f(x) - f(y)]^2 + \lambda \sum_{x \in A_y} [f(x) - f(y)]^2 \right)$$

as a lower bound for r_1 .

We now define a set of *vector-valued* functions $\{c_x(t), x \in S, t \geq 0\}$ by the condition

$$c_{xy}(t) = q_{xy}(t) - q_y, \quad y \in S.$$

The function $c_x(\cdot)$ describes the approach to equilibrium from the initial state x at time $t = 0$.

Theorem 3: For $t \geq 0$

$$\|c_x(0)\| \exp(r_{\min} t) \leq \|c_x(t)\| \leq \|c_x(0)\| \exp(r_{\max} t).$$

Proof: Since q_y and $q_{xy}(t)$ are both distributions in y , we have

$$(c_x(t), q) = \sum_y q_{xy}(t) - q_y = 0$$

so that $c_x(t) \perp q$. Also

$$\begin{aligned} \frac{d}{dt} \| c_x(t) \|^2 &= 2 \sum_y [q_{xy}(t) - q_y] \frac{d}{dt} q_{xy}(t) \mu_y \\ &= 2 \left(c_x, \frac{d}{dt} c_x \right) \\ &= 2(c_x, A c_x) \end{aligned}$$

since $(d/dt)c_x = A c_x$; that is for each $y \in S$,

$$\begin{aligned} \frac{d}{dt} c_{xy} &= \frac{d}{dt} q_{xy} = \sum_z q_{xz}(t) a_{zy} \\ &= \sum_z [q_{xz}(t) - q_z] a_{zy}. \end{aligned}$$

Hence, $\| c_x \|$ being nonzero, we find

$$2r_{\min} \leq (d/dt) \log \| c_x \|^2 \leq 2r_{\max}$$

and Theorem 3 follows by integration. The argument just given is essentially reproduced from Kramer.¹⁴

XIII. COVARIANCES OF FUNCTIONS OF z_t

For the purposes of this section it is convenient to introduce an inner product $(\cdot, \cdot)'$, closely related to but different from (\cdot, \cdot) of the previous section, and defined by

$$(r, s)' = \sum_{x \in S} r_x \bar{s}_x q_x.$$

The associated norm is denoted by $\| r \|' = (r, r)'^{1/2}$. The point of the "prime" notation is explained by the fact that the transpose A' of A is self-adjoint with respect to $(\cdot, \cdot)'$.

Remark: Where A' is the transpose of A

$$(A' r, s)' = (r, A' s)'.$$

Proof:

$$\begin{aligned} \sum_x \sum_y a_{xy} r_y \bar{s}_x q_x &= \sum_x \sum_y a_{yx} \bar{s}_x r_y q_y \\ &= \sum_y r_y \sum_x a_{yx} \bar{s}_x q_y = \sum_y r_y \overline{\sum_x a_{yx} s_x q_y}. \end{aligned}$$

Let $f(\cdot)$ be a function defined on S , and define a stochastic process f_t by the condition

$$f_t = f(z_t).$$

Theorem 4: The covariance of f_t is given by

$$R_f(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} (f, A^n f)',$$

where the vector f is defined by

$$\begin{aligned} f_x &= f(x) - \sum_{x \in S} f(x) q_x, \\ &= f(x) - m_f. \end{aligned}$$

Proof: The covariance of f_i is

$$\begin{aligned} \sum_x \sum_y q_x q_{xy}(t) f_x f_y &= (f, Q(t)' f)' \\ &= (f, (\exp tA)' f)' \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} (f, A^n f)' \end{aligned}$$

with $Q(t)'$ denoting the transpose, and not the derivative, of $Q(t)$. The covariance of f_i is thus the exponential generating function of the series of numbers

$$(f, A^n f)' \quad n = 0, 1, 2, \dots$$

These can be calculated with the help of the following results:

Lemma 4: Let the matrix elements of A^n be $a_{xy}^{(n)}$. Then

$$q_x a_{xy}^{(n)} = q_y a_{yx}^{(n)}.$$

Proof:

$$\begin{aligned} q_x a_{xy}^{(n)} &= q_x \sum_{u_1, \dots, u_{n-1}} a_{xu_1} a_{u_1 u_2} \cdots a_{u_{n-1} y} \\ &= \sum_{u_1, \dots, u_{n-1}} a_{u_1 x} a_{u_2 u_1} \cdots a_{y u_{n-1}} q_y \\ &= q_y \sum_{u_{n-1}, \dots, u_1} a_{y u_{n-1}} \cdots a_{u_2 u_1} a_{u_1 x} \\ &= q_y a_{yx}^{(n)}. \end{aligned}$$

Lemma 5: Let Q be the diagonal matrix of elements q_x , $x \in S$. Then

$$(w, A^n w)' = (A^n Qw, Qw).$$

Proof:

$$\begin{aligned} \sum_x w_x \overline{\sum_y a_{xy}^{(n)} w_y q_x} &= \sum_x (Qw)_x \sum_y \mu_y a_{xy}^{(n)} \overline{(Qw)_y} \\ &= \sum_x (Qw)_x \sum_y a_{yx}^{(n)} \overline{(Qw)_y} \mu_x. \end{aligned}$$

From the three preceding results we obtain

Theorem 5: The covariance of f_t is

$$R_f(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} (A^n Qf, Qf),$$

where the vector f is as in Theorem 4, and Q is the diagonal matrix of elements q_x , $x \in S$.

It is readily seen that A^n , $n \geq 1$, is again a self-adjoint operator with respect to (\cdot, \cdot) , and that its characteristic values are precisely the n th powers of those of A . Also, for any vector v and $n \geq 0$

$$(A^n v, v) \begin{cases} \leq 0 & \text{if } n \text{ is odd} \\ \geq 0 & \text{if } n \text{ is zero or even} \end{cases}$$

so that by the variational description of characteristic values we have

$$\left. \begin{array}{l} r_{\min}^n, \quad n \text{ odd} \\ r_{\max}^n, \quad n \text{ even} \end{array} \right\} \leq \frac{(A^n v, v)}{(v, v)} \leq \left\{ \begin{array}{l} r_{\max}^n, \quad n \text{ odd} \\ r_{\min}^n, \quad n \text{ even} \end{array} \right. \quad (10)$$

provided that $v \perp q$ (in the inequalities involving r_{\max}). Returning now to the vector Qf of Theorem 5, we find

$$\begin{aligned} \|Qf\|^2 &= \sum_x q_x^2 f_x^2 \mu_x \\ &= \sum_{x \in S} (f(x) - m_f)^2 q_x = \sigma_f^2 \end{aligned}$$

and

$$\sum_x q_x (Qf)_x \mu_x = \sum_x q_x f_x = 0,$$

so that $Qf \perp q$. Letting $v = Qf$ in (10), we obtain

$$\begin{aligned} \sigma_f^2 r_{\min}^n &\leq (A^n Qf, Qf) \leq \sigma_f^2 r_{\max}^n, \quad n \text{ odd,} \\ \sigma_f^2 r_{\max}^n &\leq (A^n Qf, Qf) \leq \sigma_f^2 r_{\min}^n, \quad n \text{ even.} \end{aligned}$$

Unfortunately, these inequalities do not give useful bounds for the covariance $R_f(\cdot)$. However, such bounds can be obtained from the formula of Theorem 5 in an elegant way by applying the spectral theorem to A .

Theorem 6: Let $\alpha_1, \dots, \alpha_k$ denote the distinct characteristic values of A , and let E_i , $i = 1, \dots, k$, denote the perpendicular projection on the subspace of all solutions $Ar = \alpha_i r$. Then the covariance $R_f(\cdot)$ of f_t is given by

$$R_f(t) = \sum_{i=0}^k (E_i Qf, Qf) e^{\alpha_i t},$$

with $1 \leq k \leq |S|$, Q the diagonal matrix of elements q_x , $x \in S$, and f given by $f_x = f(x) - m_f$.

Proof: By the spectral theorem for self-adjoint operators (Ref. 12, p. 56) we can write

$$A = \sum_{i=0}^k \alpha_i E_i$$

and

$$A^n = \sum_{i=0}^k \alpha_i^n E_i.$$

We can now calculate with formula (9) of Theorem 5:

$$\begin{aligned} R_f(t) &= \sum_{n=0}^{\infty} \frac{t^n}{n!} (A^n Qf, Qf) \\ &= \sum_{n=0}^{\infty} \sum_{i=0}^k \frac{(\alpha_i t)^n}{n!} (E_i Qf, Qf) \\ &= \sum_{i=0}^k (E_i Qf, Qf) e^{\alpha_i t}. \end{aligned}$$

This proves Theorem 6. Since we know that zero is among the characteristic values (indeed, it is a simple one), one of the α 's, say α_0 , will be zero. We may reasonably expect $R_f(\cdot)$ to approach zero for large t ; hence the constant, i.e., α_0 , term of $R_f(\cdot)$ should be zero. This can be seen as follows: the subspace associated with zero consists of vectors proportional to the equilibrium vector q , because zero is a simple characteristic value; but we have already verified that $q \perp Qf$; hence

$$(E_0 r, Qf) = 0, \quad \text{all } r.$$

Using this we prove the

Corollary 1: $R_f(t) \geq 0$ for all t , and in fact

$$0 \leq \sigma_f^2 e^{r_{\min}|t|} \leq R_f(t) \leq \sigma_f^2 e^{r_{\max}|t|}, \quad \text{all } t.$$

Proof: Since the E_i of Theorem 6 are perpendicular projections, they are linear, self-adjoint, and positive in the sense of Ref. 12, p. 140; the usual term for positive is nonnegative semidefinite. Hence

$$(E_i r, r) \geq 0$$

for any vector r . Since $(E_0 r, Qf) = 0$ if E_0 is associated with the zero characteristic value, the result follows from Theorem 6, using

$$\sum_{i=0}^k E_i = I,$$

$$\sum_{i=0}^k (E_i Qf, Qf) = \| Qf \|^2 = \sigma_f^2.$$

XIV. APPLICATIONS TO SAMPLING ERROR

Let us suppose that n samples of the process $f_t (= f(z_t))$ are observed during an interval of equilibrium of z_t at intervals τ apart, and that the normed sum

$$n^{-1} S_n = n^{-1} \sum_{j=1}^n f_{j\tau}$$

is used as an estimate of $E\{f_t\}$. We find that

$$\text{Var} \{S_n\} = \sum_{j=-n}^n (n - |j|) R_f(j\tau),$$

where $R_f(\cdot)$ is the covariance of f_t . By using the identity

$$\begin{aligned} \sum_{j=-n}^n (n - |j|) e^{-2|j|u} &= n \operatorname{ctnh} u - \frac{1 - e^{-2nu}}{2} \operatorname{csch}^2 u, \\ &= v_n(u), \end{aligned}$$

together with Corollary 1 of Section XIII, we find that

$$\sigma_f^2 v_n(-\frac{1}{2}\tau r_{\min}) \leq \text{Var} \{S_n\} \leq \sigma_f^2 v_n(-\frac{1}{2}\tau r_{\max}).$$

In a similar way, if f_t is observed continuously over an interval $(0, T)$ of equilibrium of z_t and the time average

$$M(T) = \frac{1}{T} \int_0^T f(z_t) dt$$

is used as an estimate of $E\{f_t\}$, then

$$\text{Var} \{M(T)\} = 2T^{-2} \int_0^T (T - t) R_f(t) dt,$$

and Corollary 1 gives

$$\sigma_f^2 \int_0^T (T - t) e^{r_{\min} t} dt \leq \text{Var} \{M(T)\} \leq \sigma_f^2 \int_0^T (T - t) e^{r_{\max} t} dt.$$

XV. A GENERALIZATION

As an extension of the maximum problem posed and solved in Section V we shall seek functions

$$q_x(t), \quad x \in S, \quad t_1 \leq t \leq t_2, \quad t_1 < t_2$$

such that for each t in $[t_1, t_2]$

$$\sum_{x \in S} q_x(t) = 1, \quad q_x(t) \geq 0$$

$$\sum_{x \in S} |x| q_x(t) = m(t) > 0$$

$$\int_{t_1}^{t_2} H(q(t)) dt = \text{maximum.}$$

In other words, we look for a time-dependent distribution of probability over S with prescribed mean values for the function $|\cdot|$ on S , such that the integral of the entropy functional over (t_1, t_2) is a maximum.

The Euler equations for this problem assume the trivial form (with $L_1(\cdot)$ and $L_2(\cdot)$ as Lagrange's multipliers):

$$(\partial H / \partial q_x) - L_1(t) |x| - L_2(t) = 0, \quad x \in S$$

or, writing out the H -derivative,

$$\log q_x(t) + 1 + L_1(t) |x| + L_2(t) = 0, \quad x \in S.$$

The argument of Lemma 1 following equation (4) shows that $q_x(\cdot)$ is given by

$$q_x(t) = \frac{\lambda(t)^{|x|}}{\Phi(\lambda(t))} t_1 \leq t \leq t_2$$

where $\lambda(\cdot)$ is the unique solution of the equation

$$m(t) = \frac{\sum_{x \in S} |x| \lambda(t)^{|x|}}{\Phi(\lambda(t))} = \left(u \frac{d}{du} \log \Phi(u) \right)_{u=\lambda(t)}.$$

This solution has the form of the canonical distribution at each time point in $[t_1, t_2]$, and Theorem 1 in effect is just the special case of this result that arises when $m(t) \equiv m$. It is apparent that the form of this solution does not depend on what interval $[t_1, t_2]$ was considered, so we may assume that $m(\cdot)$, and hence also $\lambda(\cdot)$ and $q(\cdot)$, are defined on the real axis.

Let us define the matrix-valued function $A(t)$ by $A(t) = (a_{xy}(t))$ where

$$a_{xy}(t) = \begin{cases} 1 & \text{if } y \in B_x \\ \lambda(t) & \text{if } y \in A_x \\ -|x| - \lambda(t) |A_x| & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases}$$

In other words $A(t)$ is obtained from the transition rate matrix A or z_t

by replacing the constant λ by the function $\lambda(\cdot)$. Then for each t

$$A(t)q(t) = 0$$

i.e.,

$$\{ | x | + \lambda(t) | A_x | \} q_x(t) = \sum_{y \in A_x} q_y(t) + \lambda(t) \sum_{y \in B_x} q_y(t).$$

Thus an analog of the statistical equilibrium equation holds at each point in time, and in this sense, a system described by $\{q(t), t_1 \leq t \leq t_2\}$ may be said to be locally in equilibrium throughout the interval (t_1, t_2) .

Let us now redefine the process z_t to be the time-dependent Markov process corresponding to the (time-dependent) transition rate matrix $A(\cdot)$. We know that if $\lambda(\cdot)$ were a constant function with the particular value $\lambda(u)$, then the process z_t would have a stationary or equilibrium distribution over S given by

$$q_x(u) \frac{[\lambda(u)]^{|\mathcal{X}|}}{\Phi(\lambda(u))}.$$

We may therefore expect that if $\lambda(\cdot)$ is not constant, but changes only *slowly* with time, and if z_0 has the absolute distribution (vector) $q(0)$, then z_t for $t > 0$ has a distribution *approximately* given by $q(t)$. Let us see in detail how this occurs.

The transition probability matrix

$$Q(t_1, t_2) = (q_{xy}(t_1, t_2)),$$

$$q_{xy}(t_1, t_2) = \Pr \{ z_{t_2} = y \mid z_{t_1} = x \},$$

is now indexed by two time parameters instead of one, because of the time-dependence of z_t . The forward Kolmogorov equation for $Q(\cdot, \cdot)$ is

$$(\partial/\partial t)Q(u, t) = Q(u, t)A(t), \quad u < t,$$

or

$$\begin{aligned} (\partial/\partial t)q_{xy}(u, t) = -[| y | + \lambda(t)s(y)]q_{xy}(u, t) \\ + \sum_{z \in A_y} q_{xz}(t) + \lambda(t) \sum_{z \in B_y} q_{xz}(t) \end{aligned}$$

with $Q(t, t) = I$. It is easily seen that

$$Q(u, t) = \exp \int_u^t Q(w) dw,$$

the exponential of a matrix being defined by the usual series in powers of the matrix. Therefore if

$$\Pr \{ z_0 = x \} = q_x(0),$$

then

$$\Pr \{z_t = x\} = \sum_{y \in S} \Pr \{z_0 = y\} q_{yx}(t),$$

and the absolute distribution of z_t is given by the vector

$$Q(0,t)q(0).$$

We now write

$$Q(0,t) = \exp tA(t) + \left\{ \exp \int_0^t A(u) du - \exp tA(t) \right\},$$

observe that

$$\int_0^t A(u) du - tA(t) = \int_0^t u dA(u)$$

and make this heuristic argument: Since dA/du is small, $Q(0,t)$ is approximately $\exp tA(t)$; however, for t fixed and u varying

$$\exp uA(t)$$

is the transition probability matrix of a process z_t obtained by setting

$$\lambda(u) \equiv \lambda(t).$$

This matrix approaches, as $u \rightarrow \infty$, the matrix each of whose rows is $q(t)$. If t is large compared to the time it takes this to happen, we may expect, by Theorem 3, that

$$\exp tA(t)$$

have rows all approximately equal to $q(t)$, so that

$$\text{distr } \{z_t\} \approx q(t).$$

The informal argument just given can be made precise. For the purposes of this last section, we again introduce the vector norm

$$\|f\| = \sum_{x \in S} |f_x|$$

and the matrix norm

$$\|M\| = \sum_{x,y \in S} |m_{xy}|.$$

Also, we use the following result:

Lemma 6: Let M, N be $|S|$ -dimensional square matrices, with

$$c = \frac{1}{2} \sup_{x,y \in S} | m_{xy} | + \frac{1}{2} \sup_{x,y \in S} | n_{xy} |$$

Then for integers $k \geq 1$

$$\| M^k - N^k \| \leq (2c | S |)^{k-1} \| M - N \|.$$

Proof: There obtains the identity

$$M^k - N^k = \frac{1}{2}(M + N)(M^{k-1} - N^{k-1}) + \frac{1}{2}(M - N)M^{k-1} + N^{k-1}.$$

If (b_{xy}) are the elements of B and $(b_{xy}^{(k)})$ are those of B^k , then

$$\sup_{x,y} | b_{xy}^{(k)} | \leq | S | \cdot \sup_{x,y} | b_{xy}^{(k-1)} | \cdot \sup_{x,y} | b_{xy} |.$$

Hence, with $k > 1$

$$\sup_{x,y} | m_{xy}^{(k-1)} + n_{xy}^{(k-1)} | \leq | S |^{k-2} (2c)^{k-1},$$

and

$$\| (M - N)(M^{k-1} + N^{k-1}) \| \leq (2c | S |)^{k-1} \| M - N \|.$$

Also,

$$\| (M + N)(M^{k-1} - N^{k-1}) \| \leq c | S | \| M^{k-1} - N^{k-1} \|.$$

Thus

$$\begin{aligned} \| M^k - N^k \| &\leq c | S | \| M^{k-1} - N^{k-1} \| + \frac{1}{2} \| M - N \| (2c | S |)^{k-1} \\ &\leq \| M - N \| (2c | S |)^{k-1} \\ &\quad \cdot \left\{ \left(\frac{1}{2}\right)^{k-1} + \frac{1}{2} \sum_{j=0}^{k-2} (c | S |)^j (2c | S |)^{-j} \right\} \\ &\leq \| M - N \| (2c | S |)^{k-1}. \end{aligned}$$

Using the lemma we find that the norm of

$$\exp \int_0^t A(u) du - \exp tA(t)$$

is at most

$$\sum_{n=1}^{\infty} \frac{\left\| \left(\int_0^t A(u) du \right)^n - t^n A^n(t) \right\|}{n!} \leq \frac{e^b - 1}{f} \left\| \int_0^t u dA(u) \right\|$$

where

$$b = 2 | S | \sup_{x,y} \left| \int_0^t a_{xy}(u) du \right| + t \sup_{x,y} | a_{xy}(t) |.$$

It can be verified that

$$\left\| \int_0^t u \, dA(u) \right\| = 2 \sum_{\substack{x,y \\ y \in A_x}} \left| \int_0^t u \, d\lambda(u) \right| \\ \leq 2t^2 \Phi'(1) \cdot \sup_{u \in [0,t]} |\lambda'(u)|.$$

Thus if $\lambda'(\cdot)$ is small on $[0,t]$ the distribution of z_t is nearly

$$(\exp tA(t))q(0)$$

(in the sense of the vector norm of this section). By Theorem 3, however, this will be nearly $q(t)$ (in the sense of the norm defined by $q(t)$) if t is large compared to the time it takes $\exp uA(t)$ to approach its limit as $u \rightarrow \infty$.

XVI. ACKNOWLEDGMENTS

The author expresses his appreciation of the help and encouragement given him in the preparation of this study by A. Descloux, J. R. Pierce, H. O. Pollak, J. Riordan, and E. Wolman.

REFERENCES

1. Beneš, V. E., Heuristic Remarks and Mathematical Problems Regarding the Theory of Connecting Systems, B.S.T.J., **41**, July, 1962, pp. 1201-1247.
2. Beneš, V. E., Algebraic and Topological Properties of Connecting Networks, B.S.T.J., **41**, July, 1962, pp. 1249-1274.
3. Lee, C. Y., Analysis of Switching Networks, B.S.T.J., **34**, November, 1955, pp. 1287-1315.
4. LeGall, P., Methode de Calcul de L'encombrement dans les Systèmes Téléphoniques Automatiques à Marquage, Ann. des Télécom., **12** (1957), pp. 374-386.
5. Jensen, A., An Elucidation of Erlang's Statistical Works Through the Theory of Stochastic Processes, in *The Life and Works of A. K. Erlang*, Trans. Danish Acad. Sciences, 1948, pp. 23-100.
6. Khinchin, A. I., *Mathematical Foundations of Statistical Mechanics*, Dover, New York, 1949.
7. Birkhoff, G., Lattice Theory, Amer. Math. Soc. Colloq. Publ. XXV, rev. ed., 1948.
8. Doob, J. L., *Stochastic Processes*, John Wiley and Sons, New York, 1953.
9. Tolman, R. C., *The Principles of Statistical Mechanics*, Oxford Univ. Press, London, 1955.
10. Syski, R., *Introduction to Congestion Theory in Telephone Systems*, Oliver and Boyd, London, 1960.
11. Beneš, V. E., Markov Processes Representing Traffic in Connecting Networks, to appear.
12. Halmos, P. R., *Finite-Dimensional Vector Spaces*, second ed., Van Nostrand, Princeton, 1958.
13. Beneš, V. E., The Covariance Function of a Simple Trunk Group, with Applications to Traffic Measurement, B.S.T.J., **40**, January, 1961, pp. 117-148.
14. Kramer, H. P., Symmetrizable Markov Matrices, Ann. Math. Stat. **30** (1959), p. 149.

Flexural Vibrations of a Propped Cantilever

By R. L. PEEK, JR.

(Manuscript received January 11, 1963)

The equations for the flexural vibrations of a propped cantilever beam have been used to compute a number of the vibration characteristics of such beams for the first five modes over the range of prop locations from 50 to 100 per cent of the length. Plots of these characteristics are included in the paper. This material has been prepared primarily for use in studies of contact spring vibration, and such application is briefly discussed. The mathematical treatment used to obtain the relations given is outlined in an Appendix.

I. INTRODUCTION

In relays and other switching apparatus, contact chatter and certain types of wear are associated with vibrations of the contact springs. As an aid in the study of these vibrations, the general theory of beam vibration has been used to develop an analytical treatment applicable to the important class of contact springs which can be considered as propped cantilever beams of uniform cross-section.

In almost all common types of switching apparatus, the contact springs are cantilever beams, clamped at the terminal end, which carry the contact at the free end. When the contact is open, the spring is usually propped or supported by a card or stud, and is therefore a propped cantilever. In some devices, the spring is supported at both the stud and contact when the latter is closed, and is then a doubly propped cantilever. In others, the spring is supported only at the contact when the latter is closed, and is therefore a singly propped cantilever in both operate and release, although the prop location differs for the two cases. Sometimes the mating contact is mounted on another spring, which constitutes a flexible prop, as contrasted with the (relatively) fixed and rigid prop provided by a card or stud.

The relations given here apply only to a uniform cantilever with a

single rigid prop. Some contact springs rigorously conform to these limitations, and the treatment is approximately applicable to a much larger number of cases. A more general treatment of relay spring vibration is given in Chapter 7 of Ref. 1. This includes an outline of approximate methods of analysis applicable to nonuniform springs and to those which are doubly propped, or supported by a compliant prop, such as a spring-mounted mating contact. The treatment given here may be used in applying these more general methods, but the present discussion is confined to the cases where it is directly applicable.

1.1 *Application to Chatter Studies*

The contact chatter of primary interest is that occurring with closed contacts, usually immediately following closure. With a fixed mating contact, the moving spring is a cantilever propped at the contact. Vibration results in modulation of the contact force and therefore of the contact resistance. If the amplitude of the force modulation exceeds the static contact force, a transient open occurs. The timing and duration of these opens can therefore be related to the force modulation and to the amplitudes and frequencies of the spring vibrations. The latter may be directly observed, or predicted from an analysis of the excitation of this vibration involved in operation.

1.2 *Application to Wear Studies*

The wear associated with vibration may occur at the contact or at a supporting or actuating card or stud which serves as the prop to a contact spring. Relative motion in the direction of the spring length results in wear. Severe wear occurs when such longitudinal motion is imposed in actuation. When this is avoided by providing purely perpendicular motion in actuation, wear may still be produced by the longitudinal component of the vibratory motion. The relations given here include those between the longitudinal amplitudes and the (normal) displacement amplitudes, or the corresponding energy content. Thus the longitudinal amplitude can be evaluated from the observed displacement amplitude, or from the estimated energy content of the spring vibration.

II. THEORETICAL FOUNDATION

The equations giving the spring vibration characteristics are derived in the Appendix to this paper. The treatment follows the usual approximate theory of beam vibration, based on the simple theory of bending,

and formally applicable only for displacements which are small compared with the spring dimensions. These formal limitations are of little consequence for the present purpose, although impact causes some deformation other than simple bending.

For springs of uniform section, flexural vibrations conform to a general differential equation (3),* having a solution of the form of (4). This represents a harmonic motion in which all points in the spring move in phase. The relative motion at different points is determined by the dynamic deflection curve X , a function of x only, where x is measured along the length of the spring. Each such deflection curve corresponds to a particular mode of vibration, having a corresponding characteristic frequency. The deflection curves for the first three modes of a propped spring (prop at 85 per cent of the length) are shown in Fig. 1. As there indicated, several modes may be present together, resulting in a configuration which is at any instant the sum of the different modes present.

The deflection curves for the several modes, and the corresponding frequencies, depend upon boundary conditions determined by the way in which the spring is supported. For a propped cantilever, the boundary conditions, and hence the deflection curves and frequencies, vary with the prop location (defined by the ratio L'/L of Fig. 1). The special cases in which L'/L is zero and unity correspond respectively to a free cantilever and an end-propped cantilever. All the relations given in the figures are shown in the form of curves in which the quantity given is plotted against L'/L over the range from 0.5 to 1.0, which covers the prop locations applying to most contact springs.

The frequency equation for any particular prop location is transcendental in form (8). The successive roots of this equation determine the frequencies and deflection curves of the several modes. These roots do not form a simple series, and the successive frequencies are not simple multiples of the fundamental. In the higher modes, however, the deflection curves approach sine curves in form (except for the end sections), and the intervals between successive frequencies are approximately equal.

From the frequency constant for a particular mode and prop location there may be determined all the constants of the corresponding deflection curve except for an undetermined multiplier (A_1 in the equations of the Appendix), which measures the amplitude or energy content T of the mode in question. As this constant determines both the energy content and the maximum deflection (or amplitude) at any point on the beam,

* Equations are cited by the numbers identifying them in the Appendix.

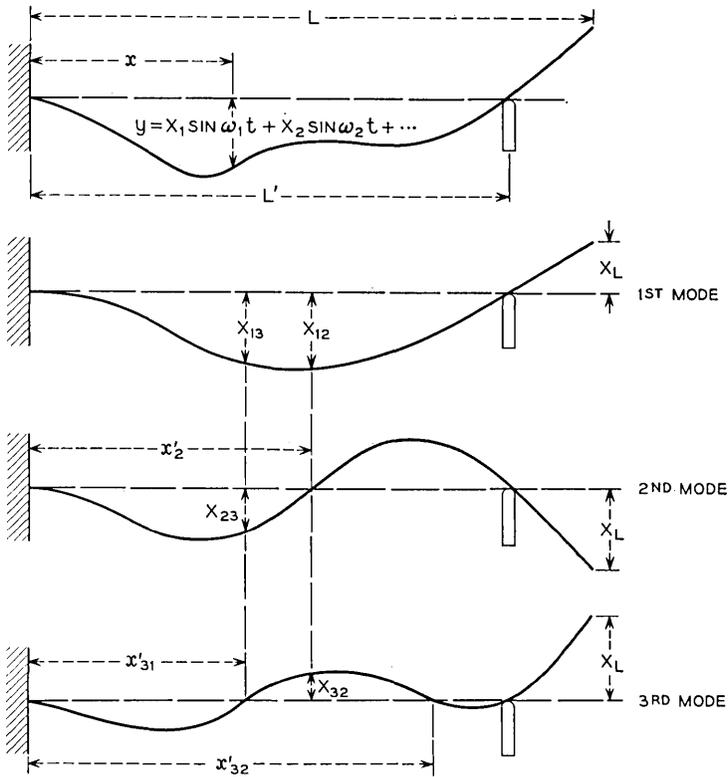


Fig. 1 — Flexural vibrations of a propped cantilever.

it may be eliminated from the equations to express the amplitude at specified points on the beam in terms of the energy content. Similarly, this amplitude constant may be eliminated from expressions for the force acting on the prop and for the longitudinal displacement there to give expressions for these quantities in terms of the energy content.

III. FREQUENCY RATIOS

The frequencies of the first five propped modes are shown in Fig. 2. These frequencies are given as multiples of f_0 , the frequency of the same beam as a free cantilever, and are shown plotted against the prop location as measured by L'/L .

The reference frequency f_0 is given by equation (12):

$$f_0 = 0.323 \sqrt{(s/m)},$$

where s is the static stiffness of the beam and m is its actual mass. Equivalent expressions are the following:

$$\text{For a circular section: } f_0 = \frac{0.1416d}{L^2} \sqrt{\frac{E}{\rho}},$$

$$\text{For a rectangular section: } f_0 = \frac{0.1636t}{L^2} \sqrt{\frac{E}{\rho}},$$

where d is the diameter of the circular section, t the thickness of the rectangular section, L is the length, and $\sqrt{E/\rho}$ is the velocity of sound in the material.

The frequencies given by these relations apply to springs of uniform cross-section. The added mass of the contact in relay springs reduces the frequency (except when propped at the contact). An approximate correction for the effect of the contact may be made by determining the effective mass m' of the spring for the mode in question by the procedure given in Section V. Then if m'' is the mass of the contact, the corrected frequency is the product of $\sqrt{m'/(m' + m'')}$ and the frequency read from Fig. 2.

IV. LOCATIONS OF NODES AND LOOPS

A node is a point of zero displacement (other than the prop location), while a loop is a point of maximum displacement. As illustrated in Fig. 1, the number of loops is the same as the order of the mode, while the number of nodes is one less than the order of the mode. Expressions for determining the locations of the nodes and loops are given in the Appendix.

Fig. 3 gives the locations of the nodes of the second and third modes.

V. RELATIONS OF AMPLITUDES TO ENERGY CONTENT

For any particular mode, the amplitude at any specified point on the spring is determined by the energy content T . Thus an estimate of T may be used to estimate the amplitude at some specified point, or the observed amplitude may be used to determine the energy content.

Fig. 4 gives the relation between the energy content T and the amplitude X_L of the free end of the spring, expressed as values of the ratio $m\omega^2 X_L^2/T$, where $\omega/(2\pi)$ is the frequency and m is the total mass of the spring. Even when a correction is made for contact mass in determining the frequency, $m\omega^2$ should be taken as the product of the mass

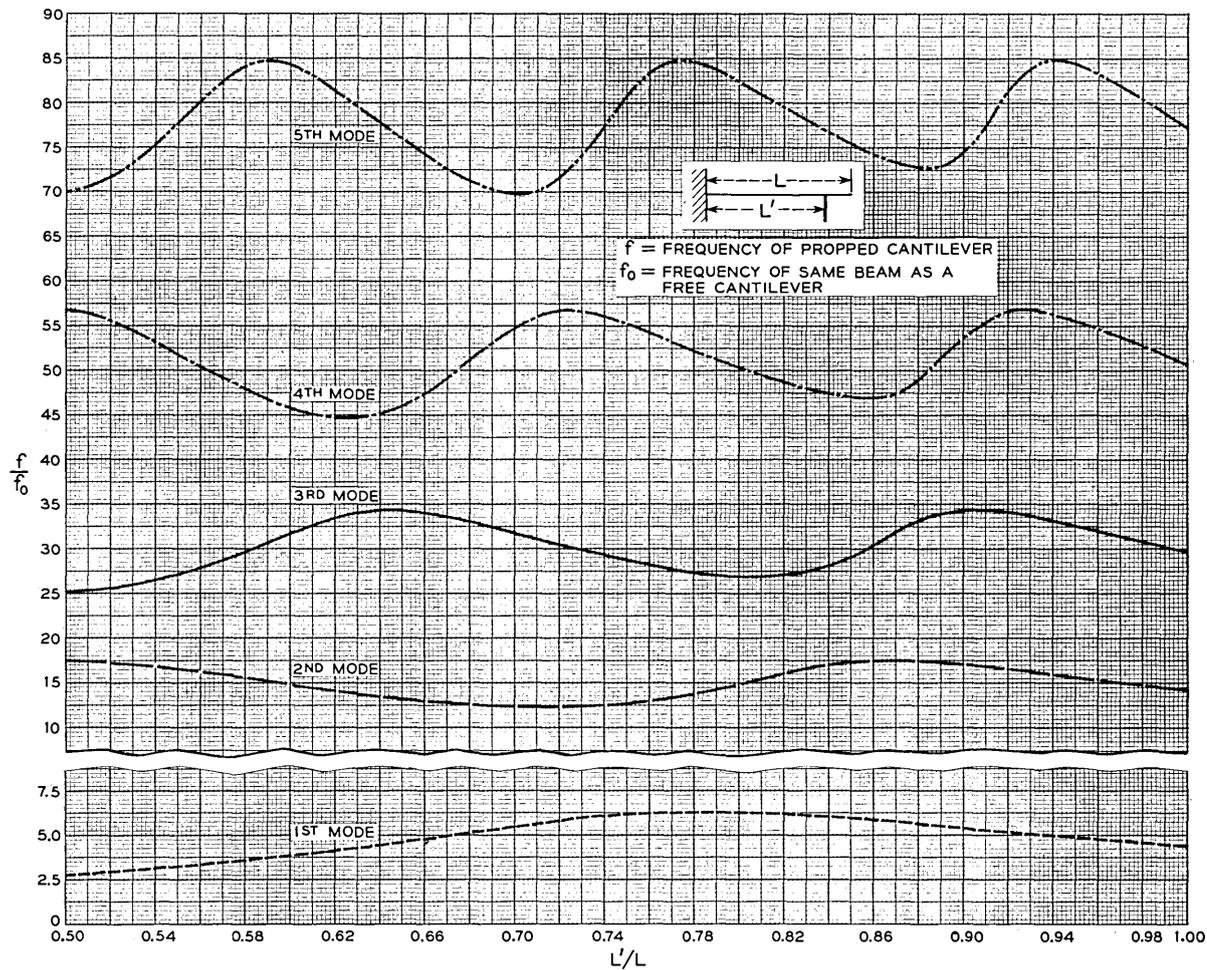


Fig. 2 — Frequency of a propped cantilever.

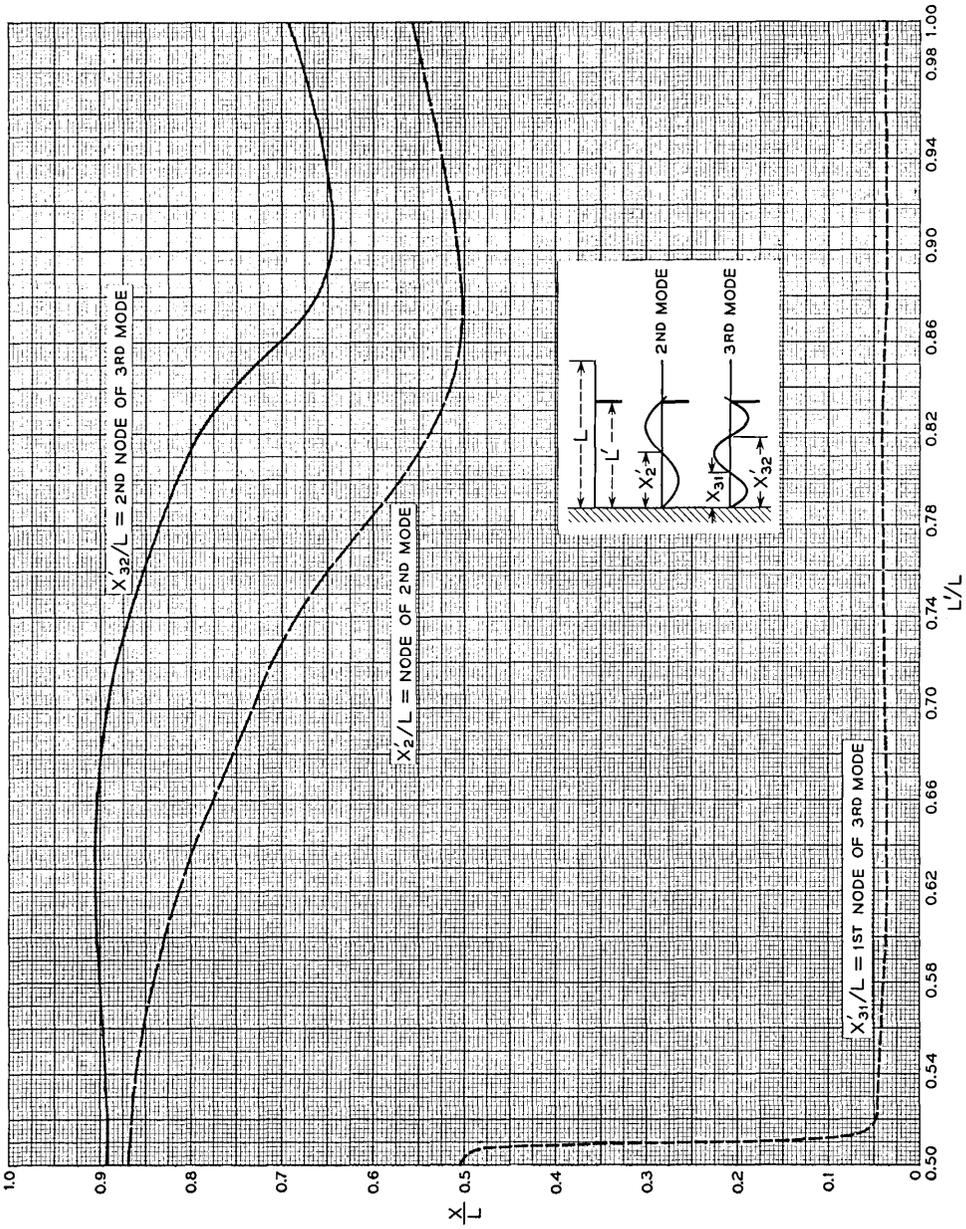


Fig. 3 — Propped cantilever: nodes of 2nd and 3rd modes.

