THE JULY-AUGUST 1975
VOL. 54 NO. 6
BELL SYSTEM
TECHNICAL JOURNAL

THE BELL SYSTEM TECHNICAL JOURNAL

THE BELL SYSTEM TECHNICAL JOURNAL is published ten times a year by the American Telephone and Telegraph Company, J. D. deButts, Chairman and Chief Executive Officer, R. D. Lilley, President, C. L. Brown, Executive Vice President and Chief Financial Officer, F. A. Hutson, Jr., Secretary. Checks for subscriptions should be made payable to American Telephone and Telegraph Company and should be addressed to the Treasury Department, Room 1038, 195 Broadway, New York, N. Y. 10007. Subscriptions $15.00 per year; single copies $1.75 each. Foreign postage $1.00 per year; 15 cents per copy. Printed in U.S.A.

THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 54 July–August 1975 Number 6

A Molded-Plastic Technique for Connecting and Splicing Optical-Fiber Tapes and Cables

By P. W. SMITH, D. L. BISBEE, D. GLOGE, and E. L. CHINNOCK

(Manuscript received September 25, 1974)

We describe a new technique for optical-fiber cable connecting and splicing. Preliminary tests with multimode fibers produced splices with an average loss of less than 0.1 dB and a peak loss of 0.18 dB.

I. INTRODUCTION

Despite the efforts of many investigators,[1–12] the problem of connecting and splicing optical-fiber cables and subgroups of fibers[13] (tapes) has remained a serious one. The continued improvement of optical fibers to the point where losses approaching 1 dB/km have now been achieved[14] has made it increasingly apparent that practical connectors and splices should have losses much lower than those initially considered.

Someda[4] demonstrated a splicing technique in which individual fibers are aligned by pressing them into a grooved substrate. Much subsequent work has involved extensions and improvements of this idea. Miller[9] used precision-grooved aluminum spacers and prepared the fiber ends by grinding and polishing. Cherin[10] used embossed grooves and devised a jig for inserting tapes with previously prepared fiber ends into these grooves. The lowest losses in splices based on this technique were obtained by Chinnock et al.,[12] who prepared the ends of the fibers using a fiber-fracture technique.[7] All these methods, however, have drawbacks resulting either from difficulties associated

with the preparation of the fiber ends or with the mechanical alignment of the previously prepared ends. These problems may make such techniques difficult to apply in the field.

In this paper we describe a new fiber-splicing technique that eliminates some of the difficulties associated with the grooved substrate type of splices, can function as a removable connector, and should be readily adaptable for field use.

II. THE MOLDED-PLASTIC SPLICING TECHNIQUE

The basic technique is illustrated in Fig. 1. The end of the tape to be spliced is prepared for molding by dissolving the plastic coating over a short region (≈ 1 cm) to expose the individual fibers. The tape with the exposed fibers is then placed in the mold as shown in Fig. 1a. The fibers are held accurately in position by means of a thin spacer plate (see insert). After a suitable plastic material is molded around the fibers, the entire assembly is removed from the mold (Fig. 1b). The fibers are exposed over a narrow region where the spacer plate held them in position in the mold. The exposed fibers are now scored, and the entire assembly is fractured by bending and applying tension in the manner previously described for single fibers.[7] The plastic material fractures in the same plane as the optical fibers, and the tape termination is now ready for splicing (Fig. 1c). To make a splice, two tapes with terminations prepared as described above are placed in an alignment channel, and a suitable index-matching epoxy is used to index-match and to hold the assembly together (Fig. 1d).

The splicing technique described above has a number of important features. Each operation is relatively simple and involves no handling of individual fibers. The fiber-breaking technique quickly produces clean ends of good optical quality. Minimal handling of the prepared ends is required. The technique can easily be adapted to make a removable connection.

To demonstrate these ideas, a mold was constructed as shown in Fig. 2. The mold was machined from brass and was made in two parts so that it could be taken apart to facilitate the removal of the molded tape ends. The spacer plate was made from 175-μm steel feeler gauge stock which was tapered to about 100 μm at the top. Figure 2b shows a close-up of the spacer plate.

Because polyester resin* is readily available, has a low initial viscosity, and shrinks little on hardening, it was used as the molding

* The polyester resin used for these experiments was No. 50111, Berton Plastics, Inc., South Hackensack, N. J.

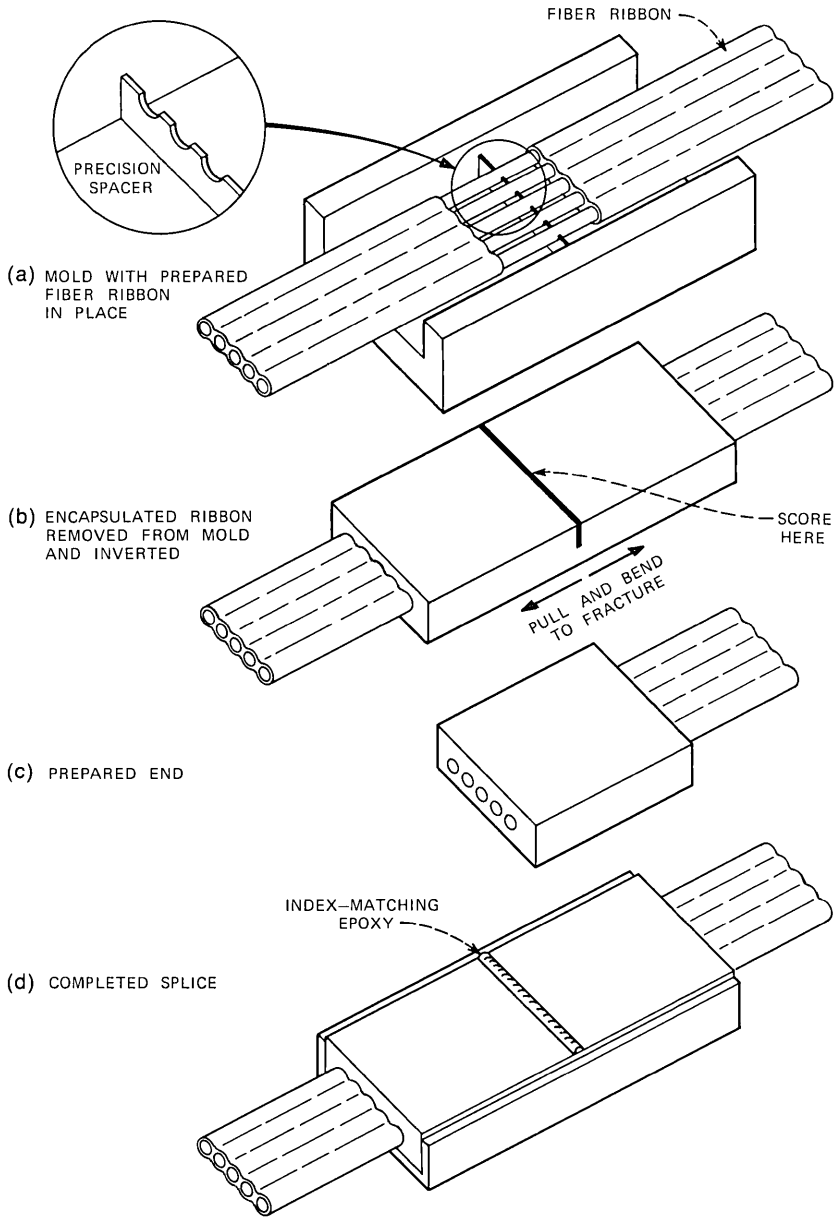


Fig. 1—The molded-plastic fiber-tape splicing technique.

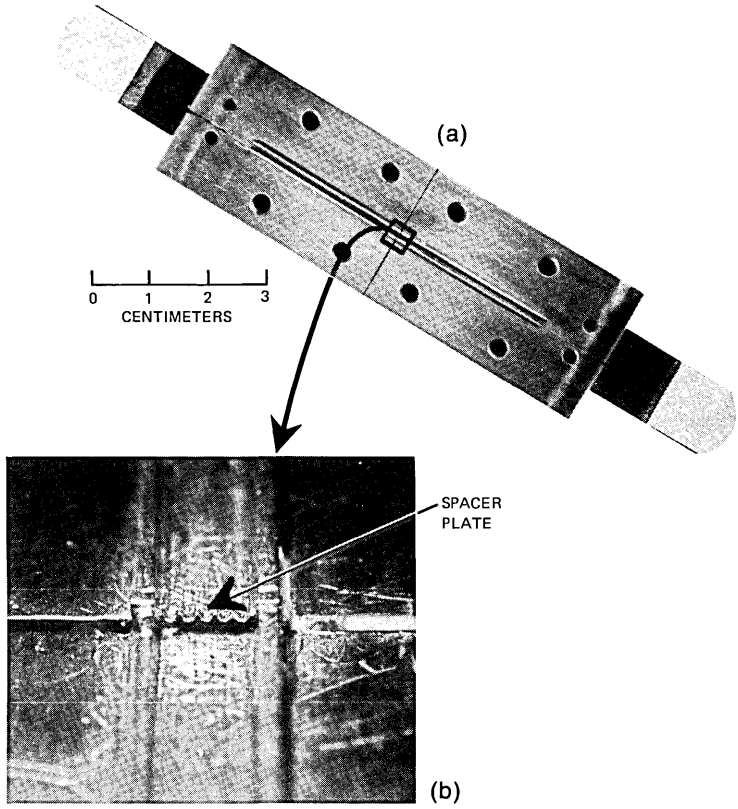


Fig. 2—Photograph of the mold used for these experiments. (a) Overall view. (b) Close-up view of precision spacer.

material, even though it required curing for several hours at approximately 50°C. Low-shrinkage plastics are available, however, that cure in a short time at room temperature.*

The optical-fiber tapes used for these splicing experiments were made from multimode silicate glass fibers of 120-μm outer diameter and 80-μm core diameter. The numerical aperture of these fibers was measured to be 0.15. The fibers were made into tapes by dipping 2-m lengths of fiber into a solution of plastic material[15†] and then slowly

* For example, Facsimile, made by Flexbar Machine Corp., Farmingdale, New York.

† The coating material used was 3M Kel-F-800, a co-polymer of vinylidene fluoride and chlortrifluoroethylene.

withdrawing the fibers. In this way, a 25-μm plastic coating was applied to the fiber. To eliminate crosstalk between fibers, a black dye was mixed with the coating material. The plastic-coated fibers were then fused into a linear array using the technique previously developed by Eichenbaum.[15]

Figure 3 shows a section of five-fiber tape with the plastic material removed in preparation for splicing. The material was easily removed by applying acetone with a cotton swab. The next step in the splicing operation involves placing the prepared tape in the mold shown in Fig. 2. To hold the fibers positively in the grooves of the spacer plate, a slight tension was applied to the tape by clamping it to the spring steel extensions on either side of the mold (see Fig. 2). Because the mold we were using was somewhat wider than the tape, we also placed pads on either side of the tape to hold the fibers straight as they passed over the spacer plate. Such pads would not be required if a narrower mold were used. Small pads of balsa wood were also used on top of the tape about 1 cm on either side of the spacer to ensure that the fibers remained pressed into the spacer grooves during the molding process. Polyester resin was then poured around the fibers, the lid of the mold

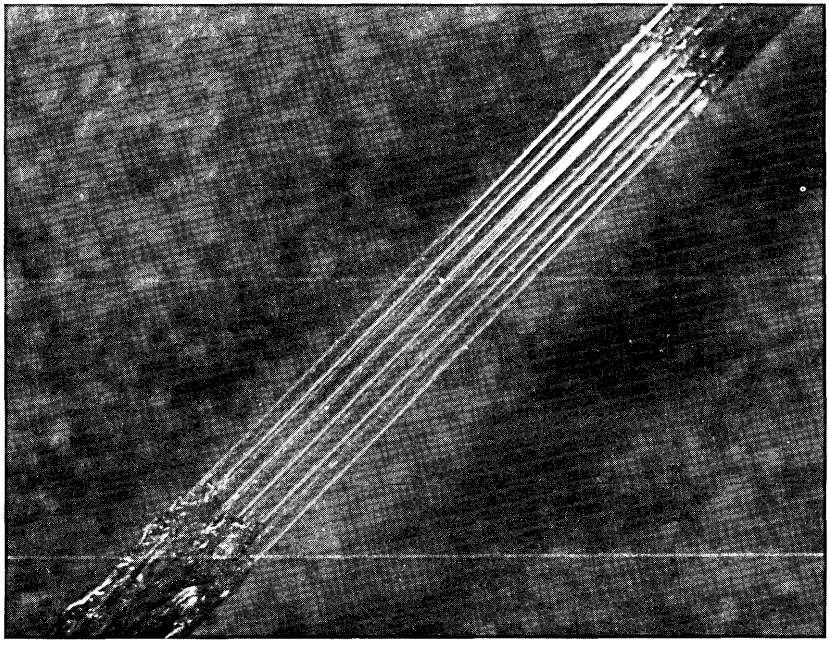


Fig. 3—Fiber tape with plastic material removed in preparation for placing in mold.

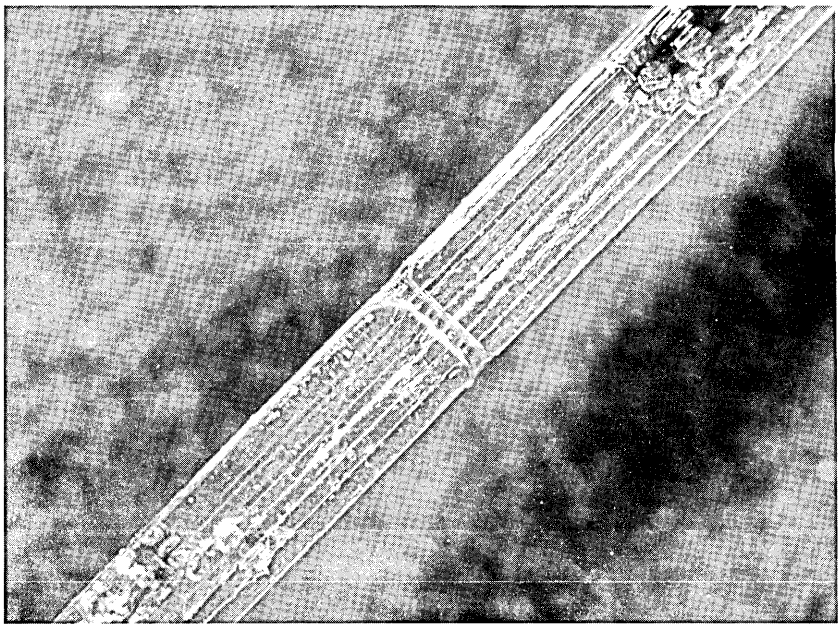


Fig. 4—Fiber tape end after molding polyester resin and before scoring fibers. Note exposed fibers.

closed, and the entire assembly allowed to cure. Figure 4 shows the molded section after removal from the mold. To facilitate removal, the brass mold was coated with mold-release compound.*

Two methods for scoring the fibers have been studied. An abrasive spray of compressed air and dental abrasive powder was found to be effective in scoring the fibers where they are exposed. Alternatively, a narrow carbide blade can be drawn across the fibers. After the fibers have been scored, the entire assembly is fractured by bending the tape and applying axial tension. This can conveniently be done by using the device described in Ref. 12. The theory of glass fracture under these conditions is derived in Ref. 7. We found that the fracture of the molded plastic always took place in the same plane as that of the fibers, and the prepared tape ends looked as shown in Fig. 5.

To complete a splice, the two ends prepared as described above are placed in a suitable alignment channel, described at the end of Section III, and an index-matching fluid is added.

* Satisfactory results were obtained with Dow Corning "Pan Shield" silicone spray. This can be removed easily with acetone.

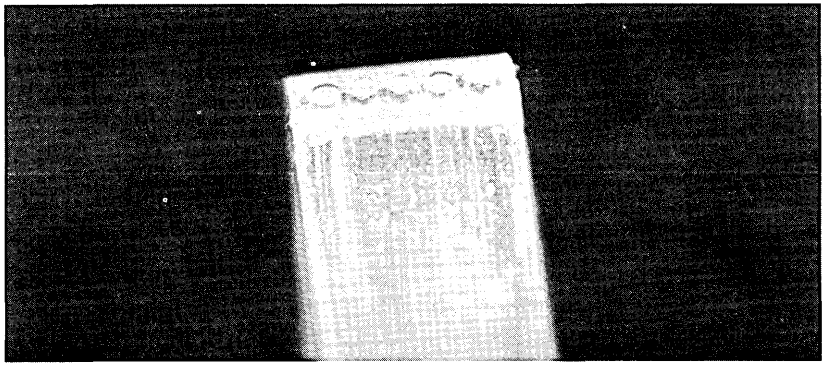


Fig. 5—Tape end after scoring and fracturing. Note the polyester resin breaks in the same plane as the optical fibers.

III. SPLICE LOSS MEASUREMENTS

The accurate measurement of small losses in good fiber-tape splices requires each fiber of the tape to be excited in the same way and with the same amount of light. Repeatable and stable focusing of the input beam on a given spot within the fiber core is a prerequisite. To accomplish this, we built the launching assembly shown in Fig. 6, which is rigidly attached to a commercially available HeNe laser. The assembly not only holds the fiber tape and shields it and the laser beam from dust and air movement, but also permits a precise visual alignment by way of a built-in microscope. The tape is epoxied to a tape holder which slides into a micropositioner from the right of Fig. 6. A three-way alignment of each fiber-front face with the focus of the launching lens is possible (only one positioning screw is shown).

The launching lens, the beam splitter, and the eyepiece form a microscope arrangement that provides a magnified view of the front end of the tape. When the beam is launched into one of the fiber cores, the entire core area lights up as a result of back-scattered light from the inside of the fiber. A cross hair built into the eyepiece facilitates the alignment of beam focus and fiber core. The beam splitter also serves as an attenuator and deflects a portion of the laser beam onto a silicon detector that provides a reference signal proportional to the input light power. As the circuit diagram of Fig. 7 indicates, both the transmitted signal and the reference signal are compared in a ratio meter where the result is displayed in digital form. This measuring technique, together with the precise visual alignment, was found to reproduce the excitation of each fiber to within ± 0.01 dB over a period of at least one hour.

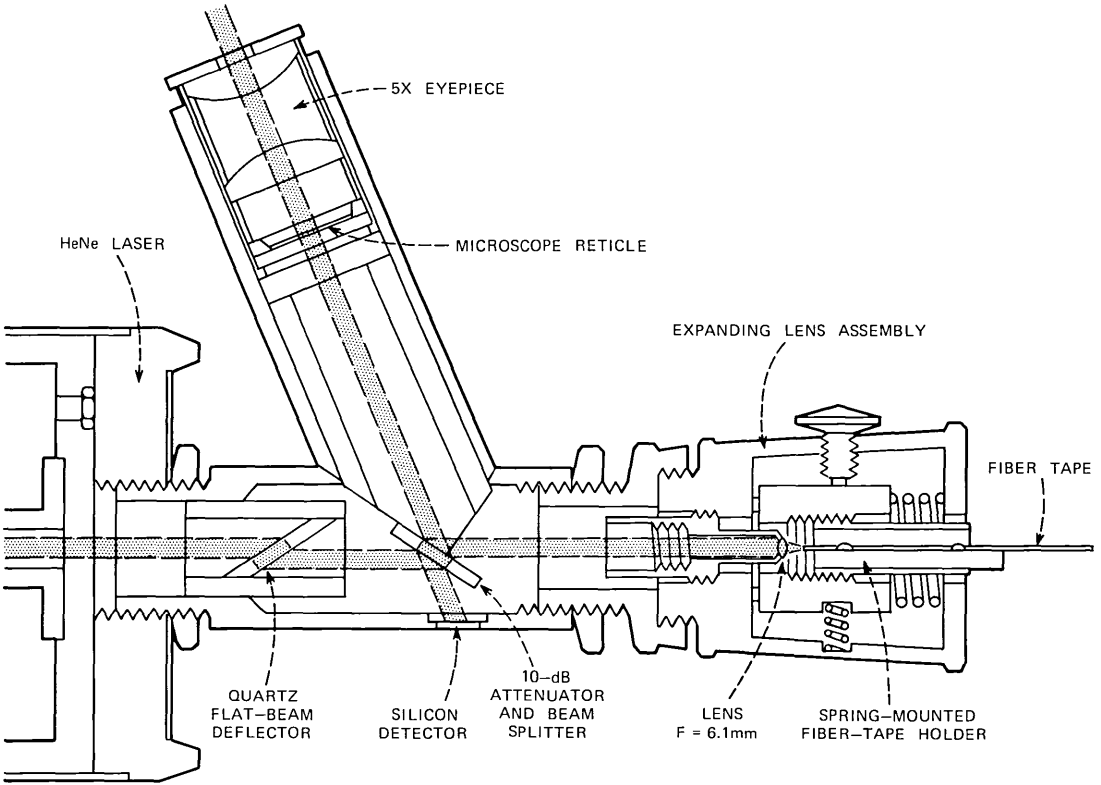


Fig. 6—Apparatus used for loss measurements.

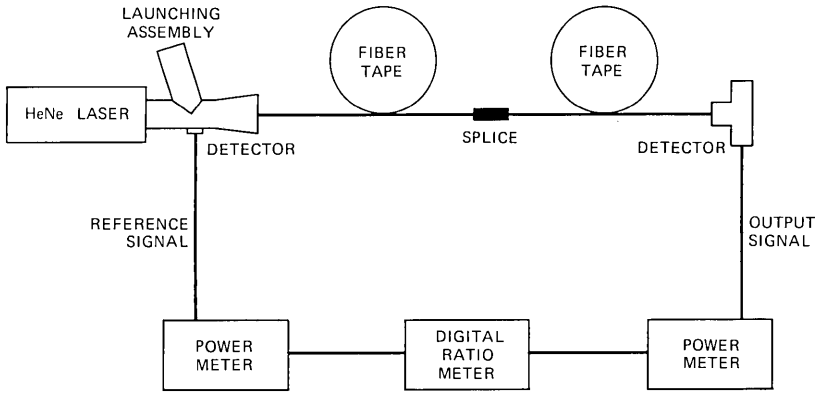


Fig. 7—Cross section of beam-launching assembly.

Radiation losses of the above order of magnitude (0.01 dB/m) can be incurred in the high-order modes as a result of minute bends along the fiber. The bends, and hence the loss, depend on the particular placement of the fiber and its holding fixtures, and can change during the breaking and splicing operation. To minimize the influence of this effect, the focal length of the launching lens is chosen to excite a numerical aperture of 0.10 out of the total numerical aperture of 0.15, so that some of the high-order modes are not excited. On the other hand, these modes may participate to some extent in the transmission (and in the splice loss), if long fiber lengths are involved. The resulting change in splice loss is expected to be small, but needs to be explored.

Our splicing tests were conducted by making two molded sections approximately 10 cm apart near the center of a 2-m length of tape. The same mold in the same orientation was used for each molded section. Prior to the splicing test, the transmission of each fiber in the unbroken tape was measured using the apparatus previously described. The two molded sections were then fractured, the center 10-cm section removed, and the two ends replaced in the mold. A drop of glycerine was added for index-matching, and the ends of both molded sections were held down with a single Teflon* blade about 1 mm in width. The transmission of each fiber was again measured.

Figure 8 is a histogram of the measured losses for tests on 20 fibers, and Fig. 9 shows the cumulative loss distribution. In two cases, either because of inadequate scoring or because of a weak region in the fiber, a fiber fractured at a point other than the desired fracture point. These cases have not been included in our data. Also omitted are three

* Registered trademark of Dupont Corporation.

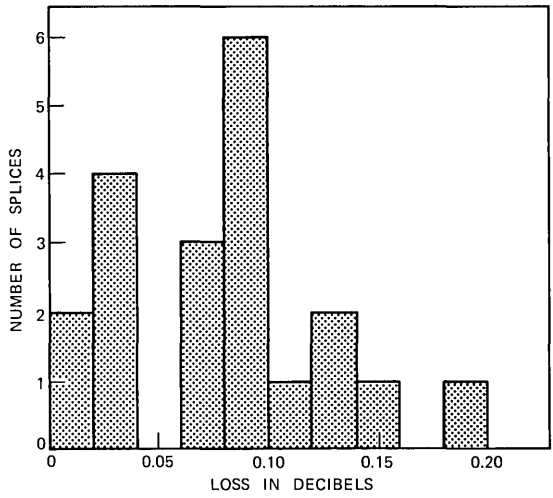


Fig. 8—Histogram of loss measurement data.

cases where fibers in one of the tapes were broken during handling. The average measuring error for these experiments was found to be 0.01 dB. No correction for this scatter was applied to the data.

Permanent splices were made using epoxy instead of glycerine as an index-matching material. This did not increase the splice loss when the

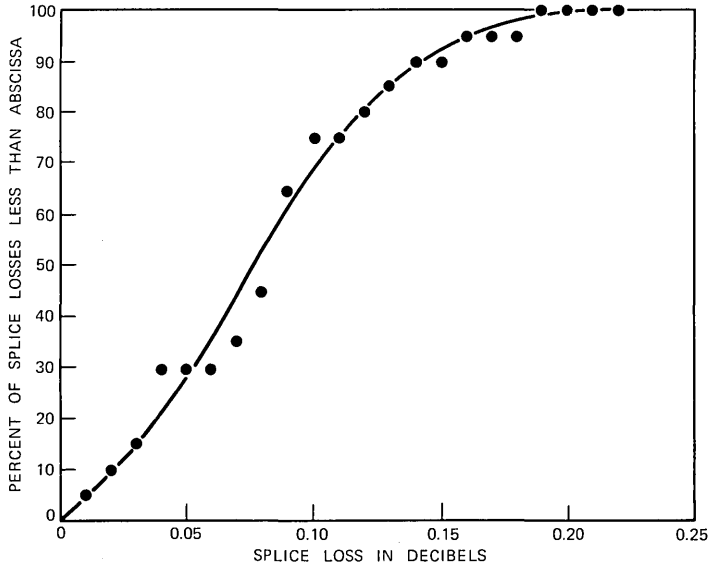


Fig. 9—Cumulative distribution of measured splice losses.

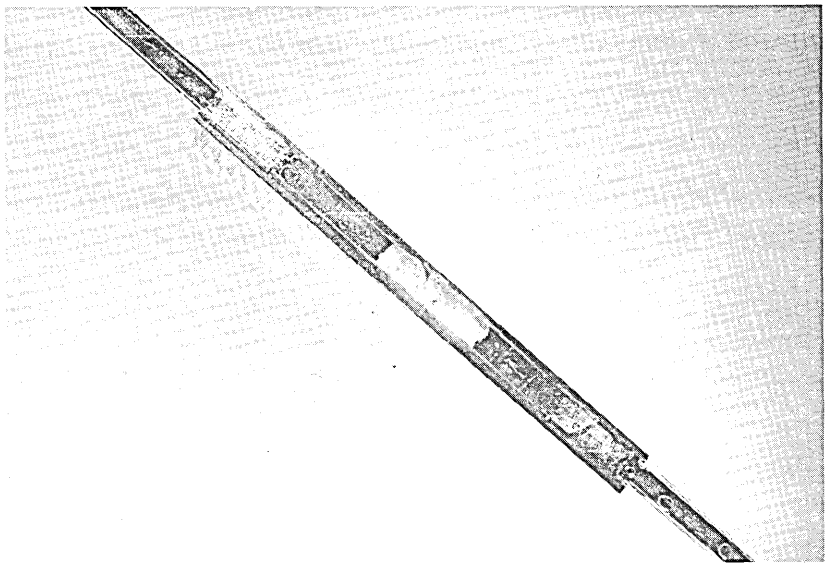


Fig. 10—Completed splice using brass alignment channel and epoxy cement both for index matching and for holding the assembly together.

mold was used as an alignment channel. To make more compact and permanent splices, we epoxied molded tape ends into either a channel made from 125-μm-thick phosphor bronze sheet or a 1-mm-by-0.3-mm milled brass channel, as shown in Fig. 10. The resulting loss increase of between 0.1 and 0.3 dB was attributed to imperfections of the mold that were reproduced in the molded ends. The use of alignment sleeves to alleviate this problem is discussed in the next section.

IV. DISCUSSION AND CONCLUSIONS

The splice losses reported in Section III are lower than those reported for other splice techniques[9,10] and are comparable with the losses we have observed using a grooved substrate to align fibers with previously prepared ends.[12] We believe, however, that the technique described in this paper has a potential for adaptation to a variety of fiber-connector problems and to field splicing of fiber cables.

There are two "precision" operations involved in our splicing technique, (*i*) the intial placing of the prepared fiber tape in the mold and (*ii*) the placing of the molded ends in the holder. Suitably designed molding tools and alignment sleeves should facilitate these operations to the point where they require no special skills. One important problem is that of maintaining a high degree of cleanliness. If the same mold is used to prepare both ends to be spliced and this mold

is also used as the alignment channel, a high-precision mold is not required. If, on the other hand, the prepared ends are to be joined in a different alignment channel, precision molds must then be used. The use of precision molds does not present a major problem, however, for the molds can be made, tested, and cleaned in advance, and these molds can, at a later time, be prepared for reuse. In this case, separate molds would be used to prepare each tape end; we thus avoid the problems inherent in the first-described technique associated with cleaning a mold in the field.

A definitive design of the alignment channel is beyond the scope of this work, since this design must depend considerably on the constraints of a specific connection: a removable or permanent connection of one fiber tape inside a building, for example, has different requirements from a cable field splice. However, some guidelines common to all designs can be extracted from our work: the reference surfaces used for the alignment of two molded terminations should be narrow and well defined to avoid misalignment as a result of mold irregularities and contamination. Thus, rather than using the channel of Fig. 10, which encloses the terminations on three sides, a sleeve that makes contact with the terminations only along their narrow sides may be more practical. More specifically, the sleeve might comprise two V-grooves designed to accept the V-shaped narrow sides of terminations molded in the shape shown in Fig. 11. The width of the V-grooves should be somewhat wider than the thickness of the terminations so that contact is only made along the grooves and not at the top and bottom surface of the molded piece. Besides minimizing and defining the alignment surfaces, this approach permits a space-saving and simple arrangement of stacks of terminations in the case of a fiber cable, as illustrated in Fig. 11. This figure shows two grooved chips spring-mounted opposite each other inside a cartridge that aligns the terminations and serves as the inside housing and protection for the cable splice.