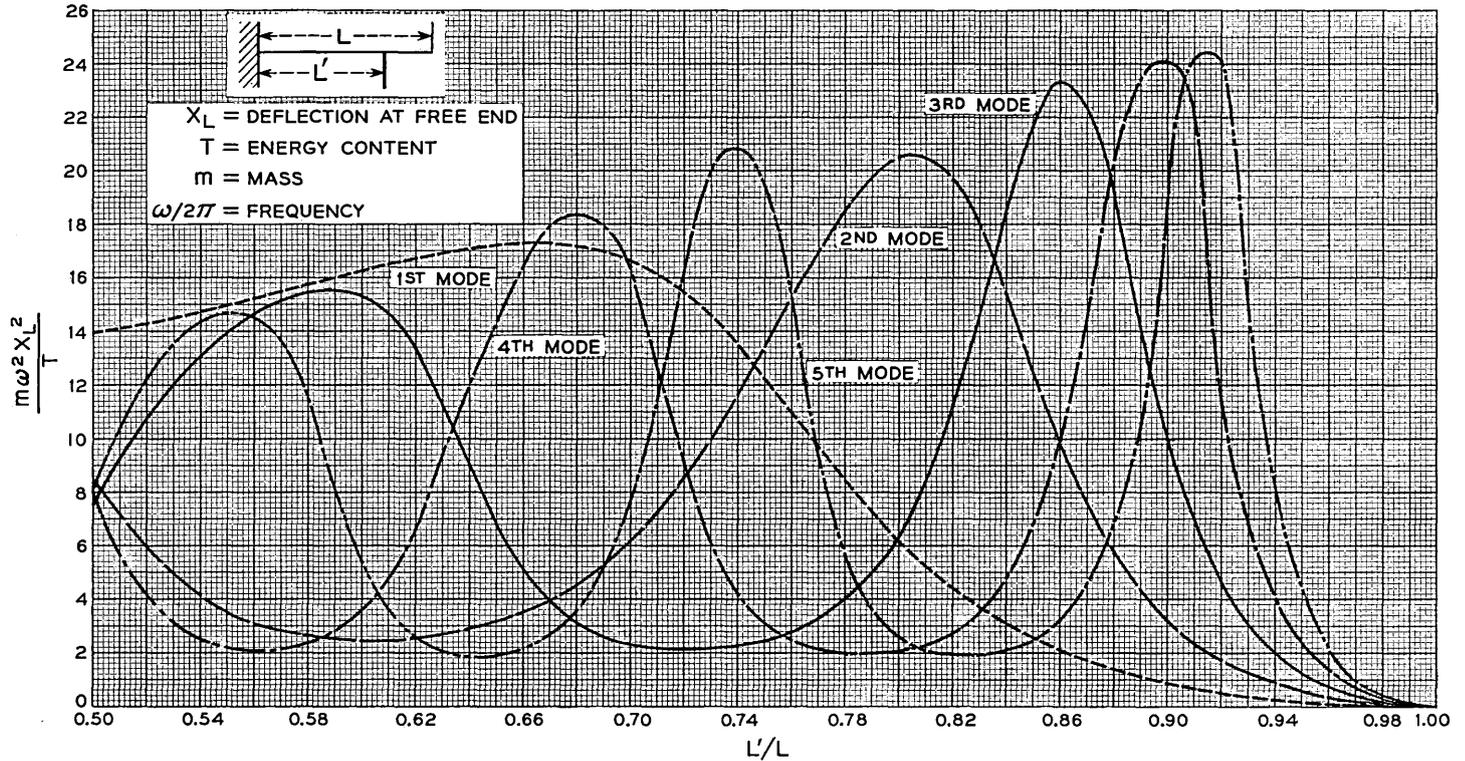


Fig. 4 — Propped cantilever: energy content in terms of end amplitude.

of the spring alone and the uncorrected frequency, without allowance for contact mass.

The effective mass m' of the spring, in terms of end motion, is the kinetic energy T divided by half the square of the end velocity ωX_L . Thus the ratio m'/m is twice the reciprocal of $m\omega^2 X_L^2/T$, given in Fig. 4, and values read from this figure may be used to evaluate m' .

These curves may be used to determine the energy content from observations of the end amplitude. When the prop is close to the free end, the end amplitude is smaller than that at or near the loops. When two or more modes are present, it is convenient to measure the amplitude at the node of one of the modes present. Values have been determined, therefore, of the ratios $m\omega_1^2 X_{12}/T$, $m\omega_1^2 X_{13}/T$, $m\omega_2^2 X_{23}/T$, and $m\omega_3^2 X_{32}/T$, which are given in Fig. 5. As shown in Fig. 1, X_{12} and X_{32} are the amplitudes of the first and third modes respectively at the node (x_2') of the second mode, while X_{13} and X_{23} are the amplitudes of the first and second modes respectively at the rear node (x_{31}') of the third mode. The locations of these nodes (x_2' and x_{31}') are given in Fig. 3.

When two or more modes are present and it is desired to determine the energy contents of the separate modes, the separate amplitudes must first be determined. This requires measuring the displacements at successive time intervals and using these successive displacements in a set of equations which can be solved for the amplitudes. If it can be assumed, for example, that only the first three modes are present, then the displacement X at a node (such as x_{31}') of the third mode is the sum of the first two modes, and is given by:

$$X = X_{13} \sin (\omega_1 t + \varphi_1) + X_{23} \sin (\omega_2 t + \varphi_2),$$

where φ_1 and φ_2 are the (unknown) phase angles of the two modes with respect to an arbitrary choice of the time origin. Let X_1 be the observed value of X at this selected time origin, and let X_2 , X_3 and X_4 be the observed values of X at the times at which $\omega_1 t$ is equal to $\pi/2$, π , and $3\pi/2$, respectively. On substituting these corresponding values of X and t in the preceding equation, there are obtained four equations in the four unknowns: $X_{13} \sin \varphi_1$, $X_{13} \cos \varphi_1$, $X_{23} \sin \varphi_2$, and $X_{23} \cos \varphi_2$. These four unknowns may be evaluated from the determinant D given by:

$$D = \begin{vmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & \sin a_1 & \cos a_1 \\ 0 & -1 & \sin a_2 & \cos a_2 \\ -1 & 0 & \sin a_3 & \cos a_3 \end{vmatrix}$$

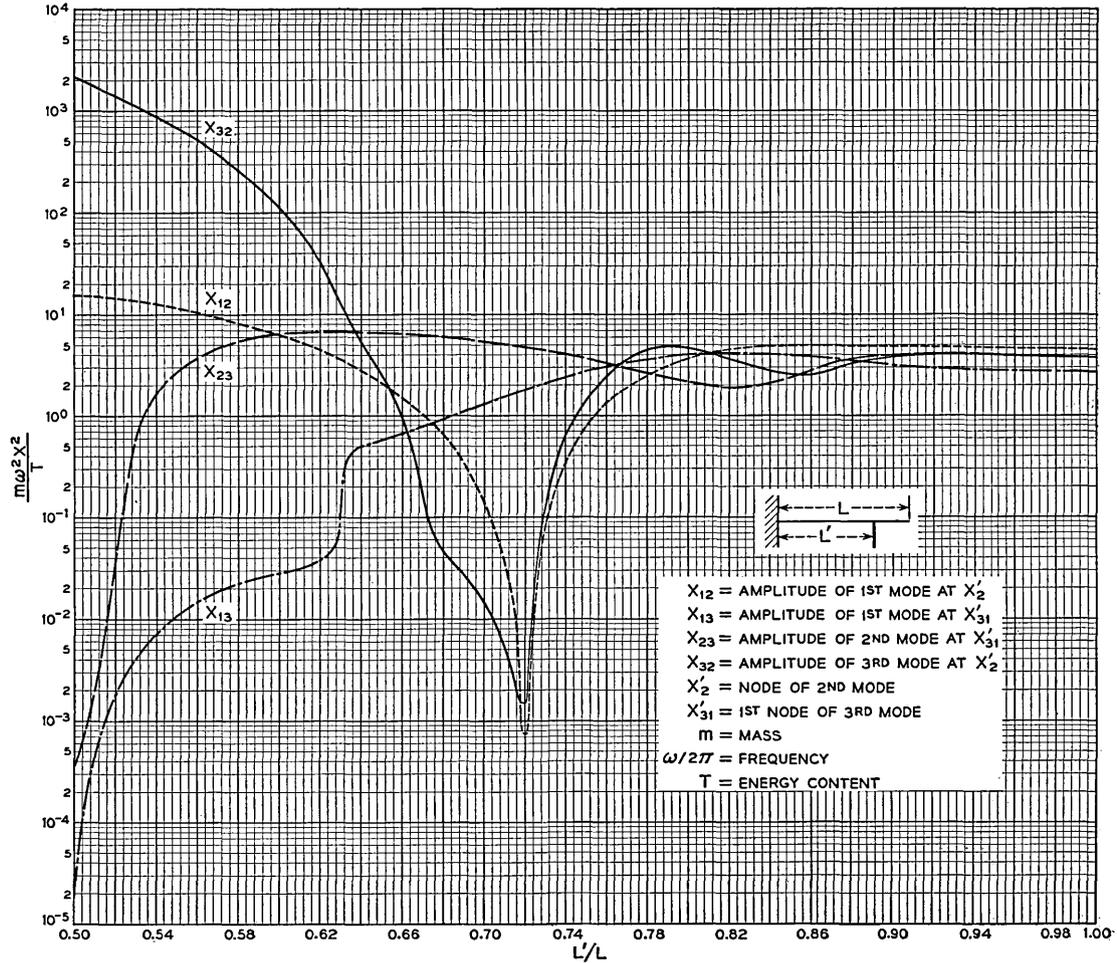


Fig. 5 — Propped cantilever: energy content in terms of amplitudes at specific locations.

where:
$$a_1 = \frac{f_2}{f_1} \cdot \frac{\pi}{2}, \quad a_2 = \frac{f_2}{f_1} \cdot \pi, \quad a_3 = \frac{f_2}{f_1} \cdot \frac{3\pi}{2}.$$

The ratio f_2/f_1 can be read from Fig. 2.

The same procedure may be employed to determine the values of X_{12} and X_{32} from observations of the deflections at the node x_2' of the second mode. (The same equations apply in this case, except that f_2/f_1 in the determinant terms is replaced by f_3/f_1 .) A check on the accuracy of the computations (or of the assumption that only the first three modes are present) is given by comparing the values for the energy content of the first mode obtained (by means of Fig. 5) from the values found for X_{12} and X_{13} : these values of T should be the same.

VI. FORCE AT PROP

A prop, or point of simple support, is taken as restraining the beam from deflection, without the application of any moment (or clamping action). Aside from the minor variation in the instantaneous point of support resulting from the finite dimensions of supporting surfaces, this condition is satisfied by the support actually provided when a spring is tensioned against a stud or contact. In general, vibration results in a force modulation $F' \sin \omega t$ corresponding to each mode present, where $\omega/(2\pi)$ is the frequency of that mode and F' is proportional to the square root of its energy content. As the sense of this force modulation varies with the phase of the mode, it alternately increases or decreases the total tension against the prop, which includes the static tension and the total force modulation of all modes present. The propped mode equations only apply rigorously when this total tension has the same sense as the static tension, as otherwise the prop is no longer effective and the spring moves away from it. (Practically, the effect of such separation may be ignored if it occurs only over a short interval of time.)

This force modulation is related to contact chatter, contact noise, and wear. When the spring is propped at the contact, an open results whenever the total tension becomes zero (or small enough to produce high contact resistance). Similarly, contact resistance variations resulting from force variations produce noise. Wear, whether at a contact or other support, such as a card, depends upon both the longitudinal motion and upon the normal force, or total tension.

If the energy content of a mode is estimated, or determined from amplitude observations, the amplitude of the force modulation for the first three modes may be determined from Fig. 6. This gives values of

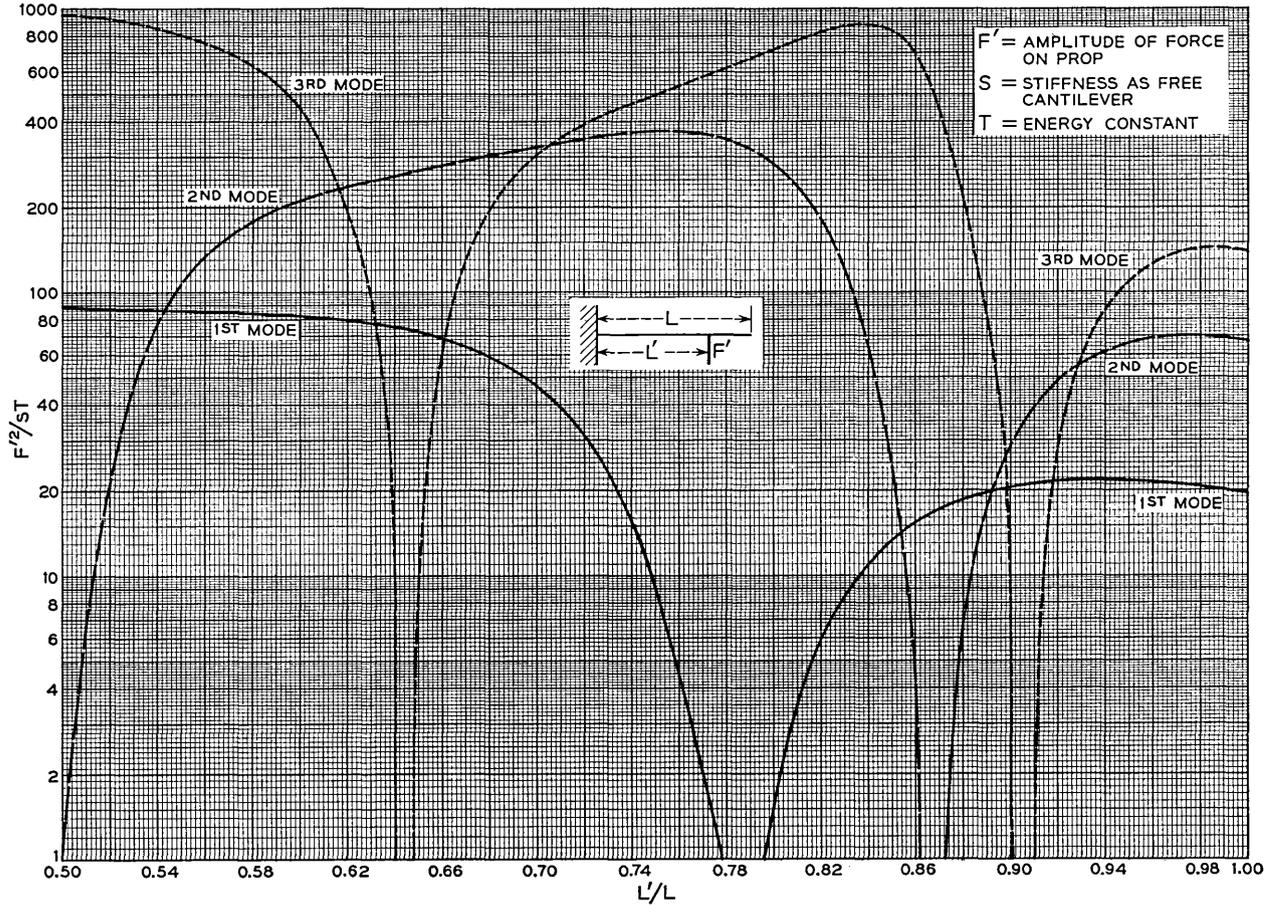


Fig. 6 — Propped cantilever: force on prop.

$F'^2/(sT)$, where T is the energy content of the mode in question and s is the static stiffness of the spring. The latter quantity may be measured directly, or computed as $3EI/L^3$, where E is Young's modulus for the spring material and I is the moment of inertia of its cross-section.

It will be seen from Fig. 6 that F'^2/sT varies greatly in magnitude with the prop location. Except for the fundamental, all modes have prop locations for which F' is zero. These correspond to the nodes of corresponding modes of a free (unpropped) cantilever, in which the spring vibrates without deflection at the prop without requiring any restraint, and hence without force modulation.

In general, and in particular for an end prop, as with a closed contact, F'^2/sT increases with the order of the mode. Hence a given energy content produces a greater force modulation, and is therefore more likely to produce chatter, the higher the order of the mode. (The open intervals, on the other hand, are necessarily shorter with the higher modes, because of the higher frequencies.) In relay spring vibration, the energy content of the fundamental is usually larger than that of the other modes, so that chatter commonly occurs at intervals equal to the period of the fundamental, but each interval may comprise a number of brief opens, corresponding to the shorter intervals in which higher modes are in phase with the fundamental. The fine chatter immediately following contact impact, however, corresponds wholly to higher modes, occurring at a time when the sense of the fundamental force modulation is the same as that of the static force.

VII. LONGITUDINAL COMPONENT OF VIBRATORY MOTION

In vibration, the deflected position of the spring defines a path from the clamp to the prop point which is necessarily longer than the distance between these points measured along the rest position of the spring. The difference between these two lengths represents a longitudinal component of the motion at the prop. This may be termed the vibratory slide, as distinguished from the slide resulting, for example, from motion of the prop point in actuation. The amplitude of this motion is of interest in connection with wear, particularly the wear of a card serving as a prop. The vibration characteristics affecting the wear are the amplitude of this vibratory slide, and the normal force on the prop, which varies with the force modulation discussed in the preceding section.

The longitudinal displacement is zero for the rest position of the spring and attains full amplitude, or maximum displacement, for full amplitude of the normal deflection in either sense. The longitudinal motion there-

fore has twice the frequency of the flexural vibration producing it. As shown in the Appendix, the longitudinal motion is harmonic, and as such has an amplitude $Z/2$ about a displacement $Z/2$ from the rest position, where Z is the full excursion, or double amplitude.

When a single mode is present alone, the longitudinal motion has only one component, with a frequency twice that of the mode producing it. Values of the ratio $m\omega^2L'Z/T$ have been determined for the first five modes, and are given in Fig. 7. It will be seen that the values of this ratio increase with the order of the mode. To obtain comparable values of Z for the same energy content T , however, these values of $m\omega^2L'Z/T$ must be divided by $(f/f_1)^2$, and when this is done it is found that the longitudinal displacement for the same energy content decreases with the order of the mode.

If two or more modes are present together, the longitudinal motion includes the motion that either would produce separately, and additional motion at frequencies, as shown in the Appendix, equal to the sums and differences of the frequencies of the modes present. Expressions for the amplitudes of the additional motions are included in the Appendix, and these were evaluated for the case where the first and second modes only are present. The additional displacement was found to be only five per cent of that produced by the first mode alone (for equal energy contents of the two modes). Thus the longitudinal motion when two or more modes are present differs little from the sum of the motions that each would produce separately for the same energy content.

For a given total energy content, therefore, the longitudinal amplitude is a maximum when only the fundamental mode is present. This, however, does not suffice to show that the wear is a maximum if all the energy is in the fundamental, rather than distributed among several modes. The wear also varies with the normal force, and it was shown in the preceding section that the force modulation for a given energy content is greater the higher the mode. It would therefore be necessary to know the relation of wear to both longitudinal motion and normal force to determine how the wear varies with the distribution of energy among the possible modes.

VIII. CONTACT WIPE

There is another type of longitudinal motion that occurs at a closed contact (end propped spring) because the contact surface is offset from the center line of the spring by a distance L'' . This results in a longitudinal motion $z' = Z' \sin \omega t$ at the contact surface, where $Z' =$

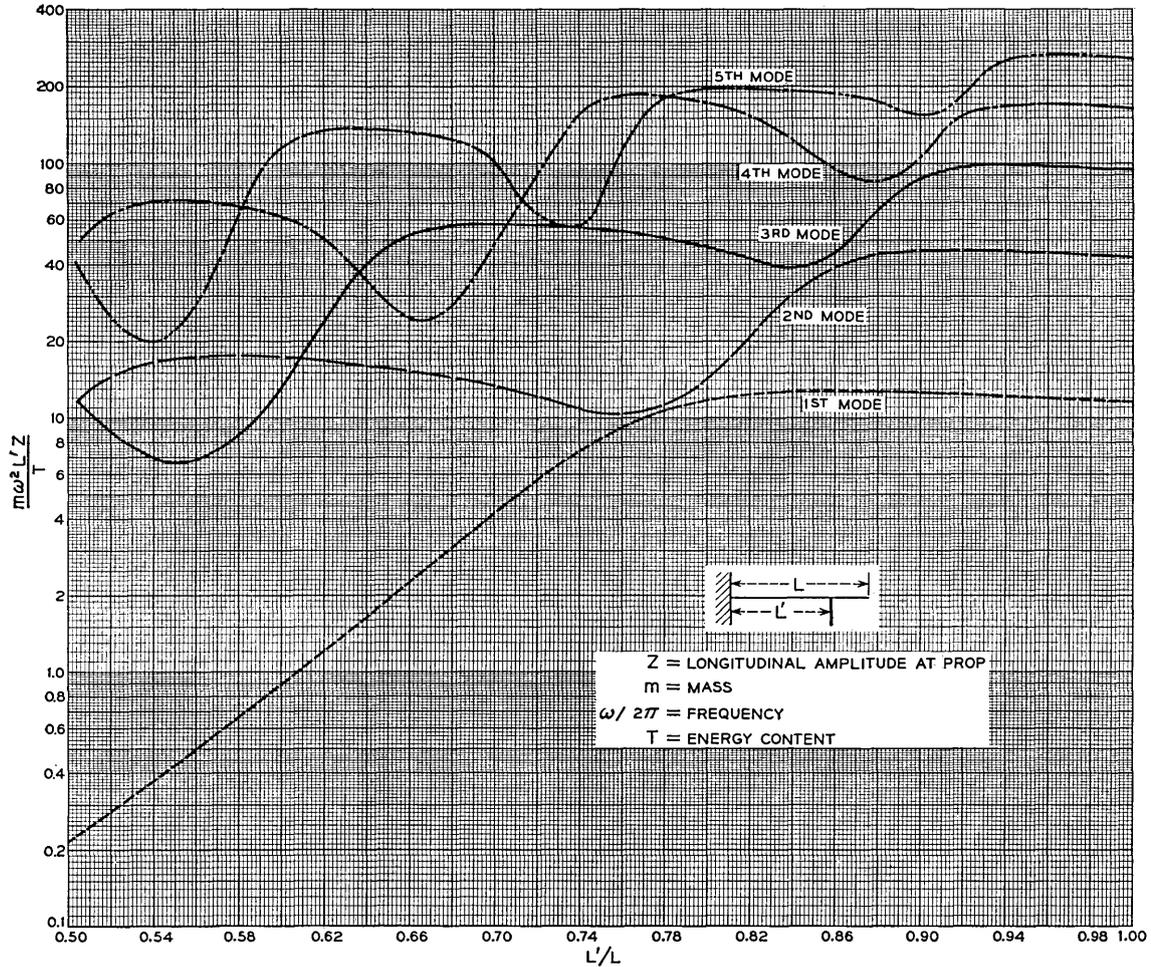


Fig. 7 — Propped cantilever: slide amplitude at prop.

$L''(dX/dx)_L$. This quantity can be evaluated from the table of the ratio $m\omega^2 L^2 (dX/dx)_L^2 / T$ given Section A.10 of the Appendix.

IX. DISCUSSION

The material contained in this paper has been prepared for reference purposes and for use in analytical or experimental studies, primarily those relating to contact springs (although it is of course applicable to any use of propped cantilever beams). Few conclusions of engineering interest may be drawn directly from this material; such value as it has must appear in its application. The possible use of the material may be indicated here by a brief discussion of its application to relay spring vibration.

Relay operation results in spring vibration, and such vibration may result in contact chatter and also in wear, particularly of such actuating members as studs or cards. The vibration amplitudes and frequencies (at least of the lower modes) are readily observed and measured, for example by the rapid record shadowgraph.² To reduce chatter and wear, information is required as to (a) the relations between the vibration characteristics and the relay design and conditions of actuation, and (b) the relations between chatter and wear and the vibration characteristics.

Such information may be obtained either by analysis or by direct experiment, but in either case the vibration characteristics are involved. In studying the excitation problem analytically, the amplitudes and frequencies must be determined from energy estimates, while an experimental study requires that the energy be evaluated from observed amplitudes.

The incidence of chatter can be determined directly from knowledge of the vibration, provided the force modulation is computed from the observed amplitudes by the relations given here. A similar analysis of wear would require knowledge of the dependence of wear (for particular materials) on both normal force and longitudinal displacement. Information as to these relations is incomplete, but if available their application would require the determination from the observed vibration of the resulting normal force variation and longitudinal displacement by means of the relations given here.

Because of the relatively large amplitudes associated with the fundamental mode, it is the most conspicuous feature of relay spring vibration. The relation given above between the normal force and the energy content shows that the force modulation for a given energy increases

with the order of the mode, indicating the considerable effect on chatter that may result from the presence of higher modes. The effect of the higher modes on wear is less well understood. The relations given here show that the diversion of energy from lower to higher modes increases the force modulation, but decreases the longitudinal displacement.

APPENDIX

Derivation of Equations

A.1 Equations of Flexural Vibration

As shown in such texts as Ref. 3, bending of a beam may be described in terms of the deflection y at a point located at a distance x , measured along the length of the beam from a clamp or other point of reference. Then dy/dx is the slope of the deflection curve, and, for the small deflections assumed in the simple theory of bending, the curvature of the neutral axis is given by d^2y/dx^2 . In this simple theory, the moment M at the point x is given by:

$$M = EI(d^2y/dx^2), \quad (1)$$

where E is Young's modulus and I is the moment of inertia of the beam's cross-section. The shearing force F is equal to dM/dx , and is therefore given by:

$$F = EI(d^3y/dx^3). \quad (2)$$

In motion, the inertia reaction of a differential element of length must equal the difference between the shearing forces at the ends of the element. Hence:

$$\frac{\partial^2 y}{\partial t^2} + \frac{EI}{\rho\alpha} \frac{\partial^4 y}{\partial x^4} = 0, \quad (3)$$

where ρ is the density of the beam and α is the area of its cross-section. Equation (3) is the general differential equation for flexural vibrations of beams of uniform section, assuming the simple theory of bending to apply. The general form of solution is given by:

$$y = X \sin(\omega t + k), \quad (4)$$

where X is a function of x only, the solution to the equation $d^4X/dx^4 = c^4X$, given by

$$X = A(\sin cx + B \cos cx + C \sinh cx + D \cosh cx), \quad (5)$$

in which c is given by:

$$c^4 = \omega^2 \rho \alpha / EI. \quad (6)$$

A.2 Modes of a Propped Cantilever

For a propped cantilever, x may be measured from the clamped end, as indicated in Fig. 1. Let L be the length of the beam, and L' the value of x for the prop. Subject to continuity, different forms of (5) apply at either side of the prop: let these be X_1 for $x < L'$, and X_2 for $x > L'$, and write (5) as:

$$X_1 = A_1 (\sin cx + B_1 \cos cx + C_1 \sinh cx + D_1 \cosh cx), \quad (5a)$$

$$X_2 = A_2 (\sin c(x - L') + B_2 \cos c(x - L') + C_2 \sinh c(x - L') + D_2 \cosh c(x - L')). \quad (5b)$$

Writing X' , X'' , X''' for the successive derivatives of X with respect to cx , the boundary conditions applying are as follows:

$$\text{For } x = 0, \quad X_1 = X_1' = 0,$$

$$\text{For } x = L', \quad X_1 = X_2 = 0, \quad X_1' = X_2', \quad X_1'' = X_2'', \quad (7)$$

$$\text{For } x = L, \quad X_2'' = X_2''' = 0.$$

Writing a for cL' , and b for $c(L - L')$, substitution of the expressions for X_1 and X_2 in the boundary conditions (7) gives the frequency equation:

$$\frac{\cos a \cdot \sinh a - \sin a \cdot \cosh a}{1 - \cos a \cdot \cosh a} = \frac{\cos b \cdot \sinh b - \sin b \cdot \cosh b}{1 + \cos b \cdot \cosh b}, \quad (8)$$

and the following expressions for the coefficients:

$$C_1 = -1,$$

$$B_1 = -D_1 = -\frac{\sin a - \sinh a}{\cos a - \cosh a},$$

$$C_2 = \frac{1 + \cos b \cdot \cosh b - \sin b \cdot \sinh b}{1 + \cos b \cdot \cosh b + \sin b \cdot \sinh b}, \quad (9)$$

$$-B_2 = D_2 = \frac{\sin b \cdot \cosh b - \cos b \cdot \sinh b}{1 + \cos b \cdot \cosh b + \sin b \cdot \sinh b},$$

$$\frac{A_1}{A_2} = \frac{\cos a - \cosh a}{1 - \cos a \cdot \cosh a} \cdot \frac{1 + \cos b \cdot \cosh b}{1 + \cos b \cdot \cosh b + \sin b \cdot \sinh b}.$$

The ratio b/a ($= L/L' - 1$) is determined by the prop location. Then the values of a which satisfy (8) determine values of c ($= a/L'$), which correspond to values of the frequency $\omega/(2\pi)$ given by (6). The successive frequencies thus determined are those of successive modes of vibration. Substitution in (9) of the value of a for any particular mode determines the coefficients of the expression for the corresponding deflection curve. The one remaining coefficient, A_1 or A_2 , measures the amplitude.

A.3 The Frequency Equation

As the mass m of the beam is equal to $\rho\alpha L$, (6) may be written in the form:

$$\omega^2 = (s/3m)(cL)^4, \quad (10)$$

where s is the static stiffness of the beam as a free cantilever, or $3EI/L^3$. As $a = cL'$, the values of a satisfying (8) for a given prop location, or value of L'/L , determine corresponding values of cL , and hence corresponding values of the frequency $\omega/(2\pi)$.

The special case in which $L' = 0$, or $a = 0$, $b = cL$, is that of a simple (unpropped) cantilever, and (8) then reduces to:

$$1 + \cos b \cdot \cosh b = 0. \quad (11)$$

The first three roots of this equation give values of b^2 , or $(cL)^2$, of 3.52, 22.0 and 61.8. For the higher roots, a good approximation to b is given by $\pi(n - \frac{1}{2})$, where n is an integer. From (10), the frequency f_0 of the fundamental cantilever mode is given by:

$$f_0 = 0.323 \sqrt{(s/m)}, \quad (12)$$

and the frequency of any other mode is given in terms of f_0 by:

$$\frac{f}{f_0} = (cL)^2/3.52 \quad (13)$$

For various values of L'/L , (8) has been solved numerically to determine the values of a and thus of cL for the first five modes. By means of (13), the resulting values of f/f_0 have been determined, and are plotted in Fig. 2.

Another special case of interest is that of an end prop, for which $L' = L$, or $a = cL$, $b = 0$. In this case, (8) reduces to:

$$\tan a = \tanh a. \quad (14)$$

The values of a (or cL) satisfying this equation are given approximately by $\pi(n + \frac{1}{4})$, where n is an integer. The value of $(cL)^2$ for the fundamental end propped mode is 15.50, giving a frequency 4.40 times f_0 , the frequency of the fundamental cantilever mode.

A.4 Nodes and Loops of Deflection Curves

On substituting in (5) the expressions for the coefficients for X_1 given by (9), the deflection curve for $x < L'$ is given by:

$$X_1/A_1 = [f_1(a) - f_1(u)] (\cosh u - \cos u), \quad (15)$$

where $u = cx$, and $f_1(u)$ is given by:

$$f_1(u) = \frac{\sin u - \sinh u}{\cos u - \cosh u}. \quad (16)$$

Thus the nodes (points of zero deflection) lying between the clamp and the prop ($x < L'$) occur at those values of u for which $f_1(u) = f_1(a)$. These values of u may thus be determined for any mode and prop location from (16) and the corresponding value of a . As $u/a = x/L'$, there may thus be determined the values of the node locations x'/L lying between the clamp and prop. The locations of the nodes for the second and third modes lying in this region are plotted against L'/L in Fig. 3.

The loops (points of maximum deflection) of the deflection curve occur at those values of u for which $dX/du = 0$. By differentiation of (15), it is found that these values of u for $x < L'$ are those for which $f_2(u) = f_1(a)$, where $f_2(u)$ is given by:

$$f_2(u) = \frac{\cosh u - \cos u}{\sinh u + \sin u}. \quad (17)$$

For any mode and prop location and the corresponding value of a there can be determined those values of u for which $f_2(u) = f_1(a)$. From these can be determined the corresponding loop locations x/L lying between the clamp and plot.

Similarly, from (5) and the coefficients of X_2 given by (9), the node locations lying beyond the prop, $x > L'$, are given by:

$$0 = B_3 \sin(u - a) + B_4 \sinh(u - a) \\ - B_5 [\cos(u - a) - \cosh(u - a)],$$

where:

$$B_3 = 1 + \cos b \cosh b + \sin b \sinh b,$$

$$B_4 = 1 + \cos b \cosh b - \sin b \sinh b,$$

$$B_5 = \sin b \cosh b - \cos b \sinh b.$$

The node locations in Fig. 3 lying beyond the prop have been determined from these equations.

Similarly, the loop locations lying beyond the prop are given by:

$$0 = B_3 \cos(u - a) + B_4 \cosh(u - a) + B_5 [\sin(u - a) + \sinh(u - a)].$$

A.5 Free-End Deflection

On substituting in (5) the expressions for the coefficients given by (9), it is found that the free-end deflection X_L , the value of X_2 for $x = L$, is given by:

$$\frac{X_L}{A_2} = \frac{2(\sin b + \sinh b)}{1 + \cos b \cdot \cosh b + \sin b \cdot \sinh b}, \tag{18}$$

or by:

$$\frac{X_L}{A_1} = \frac{2(\sin b + \sinh b)(1 - \cos a \cdot \cosh a)}{(1 + \cos b \cdot \cosh b)(\cos a - \cosh a)}. \tag{19}$$

A.6 Energy Content

The energy content of a vibrating beam is the integral over the length of $\rho\alpha\dot{y}^2 \cdot dx/2$, where \dot{y} , or dy/dt , is the maximum velocity (occurring at zero deflection, when all the energy is kinetic). Then, from (4), the energy content T is given by:

$$2T = \omega^2 \rho\alpha \int_0^L X^2 \cdot dx. \tag{20}$$

For the propped beam, the integral

$$\int_0^L X^2 \cdot dx$$

is given by:

$$\int_0^L X_1^2 \cdot dx + \int_{L'}^L X_2^2 \cdot dx.$$

As shown in Ref. 3 these two component integrals are given by:

$$\begin{aligned}
4c \int_0^{L'} X_1^2 \cdot dx &= \int_0^{L'} [3X_1 X_1''' - X_1' X_1'' \\
&\quad + cx(X_1^2 - 2X_1' X_1''' + X_1''^2)] \\
4c \int_{L'}^L X_2^2 \cdot dx &= \int_{L'}^L [3X_2 X_2''' - X_2' X_2'' \\
&\quad + c(x - L')(X_2^2 - 2X_2' X_2''' + X_2''^2)].
\end{aligned} \tag{21}$$

On substituting the boundary conditions given by (7), the expression for the energy content reduces to:

$$T = \frac{\omega^2 \rho \alpha}{8} [(L - L')X_L^2 + L'(X_L''^2 - 2X_L' X_L''')], \tag{22}$$

where $X_{L'}$, $X_{L'}''$ and $X_{L'}'''$ are the first three derivatives of X_1 with respect to cx at $x = L'$, and X_L is, as before, the value of X_2 at $x = L$. Expressions for the derivatives can be obtained by differentiation of (15), from which it is found that the second term in the bracket of (22) is equal to $4L'A_1^2 f_1^2(a)$. Then (22) reduces to:

$$\frac{T}{m\omega^2 X_L^2} = \frac{1}{8} \left[1 - \frac{L'}{L} + 4 \frac{L'}{L} \left(\frac{A_1}{X_L} \right)^2 f_1^2(a) \right], \tag{23}$$

where X_L/A_1 is given by (19), and $f_1(a)$ by (16). Values of $m\omega^2 X_L^2/T$ for the first five modes have been computed from (23) and are shown plotted against L'/L in Fig. 4.

As the effective mass m' in terms of motion at the free end is defined as the kinetic energy T divided by half the square of the end velocity, or $(\omega X_L)^2/2$, the quantity given by (23) is one half the ratio of the effective mass of the beam to its actual mass, or $m'/(2m)$.

A.7 Energy Content in Terms of Amplitude at Nodes of Other Modes

Let X_{12} be the deflection in the first mode at $x = x_2'$, the location of the node of the second mode. Using the value of a applying to the first mode (for a particular prop location L'/L), and taking $u = ax_2'/L'$, (15) may be used to determine X_{12}/A_1 . From this and the value of $T/(m\omega^2 A_1^2)$ given by (23) there may be determined the value of $m\omega^2 X_{12}^2/T$. Such values have been determined for various values of L'/L , and are plotted against L'/L in Fig. 5.

The same procedure has been used to determine the values of $m\omega^2 X_{13}^2/T$, $m\omega^2 X_{23}^2/T$ and $m\omega^2 X_{32}^2/T$ given in Fig. 5. X_{32} is the amplitude of the third mode at x_2' , while X_{13} and X_{23} are the amplitudes of

the first and second modes respectively at x_{31}' , the first (rear) node of the third mode.

A.8 Force Modulation at the Prop

The shearing force F on a section of the beam is given by (2), and its maximum value, or force amplitude, is therefore given by $c^3EI \cdot X'''$ (where, as before, $X''' = d^3X/du^3$). The discontinuity at the prop results in a difference between X_1''' and X_2''' corresponding to a force $F' \sin(\omega t + k)$ acting on the prop, where F' is given by:

$$F' = c^3EI(X_1''' - X_2''')_{(x=L')} \tag{24}$$

The force $F' \sin(\omega t + k)$ is, of course, additive to any static tension acting at the prop.

Expressions for X_1''' and X_2''' at $x = L'$ may be obtained by differentiation of (5) and substitution of the expressions for the coefficients given by (9), giving:

$$\frac{F'}{A_1c^3EI} = f_3(b) \frac{X_L}{A_1} - 2 \frac{f_3(a)}{f_2(a)}, \tag{25}$$

where $f_3(u)$ is given by:

$$f_3(u) = \frac{\sin u \cdot \sinh u}{\sin u + \sinh u}. \tag{26}$$

The quantity given by the left side of (25) may be squared and simplified as follows:

$$\left[\frac{F'}{A_1c^3EI} \right]^2 = \frac{L^3F'^2}{A_1^2EI(cL)^2} \cdot \frac{1}{EIc^4L} = \frac{3F'^2}{sA_1^2(cL)^2\omega^2m},$$

where, from (6), $\omega^2\rho\alpha$ is substituted for c^4EI , m is substituted for $\rho\alpha L$, and s for $3EI/L^3$. As before, m is the mass of the spring and s is its stiffness as a free cantilever.

It follows that $F'^2/(sT)$ is given by:

$$\frac{F'^2}{sT} = \frac{(cL)^2}{3} \cdot \left(\frac{F'}{A_1c^3EI} \right)^2 \cdot \frac{m\omega^2A_1^2}{T}. \tag{27}$$

Using (23) and (25) to evaluate the second and third terms on the right-hand side of (27), the latter equation has been used to determine values of $F'^2/(sT)$ for the first three modes for various prop locations (L'/L). The results are plotted in Fig. 6.

A.9 Longitudinal Component of Motion at the Prop

The longitudinal component of the motion at the prop is the displacement z in the direction of the spring length. This is equal to the difference between the distance from the clamp to the prop measured along the displacement curve and that measured along the rest position. As the latter distance is L' , z is given by:

$$z = L' - \int_0^{L'} \sqrt{1 + \left(\frac{dy}{dx}\right)^2} \cdot dx.$$

Neglecting quantities of smaller order, this reduces to:

$$z = \frac{1}{2} \int_0^{L'} \left(\frac{dy}{dx}\right)^2 \cdot dx. \quad (28)$$

If only one mode of vibration is present, $y = X \cdot \sin(\omega t)$, where $\omega/(2\pi)$ is the frequency of this mode. Substitution in (28) gives the following expression for z :

$$z = Z \sin^2(\omega t), \quad (29)$$

where Z is given by:

$$Z = \frac{1}{2} \int_0^{L'} \left(\frac{dX}{dx}\right)^2 \cdot dx. \quad (30)$$

Equation (29) shows that the maximum longitudinal displacement is Z . As $\sin^2(\omega t) = (1 - \cos(2\omega t))/2$, the longitudinal motion is an harmonic motion of frequency $2\omega/(2\pi)$ and amplitude $Z/2$ about a displacement $Z/2$ from the rest position. In other words, Z is the double amplitude, or full range of the longitudinal motion.

If two modes of vibration of frequencies $\omega_m/(2\pi)$ and $\omega_n/(2\pi)$ are present together, the normal displacement y is given by:

$$y = X_m \sin(\omega_m t) + X_n \sin(\omega_n t).$$

In this case, substitution in (28) shows that the longitudinal displacement z is given by:

$$z = Z_m \sin^2(\omega_m t) + Z_n \sin^2(\omega_n t) + Z_{mn} \sin(\omega_m t) \cdot \sin(\omega_n t), \quad (31)$$

where Z_m and Z_n are given by (29) for $X = X_m$ and $X = X_n$, respectively, and Z_{mn} is given by:

$$Z_{mn} = \int_0^{L'} \frac{(dX_m)}{(dx)} \frac{(dX_n)}{(dx)} dx. \quad (32)$$

Equation (31) shows that in this case the longitudinal displacement is

the sum of (i) the displacements that would be produced by each mode if it were present alone and (ii) an additional displacement having a maximum value of Z_{mn} . From the relation:

$$\sin(\omega_m t) \cdot \sin(\omega_n t) = \frac{1}{2} [\cos(\omega_m - \omega_n)t + \cos(\omega_m + \omega_n)t],$$

it follows that this additional displacement corresponds to the sum of two motions, both of amplitude $Z_{mn}/2$, having frequencies equal to the sum and difference of the frequencies of the two modes present.