This approach, together with the molding technique, seems particularly well suited to produce removable connections. In this case, the index-matching epoxy or liquid would be advantageously replaced by a gel that can easily be removed when the connection is dismantled. When such terminations are prefabricated in the factory, the technique proposed here offers the additional advantage that the terminations can be delivered unfractured, so that the end faces remain protected until they are fractured on site shortly before the connection is made.

In conclusion, we have described and demonstrated a new splicing technique for optical fiber tapes that yields an average splice loss of

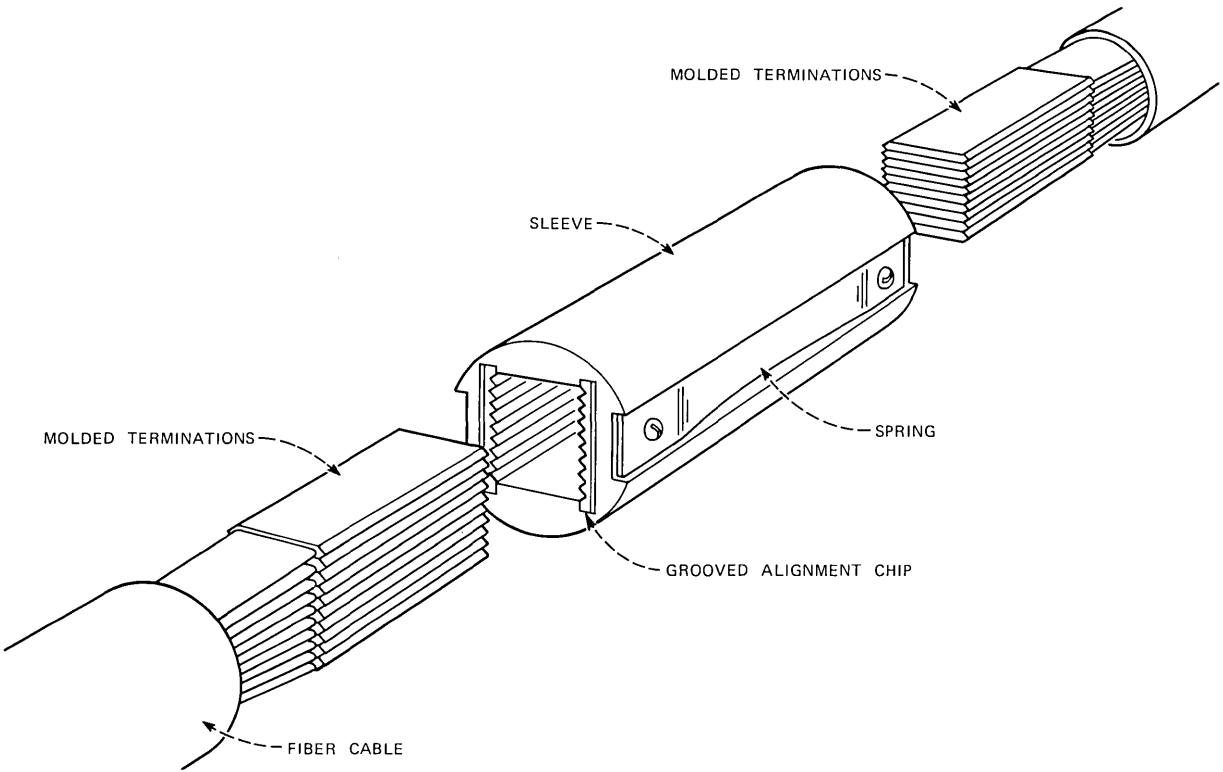


Fig. 11—Grooved alignment sleeve for cable splicing and connectors.

less than 0.1 dB. This technique involves no handling of individual fibers and alleviates some of the difficulties of fiber-end preparation and mechanical alignment of previously prepared ends that are encountered with other techniques.

An adaptation of our technique for field use would involve developing a molding technique applicable at room temperature with a quick-setting plastic encapsulating material and the design of molding tools and alignment sleeves to facilitate the joining operation.

REFERENCES

1. D. L. Bisbee, "Optical Fiber Joining Technique," B.S.T.J. *50*, No. 10 (December 1971), pp. 3153–3158.
2. R. B. Dyott, J. R. Stern, and J. H. Stewart, "Fusion Junctions for Glass Fiber Waveguides," Elec. Letters, *8*, (June 1973), pp. 290–292.
3. O. Krumpholz, "Detachable Connector for Monomode Glass Fiber Waveguides," Archiv Elektronik Vbertragungstechnik, *26* (1972), pp. 288–289.
4. C. G. Someda, "Simple Low-Loss Joints Between Single-Mode Optical Fibers," B.S.T.J., *52*, No. 4 (April 1973), pp. 583–596.
5. A. H. Cherin, E. R. Eichenbaum, and M. I. Schwartz, "Splicing Optical Fiber Ribbons, First Attempts," unpublished work.
6. A. H. Cherin, "Multi-Groove Embossed Plastic Splice Connector for Optical Fibers, A Feasibility Study," unpublished work.
7. D. Gloge, P. W. Smith, D. L. Bisbee, and E. L. Chinnock, "Optical Fiber End Preparation for Low-Loss Splices," B.S.T.J., *52*, No. 9 (November 1973), pp. 1579–1589.
8. H. W. Astle, "Optical Fiber Connector with Inherent Alignment Feature," unpublished work.
9. C. M. Miller, "A Fiber Optic Cable Connector," B.S.T.J., November 1975.
10. A. H. Cherin and P. J. Rich, "A Splice Connector for Joining Linear Arrays of Optical Fibers," in *Optical Fiber Transmission* (digest of technical papers presented at the topical meeting on optical fiber transmission, January 7–9, 1975, Williamsburg, Va.), Optical Society of America, pp. WB3-1 to WB3-4.
11. R. M. Derosier and J. Stone, "Low-Loss Splices in Optical Fibers," B.S.T.J., *52*, No. 7 (September 1973), pp. 1229–1235.
12. E. L. Chinnock, D. Gloge, P. W. Smith, and D. L. Bisbee, "Preparation of Optical Fiber End for Low-Loss Tape Splices," B.S.T.J., *54*, No. 3 (March 1975), pp. 471–477.
13. R. D. Standley, "Fiber Ribbon Optical Transmission Lines," B.S.T.J., *53*, No. 6 (July–August 1974), pp. 1183–1185.
14. W. G. French, J. B. MacChesney, P. B. O'Connor, and G. W. Tasker, "Optical Waveguides with Very Low Losses," B.S.T.J., *53*, No. 5 (May–June 1974), pp. 951–954.
15. B. Eichenbaum, "A Technique for Assembling Fiber Optic Arrays," unpublished work.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

Coupled-Mode Theory for Anisotropic Optical Waveguides

By D. MARCUSE

(Manuscript received December 11, 1974)

The well-known coupled-mode theory of waveguides is extended to include dielectric guides made of anisotropic materials. Exact coupled-wave equations for anisotropic dielectric waveguides are derived, and explicit expressions for the coupling coefficients are given. The coupling coefficients for isotropic waveguides are obtained as a special case. A simple approximation for the coupling coefficients in the case of slight anisotropy and slight departure from an ideal waveguide is presented.

I. INTRODUCTION

The theory of dielectric optical waveguides deals with electromagnetic wave propagation in optical fibers and in the waveguides used for integrated optics. Wave propagation in these structures is described in terms of normal modes.[1–3] However, normal modes preserve their identity only in perfect waveguides without irregularities of either the refractive index distributions or the waveguide geometry. Electromagnetic wave propagation in waveguides with any kind of irregularities must be described by means of coupled-mode theory.[3,4] The electromagnetic waves in imperfect waveguides are expressed as superpositions of all the modes of a perfect waveguide. The mode amplitudes are coupled together by coupling parameters that depend on the nature of the waveguide imperfections. A description of wave propagation by means of coupled-mode theory allows calculation of radiation losses caused by intentional or unintentional fluctuations of the refractive index along the axis of the waveguide or by core-cladding boundary fluctuations.[2,3] Coupling among guided modes is used to design modulators or distributed feedback circuits for lasers or to effect improvements in the multimode dispersion properties of overmoded waveguides. The coupled-mode theory is well developed for waveguides that consist of isotropic dielectric materials.[3,4] Some work has been done to extend this theory to waveguides consisting of anisotropic materials.[5–7] These waveguides are assuming increasing im-

portance in integrated optics as methods are being perfected for fabricating waveguides by diffusing different dopants (or outdiffusion of certain component atoms) into anisotropic crystals.[8–10]

This paper describes the derivation of coupled-wave equations for the modes of waveguides consisting of anisotropic materials. The coupled-wave theory is based on the definition of guided and radiation modes as solutions of Maxwell's equations for idealized structures. An orthogonality relation is derived that is needed to isolate individual terms in the infinite series expansion of the electromagnetic field. The principal result of this theory is the derivation of coupling coefficients that are important for solving coupled-mode problems. Readers not interested in the derivation should look at eqs. (46) and (48). Applications of this theory are not presented here, since they will be the subject of further publications.

II. THE FIELD EQUATIONS FOR ANISOTROPIC MEDIA

The derivation of coupled-wave equations for anisotropic dielectric waveguides follows closely the procedure used for deriving coupled-wave equations for isotropic waveguides.[3] The objective of coupled-wave theory is to construct solutions of Maxwell's equations for waveguiding structures consisting of general refractive-index distributions.

Anisotropic media are characterized by a dielectric tensor,

$$\epsilon = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{pmatrix}. \tag{1}$$

We assume that the elements of this tensor are real quantities characteristic of lossless materials. It can be shown that conservation of energy requires that the dielectric tensor form a symmetric matrix so that the following relations hold:[11]

$$\epsilon_{xy} = \epsilon_{yx}, \qquad \epsilon_{xz} = \epsilon_{zx}, \qquad \epsilon_{yz} = \epsilon_{zy}. \tag{2}$$

The magnetic properties of the medium are assumed to be the same as that of a vacuum so that we use the (isotropic) magnetic permeability constant μ_0. Maxwell's equations for anisotropic media assume the form

$$\nabla \times \mathbf{H} = i\omega\epsilon \cdot \mathbf{E} \tag{3}$$

$$\nabla \times \mathbf{E} = -i\omega\mu_0 \mathbf{H}. \tag{4}$$

It was assumed that the electric field vector $\mathbf{E}$ and the magnetic field vector $\mathbf{H}$ have the time dependence,

$$e^{i\omega t}. \tag{5}$$

The tensor notation $\epsilon \cdot \mathbf{E}$ may be expressed in component form as

$$(\epsilon \cdot \mathbf{E})_i = \epsilon_{ij} E_j. \tag{6}$$

Summation over double indices is understood, and the subscripts i and j assume the values 1, 2, and 3 that represent the x, y, and z components of the vector $\mathbf{E}$ or tensor ϵ.

Derivation of coupled-wave equations for isotropic media is facilitated by expressing the longitudinal components of $\mathbf{E}$ and $\mathbf{H}$ in terms of the transverse components.[3] This practice is preserved for our derivation of coupled equations for anisotropic media. We single out the z coordinate as the direction of the waveguide axis and express the field vectors and the differential operator ∇ as superpositions of transverse and longitudinal parts. The symbol t indicates the transverse directions x and y. Thus, we have

$$\mathbf{E} = \mathbf{E}_t + \mathbf{E}_z, \tag{7}$$

$$\mathbf{H} = \mathbf{H}_t + \mathbf{H}_z, \tag{8}$$

and

$$\nabla = \nabla_t + \mathbf{e}_z \frac{\partial}{\partial z}. \tag{9}$$

We use the notations $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ to indicate unit vectors in x, y, and z directions.

The transverse part of the vector $\epsilon \cdot \mathbf{E}$ is indicated by the notation $\epsilon_t \cdot \mathbf{E}$ or, in component notation,

$$\epsilon_x \cdot \mathbf{E} = \mathbf{e}_x(\epsilon_{xx} E_x + \epsilon_{xy} E_y + \epsilon_{xz} E_z) \tag{10}$$

$$\epsilon_y \cdot \mathbf{E} = \mathbf{e}_y(\epsilon_{yx} E_x + \epsilon_{yy} E_y + \epsilon_{yz} E_z). \tag{11}$$

The longitudinal part is

$$\epsilon_z \cdot \mathbf{E} = \mathbf{e}_z(\epsilon_{zx} E_x + \epsilon_{zy} E_y + \epsilon_{zz} E_z). \tag{12}$$

We may now separate Maxwell's equations into transverse and longitudinal parts. The transverse parts of (3) and (4) are

$$\nabla_t \times \mathbf{H}_z + \mathbf{e}_z \times \frac{\partial \mathbf{H}_t}{\partial z} = i\omega \epsilon_t \cdot \mathbf{E} \tag{13}$$

$$\nabla_t \times \mathbf{E}_z + \mathbf{e}_z \times \frac{\partial \mathbf{E}_t}{\partial z} = -i\omega \mu_0 \mathbf{H}_t. \tag{14}$$

Their longitudinal parts may be written as

$$\nabla_t \times \mathbf{H}_t = i\omega(\epsilon_z \cdot \mathbf{E}_t + \epsilon_{zz} \mathbf{E}_z) \tag{15}$$

and

$$\nabla_t \times \mathbf{E}_t = -i\omega \mu_0 \mathbf{H}_z. \tag{16}$$

The longitudinal parts of $\mathbf{E}$ and $\mathbf{H}$ follow immediately from (15) and (16),

$$\mathbf{E}_z = \frac{1}{i\omega\epsilon_{zz}} \nabla_t \times \mathbf{H}_t - \frac{1}{\epsilon_{zz}} \epsilon_z \cdot \mathbf{E}_t \tag{17}$$

and

$$\mathbf{H}_z = -\frac{1}{i\omega\mu_0} \nabla_t \times \mathbf{E}_t. \tag{18}$$

On the right-hand side of (17) and (18) appear only transverse components of $\mathbf{E}$ and $\mathbf{H}$. It is important to distinguish between the single and double subscript notation of ϵ. A double subscript, like ϵ_{zz}, indicates a single tensor element of ϵ, while a single subscript, like ϵ_z, is defined by (10) through (12). In particular, we have

$$\epsilon_z \cdot \mathbf{E}_t = \mathbf{e}_z(\epsilon_{zx}E_x + \epsilon_{zy}E_y). \tag{19}$$

We now use (17) and (18) to eliminate the z components of $\mathbf{E}$ and $\mathbf{H}$ from the transverse parts of Maxwell's equations (13) and (14),

$$-\frac{1}{i\omega\mu_0} \nabla_t \times (\nabla_t \times \mathbf{E}_t) + \mathbf{e}_z \times \frac{\partial \mathbf{H}_t}{\partial z} = i\omega\epsilon_t \cdot \mathbf{E}_t - \frac{i\omega}{\epsilon_{zz}} \epsilon_t \cdot \epsilon_z \cdot \mathbf{E}_t + \frac{1}{\epsilon_{zz}} \epsilon_t \cdot (\nabla_t \times \mathbf{H}_t) \tag{20}$$

and

$$\nabla_t \times \left[\frac{1}{i\omega\epsilon_{zz}} \nabla_t \times \mathbf{H}_t - \frac{1}{\epsilon_{zz}} \epsilon_z \cdot \mathbf{E}_t\right] + \mathbf{e}_z \times \frac{\partial \mathbf{E}_t}{\partial z} = -i\omega\mu_0 \mathbf{H}_t. \tag{21}$$

These two vector equations represent four scalar equations. Once eqs. (20) and (21) are solved, the z components of $\mathbf{E}$ and $\mathbf{H}$ can be obtained by simple differentiation from (17) and (18). We have thus achieved a simplification of the original problem by reducing the number of equations from six, in (3) and (4), to only four.

The components of the ϵ tensor are assumed to be functions of x, y, and z. The ϵ tensor defines the wave-guiding structure. Because of the z dependence of ϵ, eqs. (20) and (21) do not have mode solutions. A normal mode is defined as a solution of Maxwell's equations whose z dependence can be expressed by the simple function

$$e^{-i\beta z}. \tag{22}$$

Such solutions exist only if the dielectric tensor does not depend on the z coordinate. To construct solutions of the general eqs. (20) and (21), we consider solutions of simpler equations that are defined by a tensor $\bar{\epsilon}$ that is similar to ϵ but is independent of z. The choice of $\bar{\epsilon}$ is obviously arbitrary and is determined by convenience. Using (22), we find from (20) and (21) the following equations for the normal

modes of the waveguide structure defined by $\bar{\epsilon}$:

$$-\frac{1}{i\omega\mu_0}\nabla_t \times (\nabla_t \times \mathcal{E}_{\nu t}^{(p)}) - i\beta_\nu^{(p)}\mathbf{e}_z \times \mathcal{H}_{\nu t}^{(p)} = i\omega\bar{\epsilon}_t \cdot \mathcal{E}_{\nu t}^{(p)} - \frac{i\omega}{\bar{\epsilon}_{zz}}\bar{\epsilon}_t \cdot \bar{\epsilon}_z \cdot \mathcal{E}_{\nu t}^{(p)} + \frac{1}{\bar{\epsilon}_{zz}}\bar{\epsilon}_t \cdot (\nabla_t \times \mathcal{H}_{\nu t}^{(p)}) \quad (23)$$

and

$$\nabla_t \times \left[\frac{1}{i\omega\bar{\epsilon}_{zz}}\nabla_t \times \mathcal{H}_{\nu t}^{(p)} - \frac{1}{\bar{\epsilon}_{zz}}\bar{\epsilon}_z \cdot \mathcal{E}_{\nu t}^{(p)}\right] - i\beta_\nu^{(p)}\mathbf{e}_z \times \mathcal{E}_{\nu t}^{(p)} = -i\omega\mu_0\mathcal{H}_{\nu t}^{(p)}. \quad (24)$$

The subscript ν indicates a mode label. Equations (23) and (24) admit an infinite number of solutions with different eigenvalues (propagation constants) $\beta_\nu^{(p)}$ and different field vectors $\mathcal{E}_\nu^{(p)}$ and $\mathcal{H}_\nu^{(p)}$. Script letters indicate mode fields, while roman letters **E** and **H** are reserved for general field distributions. The modes are of two different types, guided modes whose fields are confined to the vicinity of the waveguide and radiation modes that extend to infinity in transverse direction to the guide.[2,3] Guided modes have discrete eigenvalues $\beta_\nu^{(p)}$, while the eigenvalues of radiation modes form a continuum. The superscript (p) stands for either $(+)$ or $(-)$, depending on the direction of wave propagation. A wave traveling in the positive z direction has positive (real) values $\beta_\nu^{(+)}$, a wave traveling in the negative z direction has a negative (real) value $\beta_\nu^{(-)}$. In isotropic media, we have the simple relations,

$$\beta_\nu^{(-)} = -\beta_\nu^{(+)}, \quad (25)$$

$$\mathcal{E}_{\nu t}^{(-)} = \mathcal{E}_{\nu t}^{(+)}, \qquad \mathcal{E}_{\nu z}^{(-)} = -\mathcal{E}_{\nu z}^{(+)}, \quad (26)$$

and

$$\mathcal{H}_{\nu t}^{(-)} = -\mathcal{H}_{\nu t}^{(+)}, \qquad \mathcal{H}_{\nu z}^{(-)} = \mathcal{H}_{\nu z}^{(+)}. \quad (27)$$

General anisotropic media are more complicated, so that (26) to (27) do not apply. Modes traveling in one direction may be different from modes traveling in the opposite direction.

III. ORTHOGONALITY RELATIONS

The modes of anisotropic dielectric waveguides are mutually orthogonal.[2] For the purpose of deriving orthogonality relations, it is simpler to use Maxwell's equations in the form (3) and (4) instead of the form (23) and (24). Separating the z derivatives from the ∇ operator, we write (3) for a mode labeled ν and (4) for a mode labeled μ,

$$\nabla_t \times \mathcal{H}_\nu^{(p)} - i\beta_\nu^{(p)}\mathbf{e}_z \times \mathcal{H}_\nu^{(p)} = i\omega\bar{\epsilon} \cdot \mathcal{E}_\nu^{(p)} \quad (28)$$

and

$$\nabla_t \times \mathcal{E}_\mu^{(q)} - i\beta_\mu^{(q)}\mathbf{e}_z \times \mathcal{E}_\mu^{(q)} = -i\omega\mu_0\mathcal{H}_\mu^{(q)}. \quad (29)$$

Next, we take the complex conjugate of (28), multiply the resulting equation by $\mathcal{E}_\mu^{(q)}$, multiply (29) by $-\mathcal{H}_\nu^{(p)*}$, add the two equations, and integrate over the infinite cross section ($\bar{\epsilon}$ is assumed real):

$$\int\int \{\mathcal{E}_\mu^{(q)} \cdot \nabla_t \times \mathcal{H}_\nu^{(p)*} - \mathcal{H}_\nu^{(p)*} \cdot \nabla_t \times \mathcal{E}_\mu^{(q)} + i\beta_\nu^{(p)*} \mathcal{E}_\mu^{(q)} \cdot \mathbf{e}_z \times \mathcal{H}_\nu^{(p)*} + i\beta_\mu^{(q)} \mathcal{H}_\nu^{(p)*} \cdot \mathbf{e}_z \times \mathcal{E}_\mu^{(q)}\} dxdy = -i\omega \int\int [\mathcal{E}_\mu^{(q)} \cdot \bar{\epsilon} \cdot \mathcal{E}_\nu^{(p)*} - \mu_0 \mathcal{H}_\nu^{(p)*} \cdot \mathcal{H}_\mu^{(q)}] dxdy. \quad (30)$$

The first two terms on the left-hand side of (30) can be expressed as

$$-\int\int \nabla_t \cdot (\mathcal{E}_\mu^{(q)} \times \mathcal{H}_\nu^{(p)*}) dxdy = -\int (\mathcal{E}_\mu^{(q)} \times \mathcal{H}_\nu^{(p)*}) \cdot \mathbf{n} ds. \quad (31)$$

The two-dimensional divergence theorem was used to convert the integral over the infinite cross section in the x-y plane to an integral over the infinite circle with outward normal direction $\mathbf{n}$ and line element ds. The integral on the right-hand side vanishes if at least one of the two modes is a guided mode. If both modes ν and μ are radiation modes, the integral vanishes in the sense of a delta function of nonzero argument.[2] Using this fact and a well-known vector identity, we can express (30) as

$$(\beta_\mu^{(q)} - \beta_\nu^{(p)*}) \int\int \mathbf{e}_z \cdot (\mathcal{E}_\mu^{(q)} \times \mathcal{H}_\nu^{(p)*}) dxdy = -\omega \int\int [\mathcal{E}_\mu^{(q)} \cdot \bar{\epsilon} \cdot \mathcal{E}_\nu^{(p)*} - \mu_0 \mathcal{H}_\mu^{(q)} \cdot \mathcal{H}_\nu^{(p)*}] dxdy. \quad (32)$$

Because of the symmetry of the $\bar{\epsilon}$ tensor, the following relation holds:

$$\mathcal{E}_\nu^{(p)*} \cdot \bar{\epsilon} \cdot \mathcal{E}_\mu^{(q)} = \mathcal{E}_\mu^{(q)} \cdot \bar{\epsilon} \cdot \mathcal{E}_\nu^{(p)*}. \quad (33)$$

We take the complex conjugate of (32), interchange the superscripts p and q as well as the subscripts ν and μ and, using (33), subtract the new expression from (32) with the result:

$$(\beta_\mu^{(q)} - \beta_\nu^{(p)*}) \int\int \mathbf{e}_z \cdot [\mathcal{E}_\mu^{(q)} \times \mathcal{H}_\nu^{(p)*} + \mathcal{E}_\nu^{(p)*} \times \mathcal{H}_\mu^{(q)}] dxdy = 0. \quad (34)$$

Equation (34) is the desired orthogonality relation. It is obvious that this expression holds also for isotropic media. However, in the isotropic case it is possible to use (25) through (27) to prove that each term in (34) must vanish separately.[2] For the general anisotropic case, (34) cannot be simplified further. We infer from (34) that the integral vanishes if $\beta_\mu^{(q)} - \beta_\nu^{(p)*} \neq 0$. This means that the integral vanishes even in the case $\nu = \mu$ if p and q indicate opposite signs, and a wave is orthogonal to its backward traveling counterpart (if $\beta_\nu^{(q)}$ is real) if orthogonality means vanishing of the integral in (34).

The integral in (34) expresses the total power flow if $\nu = \mu$ and $p = q$. We may therefore use the orthonormality relation,

$$\int\int \mathbf{e}_z \cdot [\mathcal{E}_{\mu t}^{(q)} \times \mathcal{H}_{\nu t}^{(p)*} + \mathcal{E}_{\nu t}^{(p)*} \times \mathcal{H}_{\mu t}^{(q)}] dx dy = 2s_\nu^{(q)} \frac{\beta_\nu^{(q)} + \beta_\nu^{(p)*}}{|\beta_\nu^{(q)}|} P\delta_{\nu\mu}, \quad (35)$$

to express mode orthogonality and normalization. The subscripts t indicating the transverse parts of the modes were added since the z components of the fields do not contribute to (35). P is a normalizing factor common to all modes that is used to adjust the arbitrary amplitudes of the normal modes. For real values of $\beta_\nu^{(q)}$, we have

$$s_\mu^{(q)} = 1. \quad (36)$$

In this case, the sign of the integral is expressed correctly by the fact that $\beta_\nu^{(q)}$ reverses its sign if q goes from $(+)$ to $(-)$. For opposite signs of p and q, the right-hand side of (35) vanishes as required by (34) if $\beta_\nu^{(q)}$ is real. For imaginary $\beta_\nu^{(q)}$, (35) vanishes for $q = p$. The orthogonality relation also holds for imaginary values of $\beta_\nu^{(q)}$. Imaginary values of the propagation constants occurs only for evanescent "radiation" modes.[2,3] In the case of imaginary $\beta_\nu^{(q)}$, the sign of the right-hand side of (35) is not certain. For this reason, we have introduced the factor $s_\mu^{(q)}$ that must be adjusted so that P is a positive real quantity. This means that $s_\mu^{(q)}$ may have to be negative, $s_\mu^{(q)} = -1$. However, this case can arise only in connection with evanescent "radiation" modes. The $\delta_{\nu\mu}$ symbol in (35) indicates Kronecker's delta if both modes are guided. When one mode is guided while the other is a radiation mode, we have $\delta_{\nu\mu} = 0$. If both modes are radiation modes, $\delta_{\nu\mu}$ must be interpreted as the Dirac delta function.