In the same way, it can be shown that if there are more than two modes present, the displacement will include the displacements produced by the separate modes, and additional displacements having maximum values Z_{mn} corresponding to each pair of modes present.

To evaluate the maximum longitudinal displacements, it is necessary to obtain expressions for Z from (30) and Z_{mn} from (32). This may be done by a procedure paralleling that used by Timoshenko (loc. cit.) to derive (21). As $c^4 x = d^4 X/dx^4$ for any mode, $c_n^4 dX_m/dx = d^5 X_m/dx^5$ and $c_n^4 dX_n/dx = d^5 X_n/dx^5$. By cross multiplication and subtraction there is obtained:

$$\begin{aligned} (c_n^4 - c_m^4) \int_0^{L'} \frac{dX_m}{dx} \frac{dX_n}{dx} dx &= \int_0^{L'} \left(\frac{dX_m}{dx} \frac{d^5 X_n}{dx^5} - \frac{dX_n}{dx} \frac{d^5 X_m}{dx^5} \right) \cdot dx \quad (33) \\ &= \int_0^{L'} \left[\frac{dX_m}{dx} \frac{d^4 X_n}{dx^4} - \frac{dX_n}{dx} \frac{d^4 X_m}{dx^4} - \frac{d^2 X_m}{dx^2} \frac{d^3 X_n}{dx^3} + \frac{d^2 X_n}{dx^2} \frac{d^3 X_m}{dx^3} \right], \end{aligned}$$

where the second equation is obtained by integration by parts.

In this equation, X_m and X_n are the expressions for the deflection curve X_1 for the portion of the beam between the clamp and the prop location, corresponding to the modes of frequencies $\omega_m/(2\pi)$ and $\omega_n/(2\pi)$. Equation (33) may be used to provide an expression for Z_{mn} by substituting the boundary conditions (7), which eliminate the first two terms, and by using the derivatives of (15) to express the remaining terms. There is thus obtained:

$$\begin{aligned} \frac{L' Z_{mn}}{A_m A_n} = \frac{4\omega_m \omega_n}{\omega_n^2 - \omega_m^2} \left\{ a_n \left[\frac{f_4(a_m) f_3(a_n)}{f_2(a_n)} - f_1(a_m) \right] \right. \\ \left. - a_m \left[\frac{f_4(a_n) f_3(a_m)}{f_2(a_m)} - f_1(a_n) \right] \right\}, \quad (34) \end{aligned}$$

where $f_1(u)$, $f_2(u)$ and $f_3(u)$ are given by (16), (17), and (26), and

$f_4(u)$ is given by:

$$f_4(u) = \frac{\sin u \cdot \cosh u - \cos u \cdot \sinh u}{\cos u - \cosh u}. \quad (35)$$

An expression for Z , the integral of (30), may be obtained from (33) by considering X as a function of c and letting $c_n - c_m$ equal δc , where δc is a small quantity. Then $c_n^4 - c_m^4 = 4c_n^3 \cdot \delta c$, neglecting quantities of smaller order, and similarly, $X_n = X_m + (dX_m/dc) \cdot \delta c$. On making these substitutions in (33) and neglecting quantities of higher order of δc than the first, this equation reduces to:

$$4 \int_0^{L'} \left(\frac{dX}{dx} \right)^2 dx = L'_0 [3cXX' + c^2xX'^2 - 2c^2xXX'' - cX''X''' + c^2xX''''],$$

where the subscripts n and m have been dropped because, for δc negligible, c and X apply to a single mode.

As before, substitution in this of the boundary conditions (7), and of the derivatives of (15), gives finally:

$$\frac{L'Z}{A_1^2} = \frac{a}{2} \left[a + \frac{f_3(a) \cdot f_4(a)}{f_2(a)} - f_1(a) \right], \quad (36)$$

where Z is the (double) amplitude of the longitudinal motion for a single mode. By evaluating the right-hand side of (36) for particular modes and values of L'/L , and dividing these by corresponding values of $T/(m\omega^2A_1^2)$ given by (23), values are obtained of the ratio $m\omega^2L'Z/T$. Values of this ratio, determined in this way, are shown plotted in Fig. 7 against L'/L for the first five modes.

The ratios given in Fig. 7 may be used to determine the maximum longitudinal displacement for a given energy content when only one mode is present. When two or more modes are present, the maximum displacement is the sum of the displacements for the individual modes and the additional term or terms Z_{mn} . The magnitudes of these additional terms depend upon the division of the kinetic energy among the modes involved. For the case of two modes present together, the one additional term present, is, from (34), proportional to $A_m A_n$, and therefore to the square root of $T_m T_n$, the corresponding energies, whose sum $T_m + T_n$ equals the total kinetic energy T . It is easily shown that for a given value of T , $T_m T_n$, and therefore $A_m A_n$, is a maximum for $T_m = T_n = T/2$. In this case,

$$\frac{m\omega_m^2 A_m A_n}{T} = \frac{\omega_m}{2\omega_n} \sqrt{\left(\frac{m\omega^2 A^2}{T} \right)_m \left(\frac{m\omega^2 A^2}{T} \right)_n}.$$

By evaluating the right-hand side of this equation [by means of (23)], and multiplying the result into the corresponding value of the right-hand side of (34), values may be obtained of the ratio $m\omega_m^2 L' Z_{mn} / T$. This was done for the case where $m = 1$ and $n = 2$, where the first two modes are present. The resulting values of $m\omega_1^2 L' Z_{12} / T$ are directly comparable with those of $m\omega_1^2 L' Z / T$ for the first mode alone, shown in Fig. 7. The values for $m\omega_1^2 L' Z_{12} / T$ were all less than 3 as compared with values for $m\omega_1^2 L' Z / T$ of about 50. It follows that the additional displacement resulting from the cross product term is minor.

A.10 *Angular End Displacement*

To estimate contact wipe for a spring propped at the contact, there is required an expression for the angular displacement at the prop, when this is at the end of the spring. This is given by:

$$\left(\frac{dX}{dx}\right)_L = cX_L'$$

As in this case, $L' = L$, and $a = cL$, this equation can, by differentiation of (15), be expressed as follows:

$$\frac{L}{A_1} \left(\frac{dX}{dx}\right)_L = \frac{2a(1 - \cos a \cdot \cosh a)}{\cos a - \cosh a} \tag{37}$$

Values of the right-hand side of this expression have been determined for the first five modes. By squaring these, and dividing them by the corresponding values of $T / (m\omega^2 A_1^2)$, there have been obtained the following values of $m\omega^2 L^2 (dX/dx)_L^2 / T$.

<i>Mode</i>	$\frac{m\omega^2 L^2}{T} \left(\frac{dX}{dx}\right)_L^2$
1	16.416
2	49.700
3	103.68
4	178.60
5	271.44

A.11 *Use of Equations*

Many of the relations given here have been expressed in numerical form, and are shown in the figures. If additional relations are required, they may be computed from the equations given in this Appendix (or from expressions derived from them).

If such computation is required, it should be noted that all relations

involved in flexural vibrations of the type considered here can be expressed as dimensionless ratios which are, directly or indirectly, dependent on the roots of the frequency equation, and hence on the appropriate values of a , a pure number, equal to cL' . The similar number b , appearing in expressions relating to the part of the beam between the prop and the free end, is equal to $c(L - L')$, and therefore equals $a(L/L' - 1)$. The frequency, $\omega/(2\pi)$, is related to a by (6) (of which (10) is an alternate form), and the frequency ratios are therefore functions of a only. Similarly, such ratios as $F'/(sT)$, $m\omega^2 X_L^2/T$, and $m\omega^2 L'Z/T$ discussed above are all functions of a only.

Values of a have been determined for the first five modes for values of L'/L in the range from 0.5 to 1.0. To use the equations in this range and for these modes, these values of a may be easily obtained from Fig. 2, as the values of f/f_0 given there are [from (13)] equal to $(cL)^2/3.52$, so they may be used to evaluate cL , and hence $a (= cL')$.

For values of L'/L outside this range, or for modes of higher order, values of a must be determined by solution of (8) for the case in question.

REFERENCES

1. Peek, R. L., Jr., and Wagar, H. N., *Switching Relay Design*, Van Nostrand, 1955.
2. Wagar, H. N., Relay Measuring Equipment, B.S.T.J., **33**, January, 1954, p. 3.
3. Timoshenko, S., *Vibration Problems in Engineering*, 3rd edition, Van Nostrand, 1955.

Numerical Computation of Phase from Amplitude at Optical Frequencies

By D. E. THOMAS

(Manuscript received January 31, 1963)

This paper presents phase tables for use in determining phase from amplitude or vice versa at optical and higher frequencies. The new tables, combined with tables previously published by the author, are believed to make possible the determination of phase from amplitude or vice versa of any minimum phase function occurring in any area of the physical sciences, and at any place in the frequency spectrum. The phase is determined by a summation process based on Bode's straight-line approximation method. The paper gives a brief historical background of the method, discusses the application of the numerical phase summation technique to optical and higher frequencies, describes the derivation of new tabulations useful at these frequencies, and gives quantitative examples of their use. A table expanding the existing tables of phase of a semi-infinite unit slope near f/f_0 equal to one is given. Additional tables of phase of a unit line segment and a new straight-line element, the unit wedge, are given. Finally, there is a brief discussion of the potential of the method in solving physical and engineering problems.

I. INTRODUCTION

The fact that nature ties the real and imaginary components of a complex variable function of frequency inextricably together, when the variable represents some physically real quantity or phenomena, has been recognized to varying degrees for nearly half a century. For example, Kramers¹ in 1927 noted the general relations between the refractive index and absorption resulting from the simple relationships to the real and imaginary parts of a complex dielectric constant. Because one of the relations was contained in an earlier paper of Kronig's,² this relationship is commonly known in the physical science world as the Kramers-Kronig relation. The awareness of the relationship between

the real and imaginary parts of the impedance of an electrical network emerged about the same time as the Kramers-Kronig work.^{3,4}

The usefulness of a quantitative solution to the general real and imaginary component relationship was soon recognized. Bode has provided us with a key to the solution of this problem (Ref. 5, Ch. XIV). He gives general integral equations relating the two components, but points out what many have since discovered, namely, that the general integrals can be readily evaluated for only the simplest of functions. Bode, however, presents a practical numerical integration technique for summing the imaginary component associated with a multiple straight-line approximation to the real component as a function of frequency (Ref. 5, Ch. XV). The accuracy of this summation is limited only by the number of straight lines used to approximate the true real component and the accuracy to which the imaginary contribution of each of the straight line segments to the total imaginary component is known. The author has published tables for accomplishing this summation at telecommunication and radio frequencies.⁵ These tables made possible the computation of the nonlinear phase from which the delay distortion (dispersion) to be expected in a projected transatlantic repeatered submarine telephone cable system was determined and the delay distortion correction required to make existing coaxial cable systems suitable for the transmission of television programs. Van Vleck utilized the Kramers-Kronig relation while studying microwave propagation during World War II to establish that a significant difference in the refractive index of the atmosphere between wavelengths of 3 cm and 1 cm would lead to an unreasonably high absorption, in contradiction with experiment.⁷

The invention of the optical maser and the availability of coherent light directed attention to the possibility of transmission of intelligence beyond the microwave frequencies to the optical frequencies. The realization of the potential usefulness of the numerical imaginary component summation at optical frequencies resulted from a discussion initiated by a colleague, W. L. Faust. This discussion concerned a proposal by Miller and Lopez⁸ that the difference in determination of the velocity of light obtained from measurements at optical frequencies and at microwave frequencies could be explained by the difference in delay time experienced by a wave reflected from a high-quality reflecting surface at optical and microwave frequencies. This is a recurrence at optical frequencies of the delay distortion problem which the earlier phase tables were computed to solve at telecommunication frequencies. These tables were, therefore, extended to make possible similar computations at optical frequencies.

A use for this extension soon arose. Bennett,⁹ in his analysis of hole burning effects in a He-Ne optical maser, needed the phase associated with an emission line, Gaussian in shape, but truncated due to an increase in RF power to the maser. The extended tables provided the answer to Bennett's problem and the motivation for writing this paper.

This paper will have two objectives. First, it will extend the numerical computation of the imaginary part from the real part or vice versa of a physical complex variable as a function of frequency from telecommunication and radio frequencies to optical and higher frequencies. Secondly, it will describe a mathematical tool which has proved extremely useful in the telecommunications field and which, it is believed, can be equally useful in the physical sciences.

II. THE NUMERICAL PHASE COMPUTATION TECHNIQUE

In all the discussion to follow, the five statements listed below will apply.

(a) Loss, attenuation, gain, or amplitude, all designated as A , and phase, designated as B , will be used interchangeably with real and imaginary parts, respectively. This is because attenuation in nepers, which is equal to \log_e of the magnitude of a complex voltage or current loss ratio, or \log_e of the amplitude of a complex variable expressed in polar form, and their associated polar angles in radians are identically and respectively interchangeable with real part and imaginary part of a complex variable expressed in rectangular coordinates in the numerical computations to be discussed. In communications problems loss in decibels and angle in degrees rather than nepers and radians respectively are in common use. However, if nepers and radians are considered as the basic units and decibels and degrees as derived units, there will be no difficulty.

(b) Since B_c , the phase at $\omega_c = 2\pi f_c$, is given by Bode (Ref. 5, p. 335) as

$$B_c = \frac{1}{\pi} \int_0^\infty \frac{dA}{d\omega} \log_e \left| \frac{\omega + \omega_c}{\omega - \omega_c} \right| d\omega \quad (1)$$

an amplitude characteristic constant from frequency $f = 0$ to $f = \infty$ contributes nothing to the phase. Therefore, a constant amplitude can be added or subtracted from any amplitude characteristic without affecting the associated phase characteristic.

(c) Since frequency, f , appears only as a ratio in (1), all frequencies can be changed in the same ratio without changing the attenuation-phase relationship in magnitude or form.

(d) All frequencies will be considered on a log frequency scale. Linear frequency scale is permitted in the narrow-band summations only because $\log f$ and f are linearly related over a very narrow band. A narrow band will be considered one in which the total frequency range of interest is less than 10^{-3} times the center frequency. All other bands will be referred to as broad bands.

(e) As seen from (1) above, the phase is determined from the integrated slope, $dA/d\omega$, of the amplitude characteristic, A . The slope of a given straight line section of a straight-line approximation to an amplitude characteristic will be designated by k , and k will be defined as $(A_n - A_{n-1})$ in nepers divided by $\log_e (f_n/f_{n-1})$ where A_n and A_{n-1} are the amplitudes at frequencies f_n and f_{n-1} respectively on the straight line section. A unit slope designated by $k = 1$ will be one in which there is a change in A of one neper between two frequencies which are in the ratio $e = 2.7183$. When A is expressed in decibels a unit slope is a change of 6.02 decibels per octave or 20 decibels per decade.

2.1 Phase Summation Using the Semi-infinite Unit Attenuation Slope

The numerical phase computation is based on a straight-line approximation to the amplitude characteristic, A . A hypothetical attenuation (real part) characteristic plotted on a log frequency scale along with its straight-line approximation is shown in Fig. 1(a). In Fig. 1(b) this straight-line approximation is in turn broken down into the sum of a series of so-called semi-infinite constant slopes of attenuation. A semi-infinite slope is an attenuation characteristic which has a constant magnitude from 0 to some frequency f and a constant slope from f to $f = \infty$. Thus, in Fig. 1(b), the first semi-infinite slope, k_1 , has the constant slope k_1 extending from f_0 to ∞ rather than terminating at f_1 as in Fig. 1(a). Beginning at f_1 a semi-infinite slope equal in magnitude to k_1 but opposite in sign adds to the $+k_1$ slope to produce the straight line segment of our amplitude approximation extending from f_0 to f_1 . This process is continued until the complete approximation is obtained. The semi-infinite unit ($k = 1$) slope of attenuation or real part is the fundamental element of all the numerical phase summations. The phase associated with a semi-infinite unit slope is given by Bode (Ref. 5, pp. 342-43) as

$$\begin{aligned} B(x_c) &= \frac{1}{\pi} \int_{x=0}^{x=x_c} \log_e \left| \frac{1+x}{1-x} \right| \frac{dx}{x} \\ &= \frac{2}{\pi} \left(x_c + \frac{x_c^3}{9} + \frac{x_c^5}{25} + \dots \right) \end{aligned} \quad (2)$$

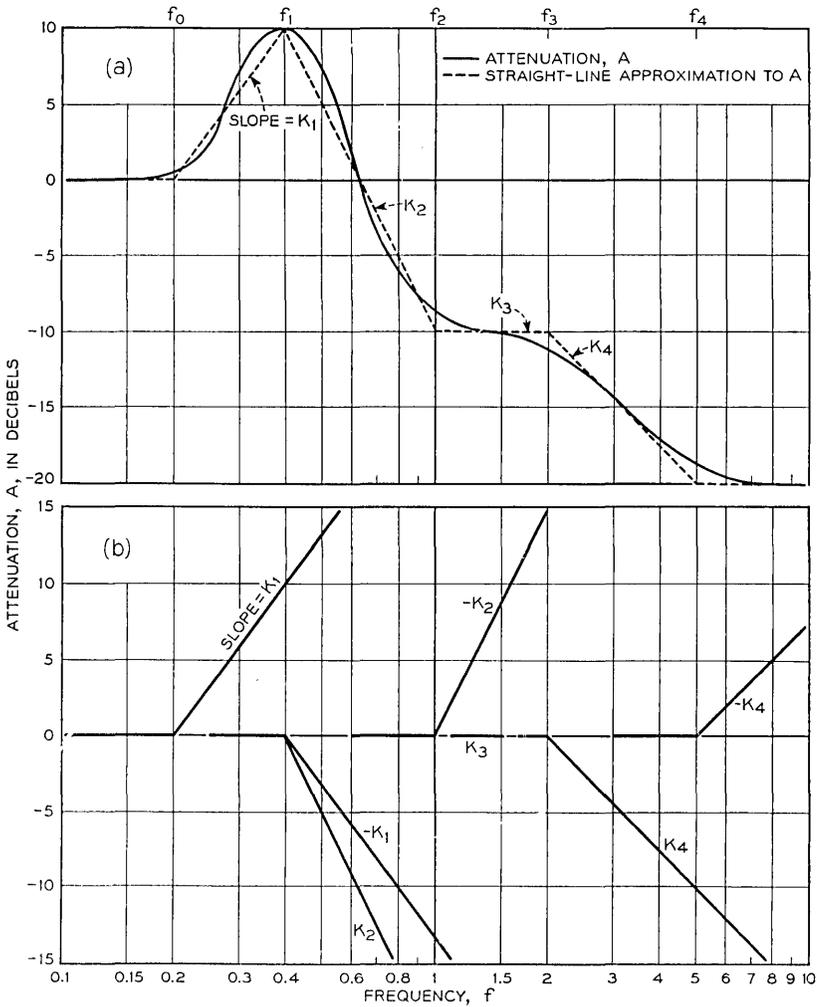


Fig. 1 — (a) Straight-line approximation to attenuation. (b) Semi-infinite slopes which add to produce straight-line approximation.

where $B(x_c)$ is the phase in radians at frequency f_c , $x = f/f_0$, $x_c = f_c/f_0$, $x_c < 1.0$, and f_0 is the frequency at which the unit slope begins.

$B(x_c)$ has a value of 0 at $x_c = 0$, increases monotonically to $\pi/4$ radian at $x_c = 1$, and to $\pi/2$ radians at $x_c = \infty$ with odd symmetry about $x_c = 1$ on a log frequency scale. From the odd symmetry of $B(x_c)$ around $x_c = 1$, $B(x'_c)$ for $f'_c > f_0$ is given by

$$B(x_c' = f_0/f_c') = \pi/2 - B(x_c = x_c'). \quad (3)$$

$B(x_c)$ is the function which was tabulated in the tables of Ref. 6.

The phase associated with a semi-infinite slope of magnitude k_n is k_n times $B(x_c)$ of (2). To get the total phase associated with the straight-line approximation and thus with the true amplitude characteristic within the limits of error of the approximation, it is only necessary to sum the phase contributions of each of the semi-infinite slopes making up the straight-line approximation. Thus, the total phase $\theta(f)$ at frequency f is given by

$$\theta(f) = k_1(\theta_0 - \theta_1) + k_2(\theta_1 - \theta_2) + \cdots + k_n(\theta_{n-1} - \theta_n) \quad (4)$$

where θ_n is the phase of a semi-infinite unit slope commencing at f_n ,

$$k_n = \frac{(A_n - A_{n-1}) \text{ in nepers}}{\log_e (f_n/f_{n-1})} = \frac{(A_n - A_{n-1}) \text{ in decibels}}{20 \log_{10} (f_n/f_{n-1})}. \quad (5)$$

A separate summation must be made for each frequency at which the total phase is desired.

Note the following:

(a) That, as expected, adding or subtracting a constant amplitude to the characteristic does not affect the phase summation of (4).

(b) That initial and final amplitudes need not be the same.

(c) The amplitude need not approach a constant at high or low frequencies but may have a constant slope extending to $f = 0$ or ∞ . This is common in electrical networks. A slope extending to ∞ is covered by $B(x_c)$ of (2). The phase of a slope extending to 0 can be read from the $B(x_c)$ tables for the constant slope extending to ∞ by reading the phase for $f/f_0 < 1$ from Table IV designated $f > f_0$ and the phase for $f_0/f < 1$ from Table III designated $f < f_0$ (see Ref. 6, B.S.T.J., p. 881).

Complete step-by-step examples of summing phase using (4) and the tables of phase of a semi-infinite unit attenuation slope are given in Ref. 6.

2.2 Phase Summation Using the Unit Attenuation Line Segment

When the value of k_n as given by (5) is substituted in (4), (4) can be rewritten as

$$\theta(f) = \sum_{n=1}^n (A_n - A_{n-1}) \frac{\theta_{n-1} - \theta_n}{\log_e (f_n/f_{n-1})} \quad (6)$$

where $(A_n - A_{n-1})$ is the change in amplitude or real part (nepers) on the straight line segment of the approximation to the amplitude characteristic between f_{n-1} and f_n , and $(\theta_{n-1} - \theta_n)/\log_e (f_n/f_{n-1})$ is the phase

contribution of a straight line segment of attenuation or real part having a one-neper change in amplitude between frequencies f_{n-1} and f_n and a constant amplitude below and above f_{n-1} and f_n , respectively. This line segment is identified in its position by the geometric mean of f_{n-1} and f_n , $\sqrt{f_n f_{n-1}}$ and by a slope parameter, a , defined as the ratio of this geometric mean frequency to f_{n-1} .

The "unit line segment" was introduced by Bode (Ref. 5, Ch. XV, Charts V-IX), who gave graphical plots of the phase associated with this element as a function of $x_c = f_c/f_0$ with (a) as a parameter. In a reasonably precise phase summation over a broad band of frequencies, using these charts involves a nonlinear interpolation between curves for different values of a . Therefore, it often proves easier to sum the phase using (4) and the semi-infinite slope charts or tables.

However, in narrow-band problems at optical frequencies, the unit line segment is extremely useful in fast and accurate phase summation. A unit line segment for use with narrow bands is illustrated in Fig. 2. By virtue of the fact that $\log_e f_{12}/f = \log_e (f + \Delta f_{12})/f = \Delta f_{12}/f_{12}$ when $\Delta f_{12} < 10^{-3} f_{12}$, to better than the accuracy to which the amplitude data is likely to be known, a linear frequency plot of amplitude may be used.

The phase of the unit line segment of Fig. 2 will be designated as Φ and will be identified in tabulation by its frequency width Δf and the

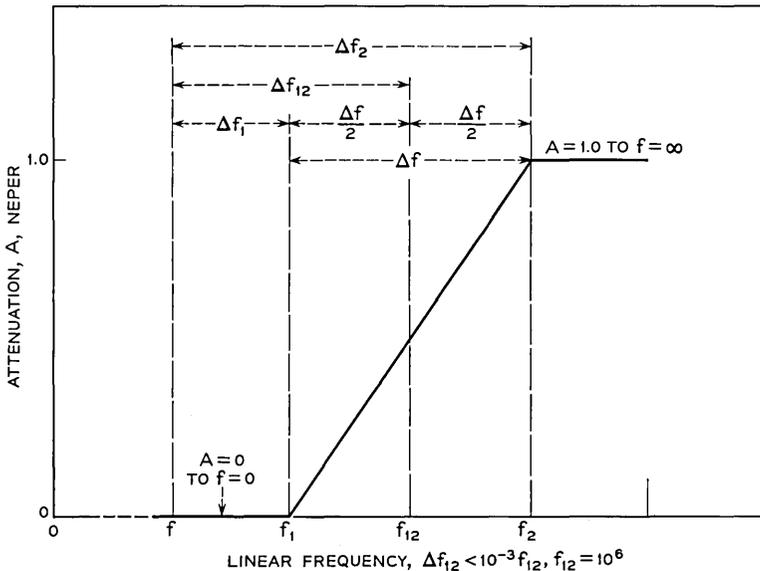


Fig. 2 — Unit attenuation line segment.

difference, Δf_{12} , between its geometric mean frequency $f_{12} = \sqrt{f_1 f_2} = (f_1 + f_2)/2$ and f .

Using unit line segments having a phase contribution of Φ , (6) can now be written

$$\theta(f) = (A_1 - A_0)\Phi_{01} + (A_2 - A_1)\Phi_{12} + \cdots (A_n - A_{n-1})\Phi_{(n-1)n} \quad (7)$$

where $\Phi_{(n-1)n}$ is the phase contribution of a unit attenuation line segment of width $\Delta f = f_n - f_{n-1}$ and a Δf_{12} of $(f_{n-1} + f_n)/2 - f$.

Φ is evaluated in the next section and tabulated in Table V for $f_{12} = 10^6$. Φ is always positive for a positive slope and increases monotonically from 0 at $f = 0$ to a maximum at $f = f_{12}$ beyond which it decreases monotonically to 0 at $f = \infty$. As a function of Δf_{12} it has even symmetry about $\Delta f_{12} = 0$ so that

$$\Phi(\Delta f_{12}) = \Phi(-\Delta f_{12}).$$

Note the restriction of Fig. 2 and of Table V that $f_{12} = 10^6$. This restriction applies *only* if the initial and final values of the amplitude of the characteristic *are not* the same. If they are not the same, the problem must be expanded or contracted about $f = 0$ to a center frequency of 10^6 by multiplying all frequencies by the ratio of 10^6 to the center band frequency. If they are the same, then the problem can be linearly expanded or contracted about its center frequency to best fit the range of Δf_{12} of Table V, and the phase will expand and contract to bear the same relationship to the magnitude. Proof that this is permissible will be given in Section 3.2. If the initial and final values are not the same, they can be made the same by truncating the main high- Q portion of the band from the rest of the band on a constant amplitude line. The phase of the truncated portion having equal initial and final amplitudes can then be summed using the permissible linear expansion or contraction of the band about its center frequency. The residue is then evaluated using the semi-infinite slope summation if the residue becomes a broadband problem. If the residue remains a narrow-band problem, the line segment summation may be used by expanding or contracting about $f = 0$ to make the center frequency equal 10^6 . In reassembling the problem and adding the phase of the two parts, the inverse frequency transformations must, of course, be made.

An example of phase summation using the unit line segment phase of Table V in (7) will be given in Section 4.2.

2.3 Phase Summation Using the Unit Wedge of Attenuation

The unit wedge of attenuation is a convenient element for very accurate narrow-band phase summation. Although it has been developed

primarily for use with an automatic computer, it is equally useful for rapid but less precise desk computer phase summation.

The summation is limited to characteristics having equal initial and final amplitudes. If they are not equal they can be made so by the division of the problem into two problems by constant amplitude truncation as discussed in Section 2.2.

Since it is assumed that $A_n = A_0$, A_n and A_0 can each be made 0 by subtracting a constant amplitude A_0 from the amplitude characteristic. Equation (7) can then be written:

$$\theta(f) = A_1(\Phi_{01} - \Phi_{12}) + A_2(\Phi_{12} - \Phi_{23}) + \dots + A_{n-1}(\Phi_{(n-2)(n-1)} - \Phi_{(n-1)n}).$$

Each of the terms of the above equation is a magnitude A_n multiplied by the phase difference of two unit line segments of the type illustrated in Fig. 2. The first line segment extends from $f = n - 1$ to $f = n$ and the second from the terminal of the first at $f = n$ to $f = n + 1$. If the widths Δf of these two line segments are equal, they produce the "unit wedge of attenuation," which is the third type of amplitude element used in the numerical phase summation. In using this element the straight-line approximation is limited to equal frequency interval segments. Therefore, the hypothetical problem of Fig. 1 is no longer useful in the discussion and a new problem shown in Fig. 3 will be used. In Fig. 3 the amplitude characteristic is plotted on a linear frequency scale consisting of equally spaced intervals between frequencies which are designated as either f or n . The straight-line approximation is now obtained by taking exact values of A at even values of n on the true

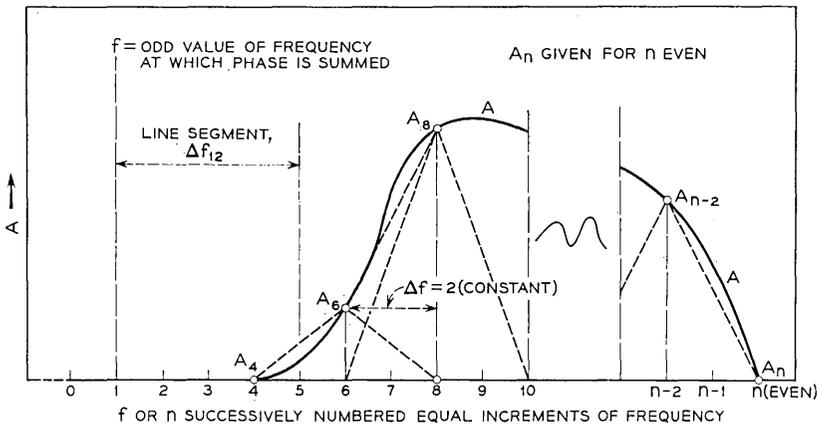


Fig. 3 — Phase summation using wedge element.

amplitude characteristic. Since the accuracy of phase summation is greatest at the midfrequency of the straight line segments approximating the amplitude characteristic, phase will be summed at odd values of n . For rapid desk computing using a less accurate approximation, it may be desirable to take A 's which lie off the true curve. This will be covered in Section 4.3.

The total phase at f associated with the full magnitude characteristic of Fig. 3 is then given by the sum of the individual phases contributed by each successive line segment from A_4 to A_6 , A_6 to A_8 , \dots A_{n-2} to A_n . Therefore, from (7)

$$\theta(f \text{ odd}) = (A_6 - A_4)\Phi_{\Delta f_{12}=5-f} + (A_8 - A_6)\Phi_{\Delta f_{12}=7-f} + \dots \\ + (A_{n-2} - A_{n-4})\Phi_{\Delta f_{12}=(n-3)-f} + (A_n - A_{n-2})\Phi_{\Delta f_{12}=(n-1)-f}$$

and since the initial value A_4 and the final value A_n are zero

$$\theta(f \text{ odd}) = A_6(\Phi_{\Delta f_{12}=5-f} - \Phi_{\Delta f_{12}=7-f}) + \dots \\ + A_{n-2}(\Phi_{\Delta f_{12}=n-3-f} - \Phi_{\Delta f_{12}=n-1-f}) \quad (8) \\ = \sum_{n \text{ even}} A_n(\Phi_{\Delta f_{12}=n-(f+1)} - \Phi_{\Delta f_{12}=n-(f+1)+2})$$

in which the Φ 's all have a Δf of 2.

Each term of (8) represents the phase due to a wedge of attenuation in the shape of an isosceles triangle having an amplitude A_n and a base width of 4 frequency intervals. The first of such amplitude elements in Fig. 3 is defined by points ($A = A_4 = 0, f = 4$), ($A_6, f = 6$), and ($A = 0, f = 8$), the second by ($A = 0, f = 6$), ($A_8, f = 8$), and ($A = 0, f = 10$), etc. These amplitudes add to approximate the true curve. When the amplitude A_n is unity, this element is called a unit wedge of attenuation, and its phase contribution is designated by Ψ . Ψ is identified by a subscript which is equal to $500 +$ its lower frequency line segment's Δf_{12} or by $500 + n - (f + 1)$. The 500 is added to $n - (f + 1)$ to avoid negative subscripts in tabulation. Equation (8) can now be written

$$\theta(f \text{ odd}) = \sum_{n \text{ even}} A_n \Psi_{500+n-(f+1)} \quad (9)$$

where

$$\Psi_{500+n-(f+1)} = \Phi_{\Delta f=2, \Delta f_{12}=n-(f+1)} - \Phi_{\Delta f=2, \Delta f_{12}=n-(f+1)+2}. \quad (10)$$

Ψ is given in Table VI for $500 + n - (f + 1)$ even from 0 to 1000. In summing phase using (9) the center of the band of the problem is placed at $n = 500$. The band is then linearly expanded or contracted about

$n = 500$ to get a maximum number of amplitude evaluations consistent with the frequency range of the phase summation desired. With Ψ tabulated for $500 + n - (f + 1)$ even from 0 to 1000, the maximum and minimum permissible values of f and n are related as follows

$$500 + f + 1 > n > f + 1 - 500.$$

Thus, for a low value of $n_{\text{even}} = L$ and a high value of $n_{\text{even}} = H$, the phase θ at f can be summed only for odd values of f between $f = H - 501$ and $f = L + 499$.

The ease and accuracy of automatic computer summation of phase using the unit wedge tables and (9) will be demonstrated in Section 4.3. A fast and good phase summation using a less accurate straight-line approximation and a desk computer will also be illustrated.

2.4 *Requirements on the Complex Variable for the Numerical Method to be Applied*

A rigorous discussion of the requirements which must be met by a complex variable if the phase computed from its amplitude characteristic is to represent its true phase is beyond the scope of this paper (see Ref. 5, Ch. XIII). Briefly, it is required that the function be an analytic function of frequency in the right half p ($p = i\omega$) plane and that its real and imaginary components be even and odd functions of frequency, respectively, on the real frequency axis.

Actually, if there is sufficient information available to rigorously determine the applicability of the method, the numerical phase summation technique will usually not be needed. Fortunately, when it is needed the phase summed by the numerical method almost always contains the desired information in spite of the fact that a portion of the phase referred to as nonminimum⁵ phase may be missed in the summation. For instance, in a long electrical, optical, acoustical, or other transmission path, where a long path is defined as one in which the length is many multiples of the wavelength of the transmitted signal, there will be an integer multiple of 2π radians which will not be included in the phase summed by the techniques described. However, the phase summed will, in general, contain all of the phase nonlinearity, and in this type of problem the nonlinear phase is usually the phase of interest. Therefore, delay distortion in television transmission lines was successfully delay distortion equalized using phase data obtained by numerical phase summation based on the loss or absorption characteristics of the lines. Also, in Bennett's He-Ne maser problem,⁹ the nonlinear phase in the truncated Gaussian line was obtained by nu-

merical phase summation in spite of the fact that an integer multiple of 2π radians in the total phase was missed in the numerical summation.

Similar situations exist with regard to nonminimum imaginary part complex variables where a portion of the imaginary component is missed in the summation. Here again, however, the minimum possible imaginary part associated with the real part which is obtained by the numerical summation, is usually of sufficient interest to make the summation valuable.

There is one important type of nonminimum phase function for which the numerical summation may not be useful. A good example of such a function is an electrical bridge having zero transmission or infinite loss at a real frequency due to bridge balance. This violates the requirement that the function be analytic in the right half p plane. In this case, the phase summed may be the true phase or it may depart radically and nonlinearly from the true phase over a wide frequency band centered at the infinite loss frequency (Ref. 6, B.S.T.J., p. 896). By analogy to the electrical case, the application of the numerical phase summation to optical or other amplitude characteristics having infinite loss obtained by interference (as in an interferometer) or multilayer reflection interference should be approached with caution if not entirely avoided.

2.5 *Computation of Amplitude from Phase*

So far only the determination of phase from amplitude has been considered. The same technique and tables can be used for the reverse computation. However, since a constant amplitude does not change the phase, the amplitude determined from a given phase characteristic must contain an additional arbitrary constant. This is taken into account by considering the attenuation determined as the difference between the true attenuation and the attenuation at either zero or infinity.

In the reverse computation the complex variable $A + iB$ is replaced by either $i\omega(A - A_\infty + iB)$ or $(A - A_0 + iB)/i\omega$. The multiplication by $i\omega$ or its reciprocal has the effect of interchanging the real and imaginary components and their even and odd symmetry characteristics. Thus in $i\omega(A - A_\infty + iB)$, the real component becomes $-\omega B$ with even symmetry and the imaginary component becomes $i\omega(A - A_\infty)$ with odd symmetry. Similarly in $(A - A_0 + iB)/i\omega$, the real component becomes B/ω with even symmetry and the imaginary component becomes $-i(A - A_0)/\omega$ with odd symmetry. The transformed variables are then in suitable form for determining B from A using the same

formula and tables and the same techniques as were described for determining A from B . It must be remembered, of course, that the values of A determined from the summation will include an arbitrary additive constant (Ref. 5, pp. 320-330).

III. COMPUTATION OF PHASE TABLES

3.1 *Semi-Infinite Unit Slope Phase Computation*

The original tables of phase of a semi-infinite slope of Ref. 6 are adequate except for the very steep slopes which may occur at microwave frequencies and frequently occur at optical frequencies. For instance, in the first optical problem to which they were applied, the delay distortion or time dispersion at the surface of a mirror,¹⁰ the critical phase values fell within the final 60 of 9,640 tabulated values of phase in radians falling in the vicinity of $x = f/f_0 = 1.0$. Therefore, the extension of the earlier tables is limited to values of $f/f_0 > 0.9999$. In this region, the best expression for obtaining the phase B is given by Bode as

$$B(x_c) + B(y_c) = \frac{\pi}{4} - \frac{1}{\pi} \log_e x_c \log_e y_c \tag{11}$$

where

$$y_c = \frac{1 - x_c}{1 + x_c}; \quad x_c = \frac{1 - y_c}{1 + y_c}.$$

From (2)

$$B(x_c) = \frac{2}{\pi} \left[\frac{1 - y_c}{1 + y_c} + \frac{1}{9} \left(\frac{1 - y_c}{1 + y_c} \right)^3 + \dots \right] = -\frac{1}{\pi} \log_e y_c \tag{12}$$

to better than 2×10^{-14} radian for $(1 - y_c) < 10^{-4}$.

Substituting (12) in (11)

$$B(y_c) = \frac{\pi}{4} + \frac{1}{\pi} \log_e y_c \log_e \frac{e}{x_c}$$

as

$$y_c \rightarrow 1.0, \log_e y_c = \frac{2(y_c - 1)}{y_c + 1}$$

and

$$B(y_c) = \frac{\pi}{4} - \frac{2}{\pi} \left[\frac{1 - y_c}{1 + y_c} \log_e \frac{e(1 + y_c)}{1 - y_c} \right]. \quad (13)$$

Equation (13) is good to 3×10^{-13} radian for $y_c > 0.9999$. This equation was used to compute $B(y_c)$ to 12 significant figures for

$$y_c = .99990(.0^41) .99998(.0^55) .999995(.0^65) \\ .999998(.0^62) .9999998(.0^61) 1.0.$$

These values were then extended by the numerical integration technique described in Ref. 6 to obtain $B(y_c)$ for

$$y_c = .999900(.0^51) .999980(.0^65) .9999980(.0^62) \\ .9999998(.0^61) 1.0.$$

These values were then graphically interpolated to obtain 11 significant figure values of $B(y_c)$ for

$$y_c = .999900(.0^51) .999980(.0^61) \\ .9999998(.0^75) 1.0.$$

These final values were rounded to 9 figures to obtain the values given in Table III. The odd symmetry of $B(x_c)$ about $x_c = 1.0$ was used to obtain Table IV in accordance with (3).*

The initial 12-figure computations were good to ± 1 in the 12th figure. The maximum error in numerically integrating and graphically interpolating to 11 significant figures is estimated to be less than 5 figures in the 11th figure. The nine-figure tables are, therefore, believed to be subject only to rounding errors in the last figure.

In order to extend the range of use of Tables III and IV of this paper, values of $B(x_c)$ for $x_c = .9970(.0001) .9999$ from the Ref. 6 tables are included.