IV. DERIVATION OF COUPLED-WAVE EQUATIONS

Any arbitrary field distribution compatible with Maxwell's equations can be expressed as the superposition of all the modes of the idealized structure defined by the dielectric tensor $\bar{\epsilon}$. Because the complete set of modes consists of a finite number of guided modes plus a continuum of radiation modes, we express the transverse parts of a general field by the expansion

$$\mathbf{E}_t = \sum_{\nu,p} a_\nu^{(p)} \mathcal{E}_{\nu t}^{(p)} + \sum_p \int_0^\infty a^{(p)}(\rho)\, \mathcal{E}_t^{(p)}(\rho) d\rho \quad (37)$$

and

$$\mathbf{H}_t = \sum_{\nu,p} a_\nu^{(p)} \mathcal{H}_{\nu t}^{(p)} + \sum_p \int_0^\infty a^{(p)}(\rho) \mathcal{H}_t^{(p)}(\rho) d\rho. \quad (38)$$

The longitudinal parts follow from (17) and (18). The superscripts assume the values (+) and (−) indicating waves traveling in positive and negative z direction. The first terms in (37) and (38) represent the contribution of the finite number of guided modes labeled ν. The second terms indicated combinations of sums and integrals. The integration ranges over the entire region of continuous-mode labels ρ and includes radiation modes with real as well as imaginary values of $\beta^{(p)}(\rho)$. The summation symbol in front of the integral sign indicates that, in addition to modes traveling in positive and negative z direction, various types of radiation modes exist and must be added to obtain the complete set of modes. For the purpose of deriving coupled-wave equations, the notation of (37) and (38) is too cumbersome. We use an abbreviated notation by omitting the integration sign, leaving it understood that the summation symbol includes summation over guided modes and summation as well as integration over radiation modes. We thus write

$$\mathbf{E}_t = \sum_{\nu,p} a_\nu^{(p)} \mathcal{E}_{\nu t}^{(p)} \tag{39}$$

and

$$\mathbf{H}_t = \sum_{\nu,p} a_\nu^{(p)} \mathcal{H}_{\nu t}^{(p)}. \tag{40}$$

$\mathcal{E}_{\nu t}^{(p)}$ and $\mathcal{H}_{\nu t}^{(p)}$ are independent of z, but $a_\nu^{(p)}$ is a function of z. Substitution of (39) and (40) into (20) and (21) and use of the mode eqs. (23) and (24) leads to

$$\begin{aligned}\sum_{\nu,p} \left(\frac{da_\nu^{(p)}}{dz} + i\beta_\nu^{(p)} a_\nu^{(p)} \right) (\mathbf{e}_z \times \mathcal{H}_{\nu t}^{(p)}) \\ = \sum_{\nu,p} a_\nu^{(p)} \left\{ i\omega(\epsilon_t - \tilde{\epsilon}_t) \cdot \mathcal{E}_{\nu t}^{(p)} - i\omega \left(\frac{1}{\epsilon_{zz}} \epsilon_t \cdot \epsilon_z - \frac{1}{\tilde{\epsilon}_{zz}} \tilde{\epsilon}_t \cdot \tilde{\epsilon}_z \right) \cdot \mathcal{E}_{\nu t}^{(p)} \right. \\ \left. + \left(\frac{1}{\epsilon_{zz}} \epsilon_t - \frac{1}{\tilde{\epsilon}_{zz}} \tilde{\epsilon}_t \right) \cdot (\nabla_t \times \mathcal{H}_{\nu t}^{(p)}) \right\}\end{aligned} \tag{41}$$

and

$$\begin{aligned}\sum_{\nu,p} \left(\frac{da_\nu^{(p)}}{dz} + i\beta_\nu^{(p)} a_\nu^{(p)} \right) (\mathbf{e}_z \times \mathcal{E}_{\nu t}^{(p)}) \\ = - \sum_{\nu,p} a_\nu^{(p)} \left\{ \nabla_t \times \left[\frac{1}{i\omega} \left(\frac{1}{\epsilon_{zz}} - \frac{1}{\tilde{\epsilon}_{zz}} \right) \nabla_t \times \mathcal{H}_{\nu t}^{(p)} \right. \right. \\ \left. \left. - \left(\frac{1}{\epsilon_{zz}} \epsilon_z - \frac{1}{\tilde{\epsilon}_{zz}} \tilde{\epsilon}_z \right) \cdot \mathcal{E}_{\nu t}^{(p)} \right] \right\}.\end{aligned} \tag{42}$$

We take the scalar product of (41) with $-\mathcal{E}_{\mu t}^{(q)*}$ and of (42) with $\mathcal{H}_{\mu t}^{(q)*}$, then we add the two equations, integrate over the infinite cross section,

and use the orthogonality relation (35). The result of this procedure is

$$2s_\mu^{(r)} \frac{\beta_\mu^{(q)} + \beta_\mu^{(r)*}}{|\beta_\mu^{(r)}|} P \left(\frac{da_\mu^{(r)}}{dz} + i\beta_\mu^{(r)} a_\mu^{(r)} \right)$$
$$= \sum_{\nu,p} a_\nu^{(p)} \int\int \Big\{ -i\omega \mathcal{E}_{\mu t}^{(q)*} \cdot (\epsilon_t - \bar{\epsilon}_t) \cdot \mathcal{E}_{\nu t}^{(p)}$$
$$+ i\omega \mathcal{E}_{\mu t}^{(q)*} \cdot \left(\frac{\epsilon_t \cdot \epsilon_z}{\epsilon_{zz}} - \frac{\bar{\epsilon}_t \cdot \bar{\epsilon}_z}{\bar{\epsilon}_{zz}} \right) \cdot \mathcal{E}_{\nu t}^{(p)} - \mathcal{E}_{\mu t}^{(q)*} \cdot \left(\frac{\epsilon_t}{\epsilon_{zz}} - \frac{\bar{\epsilon}_t}{\bar{\epsilon}_{zz}} \right) \cdot (\nabla_t \times \mathcal{H}_{\nu t}^{(p)})$$
$$- \mathcal{H}_{\mu t}^{(q)*} \cdot \nabla_t \times \left[\frac{1}{i\omega} \left(\frac{1}{\epsilon_{zz}} - \frac{1}{\bar{\epsilon}_{zz}} \right) \nabla_t \times \mathcal{H}_{\nu t}^{(p)} \right.$$
$$\left. - \left(\frac{\epsilon_z}{\epsilon_{zz}} - \frac{\bar{\epsilon}_z}{\bar{\epsilon}_{zz}} \right) \cdot \mathcal{E}_{\nu t}^{(p)} \right] \Big\} dxdy. \quad (43)$$

On the left-hand side of (43), we have used a superscript (r). This notation is necessary to distinguish between the case of real and imaginary propagation constants $\beta_\mu^{(q)}$. If $\beta_\mu^{(q)}$ is real, we have $r = q$. If $\beta_\mu^{(q)}$ is imaginary, as it is for evanescent radiation modes, we must choose for r the sign opposite to q. We now write (43) in the abbreviated form

$$\frac{da_\mu^{(r)}}{dz} = -i\beta_\mu^{(r)} a_\mu^{(r)} + \sum_{\nu,p} K_{\mu\nu}^{(r,p)} a_\nu^{(p)}. \quad (44)$$

The coupling coefficient is defined by (43). We may eliminate the transverse magnetic-mode field vector from the coupling coefficient by using (17) (applied to the mode field) and the identity (which is obtained by partial integration),

$$\int\int \mathcal{H}_{\mu t}^{(q)*} \cdot (\nabla_t \times \mathbf{F}) dxdy = \int\int (\nabla_t \times \mathcal{H}_{\mu t}^{(q)*}) \cdot \mathbf{F} dxdy. \quad (45)$$

The coupling coefficient can be expressed as

$$K_{\mu\nu}^{(r,p)} = \frac{i\omega |\beta_\mu^{(r)}|}{4s_\mu^{(r)} \beta_\mu^{(r)*} P} \int\int \Big\{ \mathcal{E}_{\mu t}^{(q)*} \cdot \left[\left(\frac{\epsilon_t \cdot \epsilon_z}{\epsilon_{zz}} - \frac{\bar{\epsilon}_t \cdot \bar{\epsilon}_z}{\bar{\epsilon}_{zz}} \right) - (\epsilon_t - \bar{\epsilon}_t) \right] \cdot \mathcal{E}_{\nu t}^{(p)}$$
$$- \mathcal{E}_{\mu t}^{(q)*} \cdot \left(\frac{\bar{\epsilon}_{zz}}{\epsilon_{zz}} \epsilon_t - \bar{\epsilon}_t \right) \cdot \left(\frac{\bar{\epsilon}_z}{\bar{\epsilon}_{zz}} \cdot \mathcal{E}_{\nu t}^{(p)} + \mathcal{E}_{\nu z}^{(p)} \right)$$
$$+ (\bar{\epsilon}_z \cdot \mathcal{E}_{\mu t}^{(q)*} + \bar{\epsilon}_{zz} \mathcal{E}_{\mu z}^{(q)*}) \cdot \left[\left(\frac{\bar{\epsilon}_{zz}}{\epsilon_{zz}} - 1 \right) \left(\frac{\bar{\epsilon}_z}{\bar{\epsilon}_{zz}} \cdot \mathcal{E}_{\nu t}^{(p)} + \mathcal{E}_{\nu z}^{(p)} \right) \right.$$
$$\left. - \left(\frac{\epsilon_z}{\epsilon_{zz}} - \frac{\bar{\epsilon}_z}{\bar{\epsilon}_{zz}} \right) \cdot \mathcal{E}_{\nu t}^{(p)} \right] \Big\} dxdy. \quad (46)$$

V. IMPORTANT SPECIAL CASES

In its complete form, (46), the coupling coefficient is very complicated. For isotropic media, where the ϵ tensor degenerates to a multiple

of the unit tensor, (46) simplifies to the exact form,

$$K_{\mu\nu}^{(r,p)} = \frac{\omega|\beta_\mu^{(r)}|}{4is_\mu^{(r)}\beta_\mu^{(r)*}P}\int\int(\epsilon - \bar{\epsilon})\left[\mathfrak{E}_{\mu t}^{(q)*}\cdot\mathfrak{E}_{\nu t}^{(p)} + \frac{\bar{\epsilon}_{zz}}{\epsilon_{zz}}\mathfrak{E}_{\mu z}^{(q)*}\cdot\mathfrak{E}_{\nu z}^{(p)}\right]dxdy, \quad (47)$$

in complete agreement with the well-known result.[12]

In many practical applications, the anisotropy of the dielectric medium is only slight, and the difference between the actual dielectric tensor ϵ and the ideal tensor $\bar{\epsilon}$ is small. In that case, (46) can be substantially simplified. A reasonable approximation of (46) for slight anisotropy and small values of $\epsilon - \bar{\epsilon}$ is

$$K_{\mu\nu}^{(r,p)} = \frac{\omega|\beta_\mu^{(r)}|}{4is_\mu^{(r)}\beta_\mu^{(r)*}P}\int\int\mathfrak{E}_\mu^{(q)*}\cdot(\epsilon - \bar{\epsilon})\cdot\mathfrak{E}_\nu^{(p)}dxdy. \quad (48)$$

Note that the whole vectors of the electric mode fields enter (48) and not just the transverse or longitudinal parts. The approximation (48) is obtained by considering off-diagonal elements and differences between diagonal elements of ϵ and $\bar{\epsilon}$ as quantities that are small of first order. Products of two first-order quantities have been neglected. For readers who did not follow the detailed derivation, we repeat here briefly the definitions of symbols appearing in (46) through (48). The symbol ω is the angular frequency of the electromagnetic field. The script symbols $\mathfrak{E}_\nu^{(p)}$ indicate the electric-field vectors of normal modes of an idealized waveguide that is defined by the dielectric tensor $\bar{\epsilon} = \bar{\epsilon}(x, y)$, the subscript ν is a mode label, and the superscript (p) stands for either $(+)$ or $(-)$, indicating the direction of wave propagation. The propagation constants $\beta_\mu^{(r)}$ of the modes are labeled in the same way as the field vectors. The superscript r is usually identical with the superscript q. Only in the case of imaginary $\beta_\mu^{(r)}$ (this case happens for coupling to a nonpropagating radiation mode and is of little practical interest) does r indicate the sign opposite to q. Likewise, $s_\mu^{(r)} = 1$ for most cases of interest. Only for imaginary values of $\beta_\mu^{(r)}$ may it become necessary to choose $s_\mu^{(r)} = -1$ to keep the power normalization coefficient P positive in (35). The asterisk indicates complex conjugation. Subscripts t and z occurring in (46) and (47) refer to the transverse and longitudinal parts of the vectors to which they are attached. Similar subscripts attached to ϵ are defined by (10) through (12) and (19). The dielectric tensor $\epsilon = \epsilon(x, y, z)$ defines the actual waveguide (in contrast to the ideal guide that is only a mathematical fiction). The integrals are extended over the infinite transverse cross section of the guide. Equation (48) assumes the same limit as (47) if the dielectric tensor degenerates into a multiple of the unit tensor since, in the spirit of the approximation (48), we must use $\bar{\epsilon}_{zz}/\epsilon_{zz} = 1$. Equation (48), even though it is only an approximation, is likely to be

of most importance in practical applications because of its simple form. For many practical problems, the approximation is justified and leads to sufficiently accurate results.

VI. CONCLUSIONS

We have derived coupled-wave equations representing exact solutions of the electromagnetic field problem for dielectric waveguides that consist of anisotropic materials whose dielectric tensor is a function of the z coordinate. The field of the general waveguide is expressed in terms of ideal modes of a hypothetical dielectric waveguide defined by a dielectric tensor whose elements are independent of the z coordinate. The main result of this paper is the expression (46) for the coupling coefficients. For many practical applications, the exact coupling coefficient can be approximated in the simple form (48).

The coupled-mode theory for anisotropic dielectric waveguides is essential for the solution of problems of mode propagation in integrated-optics guides with random or systematic irregularities. A particularly important area of applications are guides that are made anisotropic by an externally applied dc voltage or whose anisotropy is changed by such a voltage. Instead of an applied voltage, an acoustical wave may cause an anisotropic change of the refractive index of a dielectric waveguide. These cases cannot be handled by the simpler isotropic coupled-mode theory, but require the extension to anisotropic media presented here.

REFERENCES

1. N. S. Kapany and J. J. Burke, *Optical Waveguides*, New York: Academic Press, 1972.
2. D. Marcuse, *Light Transmission Optics*, New York: Van Nostrand Reinhold, 1972.
3. D. Marcuse, *Theory of Dielectric Optical Waveguides*, New York: Academic Press, 1974.
4. A. W. Snyder, "Coupled Mode Theory for Optical Fibers," J. Opt. Soc. Am., *62*, No. 11 (November 1972), pp. 1267–1277.
5. S. Wang, M. L. Shah, and J. D. Crow, "Wave Propagation in Thin-Film Optical Waveguides Using Gyrotopic and Anisotropic Materials as Substrates," IEEE J. Quant. Elec., *QE-8*, Part 2, No. 2 (February 1972), pp. 212–216.
6. A. Yariv, "Coupled Mode Theory for Guided Waves," IEEE J. Quant. Elec., *QE-9*, No. 9 (September 1973), pp. 919–933.
7. T. P. Sosnowski and G. D. Boyd, "The Efficiency of Thin-Film Optical-Waveguide Modulators Using Electrooptic Films or Substrates," IEEE J. Quant. Elec., *QE-10*, No. 3 (March 1974), pp. 306–311.
8. I. P. Kaminow and J. R. Carruthers, "Optical Waveguiding Layers in $LiNbO_3$ and $LiTaO_3$," Appl. Phys. Letters, *22*, No. 7 (April 1973), pp. 326–328.
9. J. M. Hammer and W. Phillips, "Low Loss Single-Mode Optical Waveguides and Efficient High-Speed Modulators of $LiNb_xTa_{1-x}O_3$ and $LiTaO_3$," Appl. Phys. Letters, *24*, No. 11 (June 15, 1974), pp. 545–547.
10. R. V. Schmidt and I. P. Kaminow, "Metal-Diffused Optical Waveguides in $LiNbO_3$," Appl. Phys. Letters, *25*, No. 8 (October 15, 1974), pp. 458–460.
11. M. Born and E. Wolf, *Principles of Optics*, 3rd Edition, New York: Pergamon Press, 1964.
12. Ref. 3, p. 104, Eq. (3.2-44).

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

Influences of Glass-to-Metal Sealing on the Structure and Magnetic Properties of an Fe/Co/V Alloy

By M. R. PINNEL and J. E. BENNETT

(Manuscript received December 26, 1974)

Apparent changes in the coercivity and remanence of the magnetic reed material used in the remreed sealed contact were encountered during the glass-to-metal sealing operation. By using primarily metallographic observations correlated with magnetic data, it was determined that these changes were due to a modification of the 600°C-aged microstructure caused by the time/temperature cycle experienced in sealing. The room-temperature stable precipitates that are developed by the 600°C anneal to produce magnetic hardening are dissolved, and the alloy converts to a two-phase duplex body-centered cubic (BCC) microstructure along one-half to two-thirds of the reed shank. This produces a substantial increase in coercive force and some decrease in remanence. The structure and properties of the alloy before sealing have little influence on the reaction.

I. INTRODUCTION

A recent effort in the telecommunications industry has been the development of a remanent reed, sealed contact (remreed).[1] One problem encountered in this development was an apparent change in the magnetic properties of the Remendur (Fe, Co, V alloy) reed during the glass-to-metal sealing operation. During sealing, the reed can reach peak temperatures in excess of 1050°C and may be at temperatures in excess of 900°C for as long as 8 to 10 seconds. Preliminary data supplied to the authors indicated an increase in the coercivity of reeds from the sealing operation,[2] whereas the results of Kitazawa, Oguma, and Hara showed a 22-percent decrease in coercivity at switch sealing.[3] Therefore, a description of the exact nature of this change and an explanation of the mechanism by which it occurs were sought. Also, information on the possible influence of variations in the time/temperature sealing cycle on the magnitude of this change was considered relevant to determine if improvements could be achieved by such modifications.

This paper summarizes the results of both a laboratory study on the Remendur alloy and evaluation of manufactured contacts to ascertain the mechanism for the reported property changes. The normal sealing operation was also modified by increasing the sealing speed and reducing the sealing-lamp temperature to obtain data on possible modifications to minimize the changes. Also presented are data on the influence of microstructure from the prior strand-anneal and 600°C-aging anneal-heat treatments on the stability of the aged magnetic properties.

II. EXPERIMENTAL PROCEDURES

The evaluation on actual sealed reeds was carried out on Western Electric Company assembled contacts. A group of reeds stamped from a single coil of commercially melted and processed 0.535-mm Remendur wire was aged at 610°C in hydrogen for two hours. Some reeds of this group were assembled into contacts at two separate manufacturing facilities. Samples of the aged reeds and reeds that had been removed from these assembled contacts were supplied to the authors for evaluation (Experiment A). Additionally, contacts that had been sealed at an accelerated speed were provided (Experiment B). Finally, a two-group sample was supplied consisting of six typical contacts and six contacts sealed with a 100-W reduction in power on the sealing lamps (Experiment C).

The laboratory study was carried out on a selection of coils of wire from another commercial melt of Remendur (48.7-percent Fe, 48.2-percent Co, 2.90-percent V by weight with balance impurities Mn, Ni, C, S). The wire was supplied in a 900 to 950°C strand-annealed and quenched condition. Samples from the front and back ends of every coil from this melt were examined metallographically. A variety of microstructural variations were detected in this annealed wire. These ranged from a nearly-all-α_1 phase structure produced by an anneal at the lower extreme in the temperature range to an all-α_2 phase structure produced by an anneal at the upper temperature extreme.[4,5] A sample of 15 front and back coil ends covering this range in structures of the 0.535-mm strand-annealed wire was selected. The wires were aged at 620°C for three hours in an argon/10-percent hydrogen-gas mixture. Each of these coil ends was then divided into three separate lengths of wire. Each wire length was subjected to one of three different high-temperature exposures. These exposures were carried out by inserting the wires for a fixed length of time into an open-air muffle furnace set for a fixed elevated temperature. They were gripped in asbestos-covered tongs for the insertion to prevent the tongs from acting as a heat sink. The three time-temperature cycles used on the wires are shown in Figs. 1, 2, and 3. These profiles were produced by insert-

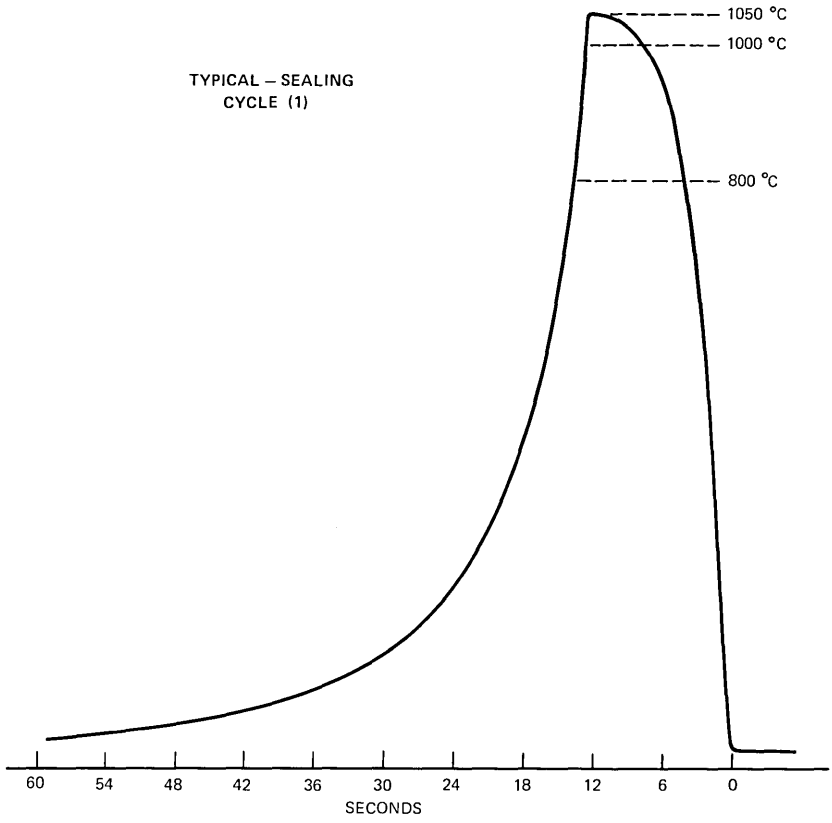


Fig. 1—Time/temperature thermocouple response for a 12-s insertion into a 1200°C furnace.

ing a chromel-alumel thermocouple made from 0.535-mm wire into the furnace in the same manner as the Remendur wire samples and monitoring the response on a strip-chart recorder. These cycles were selected to nearly duplicate typical sealing, reduced-temperature, and extended-time cycles, respectively. The typical cycle was determined from a previous study on the sealing of 237-type ferreed contacts.[6]

The production reeds and the laboratory samples in each heat-treated condition were evaluated metallographically and the magnetic parameters were determined. Metallographic preparation was routine, using a 5-percent Nital etch for 15 to 30 seconds. Structures were observed by optical light microscopy at 1500 magnification using Nomarski Differential Interference Contrast (DIC). All magnetic data were provided to the authors by E. C. Hellstrom. Coercivity and remanence were measured from full-loop traces on samples magnetized

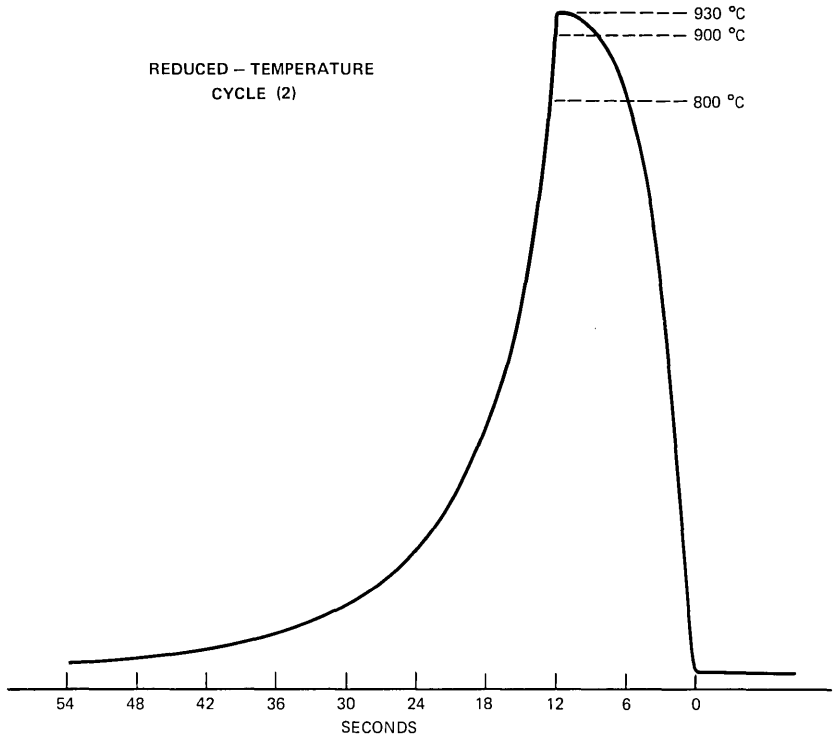


Fig. 2—Time/temperature thermocouple response for a 12-s insertion into a 1050°C furnace.

by an applied saturation field of 300 oersteds. All magnetic values presented are the average of multiple samples. The laboratory samples were 2.5 cm in length, and the production samples were actual reeds possessing their characteristic geometry (Fig. 4). The reeds were measured both as complete reeds with the search coil centered over the paddle portion and also as dissociated paddles and shanks.

III. RESULTS AND DISCUSSION

Microstructural comparisons and correlations were obtained on longitudinal sections of the shanks. Figure 5 shows the microstructures present in a typical reed shank after sealing, at approximately 1-mm increments along its length. It is apparent that, in this case, beginning at the paddle end, approximately the first 5.5 mm have undergone a structural change. This represents approximately two-thirds of the shank length. This heat-affected region is substantially larger than that actually encompassed by the glass in the seal.

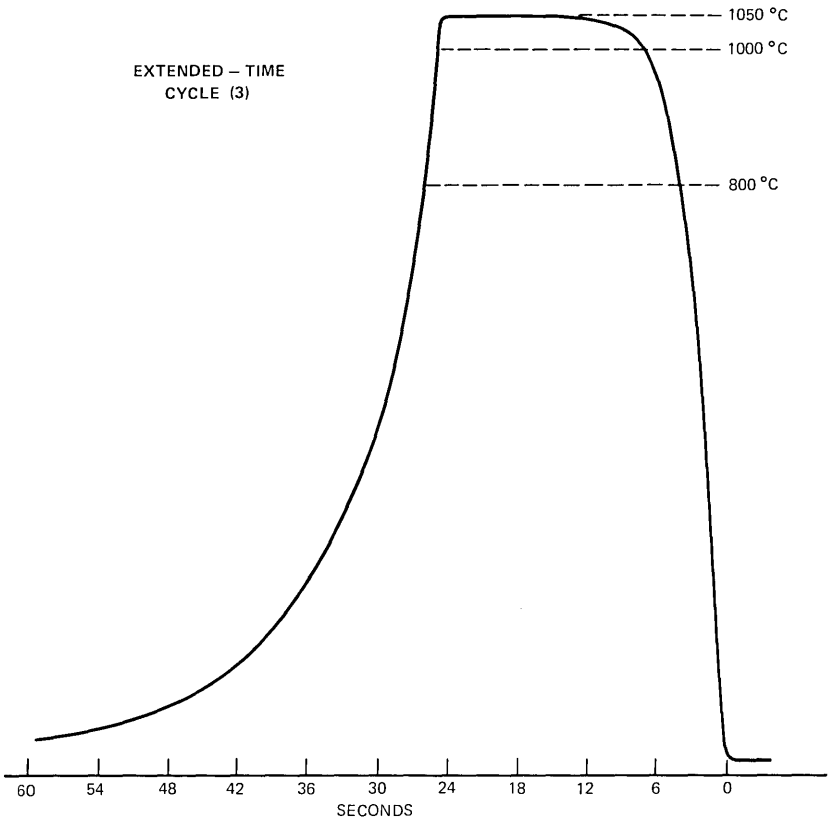


Fig. 3—Time/temperature thermocouple response for a 25-s insertion into a 1200°C furnace.

3.1 Laboratory-simulated sealing-cycle exposures

To determine the significance of this structural change, the laboratory experiment utilizing various elevated-temperature exposures was carried out. The magnetic-properties data for the 15 selected coils are summarized in Table I. Of the 15 coils selected, 11 are considered to possess a microstructure at least approximating a nearly equal distribution of the two phases, $\alpha_1 + \alpha_2$, in this duplex BCC structure.[7] This distribution is expected from an optimized-temperature strand anneal and quench.[4] In this strand-annealed state, these typical wires indicate a coercive force of 41 to 47 oersteds and a remanence of 21 to 23 maxwells for the specified measuring conditions. Three samples designated 208B, 240B, and 268B were strand-annealed at the low temperature extreme, i.e., less than about 900°C, and have a nearly single-phase structure of α_1 (BCC). They have a significantly lower co-

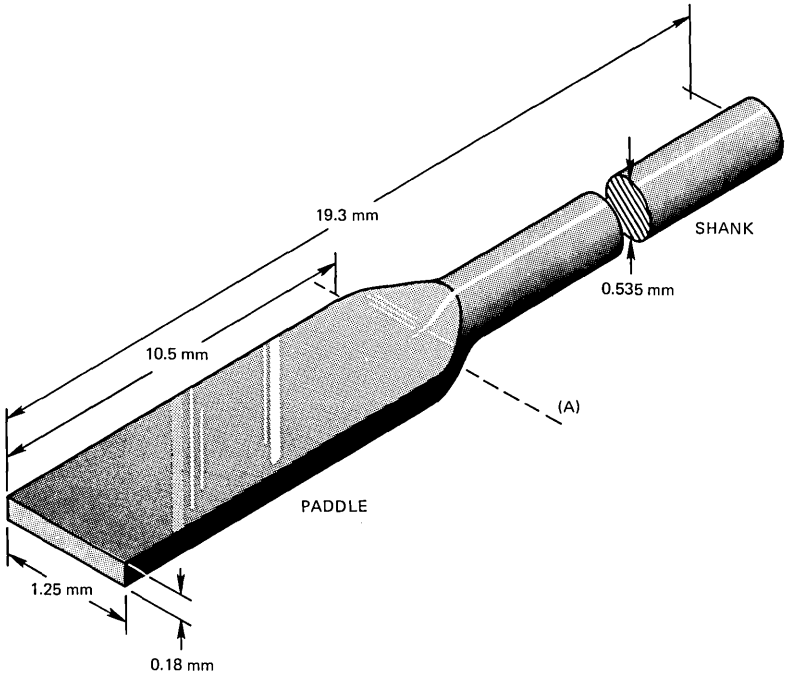


Fig. 4—Reed geometry used for magnetic evaluation.

ercive force with little difference in remanence when compared to the two-phase strand-annealed structure before aging. Finally, one sample designated 223B was strand-annealed at the upper temperature extreme, i.e., greater than about 950°C, and has a nearly single-phase structure of α_2 [the vanadium-supersaturated, FCC (face-centered cubic) to BCC transformation phase].[7] It shows not only a modestly reduced coercive force but a significantly lower remanence when compared to the two-phase strand-annealed structure. Representative microstructures for these three conditions are shown in Fig. 6, column 1.

Again referring to Table I, it appears that the 620°C-aging anneal has a tendency to normalize the magnetic properties. This is primarily a consequence of the precipitation of stable γ (FCC) phase,[5] relief of the $\gamma \rightarrow \alpha_2$ transformation strains,[7,8] and ordering of the α_1 matrix.[9] The four high- and low-temperature strand-annealed coils are still below the mean for the two-phased strand-annealed coils of 24 to 25 oersteds in coercive force and of 26 to 28 maxwells in remanence after aging. Representative microstructures in the 620°C-aged condition are also given in Fig. 6, column 2.

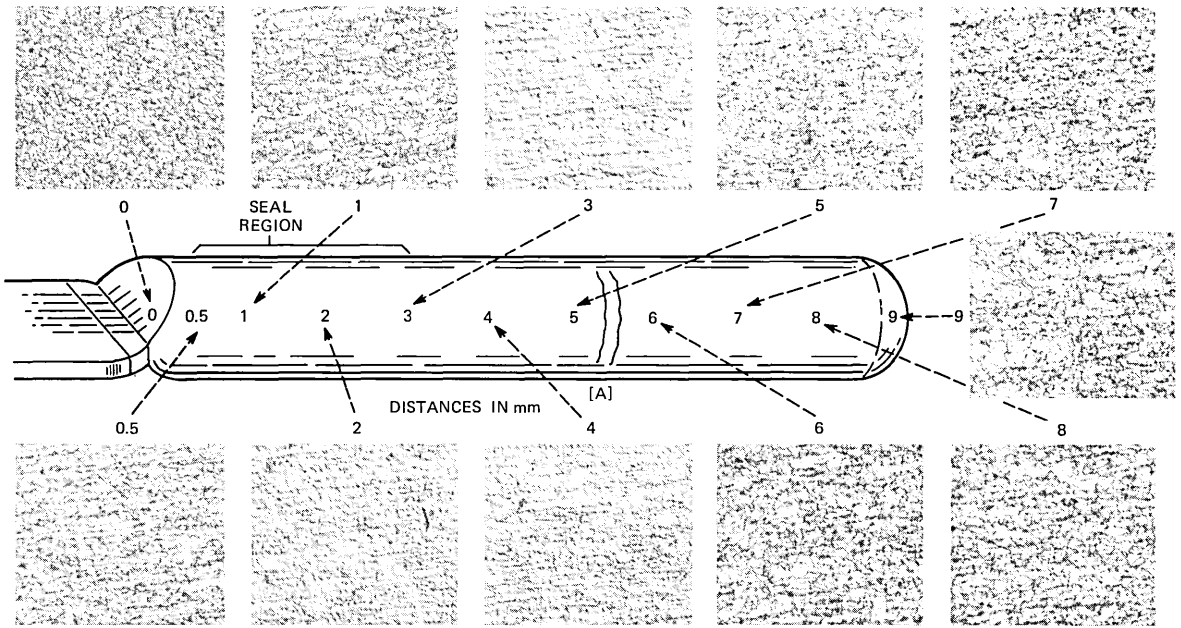


Fig. 5—Microstructures along the length of a typical sealed reed shank—etched 5-percent nital, DIC, 1500X. [A] represents the approximate end of the heat-affected zone.

Table I — Magnetic properties of Remendur wire before and after simulated sealing cycles 1, 2, and 3

Hc (oersteds)

Sample	Strand-Anneal Structure	Strand-Annealed	Aged 620°C—3 hr	Typical Cycle (1)	Low T Cycle (2)	Long t Cycle (3)
205 T	2 phase	41.4	24.3	32.8	16.6	28.6
205 B	2 phase	40.9	22.2	35.6	22.7	27.5
208 B	Primarily α_1	29.9	17.3	34.3	13.9	27.8
218 T	2 phase	46.6	25.5	35.4	—	29.1
218 B	2 phase	44.7	25.0	36.1	16.0	26.5
223 T	2 phase	42.6	23.8	37.4	19.7	28.8
223 B	Primarily α_2	35.5	22.7	33.0	14.2*	27.7*
231 T	2 phase	45.4	25.6	38.5	16.9	25.9
231 B	2 phase	47.3	27.3	38.5	18.3	28.9
236 T	2 phase	38.5	24.3	37.1	—	—
236 B	2 phase	43.5	26.4	36.3	14.5	26.3
240 T	2 phase	44.6	25.8	37.3	17.1	28.3
240 B	Primarily α_1	23.0	19.0	37.3	13.9	29.2
268 T	2 phase	46.7	27.0	38.5	17.8	27.4
268 B	Primarily α_1	21.0	19.0	37.0	—	29.5

$\emptyset_R$ (maxwells)

Sample	Strand-Anneal Structure	Strand-Annealed	Aged 620°C—3 hr	Typical Cycle (1)	Low T Cycle (2)	Long t Cycle (3)
205 T	2 phase	22.2	27.2	16.3	21.9	11.1
205 B	2 phase	22.3	23.2	18.1	24.2	11.3
208 B	Primarily α_1	22.1	21.4	16.5	20.6	11.0
218 T	2 phase	22.9	27.9	17.3	—	12.9
218 B	2 phase	22.6	27.5	18.2	22.6	11.8
223 T	2 phase	22.3	25.2	19.1	21.8	12.2
223 B	Primarily α_2	17.3	25.2	15.1	9.3*	7.6*
231 T	2 phase	22.8	28.3	20.2	22.9	11.3
231 B	2 phase	22.9	28.4	20.2	24.6	13.4
236 T	2 phase	21.9	27.7	19.6	—	—
236 B	2 phase	22.7	26.6	19.1	19.5	12.0
240 T	2 phase	22.7	25.7	19.0	23.0	11.4
240 B	Primarily α_1	20.6	20.8	19.5	19.3	12.2
268 T	2 phase	23.1	27.4	20.6	22.9	11.2
268 B	Primarily α_1	20.3	21.2	19.6	—	11.9

Material: 2.90 weight percent V (*i*) T = front of coil, (*ii*) B = back of coil, (*iii*) H_{app} = 300 oersteds, (*iv*) Values average of two or three samples, (*v*) Sample length = 2.5 cm.

* Aged at 515°C prior to elevated temperature exposure; 1.6-cm sample length.

Samples were exposed to the typical-sealing cycle given in Fig. 1, which was based on a previous experimental evaluation of the time-temperature cycle experienced by a 52-alloy reed in the sealing of a type 237 (nonlatching, Ni-Fe) reed contact.[6] This cycle creates an exposure to a peak temperature of 1050°C and temperatures in excess

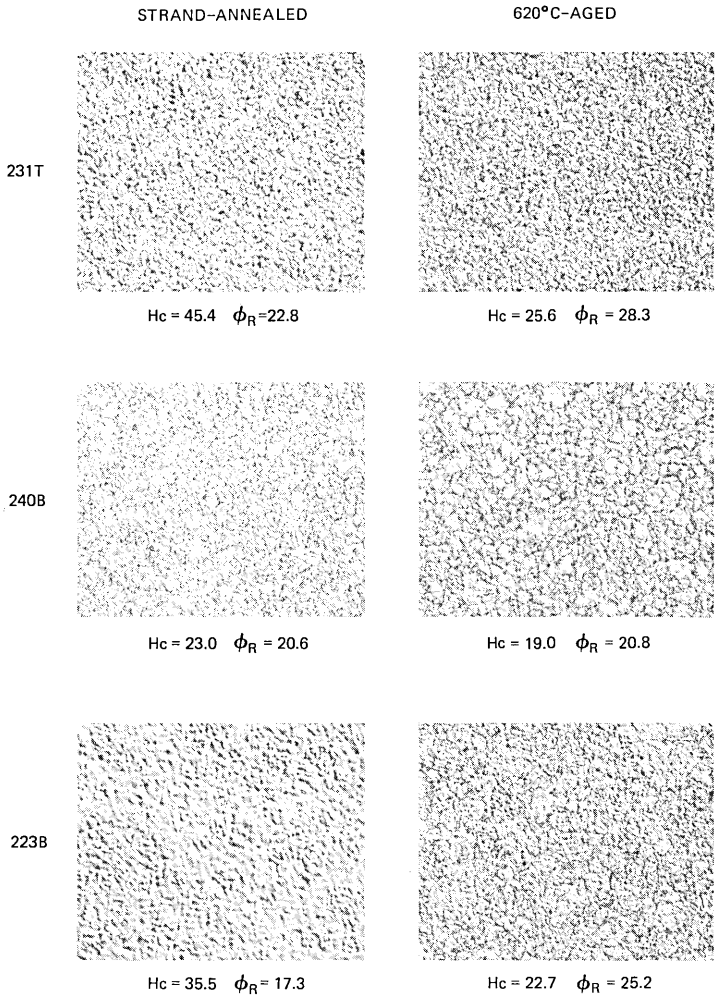


Fig. 6—Microstructures in strand-annealed and 620°C-aged conditions for two-phase (231T), primarily α_1-phase (240B), and primarily α_2-phase (223B) strand-annealed structures—etched 5-percent nital, DIC, 1500X.

of 900°C for approximately 7 seconds. This exposure produces a marked increase in coercive force into the range of 33 to 38 oersteds and a drop in remanence to 15 to 20 maxwells (Table I) for all samples. Representative microstructures given in Fig. 7, column 1, indicate that this change is a result of the material altering to the structure typically produced by the 900 to 950°C strand-anneal.[4] That is, the γ (FCC) precipitates are dissolved and portions of the matrix return to the *high-temperature* FCC phase and again undergo the FCC $\rightarrow$ BCC

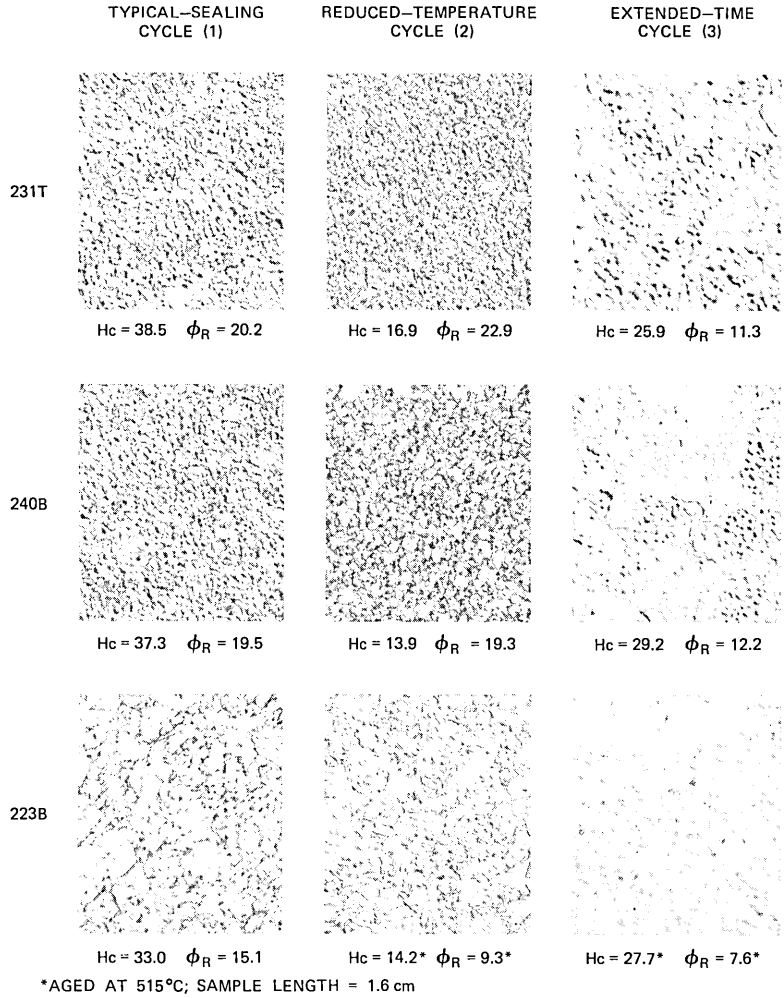


Fig. 7—Microstructures following various elevated-temperature exposures on samples shown in Fig. 6—etched 5-percent nital, DIC, 1500X.

transformation on cooling. This produces a highly strained lattice and, hence, the increased coercive force values. Similar results were obtained by Mahajan and Olsen[10] by exposing 600°C-aged Remendur to temperatures of 850, 950, and 1050°C for 5 seconds and measuring the magnetic changes as well as observing the structural variations utilizing transmission electron microscopy.

A slightly reduced temperature cycle was attempted to ascertain if a modest reduction in sealing temperatures could alleviate this change. This cycle, given in Fig. 2, had a peak temperature of 930°C and a temperature in excess of 900°C for less than 4 seconds. Magnetic-

properties results (Table I) show that the coercive force and remanence are both *reduced* by this cycle. However, the remanence decrease is small in all cases and, in general, those samples that had a two-phase strand-annealed microstructure before aging and high temperature exposure show a lesser decline in coercivity than the single-phase strand-annealed structures. Microstructural observations (Fig. 7, column 2) indicate that these changes are a result of the beginning of the dissolution of the γ (FCC) precipitate. This cycle is sufficiently low in temperature and time such that none of the matrix is returned to FCC structure at temperature and, hence, does not undergo the FCC $\rightarrow$ BCC transformation on cooling to markedly increase the magnetic hardness as observed in cycle 1.

To clearly prove that a reversion to the strand-annealed structure during sealing is the cause for the magnetic property changes observed in reeds,[2] an extended-time-at-temperature cycle was produced (Fig. 3). The peak temperature was maintained at 1050°C, but the time in excess of 900°C was increased to 20 seconds. In all cases, the structure totally converted to a single-phase α_2 matrix (Fig. 7, column 3), which is typical of a strand anneal at too high a temperature (>950 to 975°C).[4] These structures may be compared to the initial structure of coil 223B in this experiment (Fig. 6, column 1), which was stated to be a nearly-all-α_2 matrix. This heat treatment produces an extremely skewed hysteresis loop with a coercive force of 26 to 29 oersteds and a greatly reduced remanence to 11 to 13 maxwells (Table I) as measured on 1-inch samples. It is thus clear that very brief exposure to temperatures in excess of 900 to 1000°C can markedly alter the magnetic properties previously developed by the 600°C-aging heat treatment in Remendur.

3.2 Analysis of production reeds

The results of the before-and-after glass-to-metal sealing analysis on actual production reeds (Table II) correlate well with the above observations on the simulated-sealing temperature exposures. Listed are the coercive force and remanence as measured on the complete reed, with search coil on the paddle, and also as measured on the dissociated paddles and shanks both before and after exposure to sealing (Experiment A). As was previously observed by others,[2] the major effect of sealing is a marked increase in coercive force on the shank. In this instance, an average increase in the shank of about 6 oersteds from 23.5 to 29.5 occurred. As expected, no change in the paddle portion was observed, since it does not experience the peak temperature range.

The structural changes observed in the shanks of actual sealed reeds are very similar to those produced by the typical sealing cycle in the elevated-temperature-exposure experiments. The disappearance of the

Table II — Magnetic properties of Remendur reeds before and after glass/metal sealing

Experiment	Samples	Complete Reed—search coil on paddle		Shank		Paddle	
		Hc (oer.)	$\emptyset_R$ (maxwells)	Hc (oer.)	$\emptyset_R$ (maxwells)	Hc (oer.)	$\emptyset_R$ (maxwells)
A	610°C Aged reed	27.4 (3)	21.5 (3)	23.6 (3)	5.9 (3)	28.4 (3)	10.0 (3)
	610°C Sealed reed	29.7 (12)	19.3 (12)	29.0 (16)	6.7 (16)	28.7 (16)	10.0 (16)
B	Accelerated speed seal	28.3 (11)	21.6 (11)	25.9 (11)	6.1 (11)	28.9 (11)	10.3 (11)
C	Sealed reed—normal lamp power	29.5 (6)	21.1 (6)	28.2 (6)	6.4 (6)	29.2 (6)	10.3 (6)
	Sealed reed—reduced lamp power	28.0 (6)	22.0 (6)	24.9 (6)	6.1 (6)	29.1 (6)	10.2 (6)

(*i*) H applied = 300 oersteds.
(*ii*) Numbers in parentheses indicate sample size.

γ (FCC) precipitate and return to the two-phase $\alpha_1 + \alpha_2$ structure or all-α_2 structure are apparent in Fig. 5.

The results from the reeds sealed by using the accelerated speed sealing (Experiment B) and with reduced lamp power (Experiment C) show an improvement. As can be noted in Table II, the coercivity increase for the complete reed was smaller, and no drop in remanence occurred. The data on the dissociated shanks show that a coercivity increase still occurred on these shanks, but it was markedly smaller than that for the standard sealing operation. Metallographic results also showed that a structural change had taken place but was confined within a smaller percentage of the shank length. This probably accounts for the less drastically altered magnetic properties.

IV. SUMMARY AND CONCLUSIONS

The alteration of the magnetic properties of Remendur by the glass-to-metal sealing operation from those properties developed in the nominal 600°C-aging anneal is real. This is a consequence of excessive temperatures during the sealing operation causing a dissolving of the γ (FCC) precipitate and a reversion of the structure back to that developed by the typical 900 to 950°C strand anneal used in processing the wire. With the present time/temperature sealing cycle used in production, the following more specific conclusions can be listed:

(*i*) From one-half to two-thirds of the shank length, beginning at the paddle/shank transition region, is altered by the sealing operation.
(*ii*) A significant increase in coercive force occurs in the shank.

(*iii*) Laboratory results indicate that a significant drop in remanence also occurs during elevated temperature exposures, which are sufficiently high to cause a reversion to the two-phase $\alpha_1 + \alpha_2$ or all-α_2 structures.

(*iv*) Variations in the microstructure of 0.535-mm wire from the strand anneal have little effect on the subsequent structure and magnetic properties changes during sealing.

A reduced-temperature sealing cycle holds some promise for minimizing this change, but all 2.5- to 3.0-percent vanadium Remendur alloys are likely to exhibit this same behavior independent of vanadium content or prior aging heat treatment. If this properties change is considered unacceptable, a modification in the sealing cycle or a redesign of the contact to account for this change are the most promising approaches, since the materials response cannot be altered.

V. ACKNOWLEDGMENTS

The authors express their appreciation to D. M. Sutter, P. W. Renaut, L. Herring, and K. Strauss for their cooperation in supplying their data and samples to aid in this investigation. The efforts of E. C. Hellstrom in obtaining all the magnetic data cited in this investigation are also gratefully acknowledged. The excellent metallographic preparation provided by G. V. McIlhargie is noted, and, finally, the constructive comments and usual thorough review of the manuscript provided by F. E. Bader are appropriately acknowledged.

REFERENCES

1. R. J. Gashler, W. A. Liss, and P. W. Renaut, "The Remreed Network: A Smaller, More Reliable Switch," Bell Laboratories Record, *51*, No. 7 (July–August 1973), pp. 202–207.
2. D. M. Sutter and P. W. Renaut, unpublished work, 1973.
3. T. Kitazawa, T. Oguma, and T. Hara, "Miniature Semihard Magnetic Dry Reed Switch," Proc. 19th National Relay Conference, Oklahoma State University, 1971, pp. 12-1–12-9.
4. M. R. Pinnel and J. E. Bennett, "The Metallurgy of Remendur: Effects of Processing Variations," B.S.T.J., *52*, No. 8 (October 1973), pp. 1325–1340.
5. J. E. Bennett and M. R. Pinnel, "Aspects of Phase Equilibria in Fe/Co/2.5 to 3.0% V Alloys," J. of Material Sci., *9*, No. 7 (July 1974), pp. 1083–1090.
6. T. F. Strothers, "Temperature Measurement of 237 Sealed Contact During Manufacture," unpublished work, 1971.
7. S. Mahajan, M. R. Pinnel, and J. E. Bennett, "Influence of Heat Treatments on Microstructures in an Fe-Co-V Alloy," Met. Trans., *5*, No. 6 (June 1974), pp. 1263–1272.
8. M. R. Pinnel and J. E. Bennett, "Correlation of Magnetic and Mechanical Properties with Microstructure in Fe/Co/2-3 pct V Alloys," Met. Trans., *5*, No. 6 (June 1974), pp. 1273–1283.
9. A. T. English, "Long Range Ordering and Domain Coalescence Kinetics in Fe-Co-2V," Trans. AIME, *236*, No. 1 (January 1966), pp. 14–20.
10. S. Mahajan and K. M. Olsen, "An Electron Microscopic Study of the Origin of Coercivity in an Fe-Co-V Alloy," unpublished work, 1974.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

A Multibeam, Spherical-Reflector Satellite Antenna for the 20- and 30-GHz Bands

By R. H. TURRIN

(Manuscript received November 18, 1974)

A multibeam antenna for satellite communication affords reuse of frequency allocations and flexibility in traffic routing. A multibeam satellite antenna is described that employs a spheroidal reflector in a compact periscope design. The aperture is 1.5 meters in diameter radiating six beams in the 20- and 30-GHz satellite frequency bands. Electrical measurements indicate that this antenna is suitable for multibeam satellite use.

I. INTRODUCTION

A proposed domestic satellite system for the 20- and 30-GHz radio bands includes a multibeam satellite antenna.[1] The use of separate radiating beams to service specific densely populated urban areas is desirable since it permits reuse of the frequency bands and flexibility in traffic routing.

A multibeam satellite antenna that operates simultaneously in the 20- and 30-GHz bands has been constructed and evaluated electrically. The antenna is a compact periscope design[2] employing a spherical reflector 60 inches in diameter, a plane reflector, and a cluster of six feed horns. Figure 1 shows two views of the antenna system. A spherical rather than a parabolic reflector was employed to permit off-axis beam pointing with less degradation in performance than a paraboloid. The primary feature of this antenna design is the use of multiple feeds to produce separate radiation beams. The feed location format was chosen to conform with earth-station locations at approximately New York City, Denver, Atlanta, Los Angeles, Honolulu, and Puerto Rico. These locations are representative of the maximum beam-pointing requirement for domestic satellite service and the minimum beam-pointing constraint owing to physical feed separation. The synchronous orbiting satellite was assumed to be located at 100 degrees west longitude. The antenna system boresight axis that corresponds to the axis of symmetry of the spherical reflector is designed to intersect the earth at 38 degrees north latitude and 100 degrees west longitude, near the geo-

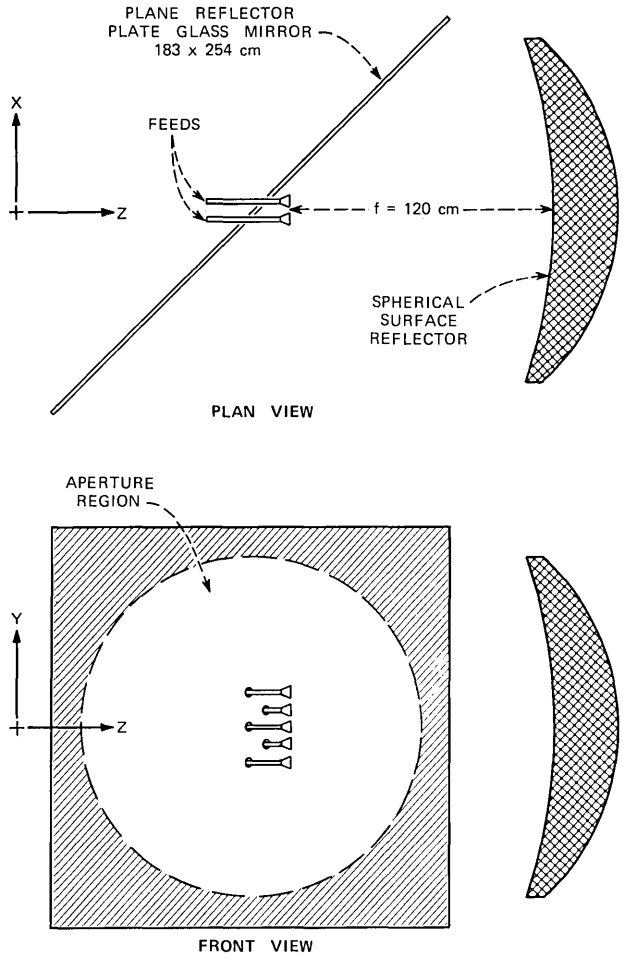


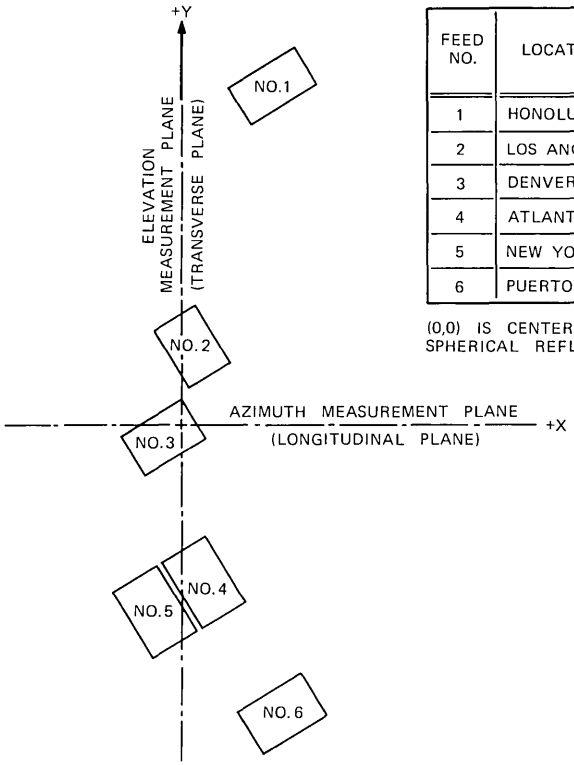
Fig. 1—Multibeam antenna.

graphic center of the contiguous United States. With these geographical constraints, the maximum beam-pointing angle off boresight is about ± 6.5 degrees in the east-west direction to include Honolulu and Puerto Rico.

Parameters of chief interest in the electrical evaluation of this antenna system are isolation between beams, absolute pointing accuracy of beams, coupling between feeds, and performance degradation because of feed-cluster blockage. Although these parameters may be evaluated analytically, the last one is especially difficult to analyze. For this reason and to demonstrate the others, a laboratory model was constructed and measurements were made.

II. CONSTRUCTION

The laboratory model was constructed to evaluate the electrical performance of the multibeam design and is not structurally compatible with space vehicle requirements. The physical size of the aperture, dictated mainly by measurement facilities, may be considered as a scale model for a larger aperture. The 150-cm-diameter spherical reflector was fabricated at the Crawford Hill laboratory of urethane foam and epoxy materials and used a deep spun-aluminum paraboloid for basic structural integrity. The surface was machined to an rms accuracy of 0.05 mm using a 244-cm radial arm end mill. The plane reflector is a 0.64-cm-thick plate glass mirror employing the metalized side as a reflecting surface. The mirror is suspended with its plane vertical to minimize warpage resulting from gravity. Six holes are bored through the plate glass to admit the feeds with a minimum loss of surface area. The feed location format shown by Fig. 2 is on a plane normal to the axis of the spheroidal reflector.



FEED NO.	LOCATION	COORDINATES (CENTIMETERS)	
		X	Y
1	HONOLULU	+4.02	+14.9
2	LOS ANGELES	+0.43	+3.46
3	DENVER	−0.79	−0.66
4	ATLANTA	+1.02	−7.00
5	NEW YORK	−1.22	−8.36
6	PUERTO RICO	+4.58	−12.70

(0,0) IS CENTER OF SPHERICAL REFLECTOR

Fig. 2—Feed format.

The feed horns were designed for dual-frequency operation with linear polarization. The polarization for the 20- and 30-GHz bands are orthogonal. Figure 3 shows details of the feed horn employing thin conductive fins that redefine the electrical aperture for the 30-GHz band. Since the 20-GHz polarization is normal to the fins, the feed-horn walls define the aperture for this band. By using the fins to selectively control the feed aperture size, the radiation patterns at both 20 and 30 GHz are more nearly the same, resulting in optimum illumination for the spheroidal reflector at both frequencies. Figures 4 and 5 show feed radiation patterns at measurement frequencies of 19 and 30.2 GHz.

The axis of each feed was aligned parallel to the axis of the spheroidal reflector aperture and the feed-phase centers were arranged in a single plane located 120 cm from the center of the spheroidal reflector. Measurements on positioning of a single movable feed showed that there is no significant difference in radiation characteristics of the antenna system for the off-axis beam-pointing directions under consideration with the feed axis oriented (*i*) along a radial direction from the center of curvature of the spheroidal surface, (*ii*) parallel to the axis, or (*iii*) directed toward the center of the spheroidal reflector to balance the amplitude distribution. Parallel mounting of the feed axes greatly simplifies the feed-line structure. It was also found that arranging the phase centers in a plane rather than in the optimum spheroidal focal surface produces an error in focal length of less than $\lambda/2$ at 30 GHz with little effect on the gain or radiation characteristics of the system.

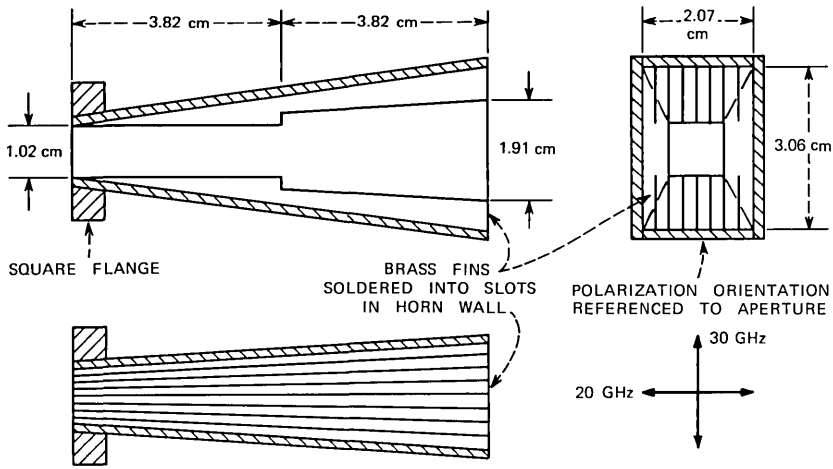


Fig. 3—Linearly polarized feed horn, dual frequency (17.7–20.2 GHz and 27.5–30.0 GHz).

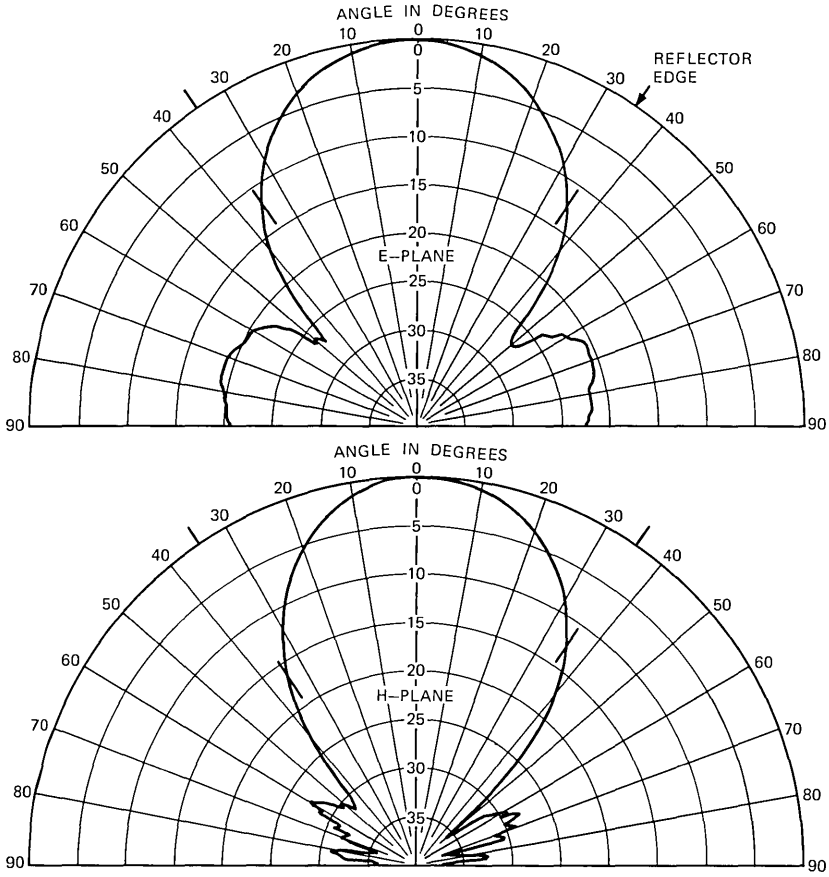


Fig. 4—Radiation patterns of the finned feed horn at 19.0 GHz.

The polarization orientations of the feeds were mandated by the physical constraint of the feeds in New York City (feed 5) and Atlanta (feed 4) feeds. These two feeds were parallel-polarized and arranged at a skew angle with respect to the measuring plane to achieve a minimum physical spacing. The remaining feeds were cross-polarized with their adjacent neighbors for minimum electrical coupling. The feed locations shown in Fig. 2 were necessary to achieve the desired beam-pointing angles for the various earth-station locations.

III. MEASUREMENTS

Electrical evaluation of the multibeam antenna was conducted at the Holmdel radio range. A variation was found of incident field over the measurement aperture of no greater than ± 0.75 dB for either 19 or 30 GHz, the measurement frequencies used.