3.2 Unit Attenuation Line Segment Phase Computation

Fig. 2 shows a unit attenuation line segment meeting the restrictions that $\Delta f_{12} < 10^{-3} f_{12}$ and $f_{12} = 10^6$. The phase associated with this amplitude element will be designated as Φ . It is determined by the difference between the phase of a positive semi-infinite slope beginning at $A = 0$ at f_1 and extending to infinity, passing through amplitude $A = 1.0$ at

* Tables of phase functions are numbered the same as tables previously mentioned (Ref. 6). Therefore Tables I and II do not appear in this paper, since angles are given in radians only. Furthermore, additional tabulations in the present paper are numbered consecutively, even though the numbers sometimes duplicate table numbers used in illustrative examples in Ref. 6.

f_2 and the phase of a semi-infinite slope of equal magnitude but opposite in sign beginning at $A = 1.0$ at f_2 . In accordance with the definition of slope given by (5), the slope k of these semi-infinite slopes will be

$$k = \frac{A = 1.0}{\log_e f_2/f_1} = \frac{1}{\log_e \frac{f_{12} + \Delta f/2}{f_{12} - \Delta f/2}}$$

or

$$1/k = \log_e \frac{1 + \Delta f/2 f_{12}}{1 - \Delta f/2 f_{12}} = \frac{\Delta f}{f_{12}} \tag{14}$$

to better than 1 in 10^{10} for the maximum value of $\Delta f = 40$ for which Φ will be tabulated. Referring to Fig. 2, Φ will therefore be given by:

$$\Phi = k[B(f/f_1) - B(f/f_2)] \tag{15}$$

$$= (10^6/\Delta f)[B(f/f_1) - B(f/f_2)], f < f_1 \tag{16}$$

$$= (10^6/\Delta f)[B(f_1/f) - B(f/f_2)], f_{12} > f > f_1 \tag{17}$$

in which the B 's are the phases of semi-infinite unit slopes of attenuation. Φ need be evaluated only for $f < f_{12}$ since $\Phi(\Delta f_{12}) = \Phi(-\Delta f_{12})$ as a result of the even symmetry of Φ about $f = f_{12}$.

Referring to Fig. 2, when $f < f_1$, f/f_1 is given by

$$f/f_1 = \frac{f_1 - \Delta f_1}{f_1} = 1 - \frac{\Delta f_1}{f_1} = 1 - \frac{\Delta f_{12} - \Delta f/2}{10^6}$$

and $B(f/f_1)$ of (16) is read from Table III for $f < f_0$. When $f > f_1$, $f_1/f = 1 - [(\Delta f/2 - \Delta f_{12})/10^6]$ and $B(f_1/f)$ of (17) is read from Table IV for $f > f_0$. Since $f < f_{12} < f_2$, $f/f_2 = 1 - [(\Delta f_{12} + \Delta f/2)/10^6]$ and $B(f/f_2)$ of (16) and (17) is always read from Table III for $f < f_0$.

Equations (16) and (17) and the approximations to f/f_1 and f/f_2 above may be used to evaluate Φ for $\Delta f_{12} < 50$, $\Delta f < 40$ to an accuracy of better than 0.0002 radian. This is sufficient since Φ is only given to 0.001 radian in Table V. These equations were therefore used to compute Φ of Table V for $\Delta f_{12} = 0$ (1) 50 for each of the following Δf 's: 2, 4, 6, 10, 20, and 40.

For $\Delta f_{12} > 50$, $\Delta f < 40$, and $f < f_1$, (15) is used to compute Φ . However, the B 's are determined from (13) with $y_e = y_1 = f/f_1$, or $y_e = y_2 = f/f_2$. Thus

$$\begin{aligned} \Phi &= k[B(y_1) - B(y_2)] \\ &= \frac{2f_{12}}{\pi \Delta f} \left[\frac{1 - y_2}{1 + y_2} \log_e \frac{e(1 + y_2)}{1 - y_2} - \frac{1 - y_1}{1 + y_1} \log_e \frac{e(1 + y_1)}{1 - y_1} \right]. \tag{18} \end{aligned}$$

The error in Φ as determined from (18) is less than 10^{-7} radian for $y_{12} = f/f_{12} > .999$, $\Delta f < 40$.

Referring to Fig. 2

$$1 - y_2 = 1 - f/f_2 = 1 - \frac{f_2 - \Delta f_2}{f_2} = \frac{\Delta f_2}{f_2}$$

$$1 + y_2 = 1 + \frac{f_2 - \Delta f_2}{f_2} = 2 \left(1 - \frac{\Delta f_2}{2f_2} \right)$$

$$1 - y_1 = \frac{\Delta f_1}{f_1}, \quad 1 + y_1 = 2 \left(1 - \frac{\Delta f_1}{2f_1} \right).$$

Substituting in (18)

$$\Phi = \frac{2f_{12}}{\pi\Delta f} \left[\frac{\Delta f_2}{2f_2 \left(1 - \frac{\Delta f_2}{2f_2} \right)} \log_e \frac{2ef_2 \left(1 - \frac{\Delta f_2}{2f_2} \right)}{\Delta f_2} - \frac{\Delta f_1}{2f_1 \left(1 - \frac{\Delta f_1}{2f_1} \right)} \log_e \frac{2ef_1 \left(1 - \frac{\Delta f_1}{2f_1} \right)}{\Delta f_1} \right] \quad (19)$$

and as shown in the Appendix, (19) can be reduced to:

$$\Phi = \frac{1}{\pi\Delta f} \left[\Delta f_2 \log_e \frac{2ef_{12}}{\Delta f_2} - \Delta f_1 \log_e \frac{2ef_{12}}{\Delta f_1} \right] - \frac{\Delta f_{12}}{2\pi f_{12}} \quad (20)$$

($\Delta f < 40$, $1000 > \Delta f_{12} > 50$).

The error term $\Delta f_{12}/2\pi f_{12}$ is only 1.6×10^{-4} radian for $\Delta f_{12} = 1000$. Since Φ in Table V is only given to 0.001 radian, this error term is dropped. Φ as given in (20) can then be further reduced, as shown in the Appendix, to

$$\Phi = \frac{1}{\pi} \left[\log_e \frac{2ef_{12}}{\Delta f_2} - \frac{\Delta f_1}{\Delta f} \log_e \frac{\Delta f_2}{\Delta f_1} \right] \quad (21)$$

$$= \frac{\log_e 10}{\pi} \left[\log_{10} e + \log_{10} \frac{2f_{12}}{\Delta f_2} - \frac{\Delta f_1}{\Delta f} \log_{10} \frac{\Delta f_2}{\Delta f_1} \right]. \quad (22)$$

Φ was computed using (22) for $\Delta f = 10$ at $\Delta f_{12} = 50(2) 80(5) 160(10) 300$, and for $\Delta f = 40$ at $\Delta f_{12} = 50(2) 100(5) 130$. Five figures to the right of the decimal were retained in spite of the fact that the error term of (20) puts an error of as much as 5 in the last figure, since in deriving unit wedge phase from this data, for Φ (Δf_{12}) differences of

$\Delta f_{12} - \Delta f_{12}' = 2$, the difference error is moved out to 3 in the 7th figure. The computed data was then graphically interpolated to give Φ for $\Delta f = 10$ at $\Delta f_{12} = 50(1) 300$ and Φ for $\Delta f = 40$ at $\Delta f_{12} = 50(1) 130$. The final data was rounded to 3 figures to the right of the decimal and is subject only to rounding errors. In tabulation, however, the $\Delta f = 10$ values are tabulated for $\Delta f \leq 20$ at $\Delta f_{12} = 50(1) 130$ and for $\Delta f \leq 40$ at $\Delta f_{12} = 130(1) 300$. This introduces a maximum error for the tabulation of 0.0015 radian for $\Delta f = 20$, $\Delta f_{12} = 50$ and 0.0011 radian for $\Delta f = 40$, $\Delta f_{12} = 130$ to give a maximum percentage error in Φ as tabulated between Δf_{12} 's of 50 and 300 of 0.05 per cent.

For $\Delta f_{12} > 300$ it is shown in the Appendix that (21) can be reduced to

$$\Phi = \frac{1}{\pi} \log_e \frac{2f_{12}}{\Delta f_{12}} + \frac{1}{24\pi} \left(\frac{\Delta f}{\Delta f_{12}} \right)^2 \tag{23}$$

$(\Delta f < 40, 1000 > \Delta f_{12} > 300)$

The error term $(1/24\pi)(\Delta f/\Delta f_{12})^2$ has a maximum at $\Delta f = 40$, $\Delta f_{12} = 300$ of 2.3×10^{-4} and can be neglected. Equation (23) can then be written for $f_{12} = 10^6$

$$\Phi = \frac{\log_e 10}{\pi} \left(6.0 - \log_{10} \frac{\Delta f_{12}}{2} \right) \tag{24}$$

$(\Delta f < 40, 1000 > \Delta f_{12} > 300).$

Equation (24) was used to compute Φ for $\Delta f \leq 40$ at $\Delta f_{12} = 300(10) 1000$ as tabulated in Table V.

One other source of error must be considered. In using the tables the actual center of the line segment being summed will not be at $f_{12} = 10^6$ but may depart from this by half the band spread of the problem. For a band of 10^3 this will be 500. Equation (23) may be used to evaluate this error. It will be given by

$$\begin{aligned} \text{Max } f_{12} \neq 10^6 \text{ error} &= \frac{1}{\pi} \log_e \frac{2f_{12}}{\Delta f_{12}} - \frac{1}{\pi} \log_e \frac{2f_{12} \pm 500}{\Delta f_{12}} \\ &= \frac{1}{\pi} \log_e \frac{2f_{12}}{2f_{12} \pm 500} = \pm \frac{500}{2\pi f_{12}} < 10^{-4} \text{ radian} \end{aligned}$$

which is acceptable for our tabulation.

Recapitulating, the Table V phase values may be considered to be reliable to ± 0.002 radian or less than 0.1 per cent, which is better than the amplitude approximations which are usually used for the unit attenuation line segment phase summation.

The reason for the restriction on center-band frequency in line segment phase summation when the initial and final amplitudes are not the same is now apparent. f_{12} appears as a factor term in all equations for Φ . Since Φ was computed for $f_{12} = 10^6$, the problem must be transformed to a center frequency of 10^6 by multiplying all frequencies by the ratio of 10^6 to the actual center frequency. This does not change the attenuation-phase relationship, since all frequency terms in the Φ expression appear as ratios.

Now consider the problem when the initial and final amplitudes are the same, as shown in the hypothetical problem of Fig. 4. Note that the entire amplitude characteristic can be constructed of trapezoidal elements by successive constant amplitude truncations. A typical element is $abcd$. Its phase at f is given by the sum of the phases of the two line segments ab and cd . Thus:

$$\Phi(f) \text{ of } abcd = A_{ab}[\Phi(\Delta f_{12}) - \Phi(\Delta f_{34})] \quad (25)$$

and from (21)

$$\begin{aligned} \Phi(\Delta f_{12}) - \Phi(\Delta f_{34}) &= \frac{1}{\pi} \left[\log_e \frac{2ef_{12}}{\Delta f_2} - \frac{\Delta f_1}{\Delta f_a} \log_e \frac{\Delta f_2}{\Delta f_1} \right. \\ &\quad \left. - \log_e \frac{2ef_{34}}{\Delta f_4} + \frac{\Delta f_3}{\Delta f_b} \log_e \frac{\Delta f_4}{\Delta f_3} \right] \\ &= \frac{1}{\pi} \left[\log_e \frac{2ef_{12}}{\Delta f_2} \frac{\Delta f_4}{2e(f_{12} + F)} \right. \\ &\quad \left. - \frac{\Delta f_1}{\Delta f_a} \log_e \frac{\Delta f_2}{\Delta f_1} + \frac{\Delta f_3}{\Delta f_b} \log_e \frac{\Delta f_4}{\Delta f_3} \right] \\ &= \frac{1}{\pi} \left[\log_e \frac{\Delta f_4}{\Delta f_2} - \frac{\Delta f_1}{\Delta f_a} \log_e \frac{\Delta f_2}{\Delta f_1} \right. \\ &\quad \left. + \frac{\Delta f_3}{\Delta f_b} \log_e \frac{\Delta f_4}{\Delta f_3} \right] - \frac{F}{\pi f_{12}}. \end{aligned} \quad (26)$$

$F/f_{12} < 10^{-3}$ by the narrow-band limitation of our problem, so the second term of (26) is less than 0.0003 radian, which is negligible. Note that f_{12} has disappeared from the first term and that the phase is now dependent only upon ratios of linear frequency intervals. Although f was chosen $< f_1$ in obtaining (26), the dependence of $\Phi(\Delta f_{12}) - \Phi(\Delta f_{34})$ on ratios of linear frequency intervals only, can be shown for all values of f . The problem can, therefore, be linearly expanded or contracted about its center frequency to best fit the range of tabulated values of Φ with-

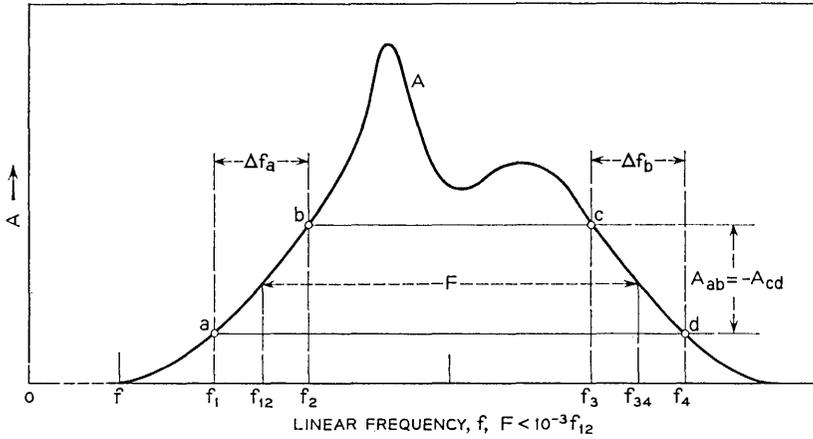


Fig. 4 — Line segment phase summation.

out changing the attenuation-phase relationship, as noted earlier in Section 2.2.

3.3 Unit Wedge Phase Computation

The phase contribution of a unit wedge of attenuation is given by (10) as

$$\Psi_{500+n-(f+1)} = \Phi_{\Delta f_{12}=n-(f+1)} - \Phi_{\Delta f_{12}=n-(f+1)+2}$$

where $\Delta f = 2$ for both Φ 's and n and f are even and odd integers respectively. If f of (10) is $n + b$ then

$$\Psi_{500-(b+1)} = \Phi_{\Delta f_{12}=-(b+1)} - \Phi_{\Delta f_{12}=-(b+1)}. \tag{27}$$

If f of (10) is $n - b$ then

$$\Psi_{500+b-1} = \Phi_{\Delta f_{12}=b-1} - \Phi_{\Delta f_{12}=b+1}. \tag{28}$$

Because of the even symmetry of Φ about $f = f_{12}$ and therefore about $\Delta f_{12} = f - f_{12} = 0$,

$$\Phi_{\Delta f_{12}=-(b+1)} = \Phi_{\Delta f_{12}=b+1}$$

and

$$\Phi_{\Delta f_{12}=-(b+1)} = \Phi_{\Delta f_{12}=b-1}.$$

Therefore, from (27) and (28)

$$\Psi_{499-b} = -\Psi_{499+b}. \tag{29}$$

$\Psi_{500+n-(f+1)}$ for $500 + n - (f + 1)$ even from 500 to 600 was computed using the 11-figure tables of $B(x_c)$ — before reduction to 9 figures for tabulation — to compute the Φ values needed in (10). The Φ values were computed in accordance with the procedure given in Section 3.2 for Φ ($\Delta f < 40$, $\Delta f_{12} < 50$). The extension of this Φ computation to $\Delta f_{12} = 100$ is permissible because of the small value of $\Delta f = 2$. The use of 11-figure tables of $B(x_c)$ good to only five in the final figure is permissible because small differences ($\Delta f = 2$) in this table are good to at least one more significant figure. The final figure in the tabulated values of Ψ depends upon differences in the 11th figure in $B(x_c)$. They are, therefore, estimated to be good to better than ± 2 in the last figure.

$\Psi_{500+n-(f+1)}$ for $500 + n - (f + 1)$ even from 600 to 1000 was computed using 5 decimal figure values of Φ computed before rounding for tabulation in accordance with the procedure given in Section 3.2 for $1000 > \Delta f_{12} > 50$. In accordance with the discussion of the reliability of these computations in Section 3.2, the resultant 5-decimal figures of Ψ are estimated to be reliable to better than ± 2 in the final figure. Values of $\Psi_{500+n-(f+1)}$ for $500 + n - (f + 1) < 500$ were obtained from the values for $500 + n - (f + 1) \geq 500$ using (29).

Table VI, giving $\Psi_{500+n-(f+1)}$ for $500 + n - (f + 1)$ even from 0 to 1000 to 5 decimal figures, was tabulated using the above data.

IV. EXAMPLES OF PHASE SUMMATION

4.1 *Semi-Infinite Unit Slope Phase Summation*

Summation of phase using the semi-infinite slope of attenuation is described in Section 2.1 and fully illustrated in Ref. 6. Therefore, an actual numerical summation is not considered necessary here.

4.2 *Unit Line Segment Phase Summation*

A part of the truncated Gaussian problem solved for Bennett⁹ will be used to illustrate unit line segment phase summation. Fig. 5(a) shows the top portion of a Gaussian amplitude characteristic, A , normalized to a peak amplitude of unity and truncated at $A = 0.712$ and $A = 0.5$.

The characteristic has a half width at half maximum of 800 mc, corresponding to the full Doppler width at half maximum for neon atoms at the temperature of the He-Ne optical maser. It has a center frequency of approximately 2.6×10^{14} cps, corresponding to the frequency of oscillation of the maser. Since the ratio of the bandwidth to

center frequency is ten orders of magnitude smaller than the narrow-band requirement, and the initial and final amplitudes of the truncated section are the same, a linear frequency summation scale was chosen for convenience in unit line segment summation as shown on Fig. 5(a).

The phase wanted is the phase due to that portion of the Gaussian lying between $A = 0.5$ and $A = 0.712$. Since the two sides of this area which are defined by the Gaussian are essentially straight lines, the characteristic was approximated by the two straight lines, ac and $c'e$, and the three constant amplitude lines $A = 0.712$ from c to c' , $A = 0.5$ from $f_T = 0$ to $f = 0$, and $A = 0.5$ from $f = 80$ to $f_T = \infty$. This approximation was then broken into four line segments, ab , bc , $c'd$, and de . The desired phase is then given by (7) as

$$\theta(f) = (A_b - A_a)\Phi_{ab} + (A_c - A_b)\Phi_{bc} + (A_d - A_{c'})\Phi_{c'd} + (A_e - A_d)\Phi_{de}$$

and since

$$\begin{aligned} (A_b - A_a) &= (A_c - A_b) = -(A_d - A_{c'}) = -(A_e - A_d) \\ &= 0.106 \text{ neper,} \end{aligned} \tag{30}$$

$$\theta(f) = 0.106[\Phi_{ab} + \Phi_{bc} - \Phi_{c'd} - \Phi_{de}] \text{ radians.}$$

From Fig. 5, $\Delta f = 6$ for all the line segments, and

ab has a center frequency of 3 and its $\Delta f_{12} = | (3 - f) |$,

bc has a center frequency of 9 and its $\Delta f_{12} = | (9 - f) |$,

$c'd$ has a center frequency of 71 and its $\Delta f_{12} = | (71 - f) |$,

de has a center frequency of 77 and its $\Delta f_{12} = | (77 - f) |$.

Table VII gives the entire tabulation and phase summation of (30). The first column gives frequency, f , at which phase is to be summed. The second column gives Δf_{12} for line ab at each value of f , and the third column gives Φ_{ab} for $\Delta f = 6$ from Table V for each value of Δf_{12} at f . This is repeated for Φ_{bc} , $\Phi_{c'd}$, and Φ_{de} . Note the orderly recurrence of values of Φ , which made for easy tabulation.

A desk computer was used to sum the four unit line segment phase contributions horizontally [$\Phi_{c'd}$ and Φ_{de} negatively from (30)] and then multiply the sum by 0.106 to get $\theta(f)$ of the last column in radians. This summed phase is plotted as B on Fig. 5(b). The precision of the summation is demonstrated by the smoothness of the data.

The ease of the computation is illustrated by the fact that the approximation and phase summation was completed in one hour.

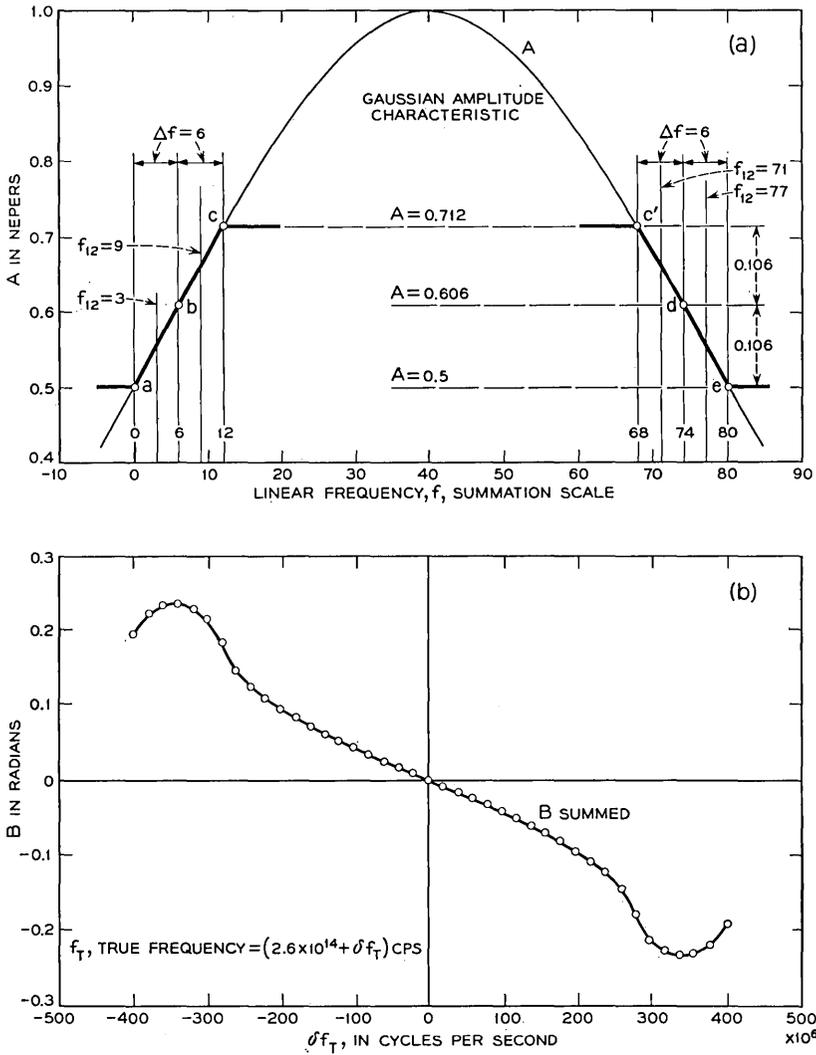


Fig. 5 — (a) Double truncated Gaussian amplitude characteristic. (b) Phase determined by unit line segment summation.

4.3 Unit Wedge Phase Summation

The quantum-mechanically derived expression for the complex dielectric constant, $\epsilon = \epsilon_1 - i\epsilon_2$, will be used to illustrate phase summation using the unit wedge for three reasons.

First, ϵ is defined by a Lorentzian whose real and imaginary parts are known. Phase summation of ϵ can therefore be checked against known data. Secondly, it is derivable from the classical equation for a damped harmonic oscillator which occurs repeatedly in science and engineering.¹¹ Finally, the real part of ϵ is summed from the imaginary part and serves to illustrate the reverse summation discussion in Section 2.5.

The formula for complex dielectric constant given by Van Vleck (Ref. 7, p. 644) can be written as the sum of two Lorentzians as follows

$$\epsilon - 1 = A - \frac{A}{2} \left\{ \frac{\nu}{(\nu - \nu_0) - i\Delta\nu} + \frac{\nu}{(\nu + \nu_0) - i\Delta\nu} \right\} \quad (31)$$

where ν = frequency (f), $\Delta\nu$ = half bandwidth at $|\epsilon - 1| = 0.5$ maximum, and A is a constant.

For a narrow band about ν_0 only the first term of (31) is important and (31) can therefore be written

$$\epsilon - 1 = A - \frac{A\nu_0}{2} \left\{ \frac{1}{(\nu - \nu_0) - i\Delta\nu} \right\}. \quad (32)$$

From (31), A is seen to be $\epsilon_0 - 1$ where $\epsilon_0 = \epsilon(\nu = 0)$. Substituting for the isolated A term in (32), and separating into real and imaginary parts (32) becomes

$$(\epsilon_1 - \epsilon_0) - i\epsilon_2 = \frac{A\nu_0}{2} \left[\frac{(\nu_0 - \nu)}{(\nu - \nu_0)^2 + \Delta\nu^2} - i \frac{\Delta\nu}{(\nu - \nu_0)^2 + \Delta\nu^2} \right]. \quad (33)$$

It is desired to obtain refraction from absorption, and the absorption term is in the imaginary part of (33). Therefore real and imaginary must be reversed by multiplying by $i\omega$ or $i2\pi\nu$ as discussed in Section 2.5. Since $2\pi\nu$ is effectively constant across a narrow band, (33) need only be multiplied by i to obtain

$$\epsilon_2 + i(\epsilon_1 - \epsilon_0) = \frac{A\nu_0}{2} \left[\frac{\Delta\nu}{(\nu - \nu_0)^2 + \Delta\nu^2} + i \frac{\nu_0 - \nu}{(\nu - \nu_0)^2 + \Delta\nu^2} \right]. \quad (34)$$

Multiplying (34) by $1/\Delta\epsilon$ where $\Delta\epsilon = A\nu_0/2\Delta\nu$, a constant which does not change the real-imaginary relationship, (34) becomes

$$\frac{\epsilon_2}{\Delta\epsilon} + i \frac{(\epsilon_1 - \epsilon_0)}{\Delta\epsilon} = \frac{\Delta\nu^2}{(\nu - \nu_0)^2 + \Delta\nu^2} + i \frac{\Delta\nu(\nu_0 - \nu)}{(\nu - \nu_0)^2 + \Delta\nu^2}. \quad (35)$$

Van Vleck plots $(\epsilon_2/\Delta\epsilon)2\pi \log_{10} e$ and $(\epsilon_1 - \epsilon_0)/\Delta\epsilon$ in his atmospheric absorption study at microwave frequencies (Ref. 7, Fig. 8.2). Equation (35) is also identical with the expression for the impedance of a parallel RLC circuit having a half width of Δf which shows the recurrence of the damped harmonic oscillator problem noted above.

The real and imaginary parts of (35), hereafter designated as A and B , respectively, were arithmetically computed to four significant figures for $\nu_0 = 10^6$, $\Delta\nu^2 = 10^3$. A and B are plotted in Fig. 6(a) on an f (also n) scale chosen for summation convenience in summing by (9). Because of the even and odd symmetry of A and B respectively about the center frequency, only half of the curves are shown.

Amplitude A data for phase summation were taken for n even from 250 to 750 from the four figure computed values of A . However A was cut off linearly from $A = 0.016$ at $n = 258$ to $A = 0$ at $n = 250$, even though A was decreasing very slowly for $n < 250$. In accordance with Section 2.3, phase can then be summed between $f = 249$ and 749 odd. Equation (9) then becomes

$$B(f = 249 \text{ to } 749 \text{ odd}) = \sum_{\substack{n=250 \\ \text{even}}}^{750} A_n \Psi_{500+n-(f+1)} \quad (36)$$

$B(f)$ of (36) was summed on the 7090 computer.

The difference between $B(f)$ summed and four-figure $B(f)$ computed from (35) are plotted as the "Error in Radians — Precision Summation" in Fig. 6(b). The maximum error between $f = 419$ and 499 is only 0.001 radian. For $f < 419$, the error gradually increases. This is due to the arbitrary cutoff of A at $n = 250$ noted above. If a correction is made for this cutoff, the error at $f = 369$ drops from point $a = +0.0022$ radian to point $b = 0.0002$ radian [see Fig. 6(b)]. This shows that the potential overall accuracy of the phase summation is equal to the accuracy of the amplitude data.

In order to illustrate the accuracy of an order of magnitude poorer approximation to A , the summation frequency scale was reduced by a factor of 10 to the scale for f or n marked "Desk Computer Summation." Now A changes an order of magnitude more in frequency interval of $\Delta f = 2$ than on the precision f or n scale. Therefore a better approximation to A is sometimes obtained by taking straight line terminal points off the true A curve. The points selected are indicated and several of the resultant line segments making up the approximation are shown in dotted lines.

The summation performed was

$$\theta(f = 491 \text{ to } 499 \text{ odd}) = \sum_{\substack{n=480 \\ \text{even}}}^{520} A_n \Psi_{500+n-(f+1)} .$$

This summation required 30 minutes with a desk computer and produced the good approximation to the true phase shown on Fig 6(a).

V. VALUE OF THE NUMERICAL PHASE SUMMATION TECHNIQUE

A knowledge of the imaginary as well as the real part of experimentally observed physical phenomena adds a new dimension to the understanding of the phenomena especially when the physical mechanisms involved are only partially understood. Consider for instance the difficulty of solving the time dispersion of reflection at the surface of a mirror as discussed in Ref. 9. This problem was easily solved using the phase tables, with no need for a quantitative knowledge of the physical mechanisms involved.

When the phenomena can be represented by a Lorentzian or Gaussian, as is often the case, the numerical solution of phase is not necessary. For instance a Lorentzian approximation to the common-base current gain of a transistor revealed that the high common-emitter current gain is obtained at the price of a corresponding loss in frequency band.¹² However this approximation was not good enough for later study of VHF transistors. Here a knowledge of the numerical relationship between amplitude and phase made possible an understanding of current gain and phase from simple amplitude measurements only. The results not only prove good for all types of junction transistors but also reveal rather than require information on the gradient of the base layer impurity distribution.¹³ And the computations of delay distortion mentioned in the introduction, although theoretically possible, would have been extremely difficult without a knowledge of the numerical computation of phase.

Finally, consider the potential range of usefulness of the phase tables. It is believed that the phase tables presented in Ref. 6 combined with the phase tables of this paper are sufficient to sum the phase of any minimum phase function occurring in any area of the physical or engineering sciences and in any part of the frequency spectrum.

VI. ACKNOWLEDGMENTS

The author is particularly indebted to Hendrik W. Bode, whose basic theoretical work and straight-line approximation method provided the

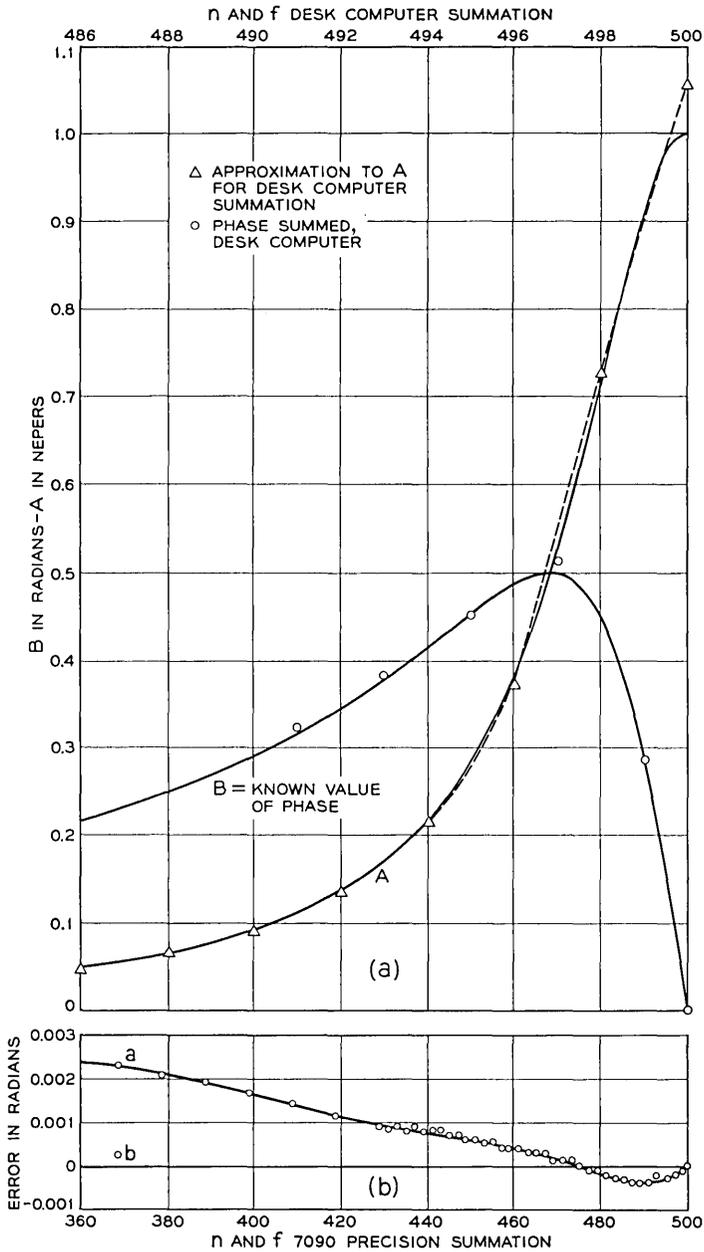


Fig. 6 — (a) Lorentzian line $A + jB = (\epsilon_2/\Delta\epsilon) + i[(\epsilon_1 - \epsilon_0)/\Delta\epsilon]$. (b) Error in precision phase summation using the unit wedge.

foundation for the entire development of the precision numerical phase summation technique. He also wishes to thank W. L. Faust, whose discussions with the author led to a realization of the potential usefulness of the method at optical frequencies; W. R. Bennett, Jr., whose He-Ne maser mode splitting analysis demonstrated the usefulness of the method; Miss Ruth A. Weiss for programming and following through computations of narrow-band summation; J. M. Klein for carrying out a large part of the numerical computations of the new tables; and D. E. McCumber and C. G. B. Garrett for helpful discussions during the writing of this paper.

APPENDIX

Equation (19) of Section 3.2 is reduced to (20) as follows: leaving out the coefficient $2f_{12}/\pi\Delta f$ and treating only the portion in brackets

$$\left[\frac{\Delta f_2}{2f_2 \left(1 - \frac{\Delta f_2}{2f_2}\right)} \log_e \frac{2ef_2 \left(1 - \frac{\Delta f_2}{2f_2}\right)}{\Delta f_2} - \frac{\Delta f_1}{2f_1 \left(1 - \frac{\Delta f_1}{2f_1}\right)} \log_e \frac{2ef_1 \left(1 - \frac{\Delta f_1}{2f_1}\right)}{\Delta f_1} \right]$$

$$= \left[\frac{\Delta f_2}{2f_{12} \left(1 + \frac{\Delta f}{2f_{12}}\right) \left(1 - \frac{\Delta f_2}{2f_{12}}\right)} \log_e \frac{2ef_{12} \left(1 + \frac{\Delta f}{2f_{12}}\right) \left(1 - \frac{\Delta f_2}{2f_{12}}\right)}{\Delta f_2} - \frac{\Delta f_1}{2f_{12} \left(1 - \frac{\Delta f}{2f_{12}}\right) \left(1 - \frac{\Delta f_1}{2f_{12}}\right)} \log_e \frac{2ef_{12} \left(1 - \frac{\Delta f}{2f_{12}}\right) \left(1 - \frac{\Delta f_1}{2f_{12}}\right)}{\Delta f_1} \right]$$

$$= \left[\frac{\Delta f_2}{2f_{12} \left(1 - \frac{\Delta f_2 - \Delta f}{2f_{12}}\right)} \log_e \frac{2ef_{12} \left(1 - \frac{\Delta f_2 - \Delta f}{2f_{12}}\right)}{\Delta f_2} - \frac{\Delta f_1}{2f_{12} \left(1 - \frac{\Delta f_1 + \Delta f}{2f_{12}}\right)} \log_e \frac{2ef_{12} \left(1 - \frac{\Delta f_1 + \Delta f}{2f_{12}}\right)}{\Delta f_1} \right]$$

$$\begin{aligned}
&= \left[\frac{\Delta f_2 \left(1 + \frac{\Delta f_1}{2f_{12}}\right)}{2f_{12}} \log_e \frac{2ef_{12} \left(1 - \frac{\Delta f_1}{2f_{12}}\right)}{\Delta f_2} \right. \\
&\quad \left. - \frac{\Delta f_1 \left(1 + \frac{\Delta f_2}{2f_{12}}\right)}{2f_{12}} \log_e \frac{2ef_{12} \left(1 - \frac{\Delta f_2}{2f_{12}}\right)}{\Delta f_1} \right] \\
&= \frac{\Delta f_2}{2f_{12}} \left(1 + \frac{\Delta f_1}{2f_{12}}\right) \left(-\frac{\Delta f_1}{2f_{12}} + \log_e \frac{2ef_{12}}{\Delta f_2}\right) \\
&\quad - \frac{\Delta f_1}{2f_{12}} \left(1 + \frac{\Delta f_2}{2f_{12}}\right) \left(-\frac{\Delta f_2}{2f_{12}} + \log_e \frac{2ef_{12}}{\Delta f_1}\right) \\
&= \left[\frac{\Delta f_2}{2f_{12}} \left\{ \log_e \frac{2ef_{12}}{\Delta f_2} + \frac{\Delta f_1}{2f_{12}} \left(1 + \log_e \frac{2f_{12}}{\Delta f_2} - 1\right) \right\} \right. \\
&\quad \left. - \frac{\Delta f_1}{2f_{12}} \left\{ \log_e \frac{2ef_{12}}{\Delta f_1} + \frac{\Delta f_2}{2f_{12}} \left(1 + \log_e \frac{2f_{12}}{\Delta f_1} - 1\right) \right\} \right] \\
&= \left[\frac{\Delta f_2}{2f_{12}} \log_e \frac{2ef_{12}}{\Delta f_2} - \frac{\Delta f_1}{2f_{12}} \log_e \frac{2ef_{12}}{\Delta f_1} \right] \\
&\quad + \left[\frac{\Delta f_2 \Delta f_1}{(2f_{12})^2} \log_e \frac{2f_{12}}{\Delta f_2} - \frac{\Delta f_2 \Delta f_1}{(2f_{12})^2} \log_e \frac{2f_{12}}{\Delta f_1} \right].
\end{aligned}$$

When the above is multiplied by the coefficient $2f_{12}/\pi\Delta f$ it becomes:

$$\begin{aligned}
&\frac{1}{\pi\Delta f} \left[\Delta f_2 \log_e \frac{2ef_{12}}{\Delta f_2} - \Delta f_1 \log_e \frac{2ef_{12}}{\Delta f_1} \right] + \frac{\Delta f_1 \Delta f_2}{2\pi\Delta f f_{12}} \log_e \frac{\Delta f_1}{\Delta f_2} \\
&= \frac{1}{\pi\Delta f} \left[\Delta f_2 \log_e \frac{2ef_{12}}{\Delta f_2} - \Delta f_1 \log_e \frac{2ef_{12}}{\Delta f_1} \right] + \frac{\Delta f_1 \Delta f_2}{2\pi\Delta f f_{12}} \log_e \frac{\Delta f_{12} - \frac{\Delta f}{2}}{\Delta f_{12} + \frac{\Delta f}{2}} \\
&= \frac{1}{\pi\Delta f} \left[\Delta f_2 \log_e \frac{2ef_{12}}{\Delta f_2} - \Delta f_1 \log_e \frac{2ef_{12}}{\Delta f_1} \right] - \frac{\Delta f_{12}^2}{2\pi\Delta f f_{12}} \times \frac{\Delta f}{\Delta f_{12}} \\
&\hspace{15em} (\Delta f < 40, \quad 1000 > \Delta f_{12} > 50) \\
&= \frac{1}{\pi\Delta f} \left[\Delta f_2 \log_e \frac{2ef_{12}}{\Delta f_2} - \Delta f_1 \log_e \frac{2ef_{12}}{\Delta f_1} \right] - \frac{\Delta f_{12}}{2\pi f_{12}} \\
&\hspace{15em} (\Delta f < 40, \quad 1000 > \Delta f_{12} > 50)
\end{aligned}$$

which is the value of Φ given in (20). The first term of (20) can be further reduced as follows

$$\begin{aligned} & \frac{1}{\pi \Delta f} \left[\Delta f_2 \log_e \frac{2ef_{12}}{\Delta f_2} - \Delta f_1 \log_e \frac{2ef_{12}}{\Delta f_1} \right] \\ &= \frac{1}{\pi \Delta f} \left[(\Delta f_1 + \Delta f) \log_e \frac{2ef_{12}}{\Delta f_2} - \Delta f_1 \log_e \frac{2ef_{12}}{\Delta f_1} \right] \\ &= \frac{1}{\pi \Delta f} \left[\Delta f \log_e \frac{2ef_{12}}{\Delta f_2} - \Delta f_1 \log_e \frac{2ef_{12}}{\Delta f_1} \times \frac{\Delta f_2}{2ef_{12}} \right] \\ &= \frac{1}{\pi} \left[\log_e \frac{2ef_{12}}{\Delta f_2} - \frac{\Delta f_1}{\Delta f} \log_e \frac{\Delta f_2}{\Delta f_1} \right] \end{aligned}$$

which is the value of Φ given in (21).