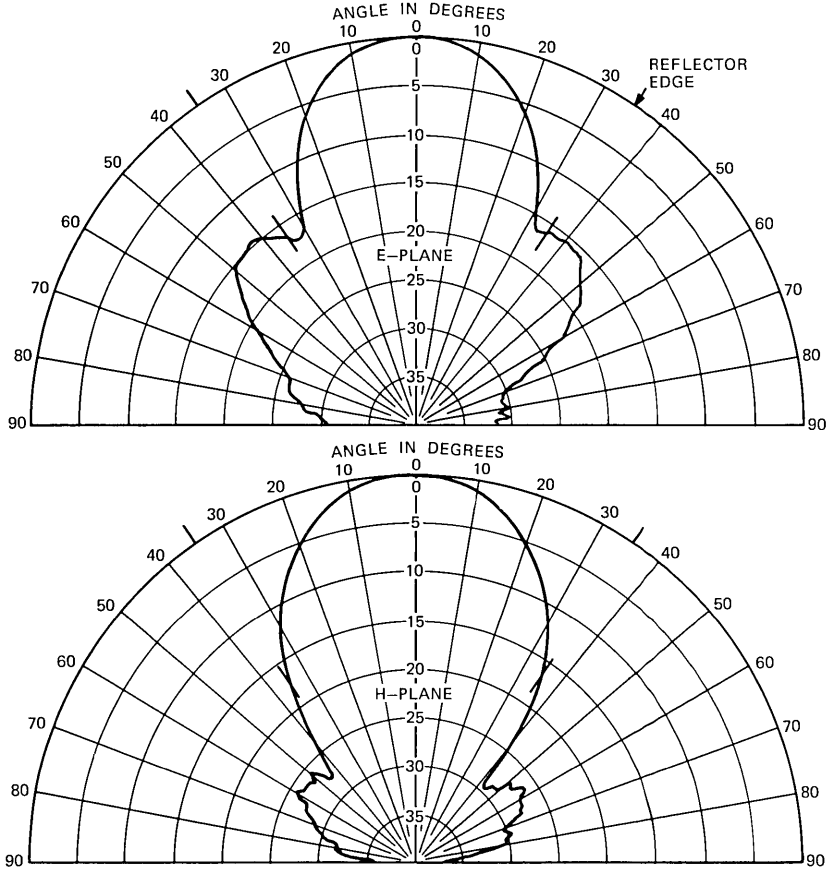


Fig. 5—Radiation patterns of the finned feed horn at 30.0 GHz.

3.1 Beam pointing

Beam-pointing accuracy is important in the design of a satellite system. Measurement of pointing accuracy for each of the six beams was accomplished by first mechanically measuring the location of each feed phase center relative to the axis of the spheroidal reflector (boresight axis) and in the X-Y plane (Fig. 1). Alignment of the boresight axis with the radio source was done by optical sighting. All measurements of beam pointing were made using synchro-type readouts with an accuracy of 0.05 degree. Angular measurements are in essentially an orthogonal coordinate system for the elevation-over-azimuth mount.

Table I shows all the measured beam-pointing data, together with computed data based on feed location, and the pointing errors. The feeds are numbered as in Fig. 2. Since the half-power beamwidth of

Table I — Measured and computed beam-pointing errors

Feed No.	Freq. (GHz)	Measured θ_x (Degrees)	Computed θ_x (Degrees)	Pointing Error (Degrees)	Measured θ_y (Degrees)	Computed θ_y (Degrees)	Pointing Error (Degrees)
1 Hon.	19.0	−1.95	−1.87	−0.08	+6.82	+6.90	−0.08
	30.2	−1.71		+0.16	+7.28		+0.38
2 L.A.	19.0	−0.35	−0.20	−0.15	+1.50	+1.61	−0.11
	30.2	−0.15		+0.05	+1.75		+0.14
3 Den.	19.0	+0.20	+0.37	−0.17	−0.45	−0.31	−0.14
	30.2	+0.45		+0.08	−0.30		+0.01
4 Atl.	19.0	−0.58	−0.48	−0.10	−3.44	−3.25	−0.19
	30.2	−0.40		+0.08	−3.40		−0.15
5 N.Y.	19.0	+0.35	+0.57	−0.22	−4.17	−3.90	−0.27
	30.2	+0.56		−0.01	−4.08		−0.18
6 P.R.	19.0	−2.19	−2.12	−0.07	−6.14	−5.89	−0.25
	30.2	−2.05		+0.07	−6.15		−0.26

each beam at both 19 and 30 GHz is about 0.65 degree, the maximum error encountered in these measurements was about one-half beamwidth. Considering the readout accuracy, mechanical errors in the structure, and boresight errors, these results are remarkably consistent; the antenna beams point in the intended directions.

3.2 Beam coupling

Beam coupling is a measure of the signal level in one specified beam when another beam is aimed toward the source. It is also a measure of the interference between any two particular ground stations. The radiation characteristics of each beam for specific off-axis angles specify the coupling. Measured coupling for all combinations of beams except the New York City-Atlanta combination was found to be less than −35 dB. In that case, the angular separation in the beams is only 1.2 degrees and the coupling is approximately −20 dB. The data were measured at specific incremental frequencies over the 2500-MHz band at both 20 and 30 GHz.

The beam coupling is strongly dependent on exact pointing directions and can be greatly improved by slight misalignment of the main beams. It may, therefore, be an engineering expedience to suffer a slight degradation in gain by beam misalignment to improve the isolation.

3.3 Radiation patterns

Studies of the radiation characteristics of this antenna with a single horn feed were made prior to installation of the six-feed cluster. These

measurements were made to determine the effects of illumination on beams pointed off-axis. Figure 6 gives measured values of relative gain and first side lobes with off-axis pointing. The asymmetric illumination with off-axis pointing produces an effect similar to coma aberration in which the first side lobe levels become increasingly unbalanced. The side lobe nearest the axis of the antenna system is always the higher in level, as expected.

The gain degradation for maximum off-axis pointing is a modest 1 dB at 19 GHz. At 30 GHz, the extrapolated degradation is approximately 2 dB. Representative measured far-field radiation patterns for the single feed are shown by Figs. 7 and 8. These are principal plane measurements, that is, patterns taken with a single feed whose polarization orientation is parallel with the measurement plane; they show the typical far-side-lobe behavior of this type of antenna.

Figures 9 and 10 are far-field radiation patterns at 19.0 GHz for feeds 1 and 3, when all other feeds are in place and terminated. These are representative of the maximum and minimum off-axis pointing. Figures 11 and 12 are for the same feeds at 30.2 GHz. These patterns illustrate the off-axis asymmetric radiation characteristics; the effect

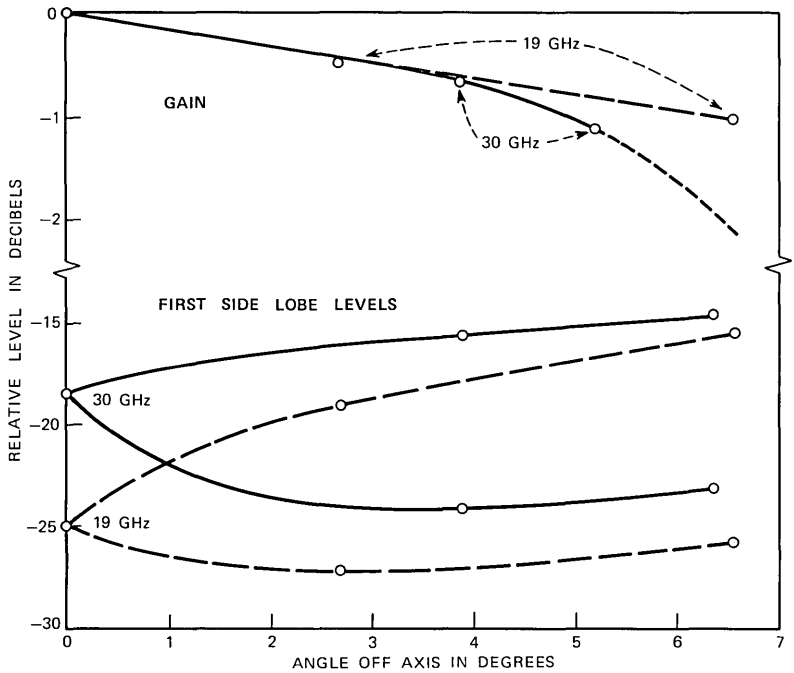


Fig. 6—Single-feed measurements of off-axis performance.

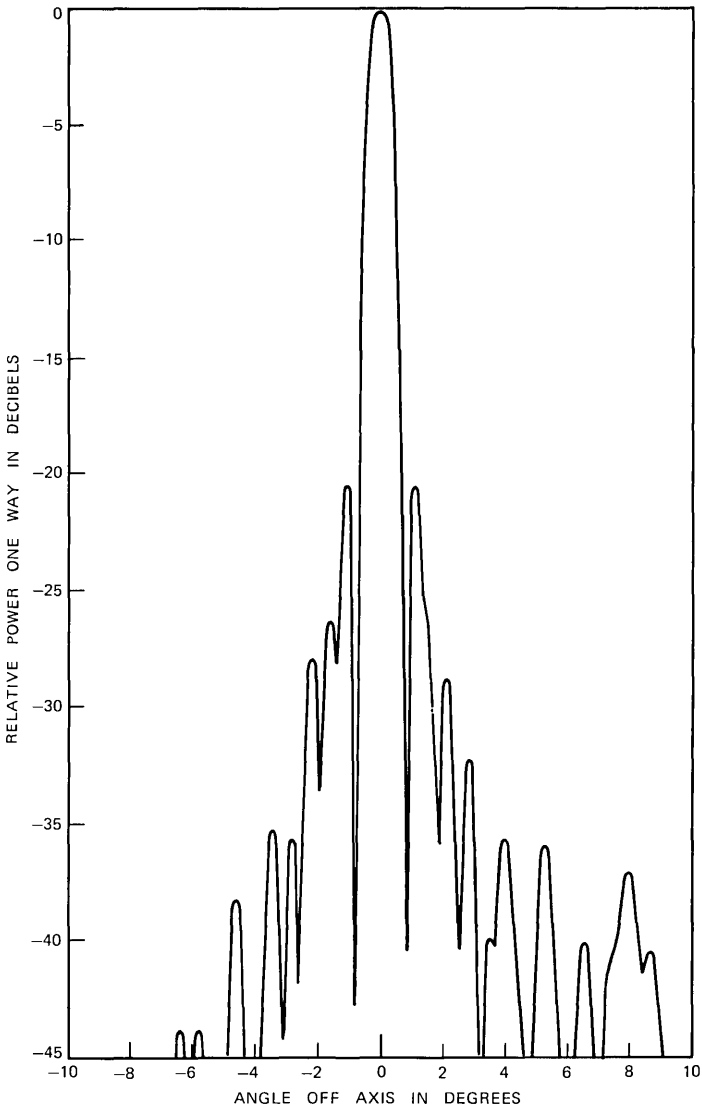


Fig. 7—Far-field radiation pattern at 30.0 GHz, single-feed centered in longitudinal plane (azimuth).

is more pronounced at the higher frequency where the relative aperture phase errors are greater. These patterns also indicate that the small blockage from the feed cluster causes negligible distortion of the beams.

The half-power beamwidths are essentially equal for both 19 and 30 GHz; this is because the feed illumination taper is somewhat more

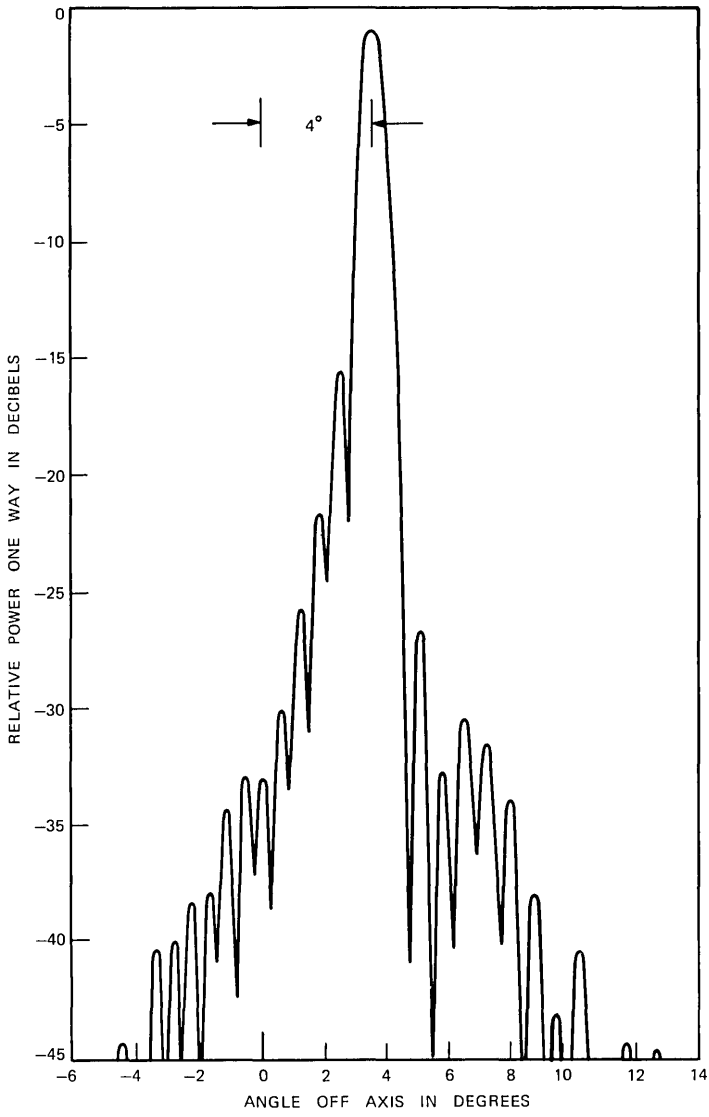


Fig. 8—Far-field radiation pattern at 30.0 GHz, single-feed offset in longitudinal plane.

severe at 30 GHz and because the reflector surface errors cause more degradation at the higher frequency.

3.4 Gain

Absolute gain measurements shown in Fig. 13 corroborate the pattern measurements; the gain falls off faster with off-axis pointing

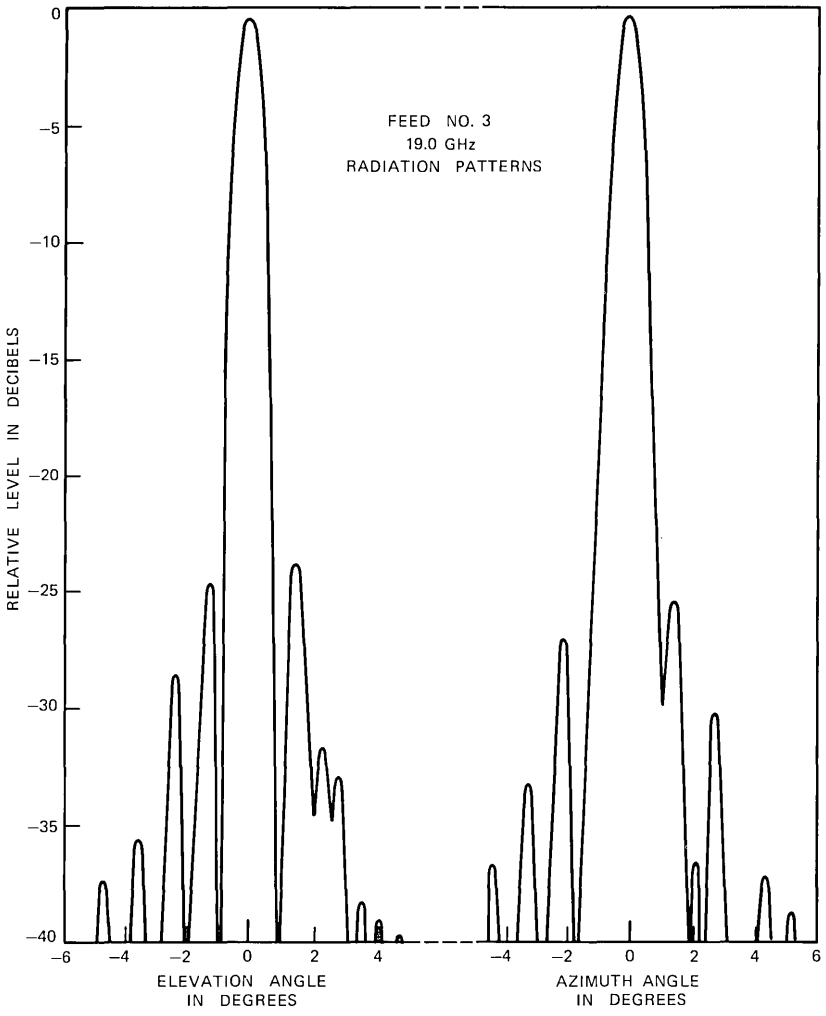


Fig. 9—Feed 3, 19.0-GHz radiation patterns.

at 30 GHz than at 19 GHz. The plotted values in Fig. 13 show the relative gain measured for each feed. These data are consistent with single feed measurements where off-axis gain degradation was somewhat more at 30 GHz than at 19 GHz. The aperture efficiency for the feed nearest center, No. 3, is 54 percent at 19.0 GHz and 35.5 percent at 30.2 GHz.

3.5 Reflection coefficient

The magnitude of the reflection coefficient over each frequency band was measured at each feed port. The maximum value is −18 dB typi-

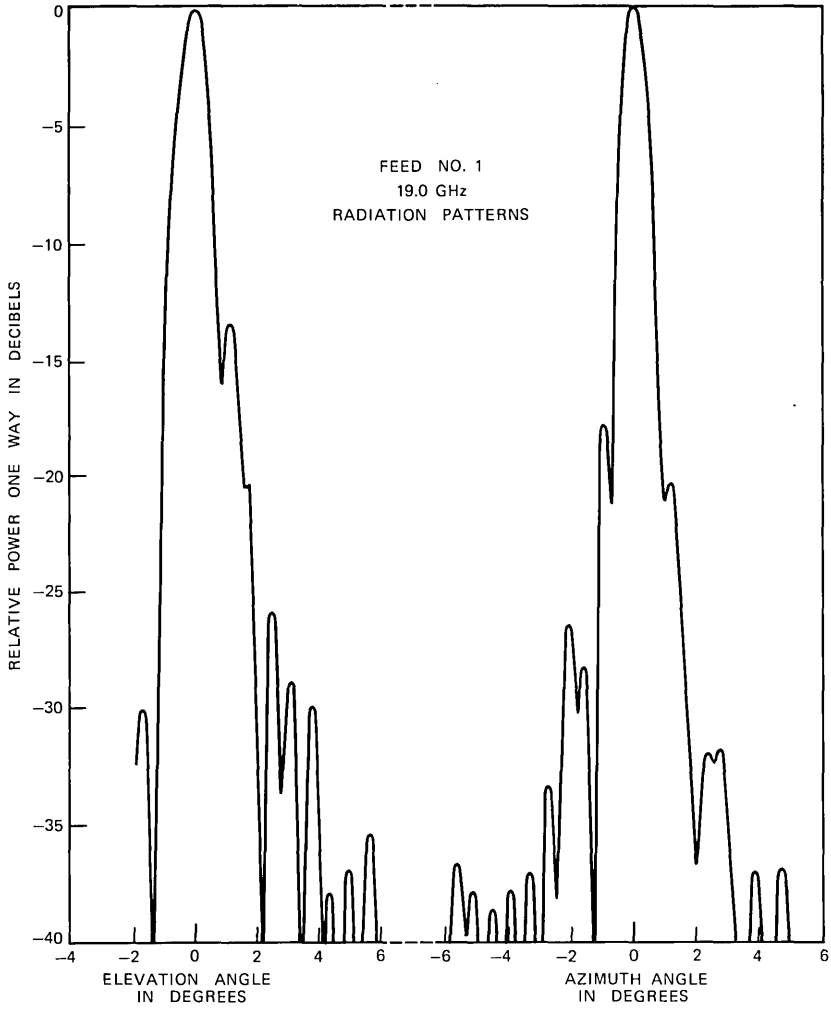


Fig. 10—Feed 1, 19.0-GHz radiation patterns.

cally for all ports over the band 18.3 to 20.3 GHz. Figure 14 shows a typical reflection coefficient measurement for the 19-GHz band. The periodicity in the reflection coefficient indicates a mismatch at the feed horn to be approximately the same magnitude as the echo from the spheroidal reflector. The same type of periodicity is observed in the reflection coefficient measurement over the 27.5- to 30.2-GHz band as shown in Fig. 15. In this case, however, the reflection coefficient is much worse with a maximum value of −12 dB toward the low-frequency end of this band.

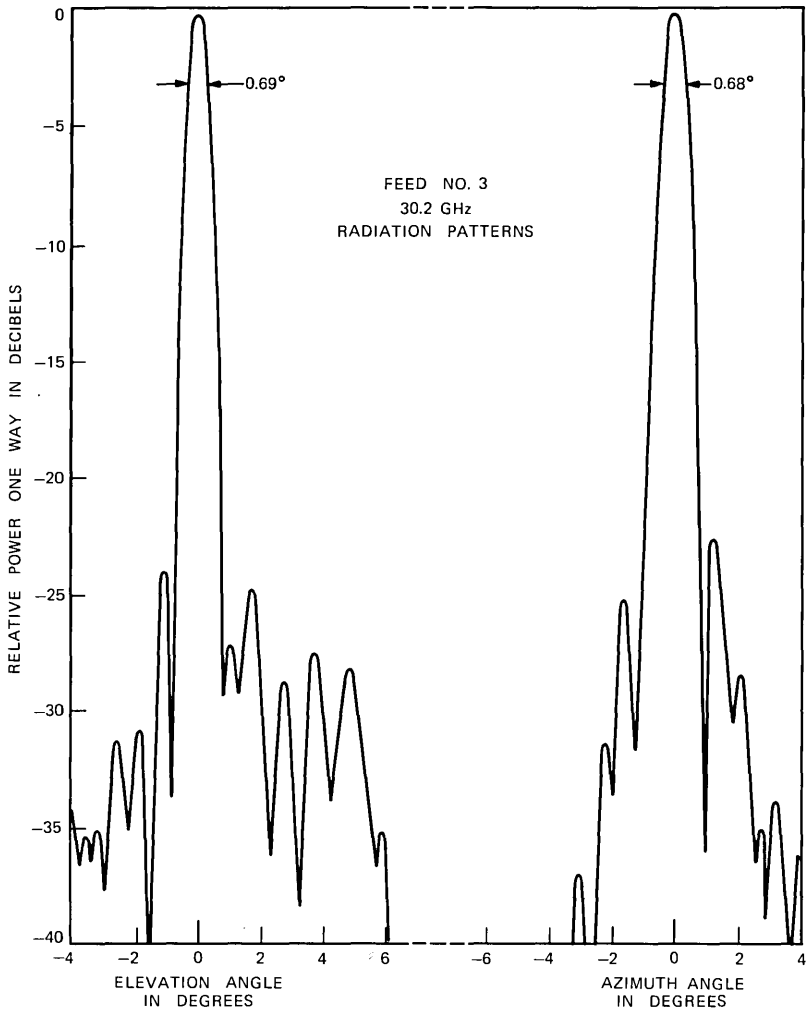


Fig. 11—Feed 3, 30.2-GHz radiation patterns.

3.6 Coupling between ports

Direct coupling between feed ports results from two causes: (*i*) reflections from the spherical reflector into other feed horns and (*ii*) direct side-to-side radiation of the feed horns. Direct measurement of all 15 feed-coupling combinations (assuming reciprocity) was made on a scanned frequency basis. Three of the feeds are essentially cross-polarized with respect to the other three, and those nine cross couplings that involved the cross-polarized condition were found to have coupling levels less than −40 dB. Of the remaining six combinations of

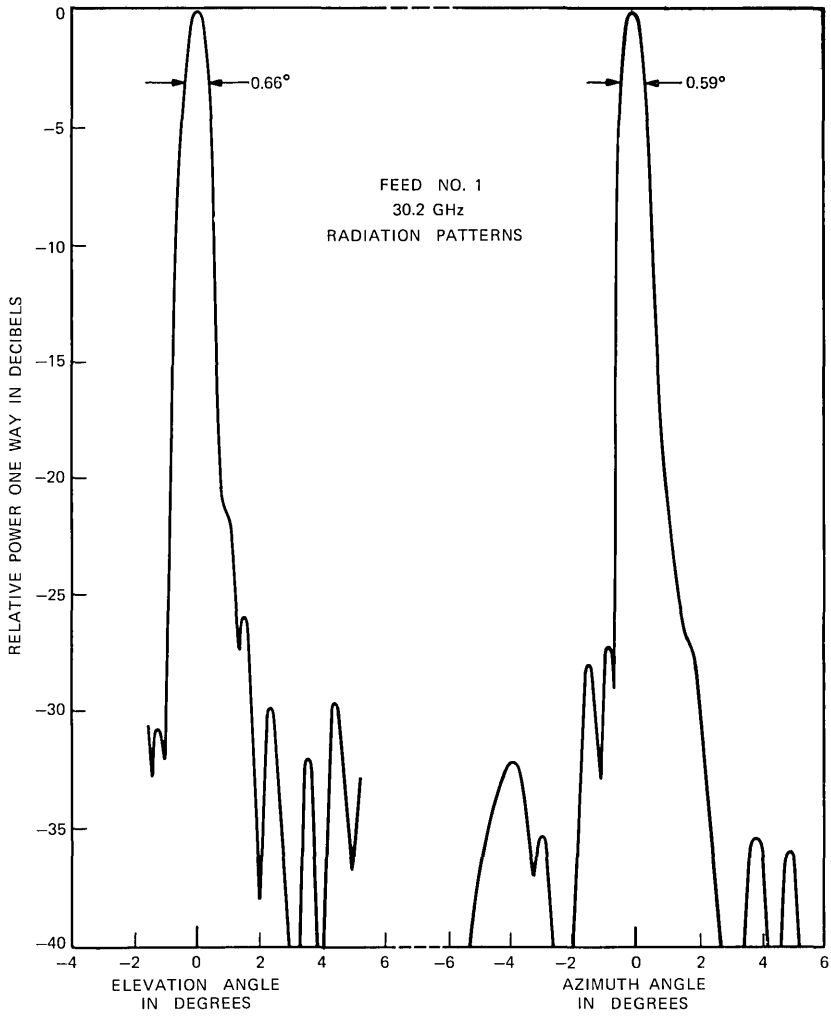
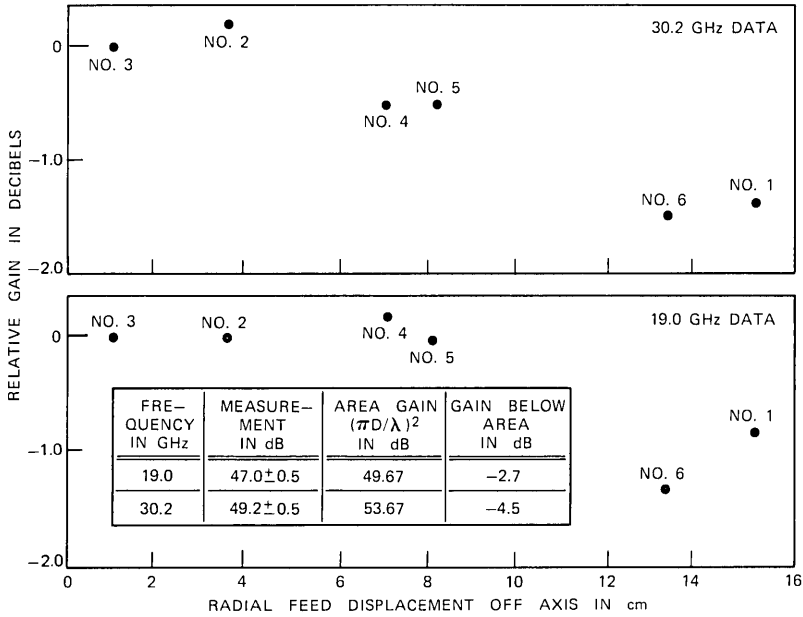


Fig. 12—Feed 1, 30.2-GHz radiation patterns.

coupling for each frequency band, the worst case is for the adjacent feeds 4 and 5 (New York-Atlanta), which show values of −24 dB at 20 GHz and −26 dB in the 30-GHz band. All others show coupling less than −28 dB, consistent with reflection coupling from the spherical surface alone. All unused ports were terminated during these measurements.

IV. CONCLUSIONS

Measurements on a compact spherical periscope antenna indicate that this design is suitable as a multibeam satellite antenna for the



FREQUENCY IN GHz	MEASUREMENT IN dB	AREA GAIN $(\pi D/\lambda)^2$ IN dB	GAIN BELOW AREA IN dB
19.0	47.0±0.5	49.67	−2.7
30.2	49.2±0.5	53.67	−4.5

Fig. 13—Absolute gain measured at feed port 3 by standard gain horn techniques.

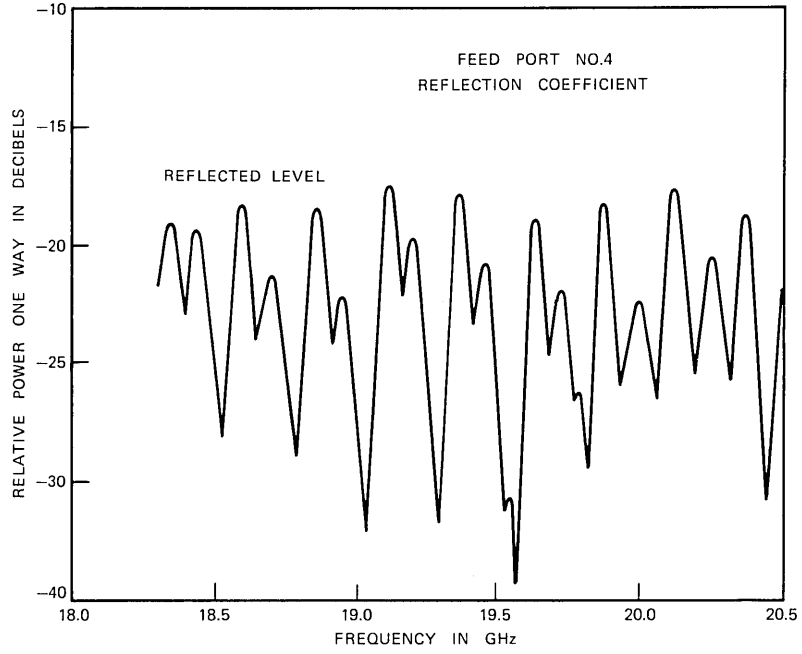


Fig. 14—Feed port 4, reflection coefficient.