Equation (21) can be still further reduced for $\Delta f_{12} > 300$ as follows. Leaving the factor $1/\pi$ and taking only the bracketed terms of (21)

$$\begin{aligned} & \log_e \frac{2ef_{12}}{\Delta f_2} - \frac{\Delta f_1}{\Delta f} \log_e \frac{\Delta f_2}{\Delta f_1} \\ &= \log_e \frac{2ef_{12}}{\Delta f_{12} \left(1 + \frac{\Delta f}{2\Delta f_{12}} \right)} - \frac{\Delta f_1}{\Delta f} \log_e \frac{\Delta f_{12} + \frac{\Delta f}{2}}{\Delta f_{12} - \frac{\Delta f}{2}} \\ &= \log_e \frac{2ef_{12}}{\Delta f_{12}} - \log_e \left(1 + \frac{\Delta f}{2\Delta f_{12}} \right) - \left(\frac{\Delta f_{12}}{\Delta f} - \frac{1}{2} \right) \left[\log_e \left(1 + \frac{\Delta f}{2\Delta f_{12}} \right) \right. \\ & \quad \left. - \log_e \left(1 - \frac{\Delta f}{2\Delta f_{12}} \right) \right] \\ &= \log_e \frac{2ef_{12}}{\Delta f_{12}} - \frac{\Delta f}{2\Delta f_{12}} + \frac{1}{2} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^2 - \frac{1}{3} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^3 + \dots \\ & \quad - \left(\frac{\Delta f_{12}}{\Delta f} - \frac{1}{2} \right) \left[\frac{\Delta f}{2\Delta f_{12}} - \frac{1}{2} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^2 + \frac{1}{3} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^3 - \frac{1}{4} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^4 \right] \\ & \quad \left[+ \frac{\Delta f}{2\Delta f_{12}} + \frac{1}{2} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^2 + \frac{1}{3} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^3 + \frac{1}{4} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^4 \right] \\ &= \log_e \frac{2ef_{12}}{\Delta f_{12}} - \frac{\Delta f}{2\Delta f_{12}} + \frac{1}{2} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^2 - \frac{1}{3} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^3 + \frac{\Delta f}{2\Delta f_{12}} + \frac{1}{3} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^3 \\ & \quad - 1 - \frac{2}{24} \left(\frac{\Delta f}{2\Delta f_{12}} \right)^2 \\ &= \log_e \frac{2f_{12}}{\Delta f_{12}} + \frac{1}{24} \left(\frac{\Delta f}{\Delta f_{12}} \right)^2 \end{aligned}$$

which when multiplied by $1/\pi$ becomes

$$\frac{1}{\pi} \log_e \frac{2f_{12}}{\Delta f_{12}} + \frac{1}{24\pi} \left(\frac{\Delta f}{\Delta f_{12}} \right)^2$$

which is the expression for Φ in (23).

REFERENCES

1. Kramers, H. A., *Atti del Congresso Internazionale dei Fisici, Como*, **2**, 1927, p. 545.
2. Kronig, R. de L., *Jour. Opt. Soc. Am.*, **12**, 1926, p. 547.
3. Carson, John R., *Electric Circuit Theory and the Operational Analysis*, New York, McGraw-Hill, 1926, p. 180. Also see Bush, Vannevar, *Operational Circuit Analysis*, New York, John Wiley, 1929, p. 180.
4. For an extensive bibliography, see Murakami, T., and Corrington, M. S., *RCA Review*, **9**, 1948, pp. 602-631.
5. Bode, Hendrik W., *Network Analysis and Feedback Amplifier Design*, New York, D. Van Nostrand, 1945.
6. Thomas, D. E., *B.S.T.J.*, **26**, July, 1947, pp. 870-899. Reprinted as Bell System Monograph B-1511; also Seven-Figure Tables of Phase of a Semi-Infinite Unit Attenuation Slope, Bell System Monograph 2550.
7. Van Vleck, J. H., *Radiation Laboratory Series*, New York, McGraw-Hill, 1948, Vol. **13**, Ch. 8.
8. Miller, Richard A., and Lopez, Adolfo, Note on the Velocity of Light, *Jour. Opt. Soc. Am.*, **49**, 1959, p. 930.
9. Bennett, W. R., Jr., Hole Burning Effects in a He-Ne Optical Maser, *Phys. Rev.*, **126**, No. 2, 1962, pp. 580-593.
10. Faust, W. L., and Thomas, D. E., to be published.
11. Wood, Robert W., *Physical Optics*, New York, MacMillan Co., 1959, pp. 487-89.
12. Thomas, D. E., *Proc. I.R.E.*, **40**, 1952, pp. 1481-83.
13. Thomas, D. E., and Moll, J. L., *Proc. I.R.E.*, **46**, 1958, pp. 1177-1184.

TABLES III AND IV — TABLES OF PHASE OF A
SEMI-INFINITE UNIT ATTENUATION SLOPE

f/f_0 or f_0/f	Table III $f < f_0$ B in Radians	Table IV $f > f_0$ B in Radians	1st Difference
.999 700	0.784 4618	0.786 3346	.0*281
710	4898	3065	282
720	5180	2783	283
730	5464	2500	284
740	5748	2216	286
.999 750	0.784 6033	0.786 1930	.0*287
760	6320	1643	288
770	6608	1355	289
780	6898	1066	291
790	7189	0775	293
.999 800	0.784 7481	0.786 0482	.0*294
810	7775	0188	295
820	8071	0.785 9893	298
830	8368	9595	299
840	8668	9296	302
.999 850	0.784 8969	0.785 8994	.0*303
860	9273	8691	306
870	9578	8385	308
880	9886	8077	311
890	0.785 0197	7766	314
.999 900	0.785 0510 79	0.785 7452 48	.0*3154
901	0542 33	7420 94	58
902	0573 91	7389 36	61
903	0605 51	7357 75	64
904	0637 15	7326 11	67
.999 905	0.785 0668 83	0.785 7294 44	.0*3171
906	0700 53	7262 73	74
907	0732 27	7230 99	77
908	0764 05	7199 22	81
909	0795 86	7167 41	84
.999 910	0.785 0827 70	0.785 7135 57	.0*3188
911	0859 58	7103 69	92
912	0891 50	7071 77	95
913	0923 45	7039 82	99
914	0955 44	7007 83	3202
.999 915	0.785 0987 46	0.785 6975 81	.0*3207
916	1019 52	6943 74	10
917	1051 62	6911 64	14
918	1083 76	6879 50	17
919	1115 94	6847 33	22
.999 920	0.785 1148 16	0.785 6815 11	.0*3226
921	1180 41	6782 85	29
922	1212 71	6750 56	34
923	1245 05	6718 22	38
924	1277 43	6685 84	42
.999 925	0.785 1309 85	0.785 6653 42	.0*3246
926	1342 31	6620 96	51
927	1374 82	6588 45	55
928	1407 37	6555 90	59
929	1439 96	6523 31	64

TABLES III AND IV — *Continued*

f/f_0 or f_0/f	Table III $f < f_0 B$ in Radians	Table IV $f > f_0 B$ in Radians	1st Difference
.999 930	0.785 1472 60	0.785 6490 67	
931	1505 28	6457 98	.0 ^s 3269
932	1538 01	6425 25	73
933	1570 79	6392 48	77
934	1603 62	6359 65	83
			87
.999 935	0.785 1636 49	0.785 6326 78	
936	1669 41	6293 86	.0 ^s 3292
937	1702 38	6260 89	97
938	1735 40	6227 87	3303
939	1768 48	6194 79	08
			13
.999 940	0.785 1801 60	0.785 6161 67	
941	1834 78	6128 49	.0 ^s 3318
942	1868 01	6095 26	23
943	1901 30	6061 97	29
944	1934 64	6028 63	34
			40
.999 945	0.785 1968 04	0.785 5995 23	
946	2001 50	5961 77	.0 ^s 3346
947	2035 02	5928 25	52
948	2068 59	5894 68	57
949	2102 23	5861 04	64
			70
.999 950	0.785 2135 93	0.785 5827 34	
951	2169 69	5793 57	.0 ^s 3377
952	2203 52	5759 75	82
953	2237 42	5725 85	90
954	2271 38	5691 89	96
			3403
.999 955	0.785 2305 41	0.785 5657 86	
956	2339 51	5623 75	.0 ^s 3411
957	2373 69	5589 58	17
958	2407 94	5555 33	25
959	2442 26	5521 00	33
			40
.999 960	0.785 2476 67	0.785 5486 60	
961	2511 15	5452 12	.0 ^s 3448
962	2545 71	5417 56	56
963	2580 36	5382 91	65
964	2615 09	5348 18	73
			82
.999 965	0.785 2649 91	0.785 5313 36	
966	2684 83	5278 44	.0 ^s 3492
967	2719 83	5243 44	3500
968	2754 93	5208 33	11
969	2790 14	5173 13	20
			30
.999 970	0.785 2825 44	0.785 5137 83	
971	2860 85	5102 42	.0 ^s 3541
972	2896 37	5066 90	52
973	2932 01	5031 26	64
974	2967 76	4995 51	75
			88
.999 975	0.785 3003 63	0.785 4959 63	
976	3039 63	4923 63	.0 ^s 3600
977	3075 77	4887 50	13
978	3112 04	4851 23	27
979	3148 46	4814 81	42
			57

TABLES III AND IV — *Continued*

f/f_0 or f_0/f	Table III $f < f_0$ B in Radians	Table IV $f > f_0$ B in Radians	1st Difference
.999 9800	0.785 3185 03	0.785 4778 24	
01	3188 69	4774 57	.0 ⁵ 0367
02	3192 36	4770 91	67
03	3196 03	4767 24	67
04	3199 70	4763 57	67
.999 9805	0.785 3203 37	0.785 4759 90	
06	3207 05	4756 22	.0 ⁵ 0367
07	3210 72	4752 55	68
08	3214 40	4748 87	68
09	3218 08	4745 19	68
.999 9810	0.785 3221 76	0.785 4741 51	
11	3225 44	4737 83	.0 ⁵ 0368
12	3229 12	4734 15	68
13	3232 81	4730 46	69
14	3236 50	4726 77	69
.999 9815	0.785 3240 18	0.785 4723 08	
16	3243 87	4719 39	.0 ⁵ 0369
17	3247 57	4715 70	69
18	3251 26	4712 01	69
19	3254 96	4708 31	70
.999 9820	0.785 3258 65	0.785 4704 61	
21	3262 35	4700 92	.0 ⁵ 0370
22	3266 05	4697 21	70
23	3269 76	4693 51	70
24	3273 46	4689 81	70
.999 9825	0.785 3277 17	0.785 4686 10	
26	3280 88	4682 39	.0 ⁵ 0371
27	3284 59	4678 68	71
28	3288 30	4674 97	71
29	3292 01	4671 26	71
.999 9830	0.785 3295 73	0.785 4667 54	
31	3299 44	4663 82	.0 ⁵ 0372
32	3303 16	4660 11	72
33	3306 88	4656 38	72
34	3310 61	4652 66	72
.999 9835	0.785 3314 33	0.785 4648 94	
36	3318 06	4645 21	.0 ⁵ 0373
37	3321 79	4641 48	73
38	3325 52	4637 75	73
39	3329 25	4634 02	73
.999 9840	0.785 3332 99	0.785 4630 28	
41	3336 72	4626 54	.0 ⁵ 0374
42	3340 46	4622 81	74
43	3344 20	4619 07	74
44	3347 95	4615 32	74
.999 9845	0.785 3351 69	0.785 4611 58	
46	3355 44	4607 83	.0 ⁵ 0375
47	3359 19	4604 08	75
48	3362 94	4600 33	75
49	3366 69	4596 58	75
			76

TABLES III AND IV — *Continued*

f/f_0 or f_0/f	Table III $f < f_0 B$ in Radians	Table IV $f > f_0 B$ in Radians	1st Difference
.999 9850	0.785 3370 45	0.785 4592 82	
51	3374 20	4589 06	.0 ⁰ 376
52	3377 96	4585 31	76
53	3381 72	4581 54	76
54	3385 49	4577 78	76
			77
.999 9855	0.785 3389 25	0.785 4574 01	
56	3393 02	4570 25	.0 ⁰ 377
57	3396 79	4566 48	77
58	3400 57	4562 70	77
59	3404 34	4558 93	77
			78
.999 9860	0.785 3408 12	0.785 4555 15	
61	3411 90	4551 37	.0 ⁰ 378
62	3415 68	4547 59	78
63	3419 46	4543 81	78
64	3423 25	4540 02	79
			79
.999 9865	0.785 3427 04	0.785 4536 23	
66	3430 83	4532 44	.0 ⁰ 379
67	3434 62	4528 65	79
68	3438 42	4524 85	80
69	3442 22	4521 05	80
			80
.999 9870	0.785 3446 02	0.785 4517 25	
71	3449 82	4513 45	.0 ⁰ 380
72	3453 62	4509 64	81
73	3457 43	4505 84	81
74	3461 24	4502 02	81
			81
.999 9875	0.785 3465 06	0.785 4498 21	
76	3468 87	4494 40	.0 ⁰ 382
77	3472 69	4490 58	82
78	3476 51	4486 76	82
79	3480 33	4482 93	82
			83
.999 9880	0.785 3484 16	0.785 4479 11	
81	3487 99	4475 28	.0 ⁰ 383
82	3491 82	4471 45	83
83	3495 65	4467 61	83
84	3499 49	4463 78	84
			84
.999 9885	0.785 3503 33	0.785 4459 94	
86	3507 17	4456 10	.0 ⁰ 384
87	3511 02	4452 25	84
88	3514 87	4448 40	85
89	3518 72	4444 55	85
			85
.999 9890	0.785 3522 57	0.785 4440 70	
91	3526 43	4436 84	.0 ⁰ 386
92	3530 29	4432 98	86
93	3534 15	4429 12	86
94	3538 01	4425 25	87
			87
.999 9895	0.785 3541 88	0.785 4421 39	
96	3545 75	4417 51	.0 ⁰ 387
97	3549 63	4413 64	87
98	3553 50	4409 76	88
99	3557 39	4405 88	88
			88

TABLES III AND IV — *Continued*

f/f_0 or f_0/f	Table III $f < f_0$ B in Radians	Table IV $f > f_0$ B in Radians	1st Difference
.999 9900	0.785 3561 27	0.785 4402 00	
01	3565 16	4398 11	.0°0388
02	3569 05	4394 22	89
03	3572 94	4390 33	89
04	3576 84	4386 43	89
			90
.999 9905	0.785 3580 74	0.785 4382 53	
06	3584 64	4378 63	.0°0390
07	3588 55	4374 72	91
08	3592 46	4370 81	91
09	3596 37	4366 90	91
			92
.999 9910	0.785 3600 29	0.785 4362 98	
11	3604 21	4359 06	.0°0392
12	3608 13	4355 14	92
13	3612 06	4351 21	93
14	3615 99	4347 28	93
			94
.999 9915	0.785 3619 93	0.785 4343 34	
16	3623 87	4339 40	.0°0394
17	3627 81	4335 46	94
18	3631 76	4331 51	95
19	3635 71	4327 56	95
			95
.999 9920	0.785 3639 66	0.785 4323 61	
21	3643 62	4319 65	.0°0396
22	3647 58	4315 69	96
23	3651 55	4311 72	97
24	3655 52	4307 75	97
			97
.999 9925	0.785 3659 49	0.785 4303 77	
26	3663 47	4299 80	.0°0398
27	3667 46	4295 81	98
28	3671 44	4291 82	99
29	3675 44	4287 83	99
			0400
.999 9930	0.785 3679 43	0.785 4283 84	
31	3683 43	4279 83	.0°0400
32	3687 44	4275 83	01
33	3691 45	4271 82	01
34	3695 46	4267 80	02
			02
.999 9935	0.785 3699 48	0.785 4263 78	
36	3703 51	4259 76	.0°0402
37	3707 54	4255 73	03
38	3711 57	4251 69	03
39	3715 61	4247 65	04
			05
.999 9940	0.785 3719 66	0.785 4243 61	
41	3723 71	4239 56	.0°0405
52	3727 77	4235 50	06
43	3731 83	4231 44	06
44	3735 89	4227 37	07
			07
.999 9945	0.785 3739 97	0.785 4223 30	
46	3744 05	4219 22	.0°0408
47	3748 13	4215 14	08
48	3752 22	4211 05	09
49	3756 32	4206 95	10
			10

TABLES III AND IV — *Continued*

f/f_0 or f_0/f	Table III $f < f_0$ B in Radians	Table IV $f > f_0$ B in Radians	1st Difference
.999 9950	0.785 3760 42	0.785 4202 85	
51	3764 53	4198 74	.0 ⁵ 0411
52	3768 65	4194 62	12
53	3772 77	4190 50	12
54	3776 90	4186 37	13
			14
.999 9955	0.785 3781 03	0.785 4182 24	
56	3785 18	4178 09	.0 ⁵ 0414
57	3789 33	4173 94	15
58	3793 48	4169 78	16
59	3797 65	4165 62	17
			17
.999 9960	0.785 3801 82	0.785 4161 45	
61	3806 00	4157 26	.0 ⁵ 0418
62	3810 19	4153 08	19
63	3814 39	4148 88	20
64	3818 60	4144 67	21
			22
.999 9965	0.785 3822 81	0.785 4140 46	
66	3827 04	4136 23	.0 ⁵ 0422
67	3831 27	4132 00	23
68	3835 51	4127 76	24
69	3839 76	4123 50	25
			26
.999 9970	0.785 3844 03	0.785 4119 24	
71	3848 30	4114 97	.0 ⁵ 0427
72	3852 59	4110 68	28
73	3856 88	4106 39	29
74	3861 19	4102 08	31
			32
.999 9975	0.785 3865 51	0.785 4097 76	
76	3869 85	4093 42	.0 ⁵ 0433
77	3874 19	4089 08	35
78	3878 55	4084 72	36
79	3882 93	4080 34	37
			39
.999 9980	0.785 3887 32	0.785 4075 95	
81	3891 72	4071 55	.0 ⁵ 0441
82	3896 14	4067 12	42
83	3900 58	4062 68	44
84	3905 04	4058 23	46
			48
.999 9985	0.785 3909 52	0.785 4053 75	
86	3914 02	4049 25	.0 ⁵ 0450
87	3918 54	4044 72	52
88	3923 09	4040 18	55
89	3927 67	4035 60	57
			60
.999 9990	0.785 3932 27	0.785 4031 00	
91	3936 90	4026 37	.0 ⁵ 0463
92	3941 57	4021 69	67
93	3946 28	4016 98	71
94	3951 04	4012 23	76
			81
.999 9995	0.785 3955 85	0.785 4007 42	
96	3960 72	4002 55	.0 ⁵ 0487
97	3965 68	3997 59	96
			0506
.999 99980	0.785 3970 74	0.785 3992 53	
985	3973 32	3989 94	.0 ⁵ 0259
990	3975 96	3987 30	64
995	3978 68	3984 58	72
			95
1.000 00000	0.785 3981 63	0.785 3981 63	

TABLE V — Φ IN RADIANS FOR LINE SEGMENT,
 A , OF 1 NEPER; $f_{12} = 10^6$ CPS

Δf_{12}	$\Delta f = 2$	$\Delta f = 4$	$\Delta f = 6$	$\Delta f = 10$	$\Delta f = 20$	$\Delta f = 40$
0	4.936	4.716	4.587	4.424	4.204	3.983
1	4.716	4.674	4.569	4.418	4.202	3.983
2	4.412	4.496	4.510	4.398	4.197	3.981
3	4.275	4.296	4.366	4.363	4.189	3.979
4	4.180	4.192	4.214	4.307	4.177	3.977
5	4.108	4.115	4.128	4.204	4.162	3.973
6	4.050	4.054	4.062	4.097	4.142	3.968
7	4.000	4.003	4.007	4.032	4.116	3.963
8	3.957	3.960	3.964	3.980	4.086	3.957
9	3.920	3.922	3.925	3.937	4.046	3.950
10	3.886	3.888	3.890	3.900	3.983	3.941
11	3.856	3.857	3.859	3.867	3.919	3.932
12	3.828	3.829	3.831	3.837	3.876	3.922
13	3.802	3.803	3.805	3.810	3.841	3.910
14	3.778	3.779	3.781	3.785	3.811	3.897
15	3.756	3.757	3.758	3.762	3.784	3.882
16	3.736	3.736	3.738	3.741	3.760	3.866
17	3.716	3.717	3.718	3.721	3.737	3.847
18	3.698	3.699	3.700	3.703	3.716	3.826
19	3.682	3.682	3.682	3.685	3.697	3.800
20	3.665	3.666	3.666	3.668	3.679	3.762
21	3.650	3.650	3.650	3.652	3.662	3.725
22	3.634	3.635	3.635	3.637	3.646	3.698
23	3.620	3.620	3.621	3.623	3.631	3.676
24	3.606	3.607	3.607	3.609	3.616	3.656
25	3.594	3.594	3.594	3.596	3.603	3.638
26	3.582	3.582	3.582	3.583	3.590	3.621
27	3.570	3.570	3.570	3.571	3.577	3.605
28	3.558	3.558	3.558	3.559	3.565	3.590
29	3.546	3.547	3.547	3.548	3.553	3.576
30	3.536	3.536	3.536	3.537	3.542	3.563
31	3.525	3.526	3.526	3.527	3.531	3.551
32	3.516	3.515	3.516	3.516	3.520	3.539
33	3.506	3.506	3.506	3.507	3.510	3.528
34	3.496	3.496	3.496	3.497	3.501	3.517
35	3.487	3.487	3.487	3.488	3.491	3.506
36	3.478	3.478	3.478	3.479	3.482	3.496
37	3.469	3.469	3.469	3.470	3.473	3.486
38	3.460	3.460	3.461	3.461	3.464	3.477
39	3.452	3.452	3.452	3.453	3.456	3.467
40	3.444	3.444	3.444	3.445	3.448	3.459
41	3.436	3.436	3.436	3.437	3.440	3.450
42	3.429	3.429	3.429	3.429	3.432	3.442
43	3.421	3.422	3.421	3.422	3.424	3.433
44	3.414	3.414	3.414	3.414	3.417	3.426
45	3.407	3.407	3.407	3.407	3.409	3.418
46	3.400	3.400	3.400	3.400	3.402	3.410
47	3.393	3.393	3.393	3.394	3.395	3.403
48	3.386	3.386	3.386	3.387	3.389	3.396
49	3.380	3.380	3.380	3.380	3.382	3.389

TABLE V — *Continued*

Δf_{12}	$\Delta f \leq 20$	$\Delta f = 40$	Δf_{12}	$\Delta f \leq 20$	$\Delta f = 40$
50	3.374	3.382	90	3.186	3.189
51	3.367	3.375	91	3.183	3.185
52	3.361	3.369	92	3.179	3.181
53	3.355	3.362	93	3.176	3.178
54	3.349	3.356	94	3.172	3.174
55	3.343	3.350	95	3.169	3.171
56	3.337	3.344	96	3.166	3.168
57	3.332	3.338	97	3.162	3.164
58	3.326	3.332	98	3.159	3.161
59	3.321	3.327	99	3.156	3.158
60	3.315	3.321	100	3.152	3.155
61	3.310	3.316	101	3.149	3.151
62	3.305	3.310	102	3.146	3.148
63	3.300	3.305	103	3.143	3.145
64	3.295	3.300	104	3.140	3.142
65	3.290	3.295	105	3.137	3.139
66	3.285	3.290	106	3.134	3.136
67	3.280	3.285	107	3.131	3.133
68	3.275	3.280	108	3.128	3.130
69	3.271	3.275	109	3.125	3.127
70	3.266	3.270	110	3.122	3.124
71	3.262	3.266	111	3.119	3.121
72	3.257	3.261	112	3.116	3.118
73	3.253	3.257	113	3.114	3.115
74	3.248	3.252	114	3.111	3.112
75	3.244	3.248	115	3.108	3.110
76	3.240	3.243	116	3.105	3.107
77	3.236	3.239	117	3.103	3.104
78	3.232	3.235	118	3.100	3.101
79	3.228	3.231	119	3.097	3.099
80	3.224	3.227	120	3.094	3.096
81	3.220	3.223	121	3.092	3.093
82	3.216	3.219	122	3.089	3.091
83	3.212	3.215	123	3.087	3.088
84	3.208	3.211	124	3.084	3.085
85	3.204	3.207	125	3.081	3.083
86	3.201	3.203	126	3.079	3.080
87	3.197	3.200	127	3.076	3.078
88	3.193	3.196	128	3.074	3.075
89	3.190	3.192	129	3.071	3.073

Δf_{12}	$\Delta f \leq 40$	Δf_{12}	$f\Delta \leq 40$	Δf_{12}	$\Delta f \leq 40$	Δf_{12}	$\Delta f \leq 40$
130	3.069	140	3.045	150	3.023	160	3.003
131	3.066	141	3.043	151	3.021	161	3.001
132	3.064	142	3.041	152	3.019	162	2.999
133	3.062	143	3.039	153	3.017	163	2.997
134	3.059	144	3.036	154	3.015	164	2.995
135	3.057	145	3.034	155	3.013	165	2.993
136	3.055	146	3.032	156	3.011	166	2.991
137	3.052	147	3.030	157	3.009	167	2.989
138	3.050	148	3.028	158	3.007	168	2.987
139	3.048	149	3.026	159	3.005	169	2.985

TABLE V — *Continued*

Δf_{12}	$\Delta f \leq 40$						
170	2.984	220	2.901	270	2.836	500	2.640
171	2.982	221	2.900	271	2.835	510	2.634
172	2.980	222	2.899	272	2.834	520	2.628
173	2.978	223	2.897	273	2.833	530	2.622
174	2.976	224	2.896	274	2.832	540	2.616
175	2.974	225	2.894	275	2.831	550	2.610
176	2.972	226	2.893	276	2.829	560	2.604
177	2.971	227	2.891	277	2.828	570	2.598
178	2.969	228	2.890	278	2.827	580	2.593
179	2.967	229	2.889	279	2.826	590	2.587
180	2.965	230	2.887	280	2.825	600	2.582
181	2.964	231	2.886	281	2.824	610	2.577
182	2.962	232	2.885	282	2.822	620	2.572
183	2.960	233	2.883	283	2.821	630	2.567
184	2.958	234	2.882	284	2.820	640	2.562
185	2.957	235	2.880	285	2.819	650	2.557
186	2.955	236	2.879	286	2.818	660	2.552
187	2.953	237	2.878	287	2.817	670	2.547
188	2.951	238	2.876	288	2.816	680	2.542
189	2.950	239	2.875	289	2.815	690	2.538
190	2.948	240	2.874	290	2.813	700	2.533
191	2.946	241	2.872	291	2.812	710	2.528
192	2.945	242	2.871	292	2.811	720	2.524
193	2.943	243	2.870	293	2.810	730	2.520
194	2.941	244	2.868	294	2.809	740	2.516
195	2.940	245	2.867	295	2.808	750	2.511
196	2.938	246	2.866	297	2.807	760	2.507
197	2.937	247	2.865	297	2.806	770	2.503
198	2.935	248	2.863	298	2.805	780	2.499
199	2.933	249	2.862	299	2.804	790	2.494
200	2.932	250	2.861	300	2.803	800	2.490
201	2.930	251	2.859	310	2.792	810	2.487
202	2.929	252	2.858	320	2.782	820	2.483
203	2.927	253	2.857	330	2.772	830	2.479
204	2.925	254	2.856	340	2.763	840	2.475
205	2.924	255	2.854	350	2.754	850	2.471
206	2.922	256	2.853	360	2.745	860	2.467
207	2.921	257	2.852	370	2.736	870	2.464
208	2.919	258	2.851	380	2.727	880	2.460
209	2.918	259	2.849	390	2.719	890	2.457
210	2.916	260	2.848	400	2.711	900	2.453
211	2.915	261	2.847	410	2.703	910	2.449
212	2.913	262	2.846	420	2.696	920	2.446
213	2.912	263	2.845	430	2.688	930	2.443
214	2.910	264	2.843	440	2.681	940	2.439
215	2.909	265	2.842	450	2.674	950	2.436
216	2.907	266	2.841	460	2.667	960	2.432
217	2.906	267	2.840	470	2.660	970	2.429
218	2.904	268	2.839	480	2.653	980	2.426
219	2.903	269	2.837	490	2.647	990	2.423

1000 | 2.419

TABLE VI — $\Psi_{500+n - (j+1)}$ IN RADIANS PER NEPER

$500 + n - (j + 1)$	Ψ	$500 + n - (j + 1)$	Ψ	$500 + n - (j + 1)$	Ψ
0	-.00128	100	-.00159	200	-.00212
2	-.00128	102	-.00160	202	-.00214
4	-.00129	104	-.00161	204	-.00216
6	-.00129	106	-.00162	206	-.00218
8	-.00130	108	-.00163	208	-.00219
10	-.00130	110	-.00164	210	-.00220
12	-.00131	112	-.00164	212	-.00222
14	-.00131	114	-.00165	214	-.00224
16	-.00132	116	-.00166	216	-.00225
18	-.00132	118	-.00167	218	-.00227
20	-.00133	120	-.00168	220	-.00229
22	-.00133	122	-.00169	222	-.00230
24	-.00134	124	-.00170	224	-.00232
26	-.00135	126	-.00171	226	-.00234
28	-.00135	128	-.00171	228	-.00235
30	-.00136	130	-.00172	230	-.00237
32	-.00136	132	-.00173	232	-.00239
34	-.00137	134	-.00174	234	-.00240
36	-.00137	136	-.00175	236	-.00242
38	-.00138	138	-.00176	238	-.00244
40	-.00138	140	-.00177	240	-.00246
42	-.00139	142	-.00178	242	-.00248
44	-.00139	144	-.00179	244	-.00250
46	-.00140	146	-.00180	246	-.00252
48	-.00141	148	-.00181	248	-.00254
50	-.00142	150	-.00182	250	-.00256
52	-.00143	152	-.00183	252	-.00258
54	-.00143	154	-.00184	254	-.00260
56	-.00144	156	-.00185	256	-.00262
58	-.00144	158	-.00186	258	-.00264
60	-.00145	160	-.00187	260	-.00266
62	-.00146	162	-.00188	262	-.00269
64	-.00146	164	-.00189	264	-.00271
66	-.00147	166	-.00190	266	-.00273
68	-.00148	168	-.00192	268	-.00275
70	-.00149	170	-.00193	270	-.00277
72	-.00150	172	-.00195	272	-.00280
74	-.00150	174	-.00196	274	-.00283
76	-.00151	176	-.00197	276	-.00286
78	-.00151	178	-.00198	278	-.00288
80	-.00152	180	-.00200	280	-.00291
82	-.00153	182	-.00201	282	-.00294
84	-.00154	184	-.00202	284	-.00297
86	-.00154	186	-.00203	286	-.00300
88	-.00155	188	-.00205	288	-.00302
90	-.00156	190	-.00206	290	-.00305
92	-.00156	192	-.00207	292	-.00308
94	-.00157	194	-.00208	294	-.00311
96	-.00158	196	-.00209	296	-.00314
98	-.00159	198	-.00210	298	-.00317

TABLE VI — *Continued*

$500 + n - (j + 1)$	Ψ	$500 + n - (j + 1)$	Ψ	$500 + n - (j + 1)$	Ψ
300	-.00320	400	-.00644	500	+.52454
302	-.00323	402	-.00657	502	+.23165
304	-.00327	404	-.00671	504	+.13096
306	-.00330	406	-.00685	506	+.09219
308	-.00334	408	-.00700	508	+.07134
310	-.00337	410	-.00717	510	+.05189
312	-.00341	412	-.00732	512	+.04916
314	-.00344	414	-.00748	514	+.04256
316	-.00348	416	-.00767	516	+.03753
318	-.00352	418	-.00785	518	+.03356
320	-.00356	420	-.00805	520	+.03036
322	-.00360	422	-.00827	522	+.02770
324	-.00365	424	-.00848	524	+.02549
326	-.00368	426	-.00872	526	+.02359
328	-.00372	428	-.00895	528	+.02197
330	-.00376	430	-.00923	530	+.02054
332	-.00381	432	-.00950	532	+.01930
334	-.00386	434	-.00979	534	+.01819
336	-.00391	436	-.01010	536	+.01721
338	-.00396	438	-.01043	538	+.01632
340	-.00401	440	-.01079	540	+.01553
342	-.00406	442	-.01116	542	+.01480
344	-.00411	444	-.01158	544	+.01415
346	-.00416	446	-.01201	546	+.01354
348	-.00422	448	-.01248	548	+.01299
350	-.00428	450	-.01299	550	+.01248
352	-.00433	452	-.01354	552	+.01201
354	-.00439	454	-.01415	554	+.01158
356	-.00445	456	-.01480	556	+.01116
358	-.00452	458	-.01553	558	+.01079
360	-.00458	460	-.01632	560	+.01043
362	-.00464	462	-.01721	562	+.01010
364	-.00471	464	-.01819	564	+.00979
366	-.00478	466	-.01930	566	+.00950
368	-.00486	468	-.02054	568	+.00923
370	-.00494	470	-.02197	570	+.00895
372	-.00502	472	-.02359	572	+.00872
374	-.00510	474	-.02549	574	+.00848
376	-.00518	476	-.02770	576	+.00827
378	-.00527	478	-.03036	578	+.00805
380	-.00536	480	-.03356	580	+.00785
382	-.00546	482	-.03753	582	+.00767
384	-.00555	484	-.04256	584	+.00748
386	-.00564	486	-.04916	586	+.00732
388	-.00573	488	-.05819	588	+.00717
390	-.00583	490	-.07134	590	+.00700
392	-.00595	492	-.09219	592	+.00685
394	-.00607	494	-.13096	594	+.00671
396	-.00619	496	-.23165	596	+.00657
398	-.00632	498	-.52454	598	+.00644

TABLE VI — *Continued*

$500 + n - (j+1)$	Ψ	$500 + n - (j+1)$	Ψ	$500 + n - (j+1)$	Ψ
600	+.00632	700	+.00317	800	+.00210
602	+.00619	702	+.00314	802	+.00209
604	+.00607	704	+.00311	804	+.00208
606	+.00595	706	+.00308	806	+.00207
608	+.00583	708	+.00305	808	+.00206
610	+.00573	710	+.00302	810	+.00205
612	+.00564	712	+.00300	812	+.00203
614	+.00555	714	+.00297	814	+.00202
616	+.00546	716	+.00294	816	+.00201
618	+.00536	718	+.00291	818	+.00200
620	+.00527	720	+.00288	820	+.00198
622	+.00518	722	+.00286	822	+.00197
624	+.00510	724	+.00283	824	+.00196
626	+.00502	726	+.00280	826	+.00195
628	+.00494	728	+.00277	828	+.00193
630	+.00486	730	+.00275	830	+.00192
632	+.00478	732	+.00273	832	+.00190
634	+.00471	734	+.00271	834	+.00189
636	+.00464	736	+.00269	836	+.00188
638	+.00458	738	+.00266	838	+.00187
640	+.00452	740	+.00264	840	+.00186
642	+.00445	742	+.00262	842	+.00185
644	+.00439	744	+.00260	844	+.00184
646	+.00433	746	+.00258	846	+.00183
648	+.00428	748	+.00256	848	+.00182
650	+.00422	750	+.00254	850	+.00181
652	+.00416	752	+.00252	852	+.00180
654	+.00411	754	+.00250	854	+.00179
656	+.00406	756	+.00248	856	+.00178
658	+.00401	758	+.00246	858	+.00177
660	+.00396	760	+.00244	860	+.00176
662	+.00391	762	+.00242	862	+.00175
664	+.00386	764	+.00240	864	+.00174
666	+.00381	766	+.00239	866	+.00173
668	+.00376	768	+.00237	868	+.00172
670	+.00372	770	+.00235	870	+.00171
672	+.00368	772	+.00234	872	+.00171
674	+.00365	774	+.00232	874	+.00170
676	+.00360	776	+.00230	876	+.00169
678	+.00356	778	+.00229	878	+.00168
680	+.00352	780	+.00227	880	+.00167
682	+.00348	782	+.00225	882	+.00166
684	+.00344	784	+.00224	884	+.00165
686	+.00341	786	+.00222	886	+.00164
688	+.00337	788	+.00220	888	+.00164
690	+.00334	790	+.00219	890	+.00163
692	+.00330	792	+.00218	892	+.00162
694	+.00327	794	+.00216	894	+.00161
696	+.00323	796	+.00214	896	+.00160
698	+.00320	798	+.00212	898	+.00159

TABLE VI — *Continued*

$500 + n - (f + 1)$	Ψ	$500 + n - (f + 1)$	Ψ	$500 + n - (f + 1)$	Ψ
900	+ .00159	940	+ .00144	980	+ .00132
902	+ .00158	942	+ .00144	982	+ .00132
904	+ .00157	944	+ .00143	984	+ .00131
906	+ .00156	946	+ .00143	986	+ .00131
908	+ .00156	948	+ .00142	988	+ .00130
910	+ .00155	950	+ .00141	990	+ .00130
912	+ .00154	952	+ .00140	992	+ .00129
914	+ .00154	954	+ .00139	994	+ .00129
916	+ .00153	956	+ .00139	996	+ .00128
918	+ .00152	958	+ .00138	998	+ .00128
920	+ .00151	960	+ .00138	1000	+ .00127
922	+ .00151	962	+ .00137		
924	+ .00150	964	+ .00137		
926	+ .00150	966	+ .00136		
928	+ .00149	968	+ .00136		
930	+ .00148	970	+ .00135		
932	+ .00147	972	+ .00135		
934	+ .00146	974	+ .00134		
936	+ .00146	976	+ .00133		
938	+ .00145	978	+ .00133		

TABLE VII — LINE SEGMENT PHASE SUMMATION — TRUNCATED GAUSSIAN SECTION

f	Line $f_{12} = 3$ ab Δf_{12}	Φ_{ab}	Line $f_{12} = 9$ bc Δf_{12}	Φ_{bc}	Line $f_{12} = 71$ $c'd$ Δf_{12}	$\Phi_{c'd}$	Line $f_{12} = 77$ de Δf_{12}	Φ_{de}	$\theta(f)$ radians
0	3	4.366	9	3.925	71	3.262	77	3.236	0.1901
2	1	4.569	7	4.007	69	3.271	75	3.244	0.2185
4	1	4.569	5	4.128	67	3.280	73	3.253	0.2294
6	3	4.366	3	4.366	65	3.290	71	3.262	0.2311
8	5	4.128	1	4.569	63	3.300	69	3.271	0.2254
10	7	4.007	1	4.569	61	3.310	67	3.280	0.2105
12	9	3.925	3	4.366	59	3.321	65	3.290	0.1781
14	11	3.859	5	4.128	57	3.332	63	3.300	0.1433
16	13	3.805	7	4.007	55	3.343	61	3.310	0.1229
18	15	3.758	9	3.925	53	3.355	59	3.321	0.1067
20	17	3.718	11	3.859	51	3.367	57	3.332	0.0931
22	19	3.682	13	3.805	49	3.380	55	3.343	0.0810
24	21	3.650	15	3.758	47	3.393	53	3.355	0.0700
26	23	3.621	17	3.718	45	3.407	51	3.367	0.0599
28	25	3.594	19	3.682	43	3.421	49	3.380	0.0504
30	27	3.570	21	3.650	41	3.436	47	3.393	0.0414
32	29	3.547	23	3.621	39	3.452	45	3.407	0.0328
34	31	3.526	25	3.594	37	3.469	43	3.421	0.0244
36	33	3.506	27	3.570	35	3.487	41	3.436	0.0162
38	35	3.487	29	3.547	33	3.506	39	3.452	0.0081
40	37	3.469	31	3.526	31	3.526	37	3.469	0.0000

$$\theta(f) = A_n (\text{Constant at 0.106 nepers})(\Phi_{ab} + \Phi_{bc} - \Phi_{c'd} - \Phi_{de}).$$

Bounds on Communication

By DAVID SLEPIAN

(Manuscript received February 6, 1963)

Six parameters of importance in many communication systems are: (a) the rate at which digital information is transmitted; (b) the bandwidth of the system; (c) the signal power of the transmitted signals; (d) the noise power of disturbances in transmission; (e) the error probability in digits recovered at the receiver output; (f) the length of time that the transmitter and receiver can store their inputs. These six parameters cannot assume arbitrary values: certain sets of values cannot be realized. In a series of curves, this paper describes the boundary between compatible and incompatible sets of parameter values. In the model studied, it is assumed that the disturbance is additive Gaussian noise with constant power density spectrum in the transmission band.