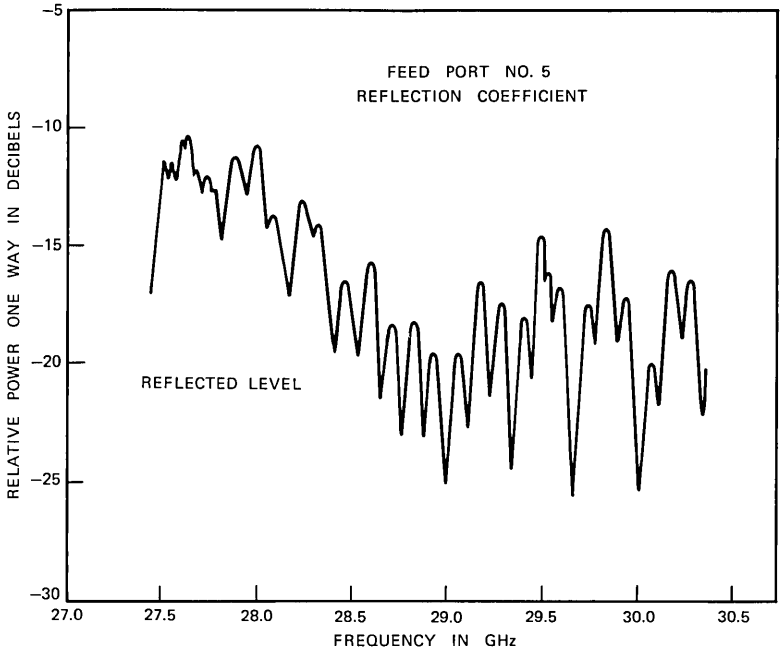


Fig. 15—Feed port 5, reflection coefficient.

20- and 30-GHz bands. While the aperture diameter, 150 cm, may be somewhat small for practical use, most measured results are equally applicable to larger aperture designs.

The only feature of the design not amenable to analysis is the blockage and scattering by the cluster of feeds. However, the measurements reported here indicate that this effect is of little consequence in the intended application. It is also clear that some limitations of the embodiment discussed here are inherent in the feed design. Minimum physical separation between feeds limits the minimum angular beam separation to about 1.2 degrees in this particular antenna, isolation between these beams being of the order of −20 dB.

V. ACKNOWLEDGMENTS

I would like to thank A. B. Crawford for inspiring the antenna design, T. S. Chu for designing the feed, and J. H. Hammond for assisting in the many range measurements.

REFERENCES

1. L. C. Tillotson, "A Model of a Domestic Satellite Communication System," B.S.T.J., *47*, No. 10 (December 1968), pp. 2111–2137.
2. A. B. Crawford and R. H. Turrin, "A Packaged Antenna for Short-Hop Microwave Radio Systems," B.S.T.J., *48*, No. 6 (July–August 1969), pp. 1605–1622.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

Polarization-Independent, Multilayer Dielectrics at Oblique Incidence

By A. A. M. SALEH

(Manuscript received January 6, 1975)

This paper gives solutions for multilayer dielectrics that yield polarization-independent operation at a given angle of incidence and at a given frequency. An exhaustive analysis is given for symmetric three-layer dielectrics surrounded by the same medium. Special solutions are also given for symmetric multilayer dielectrics with more than three layers surrounded by the same medium and for multilayer dielectrics surrounded by different media. Rules are presented for obtaining new polarization-independent solutions by cascading given solutions. Finally, for the purpose of demonstration and comparison of various solutions, different design examples are carried out for polarization-independent 3-dB beam splitters to operate at millimeter wavelengths.

I. INTRODUCTION

The state of polarization of a plane wave will, in general, be changed upon transmission through or reflection by a multilayer dielectric at oblique incidence. The reason for this depolarization is that the two eigenmodes of polarization, E-field perpendicular to the plane of incidence and E-field parallel to the plane of incidence, generally have different reflection coefficients and different transmission coefficients. This is not desirable in many applications. For example, such depolarization could cause crosstalk in dual-polarization optical or quasi-optical components such as interference filters,[1] directional couplers[2–5] and diplexers,[6] or mode conversion in beam-splitter-type hybrids[7,5] in oversized waveguides.

This paper provides solutions for multilayer dielectrics whose reflection and transmission coefficients are independent of the state of polarization of the incident wave at a given angle of incidence and a given frequency. Compromising the degree of polarization independence to allow for a broader bandwidth of operation or for variations in the angle of incidence, even though desirable in many practical applications, is not discussed.

The paper is primarily concerned with symmetric, multilayer dielectrics surrounded by the same medium. However, the case of different input and output media and antireflective coatings is also discussed. Some of the results given in this paper have been previously reported by Kard,[8] Baumeister,[9] and Rabinovitch and Pagis.[10] Their work is referred to in appropriate places in the text.

There are other types of structures, besides the ones reported in this paper, that can give polarization-independent operation at oblique incidence. For example, artificial or natural anisotropic dielectric layers or metallic-wire meshes with rectangular, rather than square, cells can, in principle, be constructed to achieve this effect.

II. BASIC EQUATIONS

Consider n adjacent isotropic dielectric layers of infinite transverse extent having relative dielectric constants of $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ and uniform thicknesses of $d_1, d_2, \cdots, d_n$ inserted in a medium of a relative dielectric constant ϵ_r. Define

$$\kappa_i \equiv \epsilon_i/\epsilon_r, \qquad i = 1, 2, \cdots, n \tag{1}$$

as the normalized dielectric constants of the layers. Let a uniform plane wave propagating in the medium be incident at an angle θ on the layers as shown in Fig. 1. The incident, transmitted and reflected waves will have the same state of polarization if the E-field is perpendicular

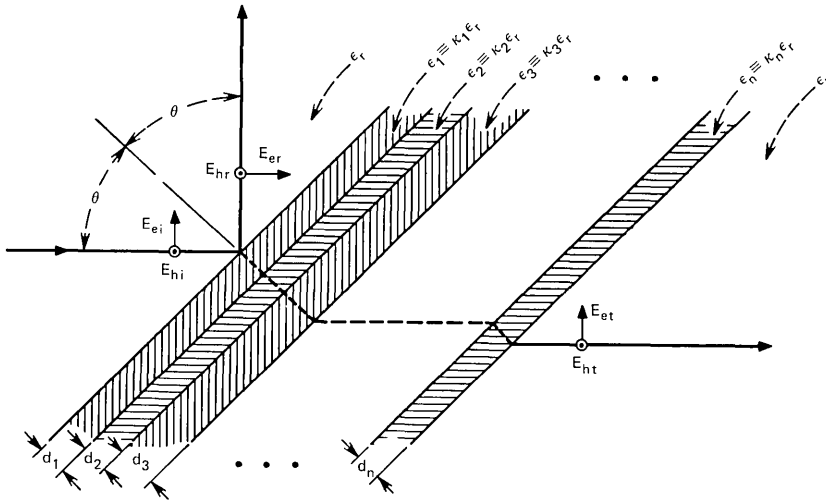


Fig. 1—n-layer dielectric inserted in a medium of a relative dielectric constant ϵ_r. The subscript h refers to the h-mode (E-field perpendicular to plane of incidence), and e refers to the e-mode (E-field parallel to plane of incidence).

to the plane of incidence (h-mode), or if the E-field is parallel to the plane of incidence (e-mode). To calculate the transmission and reflection coefficients for the h- and e-modes, one can represent each layer by an equivalent transmission line.[11,12] The impedance of the line representing the ith layer normalized to that of the line representing the medium in which the layers are inserted is

$$Z_i = Z_{ih} = \cos\theta/(\kappa_i - \sin^2\theta)^{\frac{1}{2}} \tag{2a}$$

for the h-mode, and

$$Z_i = Z_{ie} = (\kappa_i - \sin^2\theta)^{\frac{1}{2}}/(\kappa_i \cos\theta) \tag{3a}$$

for the e-mode. The corresponding electrical length is

$$\phi_i = (2\pi/\lambda_o)d_i\sqrt{\epsilon_r}(\kappa_i - \sin^2\theta)^{\frac{1}{2}} \tag{4a}$$

for both modes; λ_o being the free-space wavelength. If $\kappa_i < \sin^2\theta$, i.e., the total reflection condition is satisfied at the boundary of the ith layer, then the square root of $(\kappa_i - \sin^2\theta)$ is imaginary and one can write

$$Z_{ih} = jX_{ih}, \tag{2b}$$

$$Z_{ie} = -jX_{ie}, \tag{3b}$$

$$\phi_i = -j\alpha_i, \tag{4b}$$

where X_{ih}, X_{ie}, and α_i are positive, real quantities. In this case, coupling is accomplished across that layer by an evanescent wave.

The normalized $ABCD$ matrix corresponding to the ith layer is

$$\mathbf{M}_i = \begin{bmatrix} \cos\phi_i & j\sin\phi_i Z_i \\ j\sin\phi_i/Z_i & \cos\phi_i \end{bmatrix}. \tag{5}$$

The overall, normalized $ABCD$ matrix of the n layers is

$$\begin{bmatrix} A & jB \\ jC & D \end{bmatrix} = \mathbf{M}_1\mathbf{M}_2\cdots\mathbf{M}_n. \tag{6}$$

The reason for the j in the off-diagonal terms of the matrix of eq. (6) is that A, B, C, and D will be real quantities if the system is lossless. Henceforth, this will be assumed to be the case.

It is worth noting that reciprocity requires that

$$AD + BC = 1, \tag{7}$$

and that for a symmetric structure

$$A = D. \tag{8}$$

The overall field transmission coefficient, t, and field reflection coefficient, r, are given by

$$t \equiv E_t/E_i = 2/[A + D + j(B + C)], \tag{9}$$

$$r \equiv E_r/E_i = [A - D + j(B - C)]/[A + D + j(B + C)]. \tag{10}$$

A subscript of h or e should be added to all the variables in eqs. (5) through (10) to denote the particular mode being considered.

III. CLASSIFICATION OF THE SOLUTIONS

Our goal is to find solutions for the κ's, the d's, and θ to achieve polarization-independent operation. In applications where one is restricted to independent h- and e-modes, or to unpolarized waves, it is sufficient that

$$|t_h| = |t_e|, \qquad |r_h| = |r_e|. \tag{11}$$

To satisfy these conditions, eqs. (7), (9), and (10) require that

$$A_h^2 + B_h^2 + C_h^2 + D_h^2 = A_e^2 + B_e^2 + C_e^2 + D_e^2. \tag{12}$$

For symmetric structures, eqs. (7) and (8) reduce eq. (12) to

$$B_h - C_h = \pm(B_e - C_e). \tag{13}$$

To preserve the state of polarization of an arbitrarily polarized incident wave, condition (11) is not sufficient. Instead, it is necessary that

$$t_h = \pm t_e, \qquad r_h = \pm r_e. \tag{14}$$

The sense of polarization (i.e., the right- or left-handedness) of the transmitted wave will be unchanged if $t_h = +t_e$ and will be reversed if $t_h = -t_e$. On the other hand, the sense of polarization of the reflected wave will be unchanged if $r_h = -r_e$ and will be reversed if $r_h = +r_e$. The latter case is similar to reflection by a perfect, conducting plane.

The four combinations of signs in eq. (11) are listed in Table I where they are referred to as Cases 1 through 4. The required relations

Table I — Four cases for polarization-independent operation

Case	1	2	3	4
$t_h =$	t_e	t_e	$-t_e$	$-t_e$
$r_h =$	r_e	$-r_e$	r_e	$-r_e$
$A_h =$	A_e	D_e	$-A_e$	$-D_e$
$B_h =$	B_e	C_e	$-B_e$	$-C_e$
$C_h =$	C_e	B_e	$-C_e$	$-B_e$
$D_h =$	D_e	A_e	$-D_e$	$-A_e$

between A_h, B_h, C_h, D_h and A_e, B_e, C_e, D_e to satisfy each of these cases can be deduced from eqs. (9) and (10) and are also given in Table I.

In addition to classifying the solutions in the manner of Table I, another classification based on the numerical values of the κ's is useful. To be concise, let $\kappa_{\min}$ be the smallest value of the κ's and let it be at the mth layer, i.e.,

$$\kappa_{\min} \equiv \min (\kappa_1, \kappa_2, \cdots, \kappa_n) = \kappa_m. \tag{15}$$

If $\kappa_{\min} \geqq 1$, then none of the κ's is smaller than unity and the solution can be realized by dielectric layers, $\epsilon_i \geqq 1$, inserted in air, $\epsilon_r = 1$. This will be referred to as a Type A solution. On the other hand, if $\kappa_{\min} < 1$, the surrounding medium can no longer be air since this would require that at least ϵ_m be less than unity (recall that m is the layer with the smallest value of κ); thus, a dielectric-prism realization as shown in Fig. 2 is necessary. In this case, the mth layer can be an air separation $\epsilon_m = 1$, and thus, from eq. (1)

$$\epsilon_i = \kappa_i/\kappa_{\min} \geqq 1, \qquad \epsilon_r = 1/\kappa_{\min} > 1. \tag{16}$$

The prism realization will be referred to as a Type B solution when $1 > \kappa_{\min} \geqq \sin^2 \theta$, and a Type C solution when $\kappa_{\min} < \sin^2 \theta$. The dis-

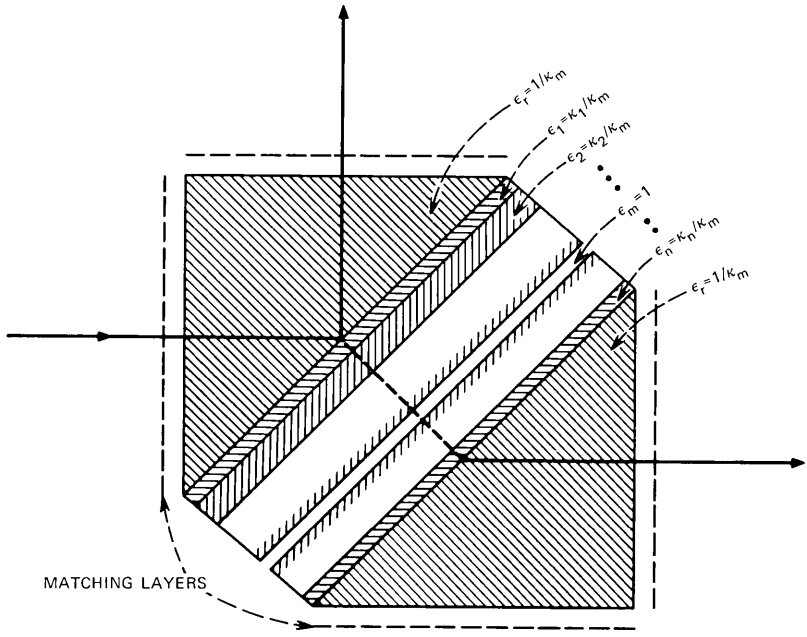


Fig. 2—A prism realization having the mth layer (the layer with the smallest dielectric constant) made of an air separation. Note that $\kappa_m = \kappa_{\min} < 1$. Matching layers are required at the external surfaces of the prisms.

Table II — The three types of solutions

Type	A	B	C
Condition	$\kappa_{min} \geqq 1$	$\sin^2\theta \leqq \kappa_{min} < 1$	$\kappa_{min} < \sin^2\theta$
Description	Dielectric slabs in air ($\epsilon_r = 1$) is possible.	Prism realization ($\epsilon_r > 1$) is necessary. No evanescent waves.	Prism realization ($\epsilon_r > 1$) is necessary. Evanescent waves exist.

tinction is made because in a Type C solution, the condition for total reflection is satisfied at the boundary of the mth layer and coupling across that layer is accomplished by an evanescent wave, as mentioned in Section II. This is not the case for a Type B solution. Table II summarizes the conditions for the three types of solution.

Dielectric-prism devices have been used for many optical[13,14] and quasi-optical[2–5,15–18] applications. When the total reflection condition is satisfied, Type C realization, the device is often said to be of the frustrated-total-reflection type.

Fresnel reflections at the outer surfaces of the prisms might cause undesirable effects.[19] However, this can be avoided by matching each of these surfaces by the familiar $\lambda/4$ transformer, or by other means.[2,20–22] In any case, for the purpose of this paper, the method of matching should be independent of the state of polarization of the incident wave. Thus, Brewster angle matching[17,18] would not be suitable here.

In the remainder of the paper, a combined reference to the four cases of Table I and the three types of Table II is adopted. For example, Case 1A refers to Case 1 of Table I and Type A of Table II.

IV. SINGLE-LAYER SOLUTIONS

It can be shown by using the procedure outlined in Section II that no solution exists for a single layer to satisfy any of the four cases indicated in Table I. However, it is possible to satisfy the weaker condition given in eq. (11) by requiring that

$$\kappa_1 \equiv \epsilon_1/\epsilon_r = \sin^2\theta/(1 + \cos^2\theta), \tag{17}$$

where the symbols are defined in Fig. 1 with $n = 1$. Substituting eq. (12) into eqs. (2) through (10), one obtains

$$t_h = t_e^*, \qquad r_h = r_e^*, \tag{18}$$

where the asterisk denotes complex conjugation.

It is clear from eq. (17) that $\kappa_1 < \sin^2 \theta$ and, hence, from Table II, the solution is of Type C, and a prism realization is necessary.

It is interesting to note that condition (17) is independent of the thickness, d_1, of the layer. Thus, varying d_1 will change the reflection and transmission coefficients without affecting the polarization independence. If $R \equiv |r|^2$ is the required power reflection coefficient, then eqs. (17) and (2) through (10), with $n = 1$, give

$$d_1 = [\lambda_o/(2\pi\sqrt{\epsilon_1})][(1 + \kappa_1)/(1 - \kappa_1)]^{\frac{1}{2}} \times \sinh^{-1}\{2[\kappa_1 R/(1 - R)]^{\frac{1}{2}}/(1 + \kappa_1)\}. \quad (19)$$

A continuous variation of d_1, and hence of R, can be accomplished most conveniently if the layer is simply an air separation, $\epsilon_1 = 1$. In this case, if the angle of incidence is $\theta = 45°$, which is convenient for mechanical considerations, eq. (17) requires the dielectric constant of the prisms to be $\epsilon_r = 3$. On the other hand, if the prisms were made of polystyrene, $\epsilon_r = 2.54$, eq. (17) requires that $\theta = 48.73°$.

V. SYMMETRIC THREE-LAYER SOLUTIONS

It was mentioned in the previous section that no single-layer solution exists that satisfies any of the four cases of Table I. The same is also true for two layers. On the other hand, one can satisfy any of the four cases with three or more layers. In this section, we give all the solutions for symmetric three-layer structures, i.e.,

$$\kappa_3 = \kappa_1, \qquad d_3 = d_1. \quad (20)$$

Even though unsymmetric three-layer structures can have polarization-independent solutions, they will not be discussed here because of the algebraic complexities involved.

Using the terminology of Section II, it can be shown that the normalized $ABCD$ parameters of a symmetric three-layer structure are given by

$$A = D = \cos\phi_2 \cos(2\phi_1) - \tfrac{1}{2}[Z_1/Z_2 + Z_2/Z_1] \sin\phi_2 \sin(2\phi_1) \quad (21)$$

$$B = Z_1 \cos\phi_2 \sin(2\phi_1) + Z_2 \sin\phi_2 \cos^2\phi_1 - [Z_1^2/Z_2] \sin\phi_2 \sin^2\phi_1 \quad (22)$$

$$C = Z_1^{-1} \cos\phi_2 \sin(2\phi_1) + Z_2^{-1} \sin\phi_2 \cos^2\phi_1 - [Z_2/Z_1^2] \sin\phi_2 \sin^2\phi_1. \quad (23)$$

Combining these equations with eqs. (2) through (4), (9), and (10), one can obtain solutions for the cases listed in Tables I and II. In Cases 3 and 4 of Table I, this process involves considerable algebraic manipulation. The results are given below. For simplicity, expressions for ϕ_1 (or α_1) and ϕ_2 (or α_2) are given instead of expressions for d_1 and

d_2. Also, we define

$$p_i \equiv (\kappa_i - \sin^2\theta)^{\frac{1}{2}}, \qquad i = 1, 2 \tag{24a}$$

$$q_i \equiv (\sin^2\theta - \kappa_i)^{\frac{1}{2}}, \qquad i = 1, 2 \tag{24b}$$

$$f(x, y) \equiv 2xy - (x + y)\sin^2\theta, \tag{25}$$

$$R \equiv |r_h|^2 = |r_e|^2, \tag{26}$$

$$S \equiv [R/(1 - R)]^{\frac{1}{2}}, \tag{27}$$

$$U \equiv [2\sqrt{R}/(1 + \sqrt{R})]^{\frac{1}{2}}, \tag{28}$$

$$V \equiv (1 - \sqrt{R})/(1 + \sqrt{R}). \tag{29}$$

In each case, necessary upper and/or lower bounds on κ_1, κ_2, θ, and R are given. Because of the algebraic complexity involved, the bounds on κ_1 and κ_2 in Cases 3 and 4 are looser than those given in Cases 1 and 2.

Case 1A:

$$\phi_1 = (2m + 1)\pi/2, \qquad m = 0, 1, 2, \cdots. \tag{30}$$

$$\kappa_2 = \kappa_1^2/[1 - (\kappa_1 - 1)^2/\cos^2\theta]. \tag{31}$$

$$\sin\phi_2 = \pm 2S/[\sqrt{\kappa_2}/\kappa_1 - \kappa_1/\sqrt{\kappa_2}]. \tag{32}$$

$$1 + U\cos\theta \leqq \kappa_1 < 1 + \cos\theta. \tag{33}$$

$$(1 + U\cos\theta)^2/V \leqq \kappa_2 < \infty. \tag{34}$$

This solution exists for all R, θ.

Case 1B:

ϕ_1, κ_2, ϕ_2 are the same as in eqs. (30) through (32).

$$\sin^2\theta < \kappa_1 \leqq 1 - U\cos\theta. \tag{35}$$

$$\sin^2\theta < \kappa_2 \leqq (1 - U\cos\theta)^2/V. \tag{36}$$

$$1 < \kappa_1/\kappa_2 \leqq \max\,[2/(1 + \sin\theta), \quad V/(1 - U\cos\theta)]. \tag{37}$$

This solution exists if and only if

$$\sin\theta < V^{\frac{1}{2}}, \tag{38a}$$

or, equivalently,

$$R < [\cos^2\theta/(1 + \sin^2\theta)]^2. \tag{38b}$$

Case 1C:

No solution exists.

Case 2A:

$$\phi_1 = (2m + 1)\pi/2, \qquad m = 0, 1, 2, \cdots. \tag{39}$$

$$\kappa_2 = \kappa_1^2. \tag{40}$$

$$\sin \phi_2 = \pm 2S \cos \theta p_1^2 p_2 / [(\kappa_1 - 1) \sin \theta]^2. \tag{41}$$

$$\kappa_1 > \kappa_a \quad (>1), \tag{42}$$

$$\kappa_2 > \kappa_a^2 \quad (>1), \tag{43}$$

where κ_a is the root larger than unity of the quartic equation

$$(W - 1)\kappa^4 + 2(2 - W \sin^2 \theta)\kappa^3 - (6 + W \sin^2 \theta \cos^2 \theta)\kappa^2 + 2(2 + W \sin^4 \theta)\kappa - (1 + W \sin^6 \theta) = 0, \tag{44}$$

where

$$W \equiv [2S \cos \theta / \sin^2 \theta]^2. \tag{45}$$

This solution exists if and only if

$$\sin \theta > U, \tag{46a}$$

or, equivalently,

$$R < [\sin^2 \theta / (1 + \cos^2 \theta)]^2. \tag{46b}$$

Case 2B:

ϕ_1, κ_2, ϕ_2 are the same as in eqs. (39) through (41).

$$\sin \theta < \kappa_1 \leqq \kappa_b \quad (<1), \tag{47}$$

$$\sin^2 \theta < \kappa_2 \leqq \kappa_b^2 \quad (<1), \tag{48}$$

where κ_b is the positive root less than unity of the quartic equation (44).

This solution exists for all R, θ.

Case 2C:

ϕ_1, κ_2 are the same as in eqs. (39) and (40).

$$\sinh \alpha_2 = 2S \cos \theta p_1^2 q_2 / [(1 - \kappa_1) \sin \theta]^2. \tag{49}$$

$$\sin^2 \theta < \kappa_1 < \sin \theta. \tag{50}$$

$$\sin^4 \theta < \kappa_2 < \sin^2 \theta. \tag{51}$$

This solution exists for all R, θ. It has been previously found by Kard.[8]

Cases 3A, 3B:

No solutions exist.

Case 3C:

$$\cos(2\phi_1) = f(\kappa_1, 1) f(\kappa_1, \kappa_2)/[(\kappa_1 - 1)(\kappa_1 - \kappa_2)\sin^4\theta]. \tag{52}$$

$$\sinh\alpha_2 = \frac{2\sqrt{\kappa_1} q_2 f(\kappa_1, 1)}{(1+\kappa_1)(\kappa_1 - \kappa_2)\sin^2\theta[Y - (\kappa_2 + X^2)\cos^2\theta]^{\frac{1}{2}}}, \tag{53}$$

$$S = \sqrt{\kappa_1}(\kappa_2 - X)\cos\theta/[Y - (\kappa_2 + X^2)\cos^2\theta]^{\frac{1}{2}}, \tag{54}$$

where

$$X \equiv f(\kappa_1, \kappa_2)/[(1+\kappa_1)\sin^2\theta], \tag{55}$$

$$Y \equiv \kappa_2(1-\kappa_2)(\kappa_1 - 1)^2/(1+\kappa_1). \tag{56}$$

$$\kappa_1 > \{1 + 3\cos^2\theta + [8\cos^2\theta(1+\cos^2\theta)]^{\frac{1}{2}}\}/\sin^2\theta. \tag{57}$$

$$\sin^2\theta\cos^2\theta/(1+\cos^2\theta) \leqq \kappa_2 \leqq \sin^2\theta/(1+\cos^2\theta). \tag{58}$$

This solution exists if and only if

$$\sin\theta > (2/S)^{\frac{1}{3}}\{[(1 + 32/27S^2)^{\frac{1}{2}} + 1]^{\frac{1}{3}} - [(1 + 32/27S^2)^{\frac{1}{2}} - 1]^{\frac{1}{3}}\}^{\frac{1}{2}}, \tag{59a}$$

or, equivalently,

$$R > 8\cos^2\theta/(8\cos^2\theta + \sin^6\theta). \tag{59b}$$

Cases 4A, 4B:

No solutions exist.

Case 4C:

$$\cosh 2\alpha_1 = (1+\kappa_1)(\kappa_2 + \kappa_1)/[(1-\kappa_1)(\kappa_2 - \kappa_1)]. \tag{60}$$

$$\tan\phi_2 = \frac{2\sqrt{\kappa_1}\,\kappa_2(1+\kappa_1) q_1 p_2}{f(\kappa_1, \kappa_2)[(1+\kappa_2)(\kappa_1^2 + \kappa_2)]^{\frac{1}{2}}}. \tag{61}$$

$$S = \sqrt{\kappa_1}(\kappa_2 - X)/[\cos\theta q_1(\kappa_2 + X^2)^{\frac{1}{2}}], \tag{62}$$

where X is given in eq. (55).

$$\kappa_1 < \sin^2\theta. \tag{63}$$

$$\kappa_2 > \sin^2\theta. \tag{64}$$

This solution exists for all R, θ.

Discussion:

The above exhausts all possible solutions for a symmetric three-layer dielectric with no polarization dependence. The choice of the particular solution to implement in a given problem depends on the desired values of R and θ, the available dielectric materials, and whether

having dielectric layers surrounded by air (Type A solution) or a prism realization (Type B or C solution) is more suitable.

Cases 1A and 2A, which are the only possible Type A solutions, require large values for the dielectric constants and/or the angle of incidence to achieve moderate or large values of R. This point is illustrated for $R = 0.5$ in the design examples given in Section VIII.

In Case 2, the required relation between κ_1 and κ_2 given in eq. (40) has an advantage for the prism realizations in Cases 2B and 2C. In these two cases, since $\kappa_2 < \kappa_1 < 1$, eqs. (16) and (40) give $\epsilon_2 = 1$, $\epsilon_1 = \kappa_1/\kappa_2$, and $\epsilon_r = 1/\kappa_2 = \epsilon_1^2$. Thus, ϵ_1 can also be used as a quarter-wave transformer to match the outer surfaces of the prisms.

As in the single-layer prism solution given in Section IV, one can vary R in Cases 1 and 2 without affecting the polarization independence by simply changing d_2. This is particularly convenient in the prism realizations in Cases 1B, 2B, and 2C with $\epsilon_2 = 1$ (air) since, in this case, d_2 can be changed continuously. Case 2C is unique in that one can obtain any value of R from zero to unity by varying d_2 from zero to infinity. In Cases 1A, 1B, 2A, and 2B, varying d_2 results in a periodic variation of R between zero and a maximum value less than unity. Varying d_1 or d_2 in Cases 3 or 4 results in a polarization-dependent operation.

Cases 1A, 2B, 2C, and 4C have solutions for all R and θ. This is not true for the remaining three cases, 1B, 2A, and 3C, as indicated by eqs. (38), (46), and (59). However, one should note that the numerical values of ϵ_r, ϵ_1, and ϵ_2, and not necessarily the limitations on R and θ, often determine whether or not any particular case can be used in practice.

In Case 4C, eq. (46) indicates that one can have $\kappa_2 = 1$. In this case, given that $\kappa_1 < 1$ from eq. (63), eq. (16) gives $\epsilon_1 = 1$ and $\epsilon_r = \epsilon_2 = 1/\kappa_1$. Thus, the prism solution can be realized by using one dielectric material with air separations. This is not true for any of the other cases.

VI. SYMMETRIC MULTILAYER SOLUTIONS

When there are more than three layers, an exhaustive analysis similar to that given in the previous section becomes quite involved. However, two relatively simple classes of solutions for a symmetric multilayer dielectric with no polarization dependence can be obtained by generalizing Cases 1 and 2 of the symmetric three-layer dielectric. First, one should note that symmetry requires that the number of layers n be odd, i.e.,

$$n = 2l + 1, \qquad l = 1, 2, \cdots. \tag{65}$$

Thus, from symmetry

$$\kappa_i = \kappa_{n+1-i}, \qquad d_i = d_{n+1-i}, \qquad i = 1, 2, \cdots, l. \tag{66}$$

Let us assume that the electrical lengths of the first, and thus also the last, l layers are odd multiples of $\pi/2$; i.e.,

$$\phi_i = (2m_i + 1)\pi/2; \qquad m_i = 0, 1, 2, \cdots, \qquad i = 1, 2, \cdots, l. \tag{67}$$

While this is a necessary condition to obtain solutions for Cases 1 and 2 of a symmetric three-layer dielectric, as indicated in the previous section, eqs. (30) and (39), it is not necessary in general. However, this assumption is employed here because it simplifies the analysis considerably.