I. INTRODUCTION

In comparing the performance of communication systems that transmit information by means of signals of limited bandwidth, six quantities descriptive of the system and its environment are of particular importance: (*i*) the rate at which the system transmits information; (*ii*) the bandwidth occupied by the transmission signals; (*iii*) a measure of the power of these signals; (*iv*) a measure of the ambient noise which perturbs the transmitted signals; (*v*) the delay time (caused by the transmitter and receiver) between the introduction of information at the input of the system and the emergence of useful information at the output of the system; (*vi*) a measure of the fidelity with which the information at the output of the system represents the information presented to the input of the system.

To compare the performance of two communication systems in a meaningful manner, it is usually necessary to consider the values of at least these six quantities for the two systems. In general, such a comparison will not yield a simple ordering of the two systems. Two systems may utilize the same bandwidth, introduce the same delay, and operate

in the same noise environment. The first system may transmit information at a greater rate with somewhat better fidelity than the second, but may require much more signal power. Which system is to be judged better then depends on external considerations such as the economics of equipment and the purpose for which communication is being established. These external considerations allow the engineer to assign relative weights or costs to the six quantities in question.

Quite apart from these costs dictated by external considerations that may vary with every conceivable usage of a communication system, it is clearly desirable to know, in the first place, what mutual values of the six quantities can ever be obtained by any means. In order to provide such quantitative information it is necessary to particularize both the model of the communication system and the six descriptive parameters.

In all that follows we shall assume that a discrete message source presents independent equiprobable decimal digits for transmission at the uniform rate R decimal digits (or dits) per second. (The output of any other discrete source having entropy rate R can be encoded into this form.) A transmitter operates on these decimal digits to produce a continuous signal of average power S lying in the frequency band $(0, W)$ cycles/second. The signal produced by the transmitter is perturbed by the addition of independent Gaussian noise of total power N and constant power spectral density N/W in the band $(0, W)$ cycles/second. A receiver operates on the perturbed signal to produce decimal digits at an average rate R symbols/second. When the receiver output symbols and the transmitter input symbols are placed in proper correspondence, the average probability, P_e , that an output symbol be different from the corresponding input symbol will be taken as the measure of fidelity with which the system operates. To perform their coding functions, the transmitter and receiver may each require the internal storage of T seconds of their inputs. We use the dimensionless parameter

$$n = 2WT$$

(that is, T measured in Nyquist intervals) as a measure of the delay or complexity of encoding associated with transmitter and receiver.

Our concern henceforth is with the six quantities R , W , S , N , n , and P_e of this model and with the determination of the boundaries of the region of compatible values for these parameters. The famous capacity formula of Shannon¹ published in 1948, $C = W \log(1 + S/N)$, provides information about this boundary when $n \rightarrow \infty$, i.e., when arbitrarily complicated receiver and transmitter coding operations are allowed. The

astonishing fact that P_e could be made arbitrarily small for certain finite nonzero values of R , W , and S/N by letting $n \rightarrow \infty$, promised the existence of most remarkable and previously unsuspected communication systems. This led Gilbert² and others to compute the values of R , W , S/N , and P_e obtainable with specific transmitters and receivers having fixed delay n and to compare these results with Shannon's formula. The results were disappointing. For all systems examined, even those permitting quite complex encodings ($n = 100$), it was found that to achieve practical values of P_e , S/N had to be at least 6 db more than that given by the capacity formula. The question arose: was this result due to the comparative poorness of the specific systems chosen, or is the approach to the ideal systems described by the capacity formula inherently very slow with increasing n ? For a fixed finite value of n , what values of R , W , S/N and P_e are theoretically attainable?

Some information on this subject for large values of n was given by Rice³ as early as 1950. The question was answered in considerable detail by Shannon in an important paper⁴ in which he presented a number of inequalities that permit rather accurate determination of the region of attainable parameter values for all values of n . Shannon's primary interest here was again in the case of large delay, and he developed asymptotic forms for his inequalities in this case. For small delay, the inequalities involve quite complicated expressions and their numerical evaluation is not a simple matter.

The present paper describes in Appendix A a technique which, by means of an electronic computer, permits highly accurate evaluation of the quantities entering these inequalities. The technique has been used to map out bounds on the compatible region of the six quantities in question over a wide range of parameter values. The results of the computations are presented here in a number of curves which cross plot the quantities in various ways which we hope will be useful to the communication engineer.* In particular, the curves show quantitatively the improvement in communication systems that can be achieved with a given degree of coding (measured by delay). Considerable improvement can be obtained with a small amount of encoding, but to approach within a few db of the capacity formula in general requires extremely complicated systems. The curves also give numerical information concerning the trade-offs of the various parameters. They should provide useful references of comparison for existing communication systems.

* An application of these curves to the problem of determining the threshold in modulation systems that expand bandwidth is given in Ref. 5.

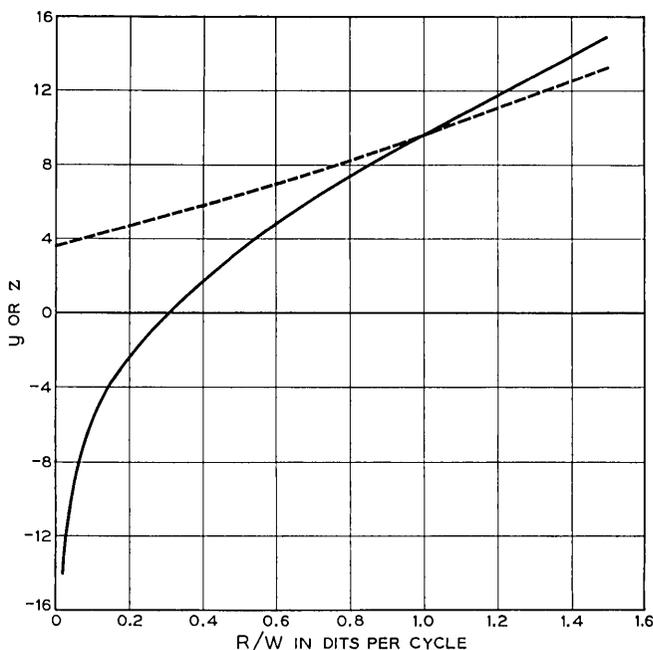


Fig. 1 — Relationship between signal parameters with arbitrarily complex encoding. Solid curve gives $y = 10 \log_{10} S/N$ vs R/W ; dashed curve gives $z = 10 \log_{10} (SW/NR)$ vs R/W .

II. IDEAL SYSTEMS — UNRESTRICTED CODING

The solid curve on Fig. 1 shows a plot of the relation

$$R = W \log_{10} (1 + S/N) \quad (1)$$

in terms of the two dimensionless quantities

$$r = R/W, \quad y = 10 \log_{10} (S/N).$$

This curve can be interpreted* as follows. For values of R , W , S and N corresponding to points above the curve, transmission with arbitrarily small positive values of P_e can be achieved by use of sufficiently complicated coding schemes (sufficiently large finite values of n). For values of R , W , S and N corresponding to points below the curve, P_e is bounded away from zero independently of n . For systems represented by these points, no amount of coding can make the error probability arbitrarily small.

* There are many subtle and thorny points in the argument that permits one to apply the capacity formula to communication systems transmitting continuous signals. Some of these points are discussed in Appendix B.

In many communication situations, the quantity

$$Z = \frac{S/N}{R/W} = \frac{S/R}{N/W}$$

is a useful system parameter. This quantity is the signal energy per dit divided by the noise power per unit bandwidth. From (1),

$$Z = (10^r - 1)/r. \quad (2)$$

The dashed curve of Fig. 1 shows a plot of

$$z = 10 \log_{10} Z$$

vs r as determined by (2). For a given value of r , values of z above the curve are attainable with arbitrarily small positive P_e and finite delay; arbitrarily small positive values of P_e cannot be obtained for z values below the curve with finite delay.

The curves on Fig. 1 describe the relations between R , W , S and N along the intersection of the planes $P_e = 0$, $n = \infty$ with the boundary of the region of mutual compatibility of the six parameters. The intersection of any two other planes, say $P_e = c_1$ and $n = c_2$, with this boundary also determines a curve in the y - r or z - r plane. Unfortunately, the exact form of these curves is not known at present.

III. FINITE n AND NONZERO P_e

To understand fully the assumptions implicit in the remaining curves to be presented here, it is necessary to recall the approach taken by Shannon in Refs. 4 and 6.

Since the signal produced by the transmitter is limited in frequency to the band $(0, W)$ cycles/second, it can (according to the sampling theorem) be thought of as generated by the application of a train of impulses as input to an ideal low-pass filter with cutoff frequency W . The impulses are spaced $1/(2W)$ seconds apart and are of varying amplitude. During a fixed time T , $n = 2WT$ such impulses are applied to the filter. During this same time T , the information source can produce one of $M = 10^{R^T}$ different messages. One method, then, of determining from the output of the information source the train of impulses to be applied to the filter is to provide a dictionary that lists for each of the possible M messages a corresponding sequence of n impulses. The transmitter examines the source output for T seconds and determines which of the M messages was produced. The dictionary is then consulted to obtain the corresponding sequence of n impulses. These impulses are applied at a uniform rate to the filter during the next T seconds. At the end of

this time, the source has produced another message from the list of M messages and the process is repeated. This method of encoding the source is known as block coding of length n .

In a block coding scheme of length n , the average power of the signal produced at the output of the filter depends on the amplitudes of the impulses listed in the encoding dictionary. It is easy to show that each word of the dictionary, i.e., each sequence of n impulses, contributes an energy $d^2/2W$ to the transmitted signal. Here d^2 is the sum of the squares of the amplitudes of the n impulses in question. Since one word is transmitted every T seconds, one method of achieving average power S for the transmitted signal is to require that $d^2 = nS$ for each word of the dictionary. We shall refer to dictionaries of this sort as equal energy block codes.

In Ref. 4, Shannon presents explicit formulae for functions $Q_n(r, Y)$ and $\bar{Q}_n(r, Y)$ which have the following significance. For the communication model under discussion, there exist transmitters and receivers using equal energy block codes of length n such that

$$P_e \leq \bar{Q}_n(R/W, S/N).$$

For every equal energy block code of length n , the system parameters satisfy the inequality

$$P_e \geq Q_n(R/W, S/N).$$

Here P_e is the probability that a transmitted word of the dictionary be decoded incorrectly. The functions Q_n and \bar{Q}_n and their numerical evaluation are discussed further in Appendix A.

Consider now a relationship such as

$$Q_{101}(R/W, S/N) = 10^{-4} \quad (3)$$

which serves to determine S/N as a function of R/W . This relation could be plotted on Fig. 1 with S/N measured in db to yield a curve lying above the solid-line capacity curve shown there. For our purposes, the vertical difference between these two curves is of primary interest. This difference is shown by the bottom solid curve of Fig. 2. Explicitly, the bottom curve of Fig. 2 is a plot of

$$y = 10 \log_{10} (S/N) - 10 \log_{10} (10^{R/W} - 1)$$

vs R/W , where S/N is given in terms of R/W by (3). The bottom dashed curve of Fig. 2 is an analogous display of the relation defined by

$$\bar{Q}_{101}(R/W, S/N) = 10^{-4}.$$

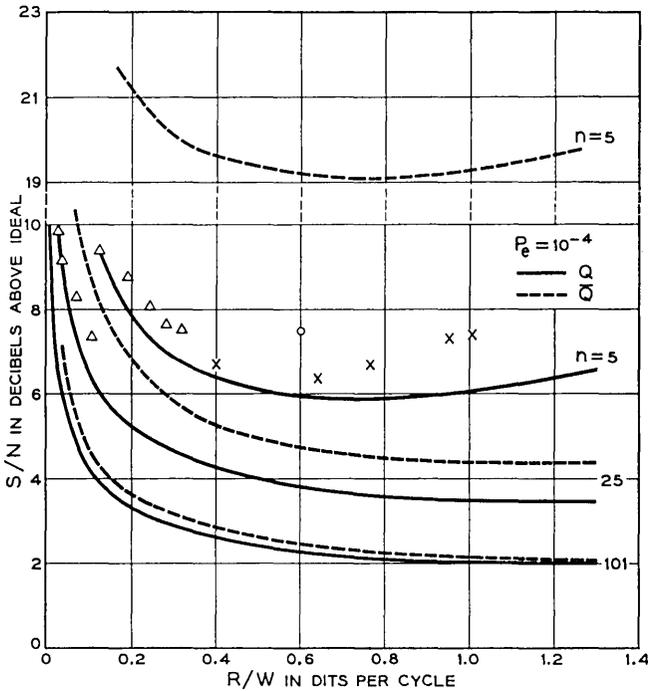


Fig. 2 — Upper and lower bounds (dashed and solid curves, respectively) on S/N needed to achieve word-error probability of 10^{-4} for various values of $n = 2WT$. Circle, triangles, and crosses give performance of some known codes.

The two bottom curves on Fig. 2 have the following significance. For a given value of R/W , there exist equal energy block codes of length 101 that will achieve an error probability of $P_e = 10^{-4}$ with as small a value of S/N as that given by the ordinate of the dashed curve. On the other hand, every equal energy block code of length 101 that achieves an error probability of 10^{-4} must operate with a value of S/N at least as large as the ordinate of the solid curve. The curves thus serve to bound the minimal signal-to-noise ratio with which an error probability of 10^{-4} can be achieved when equal energy block codes of length 101 are employed. The bounds are plotted in db above the signal-to-noise ratio given by the capacity formula, and thus measure the penalty in signal-to-noise ratio that must be paid for restricting the coding ($n = 101$).

The remaining curves on Fig. 2 give analogous results for $n = 5$ and $n = 25$. It is to be noted that the solid and dashed curves are much closer together for large n , than for small n . This effect is shown more clearly on Fig. 3, which was obtained from a cross plot of many curves

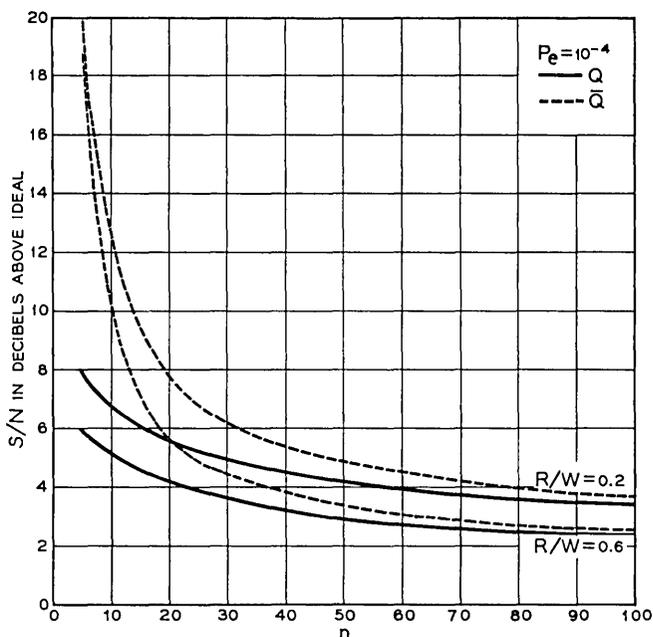


Fig. 3 — Cross-sections of Figure 2 taken for $R/W = 0.2$ and 0.6 .

of the sort shown on Fig. 2. For $n = 101$, there is little practical difference between the two bounds. For small values of n , however, the disparity is great, and the question naturally arises: does the solid curve, or the dashed curve, more nearly represent the minimal signal-to-noise ratio needed to obtain $P_e = 10^{-4}$ with an equal energy block code of fixed length n ?

We believe that the bound obtained from Q is quite close to the minimal attainable S/N even for small n . Indeed, for $n = 5$, we have been able to construct explicit equal energy block codes with a variety of rates whose parameters plot close to the top-most solid line of Fig. 2 when S/N was adjusted to guarantee an error probability not greater than 10^{-4} . The five right-most triangles in the figure locate the performance of certain block codes known as simplex codes [the $(D, D + 1)$ codes of Ref. 2]. The crosses locate the performance of certain new codes to be described in a later paper. The circle gives the performance of 5-bit PCM. The four left-most triangles locate the performance of some simplex codes of block length 25. Apart from these explicit examples that plot near the bounds obtained from Q , there are theoretical considera-

tions which show that \bar{Q} is a very weak bound for small values of n . Henceforth, in this paper we shall deal only with bounds obtained from Q and shall treat the relationship

$$Q_n(R/W, S/N) = P_e \quad (4)$$

as the defining equation of the boundary of the region of compatible values of R , W , S , N , P_e and n for equal energy block codes.

IV. DISCUSSION OF RESULTS

Figs. 4, 5 and 6 give plots of S/N vs R/W as determined from (4) for various values of P_e and n . The ordinates here, as in Fig. 2, are given in db above capacity, i.e., in db above the solid curve of Fig. 1. One advantage of this representation is that the ordinates of Figs. 4, 5 and 6 may also be interpreted as values of Z , the latter now being measured in db above the capacity value given by the dashed line of Fig. 1.

From Figs. 4, 5 and 6, it is apparent that for a fixed rate and fixed error probability modest amounts of coding (small values of n) can produce a significant reduction in signal power, but that the return for increased encoding diminishes rapidly. This is seen more clearly from the cross plot given on Fig. 7.

The improvement in performance that can be obtained by encoding can also be expressed in terms of decreased error probability for a fixed rate and signal-to-noise ratio as is shown in Fig. 8.

An interesting feature of Figs. 4, 5 and 6 is the minimum value clearly

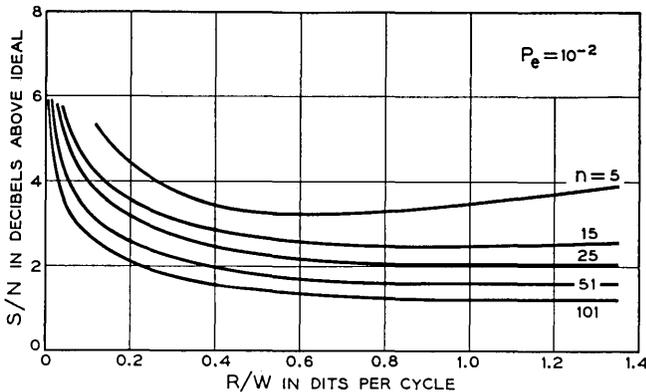


Fig. 4 — Minimum possible S/N to attain word-error probability of 10^{-2} for various values of R/W and n .

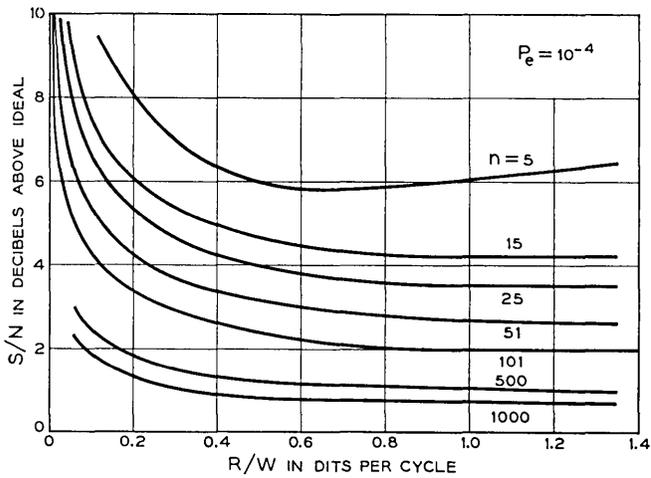


Fig. 5 — Minimum possible S/N to attain word-error probability of 10^{-4} for various values of R/W and n .

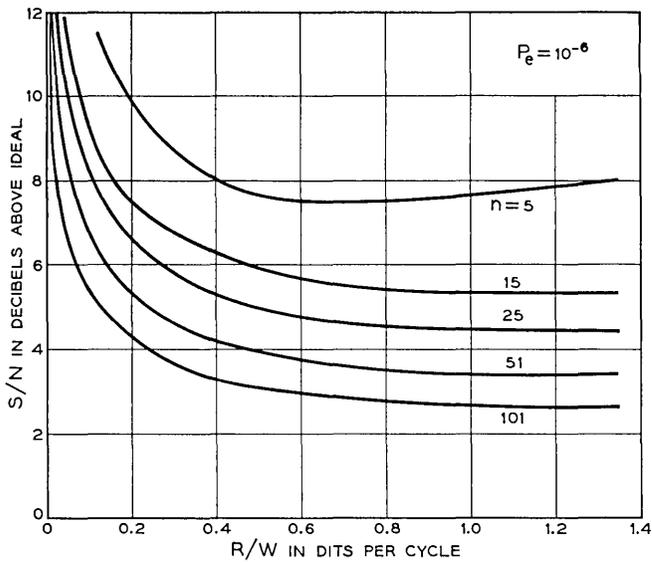


Fig. 6 — Minimum possible S/N to attain word-error probability of 10^{-6} for various values of R/W and n .

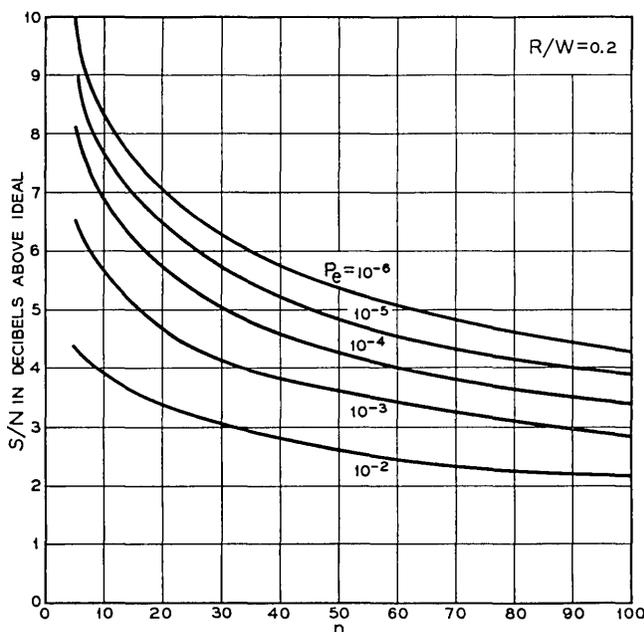


FIG. 7 — Cross-plot of Figs. 4, 5 and 6 showing (for $R/W = 0.2$) decrease in S/N needed to achieve a given word-error probability as n is increased.

evident on the $n = 5$ curves. It is not hard to show (see Appendix C) that for all values of n , the curves obtained from (4) as plotted on these figures must rise indefinitely with increasing R/W . For equal energy block codes, there is, for any fixed P_e and n , a best value of R/W in the sense of minimizing the additional signal-to-noise ratio needed above that given by the channel capacity formula. When the curves of Figs. 4, 5 and 6 are plotted on a graph such as Fig. 1 with absolute S/N as ordinate, the curves are monotone increasing but eventually for large R/W depart further and further above the capacity formula curve. This phenomenon is due to the restriction imposed here that all code words of the dictionary have the same energy, a restriction likely to be realized in practice. This point is discussed further in Appendix D.

Another way of presenting (4) that shows the departure from the ideal system of the capacity formula that results with equal energy block codes of restricted length is shown in Fig. 9. Fix P_e and n . Then from (4), a given value of $r = R/W$ determines a corresponding signal-to-noise ratio, S/N . From the capacity formula, using this value of S/N it is possible to achieve any desired P_e with a rate per bandwidth $\bar{r} =$

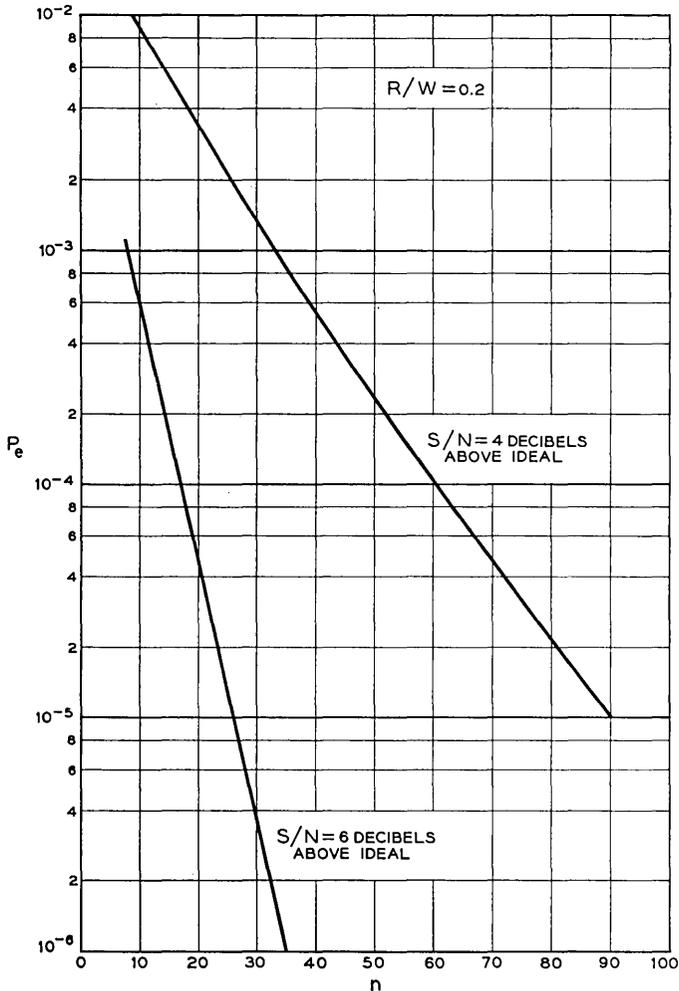


Fig. 8 — Word-error probability vs n for given R/W and S/N above ideal for best possible equal-energy codes.

$\log_{10} (1 + S/N)$ by sufficiently complex encoding. The ratio r/\bar{r} then measures the price paid in lost rate due to restricting the amount of encoding. The solid curves on Fig. 9 were obtained from Q and give upper bounds on r/\bar{r} for equal energy block codes; the dashed curves derived from \bar{Q} give lower bounds for this ratio. It can be shown (see Appendix C) that the solid curves approach $(n - 1)/n$ asymptotically with increasing R/W .

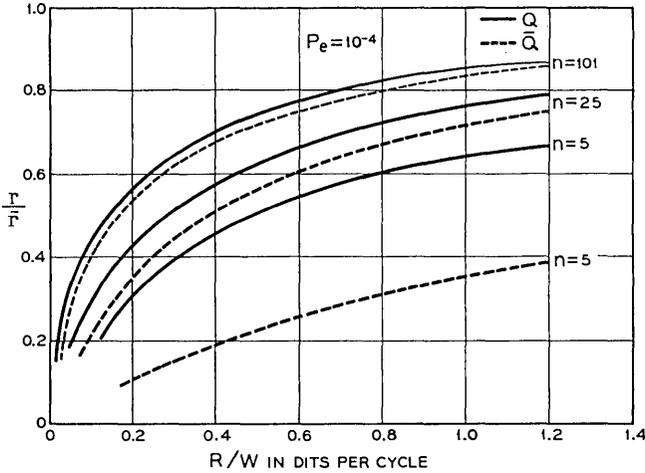


Fig. 9 — Upper and lower bounds (solid and dashed curves, respectively) on fractional loss in rate, r/\bar{r} , due to finite encoding. Loss plotted vs R/W for fixed n and P_e .

Yet another way of viewing the bounds is given on Fig. 10. Here, for a fixed signal-to-noise ratio and a fixed error probability, the improvement in signaling rate that can be obtained by increasing the length of equal energy codes is shown. It is seen, for example, that even with signal to noise ratios as high as 20 db, one cannot achieve 75 per cent of the ideal rate with equal energy codes of length less than 15 when the prescribed error probability is 10^{-6} . The $S/N = \infty$ curve is given by $r/\bar{r} = (n - 1)/n$. That this limiting curve is different from unity is again due to the fact that the bounds used here are those for equal energy codes. If restricted energy codes were used, (see Section V) the limiting curve corresponding to $S/N = \infty$ would be $r/\bar{r} = 1$.

V. CONCLUDING REMARKS

The exact computation of Q_n that was carried out here allows one to test the range of validity of Shannon's asymptotic expressions for this quantity. On plots such as Figs. 4, 5 and 6, his formula* (4) of Ref. 4 gives curves in very close agreement with those shown for $n = 101$. At $n = 25$ the error is about 0.1 db at large rates and 0.3 db at small rates. This formula was used to compute the curves for $n = 500$ and 1000 shown on Fig. 5. Although it involves only elementary functions,

* This formula contains a misprint. The printed version must be multiplied by $-G$ to be corrected.

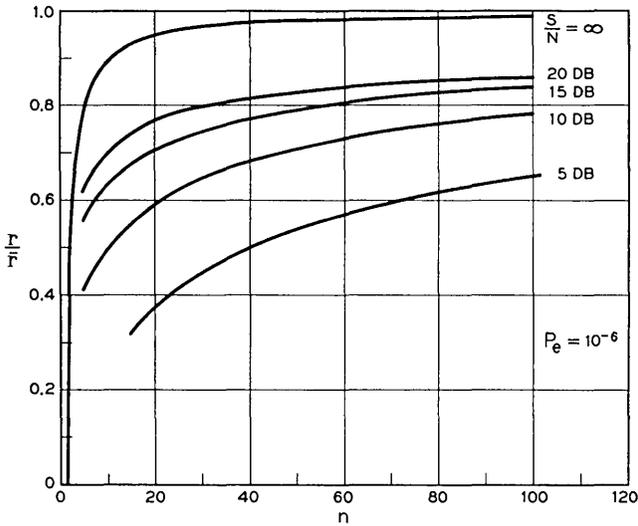


Fig. 10 — Rate loss of Fig. 9 plotted vs n for fixed S/N and P_e .

the formula is quite complicated, and for extensive computations machine methods are desirable. For moderate or small values of n , exact values of Q can be obtained by the method of Appendix A with comparable ease. Shannon's elementary asymptotic formula (73) of Ref. 4 has also been evaluated. For $n = 500$ and 1000 , it gives values that agree with the curves of Fig. 5 to about 0.1 db for $R/W > 0.5$. For small rates it gives values 0.5 db too large. The accuracy of the formula diminishes rapidly as n is decreased below 100.

The bounds presented here were obtained for communication systems using equal energy block codes of fixed length. It is, of course, possible to signal using block codes that have words of differing energy. One code of this sort of particular interest that is treated by Shannon in Ref. 4, Section XIII is the *restricted energy block code*. In these codes, each word of the dictionary contributes energy ST or less to the transmitted signal, i.e., for each code word $d^2 \leq nS$. Note that for these codes S is no longer the average signal power, but rather the maximum contribution to the signal power by any code word.

For any communication system with parameters R , W , S , N using a restricted energy block code of length n , Shannon showed that the average error probability, P_e' , for a decoded word is bounded below by

$$P_e' \geq Q_{n+1} \left(\frac{n}{n+1} \frac{R}{W}, \frac{S}{N} \right). \quad (5)$$

For any fixed value of R/W , as n becomes large this lower bound approaches the one already given for equal energy block codes, and so asymptotically (in n) one can do no better with restricted energy codes than with equal energy codes. However, for any fixed value of n , as R/W becomes large the lower bounds for the two classes of codes behave very differently, and indeed it is easy to argue that in this limit restricted energy codes are superior to equal energy codes. This point is discussed further in Appendix D.

The solid curves of Fig. 11 are those already shown in Fig. 6. The dashed curves were obtained from the lower bound (5) for restricted energy block codes. These dashed curves approach the horizontal asymptotes indicated at the right. From the figure it is seen that for $R/W < 0.6$ and $n \geq 25$ the bounds for restricted energy codes differ from those for equal energy codes by less than 0.2 db. For small values of n , the dashed curves lie below the solid ones even for small rates.

It should be pointed out in closing that the error probability P_e used throughout these calculations is the probability that a word of the block code be improperly identified when a maximum likelihood receiver is used. This is not in general the probability that an individual decoded decimal digit be in error but rather an upper bound to this quantity. For large n , a single code word is decoded into many decimal digits

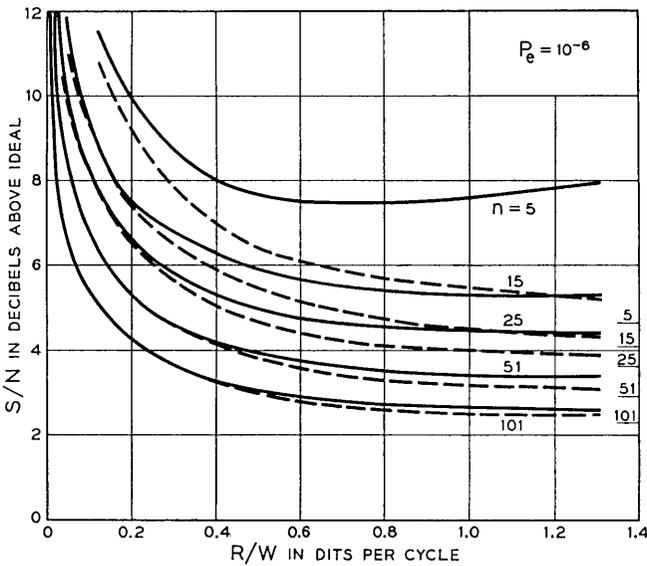


Fig. 11 — Comparison of bounds for equal-energy codes and restricted-energy codes.

The received code word may be incorrectly identified and yet decoded into a block of decimal digits many of which are correct. When large block codes are used and P_e is small, errors in the decoded stream of decimal digits are not distributed uniformly. Many successive groups of decimal digits, each containing RT digits, will be error free. Then a single block of RT digits will be produced that contains from one to RT erroneous digits. This bunching of errors may, in certain applications, be a serious drawback to the use of block coding.

VI. ACKNOWLEDGMENT

All of the key ideas contained in this paper are due to C. E. Shannon. The author's contribution has been one of exposition and the reduction to numerical form of certain inequalities contained in Ref. 4. In this latter task he was greatly aided by Miss Helen Arnold, who programmed the algorithm of Appendix A for the IBM 7090 and oversaw the computations. The author is indebted to E. Wolman for many helpful discussions of this work.

APPENDIX A

Computation of Q and \bar{Q}

Our notation is similar to Shannon's⁴ and we here adopt his geometrical point of view:

S = signal power (each signal vector is of length \sqrt{nS});

N = noise power (variance N in each dimension);

$A = \sqrt{S/N}$ = signal-to-noise "amplitude" ratio;

n = number of dimensions;

M = number of signal vectors;

$\Omega(\theta)$ = solid angle in n -space of a cone of half-angle θ , or area of unit n -sphere cut out by the cone;

$Q(\theta)$ = probability of a point X in n -space, at distance $A\sqrt{n}$ from the origin, being moved outside a circular cone of half-angle θ with vertex at the origin O and axis OX (the perturbation is assumed spherical Gaussian with unit variance in all dimensions);

θ_1 = angle such that $M\Omega(\theta_1) = \Omega(\pi)$.

Shannon shows [his equation (20)] that

$$Q(\theta_1) \leq P_e \leq Q(\theta_1) - \frac{1}{\Omega(\theta_1)} \int_0^{\theta_1} \Omega(\theta) dQ(\theta),$$

where P_e is the error probability of the best equal energy M -vector code

in n -space used with signal-to-noise ratio A . We proceed to discuss the evaluation of these quantities.

As shown by Shannon [his equation (21)]

$$\Omega(\theta) = \frac{(n-1)\pi^{(n-1)/2}}{\Gamma\left(\frac{n+1}{2}\right)} \int_0^\theta \sin^{n-2}\xi \, d\xi. \tag{6}$$

The surface $\Omega(\pi)$ of the unit n -sphere has area

$$\Omega(\pi) = \frac{n\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

A change of variable $\sin^2 \xi = t$ shows that

$$\begin{aligned} \frac{\Omega(\theta)}{\Omega(\pi)} &= \frac{1}{2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n-1}{2}\right)} \int_0^{\sin^2\theta} t^{1/2(n-1)-1} (1-t)^{1/2-1} dt \\ &= \frac{1}{2} I_{\sin^2\theta} \left(\frac{n-1}{2}, \frac{1}{2} \right), \end{aligned}$$

where $I_\alpha(p,q)$ is Pearson's incomplete beta function.⁷ Thus θ_1 is given by

$$\frac{2}{M} = I_{\sin^2\theta_1} \left(\frac{n-1}{2}, \frac{1}{2} \right). \tag{7}$$

The rate is related to θ_1 by

$$\frac{R}{W} = \frac{2}{n} \log_{10} M. \tag{8}$$

To evaluate $Q(\theta)$, it is convenient to use n -dimensional cylindrical coordinates with origin located on the axis of the cone at a distance

$$l \equiv \sqrt{nA}$$

from the vertex and within the cone. The z - or rotational axis of the coordinate system coincides with the axis of the cone and is oriented so that the vertex of the cone has z -coordinate $-l$. Denote distance from the z -axis by r . Then an element of "area" distant r from the axis and having radial dimension dr and axial dimension dz sweeps out volume

$$\frac{(n-1)\pi^{(n-1)/2} r^{n-2} dr dz}{\Gamma\left(\frac{n+1}{2}\right)},$$

when rotated about the z -axis. One has therefore

$$Q_n = Q(\theta) = \int_0^\infty \int_{-\infty}^{r^{\alpha-l}} dz \frac{\exp[-\frac{1}{2}(r^2 + z^2)]}{(2\pi)^{n/2}} \frac{(n-1)\pi^{(n-1)/2} r^{n-2}}{\Gamma\left(\frac{n+1}{2}\right)}, \quad (9)$$

where we have set

$$\alpha = \cot \theta.$$

Now set

$$c_n = \sqrt{\frac{2}{\pi}} \frac{1}{2^{(n-1)/2} \Gamma\left(\frac{n-1}{2}\right)}.$$

One then has

$$\begin{aligned} \frac{Q_n}{c_n} &= \int_0^\infty dr r \exp\left(-\frac{1}{2}r^2\right) \left\{ r^{n-3} \int_{-\infty}^{r^{\alpha-l}} dz \exp\left(-\frac{1}{2}z^2\right) \right. \\ &= -\exp\left(-\frac{1}{2}r^2\right) \left\{ r^{n-3} \int_{-\infty}^{r^{\alpha-l}} dz \exp\left(-\frac{1}{2}z^2\right) \right\} \Big|_0^\infty \\ &\quad + (n-3) \int_0^\infty dr \exp\left(-\frac{1}{2}r^2\right) r^{n-4} \int_{-\infty}^{r^{\alpha-l}} dz \exp\left(-\frac{1}{2}z^2\right) \\ &\quad + \alpha \int_0^\infty dr \exp\left(-\frac{1}{2}r^2/2\right) r^{n-3} \exp\left[-\frac{(\alpha r - l)^2}{2}\right] \\ &= (n-3) \frac{Q_{n-2}}{c_{n-2}} + \alpha J_{n-2}, \quad n > 3, \end{aligned} \quad (10)$$

on integrating by parts. Here

$$\begin{aligned} J_n &= \int_0^\infty dr r^{n-1} \exp\left[-\frac{(1+\alpha^2)r^2 - 2\alpha l r + l^2}{2}\right] \\ &= \frac{1}{1+\alpha^2} \int_0^\infty dr r^{n-2} [(1+\alpha^2)r - \alpha l] \exp\left[\frac{(1+\alpha^2)r^2 - 2\alpha l r + l^2}{2}\right] \\ &\quad + \frac{\alpha l}{1+\alpha^2} \int_0^\infty dr r^{n-2} \exp\left[-\frac{(1+\alpha^2)r^2 - 2\alpha l r + l^2}{2}\right] \\ &= \frac{\alpha l}{1+\alpha^2} J_{n-1} - \frac{1}{1+\alpha^2} r^{n-2} \exp\left[-\frac{(1+\alpha^2)r^2 - 2\alpha l r + l^2}{2}\right] \Big|_0^\infty \\ &\quad + \frac{n-2}{1+\alpha^2} \int_0^\infty dr r^{n-3} \exp\left[-\frac{(1+\alpha^2)r^2 - 2\alpha l r + l^2}{2}\right] \end{aligned}$$

$$= \frac{\alpha l}{1 + \alpha^2} J_{n-1} + \frac{n - 2}{1 + \alpha^2} J_{n-2}, \quad n > 2. \tag{11}$$

Now set

$$G_n = c_{n+2} J_n \csc \theta, \quad b_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}, \quad \xi = \frac{l}{\sqrt{2}}.$$

One has from (10) and (11)

$$\left. \begin{aligned} Q_n &= Q_{n-2} + \cos \theta G_{n-2}, & n > 3 \\ G_n &= \xi \cos \theta \sin \theta b_n G_{n-1} + \frac{n-2}{n-1} \sin^2 \theta G_{n-2}, & n > 2 \\ b_n &= \frac{n-2}{n-1} b_{n-2}, & n > 2. \end{aligned} \right\} \tag{12}$$

The initial values

$$\begin{aligned} b_1 &= \sqrt{\pi}, & b_2 &= \frac{2}{\sqrt{\pi}}, \\ G_1 &= \frac{1}{2} \exp(-\xi^2 \sin^2 \theta) \operatorname{erfc}(-\xi \cos \theta) \\ G_2 &= \frac{1}{\pi} \sin \theta e^{-\xi^2} + \frac{2\xi}{\sqrt{\pi}} \sin \theta \cos \theta G_1, \\ Q_3 &= \frac{1}{2} \operatorname{erfc}(\xi) + \cos \theta G_1, \end{aligned}$$

with

$$\operatorname{erfc}(x) \equiv \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$$

permit one to compute $Q_n(\theta)$ for odd n from the recurrence (12). Since $0 \leq \theta \leq \pi/2$, all quantities involved are positive.