Define the function g_l of the vector $\mathbf{x} = \{x_1, x_2, \cdots, x_{l+1}\}$ as

$$g_l(\mathbf{x}) \equiv x_{l+1}^{[(-1)^l]}\Big(\prod_{i\,\text{odd}\leqq l} x_i^2\Big)/\Big(\prod_{i\,\text{even}\leqq l} x_i^2\Big). \tag{68}$$

Thus, $g_1(\mathbf{x}) = x_1^2/x_2$, $g_2(\mathbf{x}) = x_1^2x_3/x_2^2$, $\cdots$, etc. Each x_i can assume the value of Z_{ih}, Z_{ie}, κ_i, or ν_i (which will be defined later). From eqs. (66) and (67) and Section II, it can be shown that the normalized $ABCD$ parameters of the symmetric multilayer dielectric are given by

$$A = D = (-1)^l \cos(\phi_{l+1}), \tag{69}$$

$$B = (-1)^l \sin(\phi_{l+1}) g_l(\mathbf{Z}), \tag{70}$$

$$C = (-1)^l \sin(\phi_{l+1})/g_l(\mathbf{Z}). \tag{71}$$

Since ϕ_{l+1} is the same for both modes, eq. (69) shows that $A_h = D_h = A_e = D_e$. Thus, from Table I, the above equations can give solutions for Cases 1 and 2 provided that the conditions for the B's and the C's are met. Before satisfying these conditions, however, we note from eqs. (2) and (3) that

$$Z_{ih}Z_{ie} = 1/\kappa_i \tag{72}$$

and define

$$\nu_i \equiv Z_{ie}/Z_{ih} = (\kappa_i - \sin^2\theta)/(\kappa_i \cos^2\theta). \tag{73}$$

As κ_i assumes the increasing values of zero, $\sin^2\theta$, 1, and $+\infty$, then ν_i assumes the increasing values of $-\infty$, 0, 1, and $\sec^2\theta$, respectively.

From eqs. (9), (10), and (65) through (73), and from Table I, one can obtain solutions for Cases 1 ($B_h = B_e$, $C_h = C_e$), and 2 ($B_h = C_e$, $C_h = B_e$). The results are given below. Because of algebraic complexity, each case has not been subdivided into Type A, B, and C solutions. Also, for the same reason, no bounds are given on the κ's, θ, or R.

Case 1:

$$g_l(\boldsymbol{\nu}) = 1. \tag{74}$$

$$\sin\phi_{l+1} = \pm 2S/\{[g_l(\boldsymbol{\kappa})]^{\frac{1}{2}} - [g_l(\boldsymbol{\kappa})]^{-\frac{1}{2}}\}. \tag{75}$$

Case 2:

$$g_l(\kappa) = 1. \tag{76}$$

$$\sin \phi_{l+1} = \pm 2S/\{[g_l(\nu)]^{\frac{1}{2}} - [g_l(\nu)]^{-\frac{1}{2}}\}. \tag{77}$$

In eqs. (75) and (77), S is the same function of R defined in eq. (27). Note that the roles of the κ's and the ν's are interchanged in the two cases.

For a solution of any particular case to be valid, the value of $\sin \phi_{l+1}$ should either be real with a magnitude not exceeding unity, which leads to a Type A or a Type B solution, or be imaginary, which leads to a Type C solution. The latter solution is not possible in Case 1 because the κ's are real positive quantities, and thus the right-hand side of eq. (75) is always real. On the other hand, since the ν's, and in particular ν_{l+1}, can be negative, a Type C solution is possible in Case 2.

As in the previous section, solutions for Cases 1A, 2B, and 2C in the present section are possible for all values of R and θ.

In Cases 2B and 2C, when κ_{l+1} does not exceed $\kappa_1, \kappa_2, \cdots, \kappa_l$, or unity, eq. (16) shows that one can have a prism realization with $\epsilon_{l+1} = 1$, $\epsilon_l = \kappa_l/\kappa_{l+1}, \cdots, \epsilon_1 = \kappa_1/\kappa_{l+1}$, and $\epsilon_r = 1/\kappa_{l+1}$. Using eq. (76), one can show that l quarter-wave layers with dielectric constants $\epsilon_1, \epsilon_2, \cdots, \epsilon_l$ will match the outer surfaces of the prisms.

In practice, it is desirable to construct the multilayer dielectric reflector with the least number of different dielectrics. Thus, let us assume that two dielectrics with normalized dielectric constants κ_1 and κ_2 are used alternatively; i.e.,

$$\kappa_i = \kappa_1, \quad i \text{ odd}, \tag{78a}$$

$$\kappa_i = \kappa_2, \quad i \text{ even}. \tag{78b}$$

Combining this assumption with those in eqs. (65) through (67), the general solutions given in eqs. (74) through (77) simplify considerably. Because of its practical importance, we restrict ourselves to Type A solution, where κ_1 and κ_2 are larger than unity. No solution exists if either κ_1 or κ_2 is equal to unity.

Case 1A:

$$\kappa_2 = \sin^2 \theta/\{1 - [(\kappa_1 - \sin^2 \theta)/\kappa_1]^{(l+1)/l} (\cos \theta)^{-2/l}\}. \tag{79}$$

$$\sin \phi_{l+1} = \pm 2S/[(\kappa_2^l/\kappa_1^{l+1})^{\frac{1}{2}} - (\kappa_1^{l+1}/\kappa_2^l)^{\frac{1}{2}}]. \tag{80}$$

$$\kappa_1 < \sin^2 \theta/[1 - (\cos \theta)^{2/(l+1)}]. \tag{81}$$

$$\kappa_1^{l+1}/\kappa_2^l < V. \tag{82}$$

This solution exists for all R, θ.

Case 2A:

$$\kappa_2 = \kappa_1^{(l+1)/l} \tag{83}$$

$$\sin \phi_{l+1} = \pm 2S/[(\nu_1^{l+1}/\nu_2^l)^{\frac{1}{2}} - (\nu_2^l/\nu_1^{l+1})^{\frac{1}{2}}]. \tag{84}$$

$$\nu_2^l/\nu_1^{l+1} < V. \tag{85}$$

This solution exists if and only if

$$\sin \theta > U, \tag{86a}$$

or, equivalently,

$$R < [\sin^2 \theta/(1 + \cos^2 \theta)]^2. \tag{86b}$$

The quantities S, U, and V used in the above equations are defined in eqs. (27) through (29). Of course, when $l = 1$, i.e., $n = 3$, the above solutions reduce to Cases 1A and 2A given in the previous section for symmetric three-layer dielectric.

VII. CASCADING SOLUTIONS

It may happen that some required values of the reflection coefficient cannot be obtained by any of the solutions given in the previous two sections because of limitations on the values of the dielectric constant available in practice, or because the resulting angle of incidence is not suitable for a particular application. In this case, the desired multilayer dielectric may be obtained by cascading two or more solutions, provided that they have the same ϵ_r and the same θ.

We now proceed to find the rules for proper cascading to maintain polarization-independent operation. Let the two solutions to be cascaded be labeled x and y. They are not necessarily symmetric or identical. Let $\phi = 2\pi d \cos \theta/\lambda$ be the electric length between x and y (see Fig. 3). Dropping the subscripts h and e, which denote the particular mode being considered, we define r_x, r_x', t_x, r_y, r_y', and t_y as the reflection and transmission coefficients of x and y in the directions specified in the figure. The transmission coefficients t_x and t_y do not depend on the direction of propagation because of reciprocity and because the medium is the same on both sides of each of x and y. Because of losslessness, the difference between r_x and r_x' or between r_y and r_y' is a phase factor. Of course, if x is symmetric then $r_x = r_x'$, and if y is symmetric then $r_y = r_y'$.

Using the method of successive reflections,[23] it can be shown that for a wave incident on the x side of the cascade, the overall field transmission coefficient, t, and field reflection coefficient, r, are given by

$$t = t_x t_y \exp(-j\phi)/[1 - r_x r_y \exp(-j2\phi)], \tag{87}$$

$$r = r_x' + t_x^2 r_y \exp(-j2\phi)/[1 - r_x r_y \exp(-j2\phi)], \tag{88}$$

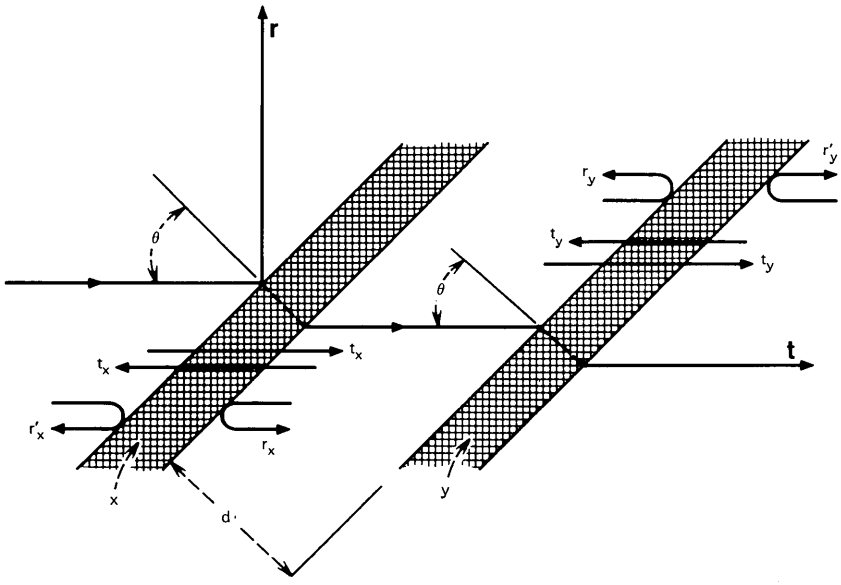


Fig. 3—Cascading of two polarization-independent solutions x and y to obtain an overall polarization-independent solution.

where we recall that a subscript of h or e has been dropped from all the t's and r's.

Since x and y are polarization-independent solutions, then, from eq. (14) or Table I, $r_{xh}r_{yh} = \pm r_{xe}r_{ye}$. Only the positive sign is acceptable, otherwise the magnitude of the denominators in eqs. (87) and (88) would be different for the h- and e-modes, which will result in a polarization-dependent operation for the cascade. Thus, the sign of r_e/r_h should be the same for all the cascaded sections; i.e., one can only have a combination of Cases 1 and 3 or Cases 2 and 4.

Let the abbreviation $i*j = k$ mean: "Cascading Cases i and j gives Case k." It can be shown from eqs. (87) and (88) and Table I that

$$1*1 = 3*3 = 1, \qquad 1*3 = 3*1 = 3, \tag{89a}$$

$$2*2 = 4*4 = 2, \qquad 2*4 = 4*2 = 4. \tag{89b}$$

Let the required overall power reflection coefficient of the cascade be

$$R \equiv |r|^2, \tag{90}$$

and let it be realized by two identical cascaded sections each having a power reflection coefficient R_x. Then, one can write

$$r_x = r_y \equiv R_x^{\frac{1}{2}} e^{j\phi_x}, \tag{91}$$

where ϕ_x is the phase angle of r_x. This can be found given the ϵ's, d's, and θ by using eqs. (1) through (10). Substituting eq. (91) in eqs. (87) and (88), and defining

$$\psi \equiv \phi_x - \phi, \tag{92}$$

one obtains

$$R = [1 + (1 - R_x)^2/(4R_x \sin^2 \psi)]^{-1}. \tag{93}$$

Thus, for a given value of R_x, one can obtain any value of R satisfying

$$0 \leqq R \leqq 4R_x/(1 + R_x)^2. \tag{94}$$

Note that the upper bound on R is larger than R_x. Conversely, for a given value of R, one can use any value of R_x satisfying

$$[1 - (1 - R)^{\frac{1}{2}}]^2/R \leqq R_x \leqq 1. \tag{95}$$

The higher the value used for R_x, the smaller is the overall bandwidth.

VIII. DIFFERENT INPUT AND OUTPUT MEDIA, AND ANTIREFLECTION COATINGS

Often, especially at optical wavelengths, the multilayer dielectric is supported by a substrate having a relative dielectric constant ϵ_s which is, in general, different from that of the input medium, ϵ_r. In this case, general analysis of polarization-independent solutions is complicated. However, it can be shown that special solutions similar to those for the symmetric $(2l + 1)$-layer dielectric given in Section VI can be easily obtained. For this purpose, let θ, as before, be the angle of incidence in the ϵ_r medium. Assume that there are $l(\geqq 1)$ layers with relative dielectric constants $\epsilon_1, \epsilon_2, \cdots, \epsilon_l$ between the ϵ_r and ϵ_s media. The normalized dielectric constants $\kappa_1, \kappa_2, \cdots, \kappa_l$ are defined by eq. (1). In addition, we define

$$\kappa_{l+1} \equiv \epsilon_s/\epsilon_r. \tag{96}$$

Further, as was done in Section VI, assume that each of the l layers is effectively a quarter-wave thick at the given angle of incidence, i.e.,

$$\phi_i = (2m_i + 1)\pi/2; \qquad m_i = 0, 1, 2, \cdots, \qquad i = 1, 2, \cdots, l. \tag{97}$$

With the above definitions and assumptions, it can be shown from eqs. (1) through (6) that the normalized input impedance as seen from the ϵ_r medium is given by

$$Z_{in} = g_l(\mathbf{Z}), \tag{98}$$

where $\mathbf{Z} = \{Z_1, Z_2, \cdots, Z_{l+1}\}$ and the function g_l is defined in eq. (68). Thus, the reflection coefficient in the ϵ_r medium is given by

$$r = (Z_{in} - 1)/(Z_{in} + 1). \tag{99}$$

A subscript of h or e, depending on the particular mode being considered, has been dropped from all the variables in eqs. (98) and (99).

It is clear that if $\kappa_{l+1} \equiv \epsilon_s/\epsilon_r > \sin^2\theta$, then Z_{l+1}, Z_{in} and r are all real quantities. In this case, it can be shown that the reflection coefficient r_s in the ϵ_s medium, and the transmission coefficient t across the l layers, are given by

$$r_s = (-1)^{l+1}r, \tag{100}$$

and

$$t = (-j)^l(-1)^{\sum_{i=1}^{l} m_i}(1 - r^2)^{\frac{1}{2}}, \tag{101}$$

where the m_i's are the integers defined in eq. (97). Since the impedances of the input and output media are different, t is defined here as the scattering transmission coefficient, which is usually referred to as S_{12}. Using the same notation, r and r_s could be referred to as S_{11} and S_{22}, respectively.

The coefficients r_s and t can not be defined if the total reflection condition, $\kappa_{l+1} \equiv \epsilon_s/\epsilon_r < \sin^2\theta$, is satisfied. In this case, Z_{l+1} is imaginary and a nonpropagating evanescent wave exists in the ϵ_s medium. It also follows that Z_{in} is imaginary, and hence, from eq. (99), $|r| = 1$.

It can be seen from eq. (99) that for $r_e = r_h$ (Case 1), $Z_{in,e} = Z_{in,h}$; and that for $r_e = -r_h$ (Case 2), $Z_{in,e} = 1/Z_{in,h}$. Thus, combining eqs. (98), (68), (72), and (73), and defining $R = |r_e|^2 = |r_h|^2$ as the power reflection coefficient, one obtains the following solutions.

Case 1:

$$g_l(\mathbf{v}) = 1. \tag{102}$$

$$R = \{[g_l(\boldsymbol{\kappa})]^{\frac{1}{2}} - 1\}^2/\{[g_l(\boldsymbol{\kappa})]^{\frac{1}{2}} + 1\}^2. \tag{103}$$

Case 2:

$$g_l(\boldsymbol{\kappa}) = 1. \tag{104}$$

$$R = \{[g_l(\mathbf{v})]^{\frac{1}{2}} - 1\}^2/\{[g_l(\mathbf{v})]^{\frac{1}{2}} + 1\}^2. \tag{105}$$

The above solution of Case 2 has been previously obtained by Baumeister[9] and by Rabinovitch and Pagis.[10] Both papers, however, overlooked the solution of Case 1.

It is observed that eqs. (102) and (104) are identical to eqs. (74) and (76) for the symmetric $(2l + 1)$-layer dielectric. This should not be surprising since the latter can be obtained by cascading, back to back, two identical solutions of the type described in the present section. In fact, this is the reason that the conditions for polarization-independent operation given in Sections V and VI for Cases 1 and 2 are independent of the thickness of the middle layer.

It is clear from eqs. (102) through (105) that if

$$g_l(\kappa) = g_l(\nu) = 1, \tag{106}$$

one obtains $R = 0$. Thus, eq. (106) gives the conditions for polarization-independent antireflection coatings at inclined incidence. It has been previously reported by Baumeister.[9] Rabinovitch and Pagis[10] have incorrectly stated that no such solution exists. It is interesting to note that the condition $g_l(\kappa) = 1$, which is independent of the angle of incidence, is the same condition on the dielectric constants that yields an antireflection coating at normal incidence.

At least two layers, $l \geqq 2$, are required to satisfy eq. (106). For $l = 2$, $\epsilon_r = 1$ and for a given finite θ, eq. (106) gives the unique solution.

$$\epsilon_1 = [\epsilon_3^{\frac{1}{2}} + \sin^2\theta + (\epsilon_3 - \sin^2\theta)^{\frac{1}{2}} \cos\theta]/(\epsilon_3^{\frac{1}{2}} + 1), \tag{107a}$$

$$\epsilon_2 = \epsilon_1\epsilon_3^{\frac{1}{2}}. \tag{107b}$$

For $l > 2$, the solution of eq. (106) is not unique, and we have extra degrees of freedom in choosing the dielectric constants of the layers. However, as pointed out by Baumeister,[9] the values of dielectric constants available in practice will often force a compromise of the effectiveness of the antireflection coatings.

IX. DESIGN EXAMPLES

For the purpose of demonstration and comparison of various solutions, let us design polarization-independent, multilayer, dielectric beam splitters with a power reflection coefficient $R = 0.5$ at 50 GHz, i.e., $\lambda_o = 6$ mm. (The frequency is relevant only for the selection of the dielectric materials.) Such 3-dB beam splitters may be used in circular-waveguide, channel-dropping filters.[7]

All the solutions considered are symmetric, and all, with the exception of the last solution, are of Type A with air as the surrounding medium; i.e.,

$$\epsilon_r = 1,$$

and thus, from eq. (1) and Table II, it follows that

$$\kappa_i = \epsilon_i \geqq 1, \qquad i = 1, 2, \cdots, n.$$

In three of the solutions, the frequency response of $|r_h|^2$ and $|r_e|^2$, in dB, and the phase difference

$$\begin{aligned} \Delta\phi &\equiv \text{phase}(r_h) - \text{phase}(r_e) \qquad (\text{mod } 180°) \\ &= \text{phase}(t_h) - \text{phase}(t_e) \qquad (\text{mod } 180°) \end{aligned}$$

are given. The second equality follows from losslessness and symmetry

Table III — Some low-loss dielectrics at millimeter wavelengths

Dielectric	Dielectric constant	Loss tangent	Frequency range of measurements	Reference No.
Kearfott High-Purity Alumina	9.4	0.00017	14 to 50 GHz	24
Polystyrene	2.54	0.0012	10 to 25 GHz	25
Teflon	2.1	0.0005	50 to 70 GHz	Unpublished report
Eccofoam PS*	1.02 to 2.0	0.0004	10 GHz	26

* This is a CO_2 foamed polystyrene whose density can be adjusted to yield any dielectric constant in the range specified. It is not clear how the Rayleigh scattering by the CO_2 bubbles will affect the loss tangent at frequencies higher than 10 GHz.

since, under these conditions, the difference between the phases of r and t is $\pm 90°$ at all frequencies as can be deduced from eqs. (8) through (10).

The dielectric materials employed in our solutions will be chosen from those given in Table III. Because of their low loss, these materials are suitable for use at millimeter wavelengths. Other suitable dielectrics can be found in Refs. 24 and 25.

9.1 Symmetric three-layer, Case 1A solution ($\epsilon_1, \epsilon_2, \epsilon_1$)

With $R = 0.5$, eqs. (33) and (34) require that $1.64 \leqq \epsilon_1 < 1.71$ and $\epsilon_2 \geqq 15.74$ for $\theta = 45°$; $1.46 \leqq \epsilon_1 < 1.5$ and $\epsilon_2 \geqq 12.34$ for $\theta = 60°$; or $1.23 \leqq \epsilon_1 < 1.26$ and $\epsilon_2 \geqq 8.90$ for $\theta = 75°$. It is difficult to realize a solution at millimeter wavelengths for θ less than about 60° because of the high required value ϵ_2. However, for $\theta \approxeq 75°$, the value of ϵ_2 can be realized by Kearfott Alumina ($\epsilon = 9.4$). The corresponding low value of ϵ_1 can be obtained by using Eccofoam PS, or foamed rubber such as Buna S Rubber whose dielectric constant can be extrapolated to be 1.26 at 50 GHz.[25] Thus, with $\epsilon_1 = 1.26$ and $\epsilon_2 = 9.4$, eq. (31) gives $\theta = 73.43°$, eqs. (4a) and (30) give $d_1/\lambda_o = 0.428 + m_1 \times 0.856$, and eqs. (4a) and (31) give $d_2/\lambda_o = 0.078 + m_2 \times 0.172$, where m_1 and m_2 are integers.

The frequency response of this solution is given in Fig. 4 for $m_1 = m_2 = 0$. The large required value of θ makes this solution undesirable.

9.2 Symmetric three-layer, Case 2A solution ($\epsilon_1, \epsilon_2, \epsilon_1$)

With $R = 0.5$, eq. (46) requires that $\theta > 65.53°$; and eqs. (40) and (42) through (44) require that $\epsilon_1 > 4.61$ and $\epsilon_2 = \epsilon_1^2 > 21.23$ for $\theta = 70°$; $\epsilon_1 > 2.09$ and $\epsilon_2 = \epsilon_1^2 > 4.38$ for $\theta = 75°$; and $\epsilon_1 > 1.58$

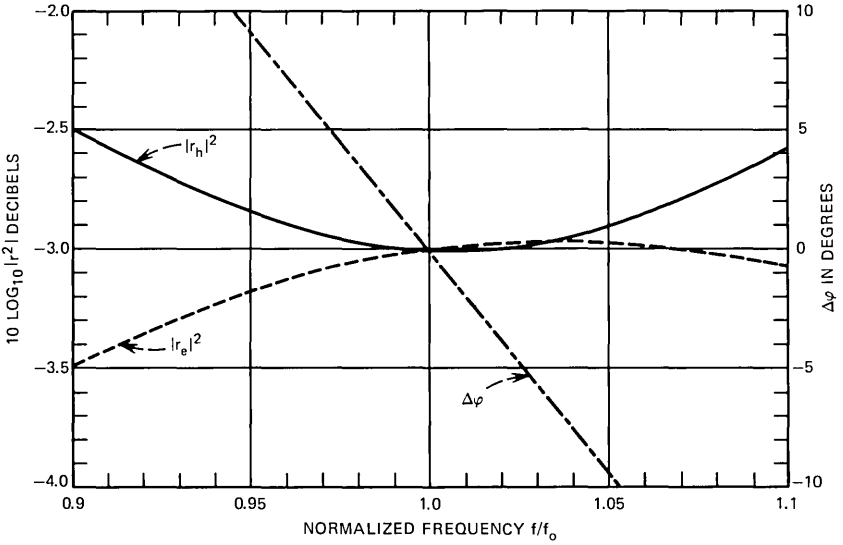


Fig. 4—Frequency response of the solution in 9.1 [three-layer, Type 1A solution with $\theta = 73.43°$, $\epsilon_1 = 1.26$ (Buna S Rubber), $\epsilon_2 = 9.4$ (Kearfott Alumina), $d_1/\lambda_0 = 0.428$, and $d_2/\lambda_o = 0.078$].

and $\epsilon_2 = \epsilon_1^2 > 2.49$ for $\theta = 78°$. In the last case, the solution can be realized by Eccofoam PS ($\epsilon = 1.59$) and polystyrene ($\epsilon = 2.54$). Just as in the solution in Section 9.1, the required large value of θ makes this solution undesirable. For symmetric multilayer, Type-2A solutions with more than three layers of two dielectrics, eq. (86) shows that θ is still required to exceed 65.53°.

9.3 Symmetric five-layer, Case 1A solution ($\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_2, \epsilon_1$)

For $R = 0.5$ and $l = 2$, i.e., n = five layers, eqs. (79), (80), and (81) can be satisfied with $\epsilon_1 = \epsilon_3 = 2.1$ (Teflon), $\epsilon_2 = 9.4$ (Kearfott Alumina), and $\theta = 46.91°$, which is a convenient angle of incidence. Further, eqs. (4a) and (67) give $d_1/\lambda_o = 0.200 + m_1 \times 0.399$ and $d_2/\lambda_o = 0.084 + m_2 \times 0.168$, and eqs. (4a) and (80) give $d_3/\lambda_o = 0.103 + m_3 \times 0.399$, where m_1, m_2, and m_3 are integers. The frequency response of this solution is given in Fig. 5 for $m_1 = m_2 = m_3 = 0$.

9.4 Symmetric seven-layer, Case 1A solution ($\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_3, \epsilon_2, \epsilon_1$)

For $R = 0.5$ and $l = 3$, i.e., n = seven layers, eqs. (79), (80), and (81) can be satisfied with $\epsilon_1 = \epsilon_3 = 2.54$ (polystyrene), $\epsilon_2 = \epsilon_4 = 9.4$ (Kearfott Alumina), and $\theta = 47.52°$. Further, eqs. (4a) and (67) give $d_1/\lambda_o = d_3/\lambda_o = 0.177 + m_1 \times 0.354$ and $d_2/\lambda_o = 0.084 + m_2 \times 0.168$,

and eqs. (4a) and (80) give $d_4/\lambda_o = 0.026 + m_4 \times 0.168$, where m_1, m_2, and m_4 are integers. This solution has a narrower bandwidth than the solution in 9.3.

9.5 Symmetric seven-layer, Case 1A solution obtained by cascading two identical, symmetric three-layer, Case 1A solutions ($\epsilon_1, \epsilon_2, \epsilon_1, 1.0, \epsilon_1, \epsilon_2, \epsilon_1$)

With $R = 0.5$, eq. (95) shows that R_x, the power reflection coefficient of each cascaded three-layer section, should not be less than 0.1716; thus let us choose $R_x = 0.1716$. In this case, eqs. (31), (33), and (34) can be satisfied with $\epsilon_1 = 1.6$ (Eccofoam PS), $\epsilon_2 = 9.4$ (Kearfott Alumina), and $\theta = 45.30°$. Further, eqs. (4a) and (30) give $d_1/\lambda_o = 0.239 + m_1 \times 0.478$ and eqs. (4a) and (32) give $d_2/\lambda_o = 0.038 + m_2 \times 0.168$.

To find the width d of the air separation, one should first find the phase ϕ_x of the reflection coefficient of each cascaded three-layer section. By substituting the values of ϵ_1, ϵ_2, d_1, d_2, and θ obtained above in eqs. (2) through (10), one finds that $\phi_x = 43.59°$. With $R = 0.5$ and $R_x = 0.1716$, eq. (93) gives $\psi = \pm 90°$ (mod 180°), and hence, the electrical length of the air spacing between the two three-layer sections is found from eq. (92) to be $\phi = 133.59° + m \times 180°$, where m is an integer. Thus, from eq. (4a) with $\epsilon_r = \kappa_i = 1$, $d/\lambda_o = 0.528$

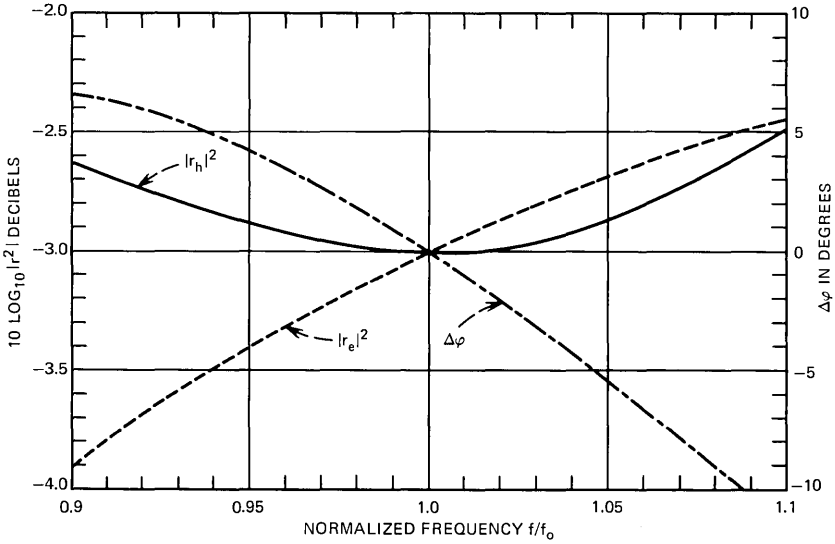


Fig. 5—Frequency response of the solution in 9.3 [five-layer, Type 1A solution with $\theta = 46.91°$, $\epsilon_1 = \epsilon_3 = 2.1$ (Teflon), $\epsilon_2 = 9.4$ (Kearfott Alumina), $d_1/\lambda_o = 0.200$, $d_2/\lambda_o = 0.084$, and $d_3/\lambda_o = 0.103$].

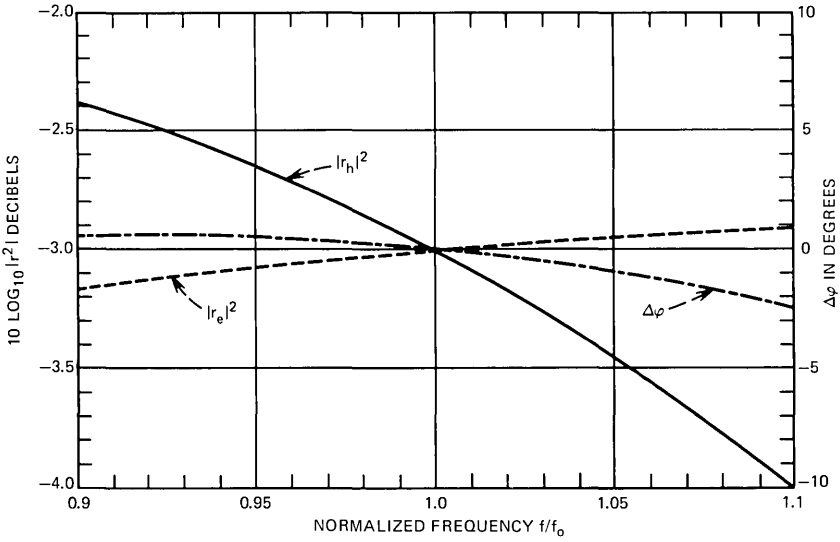


Fig. 6—Frequency response of the solution in 9.6 [three-layer, Type 2C prism solution with $\theta = 45°$, $\epsilon_r = 2.54$ (polystyrene), $\epsilon_1 = 1.59$ (Eccofoam PS), $\epsilon_2 = 1$, $d_1/\lambda_0 = 0.439$, and $d_2/\lambda_o = 0.236$].