The curves of Figs. 4, 5, and 6 were obtained as follows. With θ_1 fixed in value $Q_5, Q_{15}, Q_{25}, Q_{51}$ and Q_{101} were determined as functions of ξ by repeated application of the recurrence. A given $Q_n(\theta_1)$ was then expressed as a function of the signal-to-noise ratio, A^2 , by the relation $\xi = A\sqrt{n/2}$. Values of A^2 for which $Q_n(\theta_1)$ took the values $10^{-2}, 10^{-4}, 10^{-6}$ were determined graphically. The corresponding rate was found from (7) and (8). Repetition of the process for different values of θ_1 permits plotting the curves.

An integration by parts and (6) allow Shannon's upper bound to be

written in the form

$$\bar{Q} = \frac{M}{\sqrt{\pi}b_{n-1}} \int_0^{\theta_1} Q_n(\theta) \sin^{n-2} \theta \, d\theta. \quad (13)$$

Curves based on \bar{Q} , such as shown on Fig. 2, were obtained by using the recurrence (12) to obtain values of $Q_n(\theta)$ for a fixed ξ . The integral in (13) was evaluated numerically using a trapezoidal formula with 150 points of evaluation for the integrand. Values of ξ and θ_1 were expressed in terms of R/W and S/N as already explained.

APPENDIX B

The theorems and formulae of Shannon's Information Theory are statements about certain *mathematical* constructs. In order to make useful inferences from these formulae about *physical* communication systems, it is necessary to examine the sense in which the mathematical model approximates the behavior of the key elements of the physical system. At best, the correspondence between mathematical and physical entities is only a close approximation: the "true" theorems of the mathematical model, when stated in physical terms, are only "partial truths."

The formula

$$C = (\alpha/2) \log_{10} (1 + S/N) \text{ dits/second} \quad (14)$$

gives the capacity of the following *mathematical* channel. Real numbers are chosen at a transmitting point at the rate α numbers per second. Each number chosen is transmitted to the receiving point, but is perturbed by an additive Gaussian variate, so that the i th transmitted real number, s_i , is received as $s_i + x_i$. The x_i are assumed independent Gaussian random variables with the same variance N . The transmitted sequence satisfies the constraint

$$\lim_{K \rightarrow \infty} \frac{1}{2K} \sum_{-K}^K s_i^2 = S.$$

(The reader should consult Ref. 8, Chapter 9, for a more careful, rigorous definition of this channel and a precise mathematical interpretation of the capacity formula.)

The foregoing description of the channel is essentially that given by Shannon in Ref. 4. The channel is discrete in time; there is no mention of bandlimited continuous functions of a time variable defined on the real line. Within the mathematical theory, there is no question of the

validity of (14) for the capacity of the discrete time channel described nor of the validity of Shannon's bounds for the error probability attainable with block codes of finite length. The problem is to justify the application of these formulae derived for a discrete time mathematical channel to physical communication systems employing "continuous" signals of "bandwidth" W .

I have placed quotation marks around the words continuous and bandwidth to call attention to the fact that these two concepts have no well-accepted operational definitions in terms of experiments in the real world. They are again part of another strictly *mathematical* model that is used to describe signals of the physical world. The elements of this mathematical model are the real number continuum, functions and Fourier analysis. The correspondence between these elements and observables of the laboratory (meter readings, etc.) is again an approximation — a very good one in many circumstances, but a poor one in many others. It is meaningless to ask if the reading of a meter in the laboratory is a rational number or an irrational one, or if the trace seen on an oscilloscope is a continuous function in the sense used in the mathematical model. Within the mathematical model, there are many notions introduced for which one cannot easily find meaningful counterparts in the real world of the laboratory. The asymptotic behavior of spectra at infinity is such an example. One must be very suspicious of the utility of applying in the real world formulae derived from the mathematical models which are sensitive to assumptions about those concepts of the model that have no operationally defined counterparts in the laboratory.

It is evident that a good case for applying (14) to real communication systems can be made if one can justify the statement

"In the laboratory, using signals of duration T and bandwidth W , we can communicate about $2WT$ numbers and no more." (15)

Perhaps it would be simplest to take this statement as a basic axiom for practical communication engineering and justify it by experiment (with "bandwidth," "number," etc. suitably defined in operational terms). It is intellectually more satisfying, however, to be able to derive it from the mathematical models that have served so well to describe signals in other circumstances.

The approach taken by Shannon in Ref. 6 and paraphrased here at the beginning of Section III is one method of deriving statements in the spirit of (15) from the usual mathematical model of signals and spectra. This approach is reasonably satisfactory in justifying the fact that for very large T one can transmit $2WT$ numbers using signals of (mathe-

mathematical) bandwidth W and nominal duration T . From it one can argue rather convincingly that rates arbitrarily close to those given by the capacity formula can be achieved with arbitrarily small error probability using (mathematically) bandlimited functions for signaling. Using this approach, however, it is difficult to make a convincing argument that one cannot exceed capacity or that Shannon's bounds Q_n and \bar{Q}_n have any significance for channels employing (mathematical) bandlimited functions.

The difficulty here lies in the fact that mathematical bandlimited functions are entire functions and hence perfectly predictable for all time from knowledge over any finite interval. If one allows all the usual mathematical operations, the receiver, on the basis of observing the bandlimited signal plus noise in an arbitrarily short time interval, could extrapolate this function for all time and obtain sample values at an arbitrarily great rate.

The heart of the dilemma presented here lies in the fact that the mathematical specification that a signal be bandlimited is a statement about concepts of the model that have no well defined physical counterpart—namely, the behavior of spectra at infinity. The sampling theorem, unfortunately, requires an assumption about this nonphysically interpretable part of the mathematical model.

Yet, one feels that in the real world something like (15) holds with laboratory meanings for bandwidth. If so, this should be derivable from the mathematical model of functions and Fourier analysis without making assumptions in the model about such nonphysical entities as the behavior of spectra at infinity. A result of this sort is indeed the content of an important theorem recently published by Pollak and Landau.⁹ Their results are too complex to discuss in detail here. The main point is that within the classical model of functions and Fourier analysis they define a suitable class of functions that are "limited" in both time and frequency. The definition of this class does not entail specification of spectral behavior at infinity. The specification, when translated to physical terms, involves only an assumption about one's ability to measure energy, and the correspondence between their class and laboratory bandlimited signals defined in an operational way is easy to make. They prove that in an appropriate sense this class of functions is $2WT$ -dimensional. From this, a form of statement (15) results which is, I believe, the best justification on theoretical grounds to date of this important postulate.

Quite apart from this difficulty of justifying (15), there are, of course, many other ways in which the mathematical model only approximates

the behavior of equipment in the laboratory: measurement errors prevent one from specifying real numbers meaningfully by more than a finite number of significant figures; disturbances are not truly Gaussian; etc., etc. The attainment of arbitrarily small error by sufficient encoding in the mathematical theory entails a delicate balance between many quantities which only approximate their physical counterparts. One should not believe that real communication systems can be built which will signal at fixed rates with arbitrarily small error. Somewhere, for large enough n , the mathematical model fails to describe adequately the physical realities. How large is this n ? This is a very difficult question. My engineering judgment is that the results given on the curves of this paper for n up to 100 might conceivably be achieved with real communication systems. Until we have learned to describe and instrument optimal codes of this size, I am safe from experimental contradiction. Today, this time seems remote.

APPENDIX C

We show here that if

$$Q_n(R/W, S/N) = P_e \quad (16)$$

and

$$\bar{r} = \log(1 + S/N) \quad (17)$$

then, with n and P_e fixed ($0 < P_e < 1$),

$$\lim_{R/W \rightarrow \infty} \frac{R/W}{\bar{r}} = \frac{n-1}{n}.$$

Referring to (7) and (8) we see that if $R/W \rightarrow \infty$, then $\theta_1 \rightarrow 0$. Indeed, for small values of θ_1 , one can easily develop the incomplete beta function to obtain

$$\begin{aligned} \frac{R}{\bar{W}} = \frac{2}{n} \left[\ln(n-1) \beta\left(\frac{n-1}{2}, \frac{1}{2}\right) \right. \\ \left. - (n-1) \ln \sin \theta_1 + 0(\theta_1^2) \right] \log_{10} e. \end{aligned} \quad (18)$$

Here $\beta(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ as usual.

It is now convenient to write equation (9) as

$$\frac{Q_n(\theta_1)}{d} = \int_0^\infty dr \int_{-\infty}^{r \cot \theta_1 - \sqrt{n}A} dz r^{n-2} \exp[-(r^2 + z^2)/2]$$

where

$$d = \frac{(n-1) \pi^{(n-1)/2}}{(2\pi)^{n/2} \Gamma\left(\frac{n+1}{2}\right)}$$

and as before we adopt the abbreviation $A = \sqrt{S/N}$. The region of integration is shaded in Fig. 12.

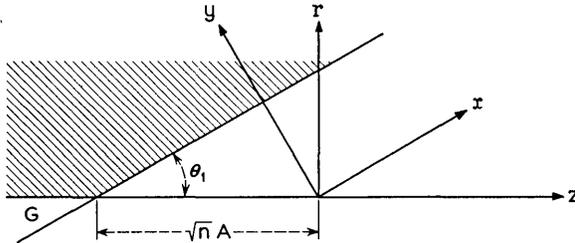


Fig. 12 — Integration region and coordinate transformation.

To investigate the behavior of Q_n as $\theta_1 \rightarrow 0$, it is convenient to transform the integral by the rotation

$$z = x \cos \theta_1 - y \sin \theta_1$$

$$r = x \sin \theta_1 + y \cos \theta_1$$

and to write the result as the integral over the region $y \geq \sqrt{n}A \sin \theta_1$ minus the integral over the region G indicated in the figure.

$$\begin{aligned} \frac{Q_n}{d} &= \int_{\sqrt{n}A \sin \theta_1}^{\infty} dy \int_{-\infty}^{\infty} dx (x \sin \theta_1 + y \cos \theta_1)^{n-2} \exp \left[-\frac{x^2 + y^2}{2} \right] \\ &\quad - \int_G \int dy dx (x \sin \theta_1 + y \cos \theta_1)^{n-2} \exp \left[-\frac{x^2 + y^2}{2} \right]. \end{aligned}$$

With $A \geq 0$, the integral over G vanishes as $\theta_1 \rightarrow 0$, so

$$\begin{aligned} \frac{Q_n}{d} &\rightarrow (\cos \theta_1)^{n-2} \int_{\sqrt{n}A \sin \theta_1}^{\infty} dy \int_{-\infty}^{\infty} dx (y + x \tan \theta_1)^{n-2} \exp \left[-\frac{x^2 + y^2}{2} \right] \\ &\rightarrow \int_{\sqrt{n}A \sin \theta_1}^{\infty} dy y^{n-2} \exp(-y^2/2) \int_{-\infty}^{\infty} dx \exp(-x^2/2) \\ &= \sqrt{2\pi} \int_{\sqrt{n}A \sin \theta_1}^{\infty} dy y^{n-2} \exp[-y^2/2] + o(\theta_1). \end{aligned}$$

One thus finds that if $A\theta_1 \rightarrow \infty$, $Q_n/d \rightarrow 0$ whereas if $A\theta_1 \rightarrow 0$, $Q_n \rightarrow 1$.

To maintain (16), therefore, we must have $A\theta_1 = \alpha + o(\theta_1)$ where $0 < \alpha < \infty$, or

$$A \sim (\alpha/\theta_1). \quad (19)$$

Equations (17) and (18) now give

$$\begin{aligned} \lim_{R/W \rightarrow \infty} \frac{R/W}{\bar{r}} &= \lim_{\theta_1 \rightarrow 0} \frac{\frac{2}{n} \left[\ln(n-1)\beta \left(\frac{n-1}{2}, \frac{1}{2} \right) - (n-1) \ln \sin \theta_1 \right]}{\ln \left(1 + \frac{\alpha^2}{\theta_1^2} \right)} \\ &= \frac{n-1}{n} \end{aligned}$$

as was to be shown.

The preceding considerations also allow one to show directly that the curves of Figs. 4, 5 and 6 rise indefinitely as $R/W \rightarrow \infty$. For a given R/W , denote by A_i^2 the corresponding signal-to-noise ratio obtained from the capacity formula, so that $R/W = \log(1 + A_i^2)$. Then $A_i^2 \sim 10^{R/W}$ as $\theta_1 \rightarrow 0$. From (18) one finds,

$$A_i^2 \sim \left[\frac{(n-1)\beta}{\sin^{n-1} \theta_1} \right]^{2/n}.$$

Using (19), one then has

$$A^2/A_i^2 \sim c/\theta_1^{2/n}$$

with c a positive constant. As $R/W \rightarrow \infty$, $\theta_1 \rightarrow 0$ and $A^2/A_i^2 \rightarrow \infty$. The logarithm of this latter ratio is plotted on Figs. 4, 5, and 6.

APPENDIX D

Each word of a block code dictionary is a sequence of n real numbers which may be regarded as a point in an n -dimensional Euclidean space. The points of an equal energy block code all lie on the surface of a hypersphere of radius \sqrt{nS} with center at the origin. The words of a restricted energy block code all lie on the surface or within such a sphere. In this geometric picture, the effect of the noise in the channel can be visualized by surrounding each word of the code by a sphere of radius \sqrt{nN} centered at the word. Due to the noise on the channel, a received word lies on the average at a distance \sqrt{nN} from the corresponding transmitted word. If the code is to have a small average error probability, the noise spheres surrounding the words of the code must not overlap too much. On the other hand, to achieve a large rate, it is necessary to have many words in the dictionary.

The volume of a sphere of radius r in n -space is proportional to r^n . The fraction of the volume of such a sphere that lies external to the concentric sphere of radius αr , $0 < \alpha < 1$ is therefore

$$\frac{r^n - (\alpha r)^n}{r^n} = 1 - \alpha^n.$$

For large enough n , then, almost all the volume of the sphere lies near its surface. For example, if $n \geq 460$, then at least 99 per cent of the volume of the sphere lies within a thin skin of the surface whose thickness is 1 per cent of the radius of the sphere.

Suppose now that N and S are fixed, and consider the problem of placing code words on or within the sphere of radius \sqrt{nS} so that the spheres of radius \sqrt{nN} surrounding each code word do not overlap appreciably. The radius of these noise spheres is a fixed fraction, $\sqrt{N/S}$, of the radius of the large sphere of radius \sqrt{nS} . As n becomes large, almost all of the volume of the large sphere lies within a skin of the surface of fractional thickness much less than $\sqrt{N/S}$. It is not surprising, then, that little is to be gained by placing code words interior to the large sphere. Indeed, Shannon's bounds prove that in the limit $n \rightarrow \infty$ restricted energy block codes give no better performance than equal energy block codes.

In contrast now consider the situation when n and S are fixed and R/W becomes large. As we seek to place more and more code words on or within the sphere of radius \sqrt{nS} , the noise power N must be continuously decreased to prevent the noise spheres surrounding the code words from overlapping. Ultimately, for large enough rates, N must be made so small that the radii of these noise spheres is very small compared to the thickness of the skin of the sphere of radius \sqrt{nS} containing most of its volume. It then becomes possible to pack appreciable numbers of code words interior to this sphere and restricted energy codes then give better performance than equal energy codes.

The asymptotic behavior of the dashed curves of Fig. 11 can readily be deduced from the bound (5) and the material of Appendix C. The curves are given by

$$P_e = Q_{n+1} \left(\frac{n}{n+1} \frac{R}{W}, \frac{S}{N} \right).$$

To maintain $0 < P_e < 1$, we find as in the derivation of (19) that

$$A \sim (\alpha/\theta_1)$$

where α is given by

$$\begin{aligned}
 P_e &= \frac{n\pi^{n/2}}{(2\pi)^{n/2}\Gamma\left(\frac{n}{2} + 1\right)} \sqrt{2\pi} \int_{\alpha\sqrt{n+1}}^{\infty} dy y^{n-1} \exp(-y^2/2) \\
 &= \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_{\frac{\alpha^2(n+1)}{2}}^{\infty} dt t^{(n/2)-1} e^{-t}.
 \end{aligned}$$

In the right member of (18), replace n by $n + 1$; in the left member, replace R/W by $[n/(n + 1)](R/W)$. There results

$$\frac{R}{W} \sim \log \frac{\left[n\beta \left(\frac{n}{2}, \frac{1}{2} \right) \right]^{2/n}}{\sin^2 \theta_1}.$$

It follows then that

$$A^2/A_i^2 \sim \frac{\alpha^2}{\left[n\beta \left(\frac{n}{2}, \frac{1}{2} \right) \right]^{2/n}}$$

so that

$$10 \log \frac{A^2}{A_i^2} \sim 20 \left\{ \log \alpha - \frac{1}{n} \log \left[n\beta \left(\frac{n}{2}, \frac{1}{2} \right) \right] \right\}.$$

This latter value is the horizontal asymptote for the dashed curves of Fig. 11.

REFERENCES

1. Shannon, C. E., A Mathematical Theory of Communication, B.S.T.J., **27**, July and October, 1948, pp. 379-424 and 623-657.
2. Gilbert, E. N., A Comparison of Signalling Alphabets, B.S.T.J., **31**, May, 1952, pp. 504-522.
3. Rice, S. O., Communication in the Presence of Noise — Probability of Error for Two Encoding Schemes, B.S.T.J., **29**, January, 1950, pp. 60-93.
4. Shannon, C. E., Probability of Error for Optimal Codes in a Gaussian Channel, B.S.T.J., **38**, May, 1959, pp. 611-656.
5. Slepian, D., The Threshold Effect in Modulation Systems that Expand Bandwidth, I.R.E. Trans. Inf. Theory, Vol. **IT-8**, No. 5, September, 1962, pp. 122-127.
6. Shannon, C. E., Communication in the Presence of Noise, Proc. I.R.E., **37**, January, 1949, pp. 10-21.
7. Pearson, K., *Tables of the Incomplete Beta-Function*, Cambridge University Press, 1934.
8. Wolfowitz, J., *Coding Theorems of Information Theory*, Springer-Verlag, Berlin, 1961.
9. Landau, H. J., and Pollak, H. O., Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—III: The Dimension of the Space of Essentially Time- and Band-Limited Signals, B.S.T.J., **41**, July, 1962, pp. 1295-1336.

Analysis of Delay in Mathematical Switching Models for Data Systems

By D. G. HAENSCHKE

(Manuscript received August 17, 1962)

Traffic delay, caused by temporary all-lines-busy conditions, is analyzed for three mathematical switching models. They are classified as "address camp-on," "retrial," and "message storage" models. The models are designed to permit a study of basic traffic theoretical problems encountered in the rapidly growing field of data communications, but they are not identical with any of the existing data switching systems. Each model assumes that a message is switched only through one switching center which must establish connections via line groups to one or more addressed receiving stations, i.e., each model contains only a single switching center. Numerical results for the average delay on all messages are obtained on the IBM 7090 computer.

I. INTRODUCTION

Switching centers can be used to link together communication lines for the transmission of data between a variety of business machines and computers. Due to randomness in the required interconnections a switching center may occasionally not find an idle line to a particular receiving station, so that a delay can occur. More than one method can be followed when a switching center finds all lines to a receiving station busy. Some switching models appear to obtain lines to the addressed receiving stations in a shorter time than other switching models. This means that with one switching model a given delay requirement can be met with fewer lines than with another switching model. This is not to say that the model which would render a given grade of service with the least number of lines also is the most desirable from an economic point of view, because delay is only one factor which enters into the choice between data switching systems. Components of a switching system, such

as lines, memory, etc. do not bear the same price tag, and minimizing the number of components of one kind does not ensure economical efficiency.

Interest in the particular traffic engineering problems of data switching has been present for at least 15 years. Yet, traffic engineering work was mainly concentrated on classical telephone trunking problems, and a variety of such fundamental problems have been worked out. Some of the data switching systems which are being studied or are now in use cannot be analyzed by standard mathematical approaches of traffic theory because the operating conditions differ from those of the mathematical models used in the analysis of classical telephone problems. The understanding of the fundamental traffic theoretical problems encountered in data switching is a prerequisite for an exact mathematical analysis of message delay in data systems. The fundamental problems need to be studied on simplified models which lend themselves best to mathematical treatment and, therefore, will not be identical with any of the present data systems in use. We have constructed for study three hypothetical models which we call "address camp-on," "retrial," and "message storage" models, and have analyzed message delay for each of them.

Message delay is defined as the delay between initial request by the switching center for a line, and the moment the message is released from the switching center for transmission. The switching center handles messages in a manner described by one of the three switching models. The delay is caused by temporary all-lines-busy conditions in the line groups which connect the switching center with the addressed receiving stations. This type of traffic delay must not be confused with the total delay from the time a message is ready at the data source and the time the message is actually received at a destination. No account is taken of messages which are switched through more than one switching center in tandem.

This study, then, shall not be looked upon as an attempt to make a choice between switching systems, since such a choice cannot be based solely on the delay performance of mathematical models. A true comparison between switching systems must include other factors, as for instance the cost of memory and logic, loading of transmitters, and loading of incoming lines, all of which are neglected here.

II. DESCRIPTION OF MATHEMATICAL SWITCHING MODELS

The following describes each of the three switching models. The description is preceded by an outline of features which are common to each model. The mathematical derivations given in the appendices and the delay curves are based on these models.

2.1 *Common Features*

Think of a data source feeding messages into a switching center that has a large number of line groups radiating from it (see Fig. 1).

Each line group connects one, and only one, receiving station with the switching center. A receiving station is capable of receiving from all lines in its group simultaneously. There might be one or more lines per group, but each group contains the same number of lines c . Full access is given to each line in a group. A message is said to have A addresses when a copy of the message must be transmitted over A different line groups to A different receiving stations. The number of addresses per message remains constant for all messages. The switching center is responsible for transmitting a copy of the message to each of the addressed receiving stations. The addresses of a message are chosen at random from a large number of possible receiving stations. This permits us to assume that all line groups are independent of each other. Messages are originated and addressed in such a way that a_I , the information load offered to a group, is the same for every group. The information load is defined as the number of first, i.e., unrepeated, message attempts which are expected to be generated during an interval equal to one average message length. First attempts are made Poisson distributed in time, meaning that the probability that exactly k first attempts are generated during an interval of length t is given by

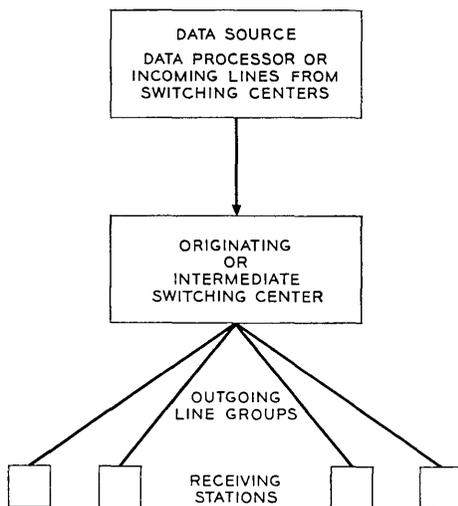


Fig. 1 — Switching model.

$$\Pr (k,t) = \frac{e^{-a_1 t} (a_1 t)^k}{k!},$$

in which t is in units of the average message length. The length of all messages is exponentially distributed with mean 1. In the mathematical derivations the average message length is taken as the unit of time. An exponential message length distribution is chosen, since it is believed that it will serve as a good approximation in a larger number of practical cases than a constant length. The instant a new or repeated message is originated, the switching mechanism begins to hunt for an idle line to each of the addressed receiving stations. No delay is imposed by the switching mechanism itself. Each new message is eventually delivered to the respective receiving stations, i.e., no messages are lost. The system is in statistical equilibrium, which is to say that the system is in the steady state such that the average number of messages in the system during any long interval of time remains constant.

When all lines in a group to one or more of the receiving stations are busy we must find a way of delaying delivery. The camp-on and storage models assume that blocked requests form queues at the switching center. The retrial model assumes that blocked requests are withdrawn from the switching center and reoffered at a later time.

The reader has doubtlessly observed the very simplified and idealized set of common features on which the switching models are based to permit mathematical analysis. The same applies to the features which are unique to each of the three models.

2.2 *Address Camp-On Model*

When a line group to an addressed receiving station is blocked, the request for service in this group will camp-on and wait in the order of arrival until a line is assigned by the mechanism which scans continuously for idle lines. The assignment of available lines to waiting requests is done on a "first come, first served" basis. When a line is assigned it is immediately made busy. The message, however, is not released from the data source until lines to all addressed receiving stations are secured. When the last line is obtained, the message content is released and transmitted simultaneously to each of the addressed receiving stations, after which the lines are released. The holding time of a line in the camp-on model is made up of the sum of two random variables: namely, the exponentially distributed message length and the time spent waiting until lines are secured to all addressed receiving stations.

2.3 *Retrial Model*

In the retrial model, a message is released from the data source only when lines are found at the switching center to all addressed receiving stations. If even one line group is blocked, the message is not transmitted to any of the addressed receiving stations and is temporarily cleared from the switching mechanism without making any lines busy. A blocked message is reoffered any number of times from the data source after a constant time interval τ , until an idle line is found simultaneously to each addressed receiving station. At a retrial of a blocked message, an attempt to seize an idle line is made in the same groups as at the previous attempt. The message delay is determined by the number of attempts made and by the length of the constant retrial interval τ . The holding time of a line in the retrial model is equal to the exponentially distributed message length.

Another way of making retrials is to let the delay in the delivery of the message content to any one addressed station be independent of the delay to the other addresses of a multiaddress message. In this case, a message having A addresses would be considered to consist of A one-address messages and the delay would be that given for the one-address case of the retrial model.

2.4 *Message Storage Model*

Message storage is analysed on a model in which requests for lines in a busy group form queues in the order of arrival. In the multiaddress case, some addresses of the message may find their line groups busy while other addresses may obtain lines to the addressed stations with no delay. The model assumes that the message is released with no delay to stations which are not blocked, and that the delivery to a blocked station is delayed only until the instant a line is found by the switching mechanism which scans continuously for idle lines. As in the retrial model, the line holding time is equal to the exponentially distributed message length.

III. METHOD OF ANALYSIS

The delay performance is analysed as messages are switched through one switching center which employs one of the three described switching models. It should be pointed out that the results obtained here apply only to the mathematical models used. All approximations mentioned in the analysis are approximations of the model to which they refer.

The mathematical analysis of the delay performance of the address camp-on model is given in Appendix A. The problem is to find the total occupancy of the outgoing line groups. In the multiaddress case, outgoing lines can be held busy in excess of the message length. This excess holding time increases the load carried from the useful information load a_I to a total load a_T . The excess holding time is the average length of time between line seizure and the time lines are found for all addresses of the message. Erlang delay probability is used by the introduction of an approximation which assumes that the total holding time of an outgoing line is exponentially distributed. No explicit expression is derived for the total load a_T . Solutions for a_T are found by solving (7) and (10) of Appendix A in an iterative computer program. The average delay follows from (11).

In Appendix B the mathematical analysis is given for the retrial model. The retrial method has been under consideration for application in both military and commercial data systems, and this method is also used in voice telephone communications. The mathematical analysis of delay in systems in which blocked attempts are reoffered is one of the fundamental traffic problems for which an exact solution is not available. The prospect of using the retrial method in data systems emphasizes the need to treat such systems analytically. The analysis given here is not exact because a number of approximations had to be introduced to obtain numerical results. Since the retrial method is a basically unsolved problem it must first be studied in its simplest form, which exists for the case of one address per message. Considerable effort, therefore, is spent in Appendix B on the discussion of the one-address case. Our approach to the retrial problem is to find approximations for the unconditional state probability of finding i lines in a group of c lines busy, $0 \leq i \leq c$. Then, approximations are found for the conditional probabilities of finding i lines busy at $t_0 + \tau$, when the state of the group is known at t_0 , $t_0 - \tau$, $t_0 - 2\tau$, etc. The delay for the one-address case follows from (23) of Appendix B. For the three-address case the delay is computed from (31) in a Markov process which is in itself an approximation of the retrial problem since it accounts only for a first-order dependency.

The basic problem in the retrial model is to find approximations of the conditional probabilities mentioned above. These are obtained by integrating a set of differential equations (16), using a line request rate $\omega(t)$ which by itself is conditioned on previous states of the line group and, therefore, is dependent on time. The line request rate $\omega(t)$ appears as a coefficient in (16). Since $\omega(t)$ can be expressed only as a function of

solutions to (16), we cannot find $\omega(t)$ explicitly, but must compute it in a long process of progressive iterations. It will become apparent from Appendix B that not all approximations made can be clearly justified, but the results obtained are sufficiently accurate for comparison with other switching models. Some of the approximations appear critical for short retrial intervals τ , particularly when c is small. The amount of effort and computer time spent on solving the retrial problem analytically is not necessarily less than the amount of effort and time spent by simulation. The problem is by no means solved, but it is hoped that by this analytical approach the way is paved toward a more complete analysis of retrial systems.

For message storage, the average delay can be determined by the well-established methods of traffic theory developed by A. K. Erlang. These are outlined in Appendix C. The average delay for the storage model is computed from (32) of Appendix C, and no approximations need to be made.

IV. RESULTS AND CONCLUSIONS

For a fixed amount of information load, each switching model produces different delays. This means that some switching models must be operated at lower occupancy than others to ensure that delays encountered will not exceed the desired maximum. The delays shown below for each switching model do not necessarily keep their relationships in respect to each other when messages are switched through several switching centers in tandem.

The results of computations for one address per message are shown in Figs. 2 and 3 for one and ten lines per group, respectively. Figs. 4 and 5 show similar results for three addresses per message and one and ten lines per group, respectively. The "information occupancy" in these figures is numerically equal to the information load offered to the line group divided by the number of lines per group, i.e., a_I/c . The term "occupancy" refers to the percentage of time a line is occupied on the average. The fraction of time a line is actually utilized for the transmission of information, then, is equal to a_I/c , so that we may also call "information occupancy" the "line utilization."

First let us discuss the address camp-on model. This method offers the advantage that error correction can be performed on multilink connections on an end-to-end basis because the message content remains in storage at the data source until a connection is set up to all addressed receiving stations. The camp-on model also is of interest because storage

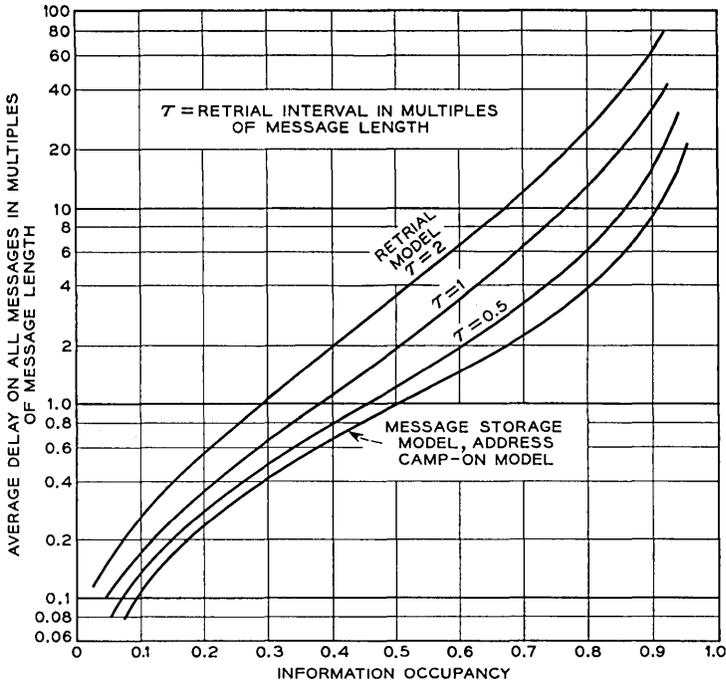


Fig. 2 — Average delay on all messages vs information occupancy; 1 line per group, 1 address per message.

at an intermediate switching center need be provided only for the address portion of a blocked message and this might have some economic advantages over other switching models. For one address per message there is, in theory, no difference in the traffic delay performance of the address camp-on and message storage models when the message is switched only once. It should be remembered that in the camp-on model an intermediate switching center keeps the incoming lines busy in excess of the message length for the duration of a delay which, for a given information load a_T , increases the actual load carried. Because we consider only single-switched messages, no account is taken here of this type of line loading.

As was mentioned before, the total holding time of an outgoing line in the address camp-on model is made up by the excess holding time, which is the time spent waiting for other addresses to find lines, and by the actual message length. In Fig. 6 we show the total occupancy a_T/c versus the actual information occupancy or line utilization a_I/c for three

addresses per message. We see that the total occupancy approaches 100 per cent at a surprisingly low information load. This is due to the fact that the excess holding time increases the load on the outgoing line groups, which in turn increases delays and thus brings about longer excess holding times. This makes the camp-on model unusable beyond certain intolerably low levels of line utilization. For instance, in the three-address case, line utilization must be limited to about 14 per cent or 60 per cent for line group sizes of $c = 1$ or $c = 10$, respectively. It can be seen in Fig. 6 that beyond this point the total occupancy blows up and with it the delay imposed on a message. A similar result was

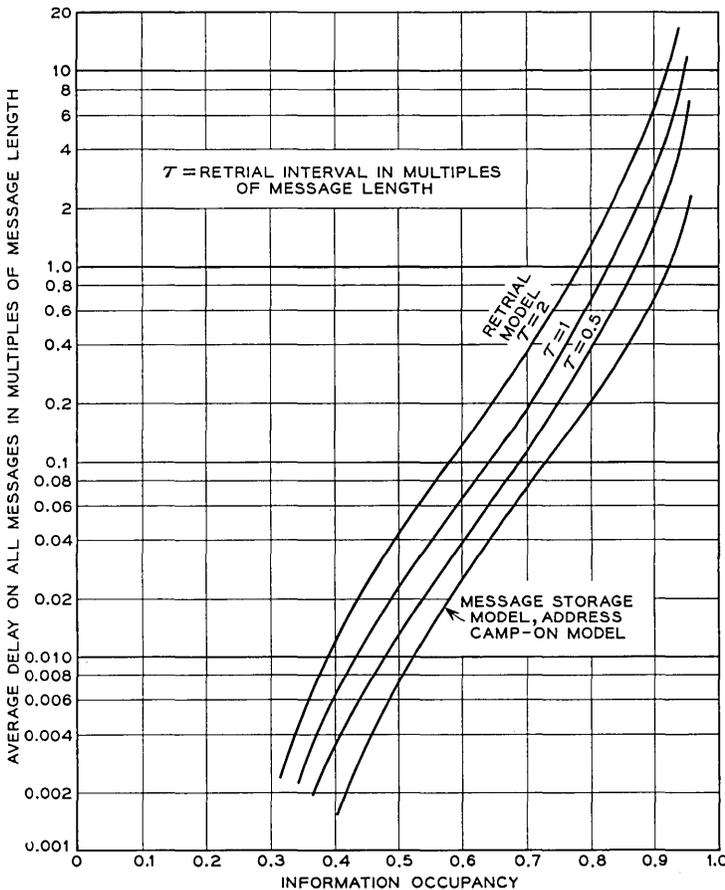


Fig. 3 — Average delay on all messages vs information occupancy; 10 lines per group, 1 address per message.

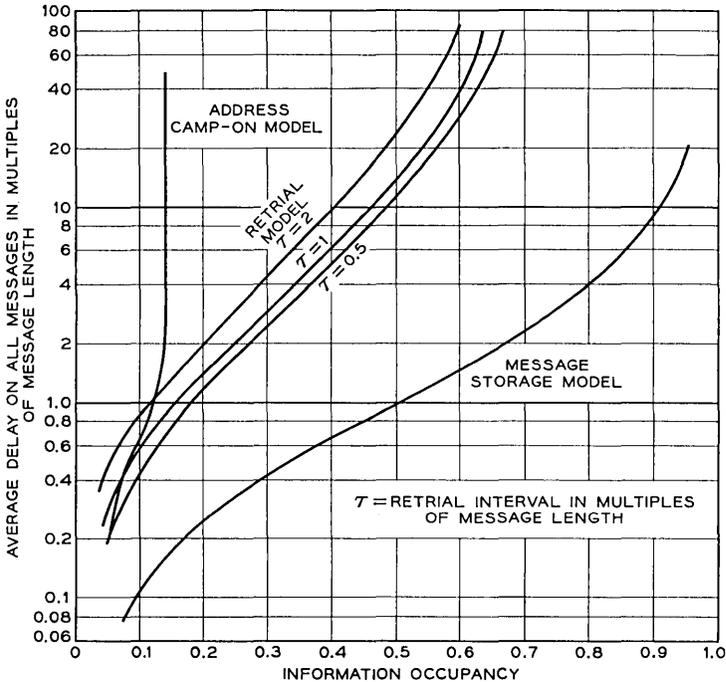


Fig. 4 — Average delay on all messages vs information occupancy; 1 line per group, 3 addresses per message.

obtained by Weber¹ in a different approach. A simulation made by Weber¹ for $c = 1$ shows close agreement with the results derived here. For example, the simulation shows a maximum utilization of about 14 per cent for $c = 1$ and three addresses per message, which is the same as derived here analytically. This indicates that considerable confidence may be placed in the approach presented in Appendix A.

From the delay performance of the address camp-on model it is concluded that any switching method in which delays become a substantial part of the line holding time will require a relatively large number of lines to provide adequate service. By the same token we may conclude that even more lines will be required when the message is switched more than once, i.e., through more than one switching center.

Next we turn our attention to the retrial model. The performance of the retrial model as a function of the retrial interval, τ , is of interest. We observe in Figs. 2-5 that when one doubles τ the delay is less than doubled. On the one hand, we expect longer retrial intervals to cause

longer delays. On the other hand, it can be shown that the probability that the message succeeds on a retry increases with increasing length of τ . Shorter retry intervals result in smaller chance for success than longer retry intervals, but the fact that in any given time there are more attempts made with short retries than with long retries makes the average delay a monotone increasing function of τ . For large values

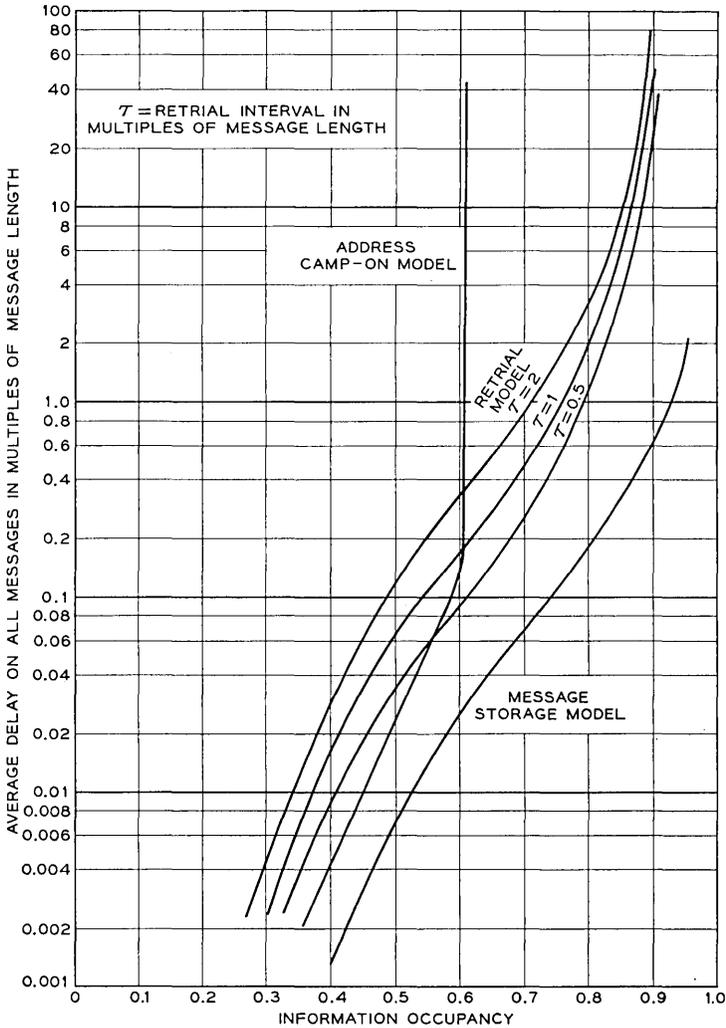


Fig. 5 — Average delay on all messages vs information occupancy; 10 lines per group, 3 addresses per message.

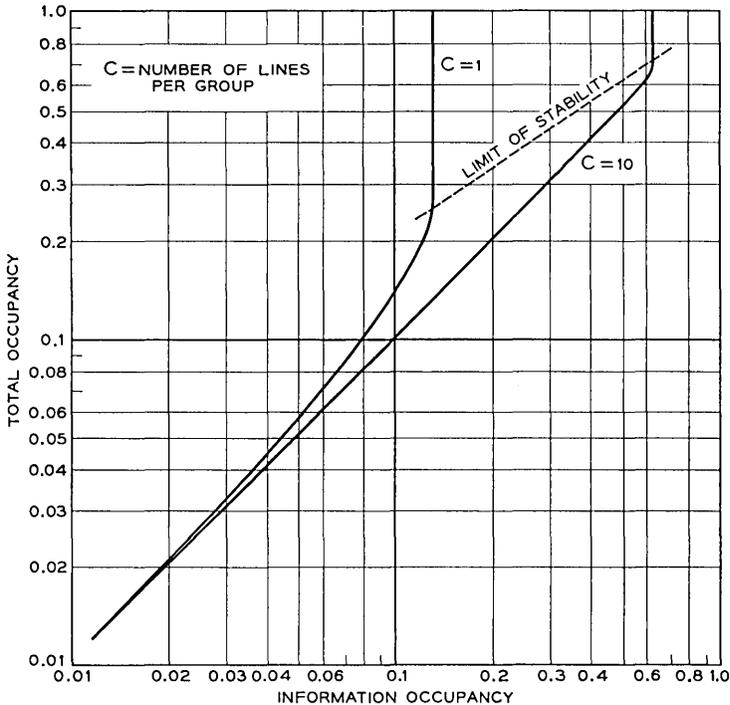


Fig. 6 — Total occupancy vs information occupancy for address camp-on model; 3 addresses per message.

of τ , say at least twice the message length, the rate of increase of the delay will be almost proportional to the rate of increase in τ . This is so because the rate of change of the probability of being blocked again becomes smaller with increasing values of τ . With τ approaching 0, the average delay with the retrial model should approach the delay for the message storage model in the case of one address per message.

The delay in the retrial model increases with the number of addresses per message. A comparison between Figs. 2 and 4 shows that this increase is quite significant when the number of lines per group is small. This increase can, of course, be avoided when a multiaddress message is broken up into several one-address messages, as suggested earlier in Section 2.3. Stations not blocked would then receive the message content independent of the availability of lines to the other stations. If this mode of operation is used for retrials, the retrial curves given in Figs. 4 and 5 are to be replaced by those for the one-address case shown in Figs. 2 and 3, respectively. Such a change of the retrial model would bring the av-

erage delay for very short retrial intervals close to that of the message storage model, independent of the number of addresses.

The retrial model has the advantage that no storage has to be provided in a switching center except at the data source. However, provision must be made to instruct the data source to reoffer the message when an all-lines-busy condition is encountered at a switching center. The retrial method is particularly well suited for error correction by retransmission from the data source on request from a terminating station which detected an error.

Since the analysis of the retrial model is based on approximations, we do not expect absolute accuracy of the curves derived. The retrial model has been simulated by others for some special cases and it is interesting to compare the results. This is done in Table I, in which time is expressed in units of the average message length. We observe some disagreement between analysis and simulation for large retrial intervals.

Finally, let us look at the message storage model. This method provides an efficient use of lines, even if the line groups are small, and delay is independent of the number of addresses per message. It requires, however, that considerable storage be provided because each switching center must be arranged to permit full message storage to allow for delays exceeding a message length. Provision must also be made for transmission of a copy of the message to each addressed station independent of the transmission to any other station.

According to the model of the message storage method, a message is delayed only until the very instant a line is found. From a practical point of view this means that the line-hunting mechanism should be activated as soon as the address is decoded. On the other hand, message storage may be operated so that the message is stored completely at the

TABLE I—COMPARISON: ANALYSIS VERSUS SIMULATION,
ONE ADDRESS PER MESSAGE

Retrial Interval	No. of Lines	Occupancy a_l/c	Average Delay on All Messages		
			Analysis	Simulation 1	Simulation 2
0.5	1	0.7	3.25	*	3.65
0.5	10	0.7	0.11	*	0.14
1.0	1	0.5	1.94	1.84	1.76
1.0	10	0.9	3.12	2.61	1.92
2.0	1	0.7	12.65	*	8.45
2.0	10	0.7	0.36	*	0.26

* Not available.

switching center before the line hunting starts. In this case, the amount of time needed for full message storage must be added to the delay. The latter mode of operation adversely affects the delay performance of the storage model, particularly when the message is long in comparison with the delay that can be tolerated.

The delay performance for the storage model will become considerably worse than shown when the line back to the originating station cannot be released as soon as the message has been transmitted over it. For multiswitched messages the release of lines between switching centers would ensure that the line holding time is not increased by the delay.

The curves given for the storage model can be considered accurate because the validity of Erlang delay formulas has long been observed.

V. ACKNOWLEDGMENTS

The author wishes to acknowledge the valuable assistance received from D. L. Clark, both in computer programming and in the mathematical derivations. Also acknowledged are suggestions made by E. Wolman in a critical review of this paper, and the encouragement received from E. E. Schwenzfeger during the preparation period.

APPENDIX A

Mathematical Analysis of Address Camp-On Model

Let the excess holding time E_A be the average time between seizure of a line and the time lines are found for A addresses of a given message. Further, let $Z_{K,A}$ be defined as the average time between initial request for lines and the time K out of A addresses have seized lines. At the time A addresses have seized lines, the message is ready to be transmitted. The average delay on all messages, d_A , is defined as the average time between initial request and the time lines have been seized by all A addresses, as illustrated in Fig. 7.

The average is a linear operator, and one obtains for the expected excess holding time

$$E_A = \frac{1}{A} \sum_{K=1}^A W_{K,A}.$$

And since

$$W_{K,A} = d_A - Z_{K,A}$$

it follows that

$$E_A = d_A - \frac{1}{A} \sum_{K=1}^A Z_{K,A}. \tag{1}$$

The term $\frac{1}{A} \sum_{K=1}^A Z_{K,A}$ is recognized as the average time between request for and seizure of a line for any given single address.

We define $Q(t)$ as the probability that the delay is less than or equal to t between the time of a request for a line by a given single address and the time of line seizure, and obtain

$$\frac{1}{A} \sum_{K=1}^A Z_{K,A} = \int_{t=0}^{\infty} t dQ(t).$$

The average excess holding time for the A -address case follows from (1) as

$$E_A = \int_{t=0}^{\infty} t d[Q(t)]^A - \int_{t=0}^{\infty} t dQ(t). \tag{2}$$

The only approximation in the analysis of the address camp-on model is the assumption that the holding time of a line is exponentially distributed, so that

$$Q(t) = 1 - \delta e^{-\varphi t}. \tag{3}$$

The holding time is made up of two random variables, namely the message length and the excess holding time. The approximation made in (3) implies that the sum of these two random variables is exponentially distributed. That this, indeed, is a reasonable assumption is confirmed

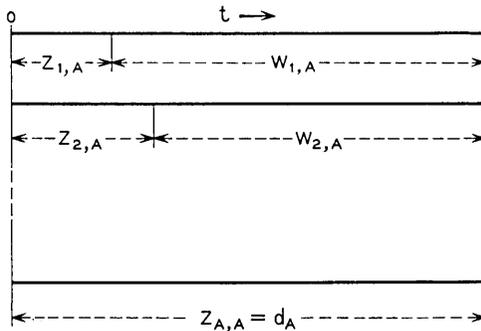


Fig. 7 — Line seizure sequence.

by the close agreement of the delay derived here with the delay derived by simulation.¹

Substitution of (3) in (2) gives

$$E_A = A\delta\varphi \int_{t=0}^{\infty} e^{-\varphi t} t(1 - \varphi e^{-\varphi t})^{A-1} dt - \varphi\delta \int_{t=0}^{\infty} te^{-\varphi t} dt.$$

The above reduces with the binomial expansion for $(1 - \delta e^{-\varphi t})^{A-1}$ to

$$E_A = -\frac{A}{\varphi} \sum_{K=0}^{A-1} \binom{A-1}{K} \frac{(-\delta)^{K+1}}{(K+1)^2} - \frac{\delta}{\varphi}. \quad (4)$$

For the case of exponential line holding time with mean \bar{t} and service of requests for lines in the order of arrival as is the case here, we must substitute in (4) according to Erlang²

$$\delta = 1 - Q(0) \quad (5)$$

$$\varphi = \frac{c - a_T}{\bar{t}} = \frac{c - a_T}{a_T/a_I}, \quad (6)$$

in which c is the number of lines per group, a_T the average number of requests per line holding time or the total load offered per line group, and a_I the average number of requests per message length or the information load. $Q(0)$, the probability of no delay, is given by Erlang² as

$$Q(0) = 1 - \frac{\frac{a_T^c e^{-a_T}}{c!} \frac{c}{c - a_T}}{1 - \sum_{i=c}^{\infty} \frac{a_T^i e^{-a_T}}{i!} + \frac{a_T^c e^{-a_T}}{c!} \frac{c}{c - a_T}}. \quad (7)$$

The unit of time being the average message length renders for \bar{t} , the average line holding time,

$$1 + E_A = \bar{t} \quad (8)$$

so that

$$E_A = \frac{a_T}{a_I} - 1. \quad (9)$$

Substitution of (5), (6) and (9) into (4) brings

$$\frac{a_T}{a_I} - 1 + \frac{a_T}{a_I} \frac{1 - Q(0)}{c - a_T} = -\frac{a_T}{a_I} \frac{A}{c - a_T} \cdot \sum_{K=0}^{A-1} \binom{A-1}{K} \frac{[Q(0) - 1]^{K+1}}{(K+1)^2}. \quad (10)$$

With c , a_I and A given, we can now compute a_T , the total load, itera-

tively from (7) and (10) above. Then, with a_T known, we find the average delay on all messages from (1) and (9) as

$$d_A = \frac{a_T}{a_I} - 1 + \frac{1 - Q(0)}{c - a_T} \frac{a_T}{a_I}, \tag{11}$$

which we recognize as the left-hand member of (10).

It is interesting to take note of the fact that there are generally two values of a_T which fulfill (10). Thus, one could conclude that the system can operate in two modes, one implying a shorter delay than the other. More than one steady state of operation has been observed by others^{1,5} in similar traffic studies. However, it appears questionable that the heavy delay mode is stable since the larger of the two a_T which fulfill (10) decreases with increasing a_I , which is physically unreasonable.

APPENDIX B

Mathematical Analysis of Retrial Model

When a newly offered message makes its first trial to seize one of c lines in a group, let $S_i, 0 \leq i \leq c$, be the unconditional probability that the group is in state i . A line group is said to be in state i when i out of all c lines in the group are busy. The message is reoffered until a line is available to each of A addressed receiving stations; therefore, no messages are lost. The load carried on each group equals a_I , the information load offered. For the special case of $c = 1$, we obtain

$$S_0 = 1 - a_I \tag{12}$$

and

$$S_1 = a_I. \tag{13}$$

For $c > 1$, i.e., for more than one line per group, S_i depends not only on a_I , but also on the procedure by which lines are made busy. By procedure is meant the type of distribution of the length of the intervals between line requests, and whether unsuccessful attempts form queues or are withdrawn. For Poisson input at the rate of a_0 and withdrawal of blocked attempts, Erlang loss probability³ gives

$$S_i = \frac{a_0^i / i!}{\sum_{x=0}^c a_0^x / x!} \tag{14}$$

in which

$$a_0 = \frac{a_I}{1 - S_c}. \tag{15}$$

In the switching model considered here, the total input to each line group, i.e., the total load offered, is not Poisson distributed and its magnitude differs from a_0 . This is so because repeated attempts are blocked with a probability which is larger than S_c and because for $A > 1$ blocking to any one of the A addressed receiving stations will cause a retrial. In order to compute S_i , it is assumed that for a sufficiently large retrial interval τ , the total input will at least resemble a_0 in distribution and magnitude. This approximation may be justified for light line occupancies but it becomes increasingly unrealistic with increasing line occupancies. The approximation for a_0 , therefore, is used here only to compute values for S_i when $c \geq 2$. For $c = 1$ the values for S_i are exactly determined by (12) and (13) above. For $c > 1$ we solve (14) and (15) iteratively with $i = c$ to obtain a_0 and then solve (14) to obtain approximations for S_i when $c \geq 2$. The approximations for S_i so obtained are used for the unconditional state probability both in the one and three-address cases.

Next, we will consider conditional probabilities which take into account past known states of a line group. Let it be known that at a given time t_0 there are j lines busy in a group, $0 \leq j \leq c$; what then is the probability that at $t_0 + t$ there are i lines busy? This conditional probability is called $X_{i,j}(t)$. For Poisson input and exponential line holding time, $X_{i,j}(t)$ is given by a well-known set of first-order differential equations.⁴ Here, now, we must take into account that the superposition of first and repeated attempts results in an input which is not Poisson. We let $\omega(t)$, $t_0 \leq t \leq t_0 + \tau$, be the instantaneous line request rate or the density of requests. As was said above, $X_{i,j}(t)$ is conditioned on state j of the group at t_0 . Consequently $\omega(t)$ depends also on the state of the group at t_0 , and this important point should be kept in mind, particularly since the notation does not always remind the reader of this condition.

Assume for the time being $\omega(t)$ is known for every value of t in the interval $(t_0, t_0 + \tau)$. The differential equations defining $X_{i,j}(t)$, $t_0 \leq t \leq t_0 + \tau$, are

$$\begin{aligned} X_{0,j}'(t) &= -\omega(t) \cdot X_{0,j}(t) + X_{1,j}(t) \\ X_{i,j}'(t) &= \omega(t) \cdot X_{i-1,j}(t) - [i + \omega(t)] \cdot X_{i,j}(t) + \\ &\quad (i + 1) \cdot X_{i+1,j}(t) \end{aligned} \quad (16)$$

for $0 < i < c$

$$X_{c,j}'(t) = \omega(t) \cdot X_{c-1,j}(t) - c \cdot X_{c,j}(t).$$

The condition that j lines are busy at time t_0 is taken into account by

$$X_{i,j}(t_0) = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j, \end{cases} \tag{17}$$

and, as said before, by $\omega(t)$. For the special case $t = t_0 + \tau$, the argument is dropped and $X_{i,j}(t_0 + \tau)$ is abbreviated to $X_{i,j}$. The system (16) can be solved with Laplace transforms for $c \leq 2$, and by numerical integration for $c \geq 3$.

Two cases are considered in the following: one in which each message has one address, i.e., the case $A = 1$; the other in which each message has three addresses, i.e., the case $A = 3$.

B.1 One Address per Message

In order to compute delay for the case $A = 1$, we must know the probability of finding the line group in state c at $t_0 + \tau$, given a state c at $t_0, t_0 - \tau, t_0 - 2\tau$, etc. In other words, we must know the probability that a message is blocked twice, three times, four times, etc. To simplify the notations for the case $A = 1$, we write C_i for the conditional probability that all c lines of a group are busy at $t_0 + \tau$, given a state c at $t_0, t_0 - \tau, \dots, t_0 - i\tau$. C_i , therefore, denotes the probability that a one-address message is blocked $i + 2$ times in a row. C_0 is identical with $X_{c,c}$ and can be computed by (16) provided $\omega(t)$ is known. It will help to keep matters clear if, for the case $A = 1$, $\omega(t)$ is subscripted so that $\omega_i(t)$ refers to the condition that at $t_0, t_0 - \tau, \dots, t_0 - i\tau$ all c lines of the group are known to be busy. For instance, the line request rate used in (16) to compute C_0 is called $\omega_0(t)$. Values for $C_i, i \geq 1$, are computed from (16) in the same manner as $X_{c,c}$, except that $\omega_i(t)$ is conditioned as indicated later. Hence, the numerical values for $X_{c,c}$ obtained from (16) with $\omega_i(t), i \geq 0$, are equal to C_i .

Let us now discuss the procedure by which $\omega_i(t)$ is obtained for the case $A = 1$. We will find functions L_i which are conditioned on a state c at $t_0, t_0 - \tau, \dots, t_0 - i\tau$, such that

$$\omega_i(t) = L_i(C_0, C_1, C_2, \dots, t). \tag{18}$$

This means that we cannot obtain an explicit expression for $\omega_i(t)$ since $\omega_i(t)$ is needed to compute C_i from (16). But with (18) we come into a position which allows us to assume values for C_i , compute $\omega_i(t)$ from (18) and then use the so-computed $\omega_i(t)$ to obtain C_i from (16). Through successive iterations stable solutions are obtained for C_i such

that (19) is satisfied

$$S_c < C_0 < C_1 < C_2 < \dots < 1. \quad (19)$$

The above approach, of course, is extremely tedious when C_i must be computed for large numbers of i . The problem is simplified when one assumes for large enough i that $C_i = C_{i+1}$. Computations have shown that for retrial intervals $\tau \geq 0.5$ we may reasonably approximate $C_i = C_{i+1}$ when $i \geq 4$. This approximation, consequently, is used in the one-address case. It is mainly for this approximation that the analysis for the case $A = 1$ is limited to retrial intervals $\tau \geq 0.5$.

Before we define the function L_i in (18), we will give the method by which the successive iterations are performed by a computer program to compute C_i . We start by iterating for a stable value of C_0 with $C_i = C_0$ for $i \geq 1$. Next, we iterate for a stable value of C_1 with C_0 fixed and $C_i = C_1$ for $i \geq 2$. Now we go back and iterate for a new value of C_0 with C_1 fixed and $C_i = C_1$ for $i \geq 2$. This process is continued until no further changes in C_0 and C_1 are detected. Continuing one step further, we iterate for a stable value of C_2 with C_0 and C_1 fixed and $C_i = C_2$ for $i \geq 3$. Again, we back up and search for a new value of C_0 with C_1 and C_2 fixed, then search for a new value of C_1 with C_0 and C_2 fixed and finally search for a new value of C_2 with C_0 and C_1 fixed, all with $C_i = C_2$ for $i \geq 3$. We proceed in steps in the manner described above until finally no changes are detected in C_0, C_1, C_2, C_3, C_4 with $C_i = C_4$ for $i \geq 5$.

With the approximation $C_i = C_4$ for $i \geq 5$, we can write for (18)

$$\omega_i(t) = L_i(C_0, C_1, C_2, C_3, C_4, t). \quad (20)$$

Line requests are made by first and repeated attempts. First attempts arrive independent of time with a density v . Repeated attempts arrive with a density $u_i(t)$, in which i refers to the condition that all lines are busy at time $t_0, t_0 - \tau, \dots, t_0 - i\tau$ and t is some time such that $t_0 \leq t \leq t_0 + \tau$. With these definitions we substitute for $L_i(C_0, C_1, C_2, C_3, C_4, t)$ in (20)

$$\omega_i(t) = v + u_i(t). \quad (21)$$

The density of first attempts, according to definition, is numerically equal to the information load offered or

$$v = a_I.$$

The density of repeated attempts in the interval $(t_0, t_0 + \tau)$ is derived from first attempts which are made before t_0 and are blocked. For in-

stance a k th attempt, $k \geq 2$, occurs at t , $t_0 \leq t \leq t_0 + \tau$, if the attempt occurs first at $t - (k - 1)\tau$ and all lines are busy at $t - (k - 1)\tau$, $t - (k - 2)\tau, \dots, t - \tau$. As an abbreviation we write $\Pr(t_x | t_1, t_2, t_3, \dots)$ for the probability that all lines are busy at t_x , conditioned on all lines busy at t_1 and t_2 and $t_3 \dots$. As before, let t be an instant in time such that $t_0 \leq t \leq t_0 + \tau$. With the condition that all lines are busy at $t_0, t_0 - \tau, \dots, t_0 - i\tau$, we obtain for the density of repeated attempts at t

$$\begin{aligned}
 u_i(t) = & a_i [\Pr(t - \tau | t_0, t_0 - \tau, \dots, t_0 - i\tau) \\
 & + \Pr(t - 2\tau | t_0, t_0 - \tau, \dots, t_0 - i\tau) \\
 & \cdot \Pr(t - \tau | t_0, t_0 - \tau, \dots, t_0 - i\tau, t - 2\tau) \\
 & + \Pr(t - 3\tau | t_0, t_0 - \tau, \dots, t_0 - i\tau) \\
 & \cdot \Pr(t - 2\tau | t_0, t_0 - \tau, \dots, t_0 - i\tau, t - 3\tau) \\
 & \cdot \Pr(t - \tau | t_0, t_0 - \tau, \dots, t_0 - i\tau, t - 3\tau, t - 2\tau) + \dots].
 \end{aligned}
 \tag{22}$$

We are left with the problem of expressing $\Pr(t_x | t_1, t_2, t_3, \dots)$ in the above as functions of C_0, C_1, C_2, C_3 and C_4 . Assume that symmetry exists such that for any positive length of time l

$$\begin{aligned}
 \Pr(t_x + l | t_x, t_x - \tau, \dots, t_x - k\tau) \\
 = \Pr(t_x - l | t_x, t_x + \tau, \dots, t_x + k\tau).
 \end{aligned}$$

In the above it is assumed that traffic congestion builds up to an all-lines-busy condition at $t_x, t_x + \tau, \dots, t_x + k\tau$ in the same manner as it subsides after $t_x + k\tau$. This assumption may not be exact for the retrial system but this concept is used here since it is expected to give a good enough approximation for the following reason.

If a group is busy, say, at t_0 , then it must be expected that part of the traffic which contributes to the congestion at t_0 is reoffered traffic. The fact that congestion occurs at t_0 implies that all lines were busy at $t_0 - \tau, t_0 - 2\tau$, etc., with a larger probability than indicated by the unconditional state probability S_c . As an approximation to the function by which traffic is expected to build up we construct linear functions in time. For example, we assume that the probability of blocking at some time $t_x < t_0$ builds up to an all-lines-busy condition at $t_0 - \tau$ and t_0 as shown in Fig. 8. Also, we assume independence of events that are not really independent. For instance, we assume that blocking between $t_0 - \tau$ and t_0 occurs with a probability $W_{1,1}(t)$ as defined below. Similarly, independence is assumed between events occurring with probability $N_i(t), M_i(t)$ or $W_{i,j}(t)$ and the event which causes a repeated attempt

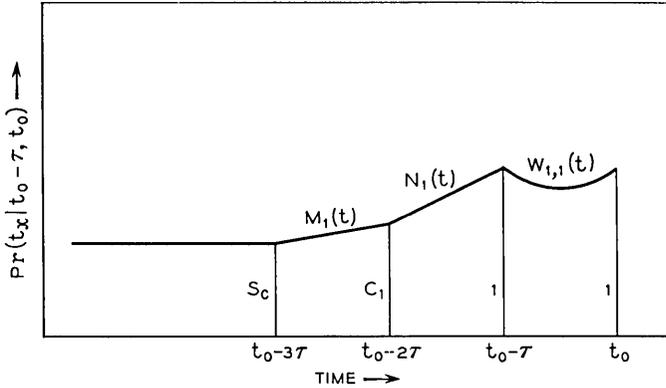


Fig. 8 — Sample of build-up function.

at some time prior to t_0 . The approximations made in the expressions below may account for some of the differences which are observed between theory and simulation.

A first attempt, made at some time $t - j\tau, t_0 \leq t \leq t_0 + \tau, j \geq 1$, is blocked in the nomenclature of (22) with a probability

$$\Pr(t - j\tau | t_0, t_0 - \tau, \dots, t_0 - i\tau)$$

for which we approximate

$$N_i(t) = \frac{1}{\tau} [t + C_i(\tau - t)] \quad \text{for } j = i + 1, i \geq 0$$

$$M_i(t) = \frac{1}{\tau} [C_i t + S_c(\tau - t)] \quad \text{for } j = i + 2, i \geq 0$$

$$S_c \quad \text{for } j \geq i + 3, i \geq 0$$

and with $N_i^*(t) = (1/\tau)[t(C_i - 1) + \tau]$

$$W_{i,j}(t) = 1 - [1 - N_{j-1}(t)][1 - N_{i-j}^*(t)]$$

$$\text{for } 1 \leq j \leq i, i \geq 1.$$

A k th repeated attempt made at some time $t - j\tau, t_0 \leq t \leq t_0 + \tau, j \geq 1$, is blocked in the nomenclature of (22) with a probability

$$\Pr[t - j\tau | t_0, t_0 - \tau, \dots, t_0 - i\tau,$$

$$t - (j + 1)\tau, t - (j + 2)\tau, \dots, t - (j + k)\tau]$$

for which we approximate

$$\begin{aligned}
 1 - (1 - C_{k-1}) [1 - N_i(t)] & \quad \text{for } j = i + 1, i \geq 0 \\
 1 - (1 - C_{k-1}) [1 - M_i(t)] & \quad \text{for } j = i + 2, i \geq 0 \\
 C_{k-1} & \quad \text{for } j \geq i + 3, i \geq 0
 \end{aligned}$$

and finally

$$1 - (1 - C_{k-1}) [1 - W_{i,j}(t)] \quad \text{for } 1 \geq j \geq i, i \geq 1.$$

We make use of the above expressions as shown in (22) to obtain $u_i(t)$. The subscript i of $u_i(t)$ corresponds with the value of i in the above approximations and refers to the condition that all lines are busy at $t_0, t_0 - \tau, \dots, t_0 - i\tau$. For every such $i, i = 1, 2, 3$ and 4 , we have, according to (21), a $L_i(C_0, C_1, C_2, C_3, C_4, t)$ and an $\omega_i(t)$, and can perform the iterations outlined before to compute C_0, C_1, C_2, C_3 , and C_4 .

Continuing in the analysis of the one-address case, we will now evaluate the delay. The probability that a one-address message is delayed exactly $i\tau$ is given by $D_1(i)$ which is

$$\begin{aligned}
 D_1(0) &= 1 - S_c \\
 D_1(1) &= S_c(1 - C_0) \\
 D_1(i) &= S_c(1 - C_{i-1}) \prod_{j=0}^{i-2} C_j, \quad i \geq 2.
 \end{aligned}$$

$H_1(i)$, the probability that the delay is greater than $i\tau$ for the one-address case, is

$$H_1(i) = 1 - \sum_{k=0}^i D_1(k), \quad i \geq 0$$

which reduces to

$$\begin{aligned}
 H_1(0) &= S_c \\
 H_1(i) &= S_c \prod_{j=0}^{i-1} C_j, \quad i \geq 1.
 \end{aligned}$$

The average delay is obtained as the summation of all possible delay values multiplied by their respective probability of occurrence and is given by

$$d_1 = \tau \sum_{i=1}^{\infty} i D_1(i)$$

which reduces to

$$d_1 = \tau S_c [1 + \sum_{i=0}^{\infty} \prod_{j=0}^i C_j]. \quad (23)$$

The above formula is used to compute the average delay on all messages for the case $A = 1$. The unconditional probability of finding all lines busy, represented by S_c , and the conditional blocking probabilities C_i are approximated by the methods outlined before.

B.2 Three Addresses per Message

The delay for the case $A = 3$ is computed in a Markov process. This means that we are considering only a first-order dependency, since we make the assumption that the conditional probability $X_{i,j}$ of finding i lines of a group busy at $t_0 + \tau$ depends only on state j of that group at t_0 . This and the following approximations appear justifiable in the multi-address case when considering the multitude of factors which determine the line request rate between t_0 and $t_0 + \tau$. The principal assumption for the case $A = 3$ is that ω , the sum of the densities of line requests of first and repeated attempts during t_0 and $t_0 + \tau$, is nearly Poisson distributed and therefore independent of any states at or before t_0 . Recall that for $A = 1$ we have been concerned only with the conditional probability of state c at $t_0 + \tau$ given also a state c at t_0 or at t_0 and $t_0 - \tau$, etc. For the case $A = 3$, however, we are concerned with an $X_{i,j}$ for all values of i and j , $0 \leq i \leq c$, $0 \leq j \leq c$, as will become apparent later. For a known ω , we obtain $X_{i,j}$ from (16). The condition that j trunks are busy, now, is accounted for only by the initial condition as given in (17). The density of line requests ω is obtained similarly to (21) as the sum of the densities of first attempts a_r , and of repeated attempts u which, according to our assumptions for $A = 3$, are time independent. It is obvious that in the case of A addresses per message, $A > 1$, a line request in a given group is made only when the condition is fulfilled that the remaining $A - 1$ groups are not busy. Since independence is assumed we can set $(1 - S_c)^2$ for this condition in the three-address case and obtain

$$\omega = (a_r + u)(1 - S_c)^2 \quad (24)$$

for the density of line requests in the interval $(t_0, t_0 + \tau)$ in any given line group. The expression given in (24) above, of course, is an approximation since in reality u is dependent on the state of the group at t_0 and since independence is assumed between the event causing a repeated

attempt during $(t_0, t_0 + \tau)$ and the event causing all lines to be busy in the other two addressed line groups. But, as said before, these dependencies are believed to be noncritical for the multiaddress case, so that ω is considered to be independent of time.

As in the one-address case, we are left with the problem of defining u , the density of repeated attempts, which is expressed below as a function of $X_{c,c}$. We obtain u by the following approach. Let G be the probability that a first attempt is blocked at some time prior to t_0 . This probability is approximated by

$$G = 3S_c(1 - S_c)^2 + 3S_c^2(1 - S_c) + S_c^3. \tag{25}$$

The above is an approximation because it assumes independence between the event which causes a group to be in state j , $0 \leq j \leq c$, at t_0 and the event causing all lines of a group to be busy at some time prior to t_0 . For the probability that a k th attempt, $k \geq 2$, is blocked prior to t_0 we approximate

$$R = X_{c,c}(1 - S_c)^2 + 2S_c(1 - X_{c,c})(1 - S_c) + 2X_{c,c}S_c(1 - S_c) + S_c^2(1 - X_{c,c}) + X_{c,c}S_c^2 \tag{26}$$

for which it is assumed that at the $k - 1$ st attempt one line group was in state c , i.e., busy, but without having made any assumptions about the state of the remaining two groups. The expression given for R in (26) is an approximation since, as before, the known state of a group at t_0 is ignored and since only a first-order dependency is considered, as mentioned earlier. The density of repeated attempts is obtained similarly to the one-address case by considering all attempts which were blocked prior to t_0 so that

$$u = a_I G + a_I GR + a_I GR^2 + \dots$$

or

$$u = a_I G \frac{1}{1 - R}. \tag{27}$$

Substitution of (27) in (24) gives

$$\omega = a_I \left(1 + \frac{G}{1 - R} \right) (1 - S_c)^2. \tag{28}$$

R in (28) above is a function of $X_{c,c}$. This means that we cannot find ω explicitly since ω is needed to compute $X_{c,c}$ as outlined in (16). To find ω we again must iterate by assuming a value for $X_{c,c}$ in (26), recompute

$X_{c,c}$ from (16) and then use the recomputed $X_{c,c}$ in (26). After having found a stable value for ω we can, from (16), readily compute $X_{i,j}$ for all $0 \leq i \leq c, 0 \leq j \leq c$.

In order to compute the delay for the three-address case we consider $P_{i,j,h}(k)$, which is defined as the probability of finding the three groups in state i, j and h respectively at the $k + 1$ st attempt, $k \geq 0$. $P_{i,j,h}(k)$ is obtained recursively by finding all possible ways in which the states of the three groups have changed to states i, j and h at the $k + 1$ st attempt when at least one group was busy at the k th attempt. Using the approximation of a first-order dependency, as mentioned before, we get for $k \geq 1$

$$\begin{aligned} P_{i,j,h}(k) &= \sum_{r=0}^c \sum_{s=0}^c P_{c,r,s}(k-1) \cdot X_{i,c} \cdot X_{j,r} \cdot X_{h,s} \\ &+ \sum_{q=0}^{c-1} \sum_{s=0}^c P_{q,c,s}(k-1) \cdot X_{i,q} \cdot X_{j,c} \cdot X_{h,s} \\ &+ \sum_{q=0}^{c-1} \sum_{r=0}^{c-1} P_{q,r,c}(k-1) \cdot X_{i,q} \cdot X_{j,r} \cdot X_{h,c}. \end{aligned} \quad (29)$$

At the first attempt,

$$P_{i,j,h}(0) = S_i S_j S_h.$$

A three-address message succeeds at the k th attempt, $k \geq 0$, when at the k th attempt all three groups are in states other than c . The probability of a delay of exactly $k\tau$, then, is given for the three-address case by

$$D_3(k) = \sum_{i=0}^{c-1} \sum_{j=0}^{c-1} \sum_{h=0}^{c-1} P_{i,j,h}(k). \quad (30)$$

$H_3(k)$, the probability that the delay is greater than $k\tau$ for the three-address case, is

$$H_3(k) = 1 - \sum_{i=0}^k D_3(i), \quad k \geq 0.$$

As in the one-address case, we find the average delay on all messages for the three-address case by summing over all possible delay values multiplied by their respective probability of occurrence

$$d_3 = \tau \sum_{k=1}^{\infty} k \cdot D_3(k)$$

or with (30)

$$d_3 = \tau \sum_{k=1}^{\infty} \sum_{i=0}^{c-1} \sum_{j=0}^{c-1} \sum_{h=0}^{c-1} k \cdot P_{i,j,h}(k). \tag{31}$$

Recall that $P_{i,j,h}(k)$ is obtained recursively as shown in (29). The unconditional state probability S_i and the conditional probability $X_{i,j}$ which are both needed in (29) are approximated as described earlier.

APPENDIX C

Mathematical Analysis of Message Storage Model

The delay for the message storage model is computed by well-known methods of traffic theory and is given here only for reasons of completeness. The delay in the delivery of a copy of the message to a given station is — according to the switching model — independent of the delay in the delivery to any other station. Delayed messages form queues in the order of arrival. An analysis for queued service and exponential line holding time was made by A. K. Erlang.

According to Erlang² we find for the average delay on all messages to any given receiving station

$$d = F(0) \cdot \frac{1}{c - a_I} \tag{32}$$

with

$$F(0) = \frac{\frac{a_I^c e^{-a_I}}{c!} \frac{c}{c - a_I}}{1 - \sum_{i=c}^{\infty} \frac{a_I^i e^{-a_I}}{i!} + \frac{a_I^c e^{-a_I}}{c!} \frac{c}{c - a_I}}.$$

The delay distribution, expressed as the probability that the delay is greater than t , is computed from

$$F(t) = F(0) \cdot e^{-(c-a_I)t}.$$

The curves for the message storage model are calculated from (32). One should, however, bear in mind that in certain specialized applications of data communication a copy of the message must sometimes have been delivered to all addressed receiving stations before the message is of use to any one station. One would then be interested in the average delay until a line is found to the receiving station with the longest delay of all

stations addressed by the message. For A addresses per message this delay is given by

$$d_A = \int_0^{\infty} t d[1 - F(t)]^A,$$

which for $A = 3$ reduces to

$$d_3 = \frac{F(0)}{c - a_t} \left\{ 3 - \frac{3}{2} F(0) + \frac{[F(0)]^2}{2} \right\}.$$

REFERENCES

1. Weber, J. H., unpublished work.
2. Molina, E. C., Application of the Theory of Probability to Telephone Trunking Problems, B.S.T.J., **6**, July, 1927, p. 469 and p. 471.
3. Wilkinson, R. I., Theories for Toll Traffic Engineering in the U.S.A., B.S.T.J., **35**, March, 1956, p. 426.
4. Brockmeyer, E., Halstrom, H. L., and Jensen, A., The Life and Works of A. K. Erlang, Transactions of the Danish Academy of Technical Sciences, No. 2, 1948, p. 33.
5. Helly, W., unpublished work.

Contributors to this Issue

VACLAV E. BENEŠ, A.B., 1950, Harvard College; M.A. and Ph.D., 1953, Princeton University; Bell Telephone Laboratories, 1953 —. Mr. Beneš has been engaged in mathematical research on stochastic processes, traffic theory, and servomechanisms. In 1959–60 he was visiting lecturer in mathematics at Dartmouth College. Member American Mathematical Society, Association for Symbolic Logic, Institute of Mathematical Statistics, Society for Industrial and Applied Mathematics, Mind Association, Phi Beta Kappa.

D. G. HAENSCHKE, Dipl. Ing., 1953, Technical University of Vienna; Institute for Telecommunication, Vienna University, 1953–1955; Bell Telephone Laboratories, 1955—. At Bell Laboratories Mr. Haenschke was concerned with air traffic control communication problems, including the integration of civil and military air traffic control systems. Later he was engaged in planning for the data switching and communications aspects of the Nike-Zeus anti-missile system, and more recently his work has been on long range systems studies for direct distance dialing (DDD) service. Member, IEEE.

ARTHUR E. KERWIEN, B.S. in Physics, 1926, Union College; M.S., 1947, Stevens Institute of Technology; Bell Telephone Laboratories, 1926 —. He first engaged in research on radio transmission and transmitter circuit problems, including applications of negative feedback in high-frequency radio transmitters. During World War II he was concerned with special research and development problems in radar and countermeasures. He later worked on development of SSB radio transmitters and receivers for Overseas Radiotelephone Services and on circuit problems in development of the personal radio signaling receiver. He is currently involved in development problems of existing and new microwave radio relay systems. Senior member, IEEE.

ROBERT LEE PEEK, Jr., A.B., 1921, Columbia College; Met. E., 1923, Columbia University School of Mines; Bell Telephone Laboratories, 1924 —. His early work was in materials analysis and testing. Since 1936

he has been chiefly engaged in switching apparatus development projects, including the wire-spring relay, sealed contacts, and ferreeds. During World War II Mr. Peek worked on underwater ordnance and magnetostriction sonar. In 1951 he was placed in charge of analytical and exploratory studies of electromagnetic devices, and he is currently a department head responsible for consulting engineering in switching apparatus. Member, IEEE.

DAVID SLEPIAN, University of Michigan, 1941-43; M.A., 1947, and Ph.D., 1949, Harvard University; Bell Telephone Laboratories, 1950 —. He has been engaged in mathematical research in communication theory, switching theory, and theory of noise, as well as various aspects of applied mathematics. He has been mathematical consultant on a number of Laboratories' projects. During the academic year 1958-59, he was Visiting Mackay Professor of Electrical Engineering at the University of California at Berkeley. Member A.A.A.S., American Mathematical Society, Institute of Mathematical Statistics, IEEE, Society of Industrial and Applied Mathematics, U.R.S.I. Commission 6.

LEON H. STEIFF, B.S.M.E., 1946, Northeastern University; Bell Telephone Laboratories, 1956 —. His first assignment was equipment design for P1 carrier, the first Bell System fully transistorized carrier system. He has since worked on wideband data transmission system development and the 150-mc pocket radio receiver. He currently supervises a group engaged in mechanical design of short-haul carrier systems; the group is also responsible for introducing new materials and techniques for Bell System use. Member, A.S.M.E.; Registered Professional Engineer, Massachusetts.

DONALD E. THOMAS, B.S. in E.E., 1929, Pennsylvania State University; M.A., 1932, Columbia University; Bell Telephone Laboratories, 1929 —. His early work was in development of repeatered submarine telephone cable systems. He later turned to sea and airborne radar development. During World War II, he served in the Signal Corps and the Air Force, and was a member of the Joint and Combined Chiefs of Staff Committees on Radio Countermeasures. When he returned in 1946, he took part in development and installation of the first deep-sea repeatered submarine telephone cable system. From 1950 to 1955 he was engaged in development and evaluation of new semiconductor devices. Since 1955 he has been interested in electrical feasibility studies of new solid-state devices and the application of electronic techniques to basic solid-state and optical maser research. Senior member, IEEE; member, Tau Beta Pi and Phi Kappa Phi.