$+ m \times 0.711$. This solution has a narrower bandwidth than the solution in 9.3.

9.6 *Symmetric three-layer, Case 2C prism solution* (ϵ_r, ϵ_1, **1.0**, ϵ_1, ϵ_r)

Choosing $\theta = 45°$ and $\epsilon_2 = 1.0$, eqs. (16), (40), (50) and (51) give $1.414 < \epsilon_1 < 2$ and $2 < \epsilon_r = \epsilon_1^2 < 4$. Thus, the solution can be obtained with $\epsilon_1 = 1.59$ (Eccofoam PS) and $\epsilon_r = \epsilon_1^2 = 2.54$ (polystyrene). Further, eqs. (4a) and (39) give $d_1/\lambda_o = 0.439 + m \times 0.879$, where m is an integer, and with $R = 0.5$, eqs. (4a), (4b), and (49) give $d_2/\lambda_o = 0.236$. The frequency response of this solution is given in Fig. 6. for $m = 0$ and under the assumption that the outer surfaces of the prisms are perfectly matched. As mentioned at the end of Section V, the matching can be accomplished by using a quarter-wave layer having a dielectric constant $\epsilon_1 = 1.59$ (Eccofoam PS).

X. ACKNOWLEDGMENT

The author wishes to thank J. A. Arnaud for useful suggestions.

REFERENCES

1. M. Born and E. Wolf, *Principles of Optics*, Oxford: Pergamon Press, 1965.
2. A. F. Harvey, "Optical Techniques at Microwave Frequencies," Proc. IEE, *B106* (March 1959), pp. 141–157.
3. R. Trembley and A. Boivin, "Concepts and Techniques of Microwave Optics," Applied Optics, *5*, No. 2 (February 1966), pp. 249–278.
4. R. Levy, "Directional Couplers," in *Advances in Microwaves, 1*, L. Young, Ed., New York: Academic Press, 1966, pp. 196–205.

5. J. J. Taub, H. J. Hindin, O. F. Hinckelmann, and M. L. Wright, "Submillimeter Components Using Oversized Quasi-Optical Waveguide," IEEE Trans. Microwave Theory Tech., *MTT-11*, No. 5 (September 1963), pp. 338–350.
6. M. Koyama and S. Shimada, "The Quasi-Optical Filters Used for the Domestic Satellite Communication System," Electr. and Commun. in Japan, *56-B*, No. 3 (March 1973), pp. 74–81.
7. E. A. J. Marcatili and D. L. Disbee, "Band-Splitting Filter," B.S.T.J., *40*, No. 1 (January 1961), pp. 197–212.
8. P. G. Kard, "On the Elimination of the Doublet Structure of the Transmission Band in a Total-Reflection Light Filter," Optics and Spectroscopy, (Translation of Optika i Spektroskopiia), *VI*, No. 3 (March 1959), pp. 244–246.
9. P. Baumeister, "The Transmission and Degree of Polarization of Quarter-Wave Stacks at Non-Normal Incidence," Optica Acta, *8*, No. 2 (April, 1961), pp. 105–119.
10. K. Rabinovitch and A. Pagis, "Polarization Effects in Multilayer Dielectric Thin Films," Optica Acta, *21*, No. 12 (December 1974), pp. 963–980.
11. R. B. Adler, L. J. Chu, and R. M. Fano, *Electromagnetic Energy Transmission and Radiation*, New York: John Wiley, 1965, pp. 354–362.
12. J. R. Wait, *Electromagnetic Waves in Stratified Media*, New York: Pergamon Press, 1962, pp. 10–15.
13. P. Leurgans and A. F. Turner, "Frustrated Total Reflection Interference Filters," J. Opt. Soc. Am., *37*, No. 12 (December 1947), p. 983.
14. M. Born and E. Wolf, Ref. 1, pp. 350–351.
15. H. D. Raker and G. R. Valenzuela, "A Double-Prism Attenuator for Millimeter Waves," IRE Trans. Microwave Theory Tech., (Correspondence), *MTT-10*, No. 5 (September 1962), pp. 392–393.
16. H. J. Hindin and J. J. Taub, "Oversized Waveguide Directional Coupler," IRE Trans. Microwave Theory Tech. (Correspondence), *MTT-10*, No. 5 (September 1962), pp. 394–395.
17. T. Makimoto and T. Sueta, "Prism Directional Couplers Making Use of Brewster Angle Transmission," *Memoirs of the Institute of Scientific and Industral Research*, Osaka University, Japan, *21*, 1964.
18. T. Sueta, N. Kumagai, and S. Kurazono, "Wideband Quasi-Optical Prism Components," Elect. and Commun. in Japan, *49*, No. 3 (March 1966), pp. 100–106.
19. R. G. Fellers and J. Taylor, "Internal Reflections in Dielectric Prisms," IEEE Trans. Microwave Theory Tech., *MTT-12*, No. 6 (November 1964), pp. 584–587.
20. R. E. Collin and J. Brown, "The Design of Quarter-Wave Matching Layers for Dielectric Surfaces," Proc. IEE, *103C* (September 1955), pp. 153–158.
21. T. Morita and S. B. Cohn, "Microwave Lens Matching by Simulated Quarter-Wave Transformers," IRE Trans. Antennas and Propagation, *AP-4*, No. 1 (January 1965), pp. 33–39.
22. E. M. T. Jones and S. B. Cohn, "Surface Matching of Dielectric Lenses," J. of Appl. Phys., *26*, No. 4 (April 1955), pp. 452–457.
23. M. Born and E. Wolf, Ref. 1, pp. 323–329.
24. Lab. for Insulation Research, M.I.T., *Tables of Dielectric Materials, 5*, April 1957.
25. A. R. Von Hipple, ed., *Dielectric Materials and Applications*, M.I.T. Press and John Wiley, New York, 1954.
26. Emerson and Cuming, Inc., Dielectric Materials Division, "Eccofoam PS—Low-Loss Adjustable Dielectric Constant Foam," Technical Bulletin No. 6-2-4, Canton, Massachusetts, June 15, 1970.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

A Method for Calculating Rain Attenuation Distributions on Microwave Paths

By S. H. LIN

(Manuscript received January 9, 1975)

An engineering method is proposed for calculating rain attenuation distributions for frequencies greater than 10 GHz and for paths of arbitrary length. The technique is based upon the observed approximate lognormality of rain attenuation and rain rate statistics within the range of interest; it reflects local meteorology through incorporation of the observed point rain rate distribution. Some important parameters in the resulting formulas are determined empirically from experimental data. Sample calculated results agree well with available experimental data from Georgia, New Jersey, and Massachusetts. This new technique may prove useful for engineering radio paths at frequencies above 10 GHz. Sample calculations of expected outage probability are given for 11- and 18-GHz radio links at Atlanta, Georgia, as a function of repeater spacing and transmission polarization.

I. INTRODUCTION

An important problem in designing radio relay systems at frequencies above 10 GHz is the radio outage caused by rain attenuation. Determination of the appropriate radio repeater spacings for economic and reliable operation requires a knowledge of the probability distribution of rain attenuation as a function of repeater spacing at various geographic locations. Most available data on rain rate statistics are measured by a single rain gauge at a given geographic location. A procedure for calculating a rain attenuation distribution from a point rain rate distribution is, therefore, needed.

The results of rain gauge network measurements[1–5] indicate, however, that the measured short-term distributions of point rain rate vary significantly from gauge to gauge. For example, at Holmdel, New Jersey, there was considerable variation[1] among the measured point rain rate distributions obtained from 96 rain gauges located in a grid with 1.3-km spacing over a 6-month period. Among these 96 distributions, the incidence of 100-mm/h rains is higher by a factor of

5 for the upper quartile gauges than for the lowest quartile (see Fig. 32, Ref. 1). Data from a rain gauge network in England[5] indicate that even with a 4-year time base and averaging over observations by four gauges with 1-km gauge spacing, the four-gauge-average rain rate incidence can differ by a factor of 3 for rain rates above 80 mm/h, depending on which four gauges are chosen for averaging. This means that, on a short-term basis, the relationship between the path rain attenuation distribution and the rain rate distribution measured by a single rain gauge is *not unique*. The prediction of a path rain attenuation distribution from a point rain rate distribution is, therefore, meaningful only if the time base is sufficiently long to yield stable, representative statistics. Accordingly, knowledge of the long-term statistical behavior of point rain rate and path rain attenuation is essential for radio path design.

The available experimental rain data (Appendix B, Figs. 10 to 13, and Ref. 6) indicate that the long-term distributions of point rain rate R and rain attenuation α are approximately lognormal within the range of interest to designers of radio paths using frequencies above 10 GHz. This paper describes a method for calculating rain attenuation distributions based upon this lognormal hypothesis.

A lognormal distribution is uniquely determined by three parameters (see Section 2.1). A set of equations are derived to relate the lognormal parameters of attenuation α to those of point rain rate R. Thus, given a long-term, representative distribution of point rain rate at a given geographic location, the rain attenuation distribution for any path length of interest can be calculated. The method is outlined as follows.

The available theory[7–13] for converting rain rate R into rain attenuation gradient β in dB/km is appropriate to spatially uniform rain rates, whereas actual rainfalls are usually not uniform over an entire radio path. To apply the uniform rain theory, the radio path volume is divided into small incremental volumes ΔV, in which the rain rate is approximately uniform. The rain rate R in each small volume ΔV is associated with a corresponding attenuation gradient β by the uniform rain theory. The total path attenuation α is then the integral of β over the path volume.

If the spatial distribution of the attenuation gradient β were uniform, the path attenuation at a given probability level would increase linearly with path length. On the other hand, if the spatial distribution of βs were not uniform and the time fluctuations of the βs were statistically independent, the incremental attenuation contributed by each ΔV would sum on an rms basis. Intuitively, we feel that the attenuation gradients at two different positions are partially corre-

lated, with a correlation coefficient that is a decreasing function of the spacing between the two positions. This behavior is described by introducing a spatial correlation function for β. This spatial correlation function is used in the calculation of lognormal parameters of path attenuation α from those of attenuation gradient β.

In this formulation, the appropriate incremental sampling volume ΔV is of the order of 1 m^3 and the corresponding appropriate rain gauge integration time about 2 s, requiring, therefore, 2-s point rain rate distributions (see Section 2.3).

Tables II and III and Figs. 5 to 8, discussed in Section IV, present comparisons of calculated and measured attenuation distributions. The satisfactory correspondence appears to validate the method of calculation. Section V and Fig. 9 present the calculated results for the outage probabilities of 11- and 18-GHz radio links in Georgia. Section VI discusses some qualifications to the methodology.

Supporting material and mathematical derivations are given in Appendices A to D. Appendix E lists symbols and their definitions.

II. BASIC DEFINITIONS AND FORMULATION

2.1 Lognormal distributions of attenuation and rain rate

The equations approximating the rain attenuation distribution and point rain rate distribution are:

$$P[\alpha(L) \geqq A] \simeq P_0(L) \cdot \tfrac{1}{2}\mathrm{erfc}\left[\frac{\ln A - \ln \alpha_m}{\sqrt{2}S_\alpha}\right] \tag{1}$$

and

$$P(R \geqq r) \simeq P_0(0) \cdot \tfrac{1}{2}\mathrm{erfc}\left[\frac{\ln r - \ln R_m}{\sqrt{2}S_R}\right], \tag{2}$$

where erfc ($\sim$) denotes the complementary error function, ln ($\sim$) denotes natural logarithm, S_α and S_R are the standard deviations of $\ln \alpha$ and $\ln R$, respectively, *during the raining time*, α_m and R_m are the median values of α and R, respectively, during the raining time, $P_0(L)$ is the probability that rain will fall on the radio path of length L, and $P_0(0)$ is the probability that rain will fall at the point where the rain rate R is measured. The definition and the determination of raining time, and hence $P_0(L)$ and $P_0(0)$, are discussed in Section 2.5. The measured distribution of point rain rate is a function of rain gauge integration time T.[14–20] The appropriate integration time is about 2 s in our formulation as discussed in Section 2.3.

2.2 Radio path definition

C. L. Ruthroff[21] has defined a "radio path," giving a physical as well as mathematical representation in the spatial volume significant

to propagation considerations (Fig. 1). In essence, the "radio path" corresponds to the first Fresnel zone, a prolate ellipsoid of revolution terminated at the ends by the transmitting and receiving antennas.

Since the first Fresnel zone is circularly symmetric with respect to the path axis connecting the transmitting and receiving antennas, we adopt a cylindrical coordinate system (Fig. 1) coaxial with the path axis (z-axis) in the formulation.

The radius $h(z)$ and the circular cross section $Q(z)$ (see Fig. 1) of the radio beam at a distance z from the transmitter are

$$h(z) = \left[\frac{\lambda \cdot z(L - z)}{L} \right]^{\frac{1}{2}} \tag{3}$$

and

$$Q(z) = \pi h^2(z), \tag{4}$$

where λ is the radio wavelength and L is the distance between transmitter and receiver. For example, at 18 GHz on a 5-km path, the average beam radius, the average beam cross section, and the radio path volume are about 3 m, 30 m^2, and 150,000 m^3, respectively.

2.3 Path integral formulation for rain attenuation

The spatial distribution of actual rainfall is usually nonuniform. The rain density, the point rain rate R, and the corresponding (point) rain attenuation gradient β are all functions of position (ρ, ϕ, z) and of time (t). The total rain attenuation α in dB incurred on a radio path of length L (Fig. 1) is calculated by integrating the incremental attenuation $d\alpha$ along the path

$$\alpha(t) = \int_0^L \frac{d\alpha}{dz}\, dz = \int_0^L \beta_q(z, t) dz \tag{5}$$

$$= \int_0^L \frac{1}{Q(z)} \int_{\phi=0}^{2\pi} \int_{\rho=0}^{h(z)} \beta(\rho, \phi, z, t) \rho d\rho d\phi dz, \tag{6}$$

where

$$\beta_q(z, t) = \frac{1}{Q(z)} \int_{\phi=0}^{2\pi} \int_{\rho=0}^{h(z)} \beta(\rho, \phi, z, t) \rho d\rho d\phi \tag{7}$$

is the average value of $\beta(\rho, \phi, z, t)$ over the radio beam cross section $Q(z)$ at a distance z from the transmitter, and

$$d\alpha(z, t) = \beta_q(z, t) dz \tag{8}$$

is the incremental attenuation experienced in the incremental segment dz at a distance z from the transmitter.

To shorten the notations in the following equations, we use the vector $\mathbf{s}$ to denote the position (ρ, ϕ, z) and dv to denote $\rho d\rho d\phi dz$ in the volume integration.

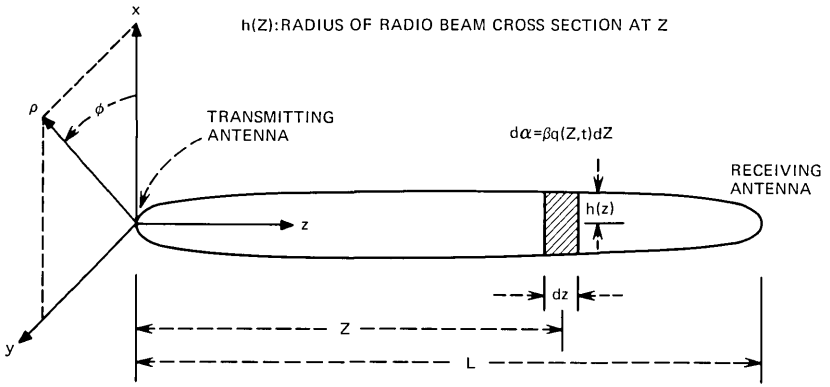


Fig. 1—Configuration of a radio path.

The rain density and the rain attenuation gradient β in dB/km are only meaningful with respect to a sampling volume ΔV large enough to contain sufficient raindrops to yield stable volume average quantities. The results of measurements of rain density and rain rate by a photographic method[22] indicate that the typical rain density at 1-mm/h rain rate varies from 50 to 100 raindrops per m^3, depending on geographic location. This means, on the average, a 0.1-m^3 volume contains only 2 raindrops at 0.25-mm/h rain rate. Such a small sampling volume will not measure the rain rate in the conventional sense, but rather will "see" individual raindrops. Thus, for a meaningful measurement of rain rate below 1 mm/h, the sampling volume should be at least 1 m^3.

On the other hand, the available theoretical results[7–13] relating β and R assume that the rain density is uniform within the volume of interest. To use the available functional relationship $\beta(R)$ to convert the statistics of R into that of β, the sampling volume ΔV must be sufficiently small so that the rain density (and the rain rate) is approximately uniform within ΔV. The observations by a capacitor flow rain gauge[23,24] and by a raindrop photographic method[25] indicate that heavy rain has fine scale structure on the order of 1 m. This means ΔV should not be much larger than 1 m^3.

These two constraints indicate that the sampling volume should be on the order of 1 m^3. We have chosen*

$$\Delta V \approx 1 \text{ m}^3 \tag{9}$$

in our formulation.

* A choice of ΔV somewhat different from 1 m^3 is also possible. Since the spatial correlation coefficient of β depends on the sampling volume, the use of a slightly different ΔV would result in a slightly different characteristic distance G, defined in eq. (35) and determined in Section 4.1. For example, a larger sampling volume, with more smoothing effect, will result in a larger characteristic distance G for β.

Therefore, in this paper, $\beta(\mathbf{s}, t)$ is defined as the rain attenuation gradient at time t, owing to rain in a 1-m^3 sampling volume, ΔV, centered at $\mathbf{s}$, and $\beta_q(z, t)$ is the average value of $\beta(\mathbf{s}, t)$ over the path cross section $Q(z)$.

If A_g is the area of the collecting aperture of a rain gauge and V_R the average descent velocity of rainfall, then the appropriate rain gauge integration time $T_{\Delta V}$ to measure rain rate in a 1-m^3 sampling volume ΔV is

$$T_{\Delta V} \simeq \frac{\Delta V}{A_g \cdot V_R}. \tag{10}$$

For example, if $A_g = 0.073$ m^2 (i.e., 12-in. diameter), then $T_{\Delta V}$ is about 2 s, assuming $V_R \cong 7$ m/s. In the measurements of raindrop size distributions, Laws and Parsons[26] have also used an integration time in the order of seconds during heavy rain. The Laws-and-Parsons raindrop size distribution is the basis of most uniform rain theories for converting R into β. We therefore define $R(\mathbf{s}, t)$ as the point rain rate measured by a rain gauge with integration time $T_{\Delta V}$, located at $\mathbf{s}$. The shape of the 1-m^3 sampling volume defined by the rain gauge is cylindrical and is considerably different from that of the incremental ΔV. We assume that the long-term distributions of rain rates for these two different shapes of 1-m^3 sampling volume are approximately the same.

Based upon these definitions, we postulate that the long-term probability distribution of $R(\mathbf{s}, t)$ can be converted into the long-term probability distribution of $\beta(\mathbf{s}, t)$ by a relationship discussed in the next section.

The integration times of most available point rain rate data are longer than the $T_{\Delta V}$ ($\cong$2 s) required by this formulation. The dependence of point rain rate distribution on the rain gauge integration time in the range

$$1.5 \text{ s} \leqq T \leqq 120 \text{ s} \tag{11}$$

has been determined by Bodtmann and Ruthroff[15] for a 2-year (1971-1972) measurement at Holmdel, New Jersey. By using this experimental result and interpolation, we convert the available point rain rate distribution with T in the range (11) into a 2-s point rain rate distribution.*

2.4 Average relationship between rain rate and rain attenuation gradient

The instantaneous relationship obtaining between the point rain rate $R(\mathbf{s}, t)$ and the corresponding rain attenuation gradient $\beta(\mathbf{s}, t)$

* This relationship has not yet been demonstrated to be geographically independent.

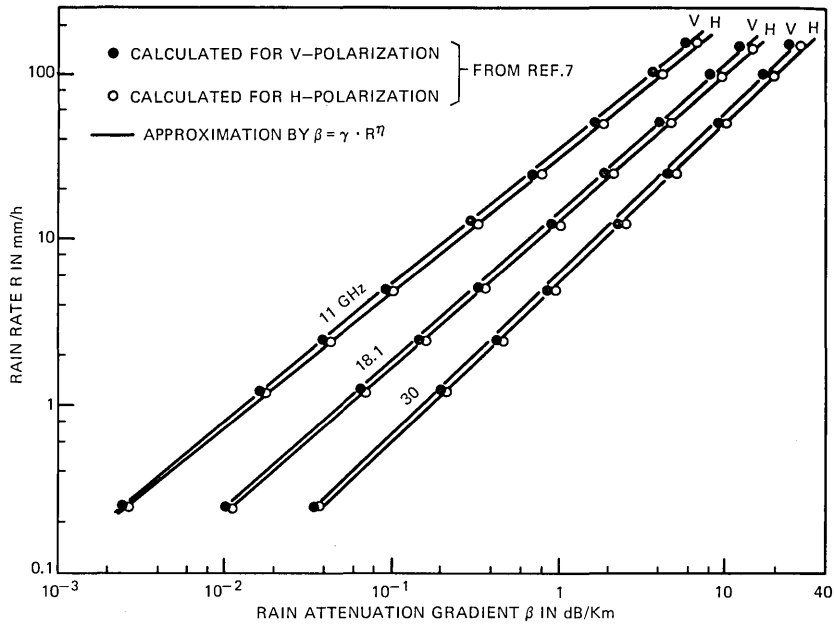


Fig. 2—Theoretical relationship between rain rate and rain attenuation gradient.

depends upon the particular distribution of raindrop sizes, shapes, and orientations, the speed and local direction of the wind, and the rain temperature. The average relationship, assuming uniform rain density, spherical raindrops, and Laws-and-Parsons drop-size distribution, has been calculated by Ryde and Ryde,[11–13] Medhurst,[9] and Setzer.[10] Recently, Morrison, Cross, and Chu,[7,27] and Oguchi[8] have refined these calculations by including the effects of nonspherical raindrops. Figure 2 shows this theoretical relationship* for transmission frequencies of 11, 18, and 30 GHz.

Many authors have pointed out that this average relationship between the rain rate R and rain attenuation gradient β can be approximately described by

$$\beta = \gamma(\lambda) \cdot R^{\eta(\lambda)}, \tag{12}$$

where $\gamma(\lambda)$ and $\eta(\lambda)$ depend upon the radio wavelength λ and the polarization of the radio signal. Table I lists the estimated $\gamma(\lambda)$ and $\eta(\lambda)$ for 11, 18, 30, and 60 GHz.

* In Fig. 2, the average of the absolute value of raindrop canting angle is assumed to be 25 degrees. This angle has been found representative by Chu (Refs. 28 and 29) in comparisons of calculated results with experimental observations (Refs. 30 and 31) of the differential rain attenuation experienced by horizontally and vertically polarized signals on the same radio path.

Table I — Parameters relating rain rate R and rain attenuation gradient β

$\beta = \gamma \cdot R^{\eta}$ β in dB/km R in mm/hour

Frequency (GHz)	γ		η	
	V-Pol	H-Pol	V-Pol	H-Pol
11	0.013	0.015	1.22	1.23
18.1	0.05	0.054	1.11	1.14
30	0.15	0.17	1.04	1.04
60*	0.7*	0.7*	0.814*	0.814*

* The 60-GHz parameters are estimated from results in Ref. 10 in which only spherical raindrops are considered.

Taking logarithms of both sides of eq. (12) yields

$$\ln \beta = \ln \gamma + \eta \cdot \ln R. \tag{13}$$

From this equation, if the distribution of point rain rate R is approximately lognormal in the range of interest, then the distribution of attenuation gradient β will also be approximately lognormal (see Appendix B). The distribution of β can therefore be written as:

$$P(\beta \geqq B) \simeq P_0(0) \cdot \tfrac{1}{2}\mathrm{erfc}\left[\frac{\ln B - \ln \beta_m}{\sqrt{2} \cdot S_\beta}\right], \tag{14}$$

where β_m is the median value of β during the raining time and S_β is the standard deviation of $\ln \beta$ during the raining time. Furthermore, eqs. (12) and (13) imply the relationships

$$S_\beta = \eta \cdot S_R \tag{15}$$

and

$$\beta_m = \gamma \cdot R_m^{\eta}. \tag{16}$$

Equations (15) and (16) allow us to convert the lognormal distribution (2) of R into the lognormal distribution (14) of β, and vice versa.

2.5 *Rainfall probability $P_0(0)$ and raining time*

In principle, the probability of raining, $P_0(0)$, is obtained as the limit

$$\lim_{\epsilon \to 0^+} P(R \geqq \epsilon) \equiv P_0(0). \tag{17}$$

An instant t is considered to be raining time if the condition

$$\lim_{\epsilon \to 0^+} R(t) > \epsilon \tag{18}$$

is satisfied. The lower cutoff threshold in most presently available rain

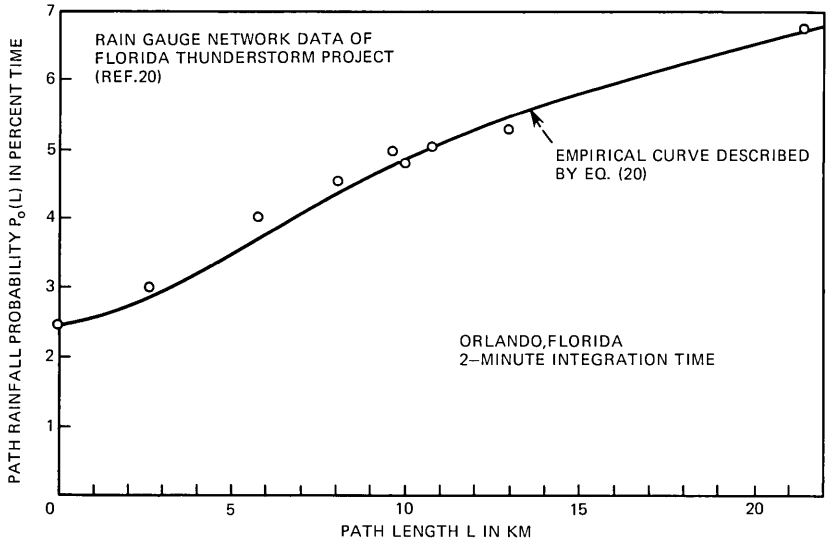


Fig. 3—Rain-gauge network data on path length dependence of rainfall probability $P_0(L)$.

rate data is about 0.25 mm/h. Therefore, in practice, we approximate 0^+ in definitions (17) and (18) by 0.25 mm/h. The rationale for this approximation is twofold.

(*i*) Rain rates below 0.25 mm/h have practically no significant effects on radio communication links at frequencies below 60 GHz.

(*ii*) Rain rates below 0.25 mm/h cannot be measured accurately by most existing rain gauges with standard recording strip charts.

At the present time, the probability $P(R \geqq 0.25 \text{ mm/h} \mid T = 1 \text{ min})$ is available at only a few locations.[14,18,20] For most locations, we can obtain $P(R \geqq 0.25 \text{ mm/h})$ with 1-h integration time from the Weather Bureau hourly precipitation data.[32] The experimental results on the effect of rain gauge integration time T on $P_0(0)$ in Florida[14,20] and Japan[18] indicate that

$$P(R \geqq 0.25 \text{ mm/h} \mid T \leqq 1 \text{ min}) \simeq 0.5 \cdot P(R \geqq 0.25 \text{ mm/h} \mid T = 1 \text{ h}). \quad (19)$$

Therefore, we use Weather Bureau data and approximation (19) to estimate $P_0(0)$ at several locations of interest where direct measurement of $P_0(0)$ with 1-min integration is not available.

Intuitively, we expect the probability $P_0(L)$ of rainfall on a radio path of length L to increase with L, since a longer path has a higher chance of intercepting rain of limited extent. From the rain gauge network data (Fig. 3) of the Florida Thunderstorm Project,[14] we obtain the empirical formula

$$P_0(L) \cong 1 - \frac{1 - P_0(0)}{\left[1 + \frac{L^2}{21.5}\right]^{0.014}}, \tag{20}$$

where L is in units of kilometers and

$$P_0(0) = \lim_{L \to 0} P_0(L) \tag{21}$$

is the point rain probability that depends on geographic location. A theoretical consideration leading to the empirical form (20) is discussed in Appendix C.

III. OUTLINE FOR CALCULATING RAIN ATTENUATION DISTRIBUTION

3.1 Path length dependence of median attenuation α_m and standard deviation S_α

From the lognormal approximations (1) and (14), it can be shown[33] that:

$$S_\alpha^2 = \ln\left\{1 + \frac{\sigma_\alpha^2}{\bar{\alpha}^2}\right\} \tag{22}$$

$$S_\beta^2 = \ln\left\{1 + \frac{\sigma_\beta^2}{\bar{\beta}^2}\right\} \tag{23}$$

$$\alpha_m = \bar{\alpha} \cdot \exp\left[\frac{-S_\alpha^2}{2}\right] \tag{24}$$

$$\beta_m = \bar{\beta} \cdot \exp\left[\frac{-S_\beta^2}{2}\right], \tag{25}$$

where

$$\bar{\alpha} = E_L\{\alpha(t)\} \tag{26}$$

$$\bar{\beta} = E_0\{\beta(t)\} \tag{27}$$

$$\sigma_\alpha^2 = E_L\{\alpha^2(t)\} - \bar{\alpha}^2 \tag{28}$$

and

$$\sigma_\beta^2 = E_0\{\beta^2(t)\} - \bar{\beta}^2. \tag{29}$$

$E_L\{\sim\}$ denotes a statistical (time) average under the condition that rain is falling on the radio path of length L, $E_0\{\sim\}$ denotes statistical (time) average under the condition that rain is falling at the location

of interest. In eqs. (27) and (29), we assume that the long-term, large-sample, conditional statistical average $\bar{\beta}$ and variance σ_β^2 are independent of position **s** in or near the radio path of interest. Therefore, we omit the position **s** argument in these equations.

By using eqs. (6), (22), (23), (24), and (25), we can derive formulas for the dependence of $S_\alpha(L)$ and $\alpha_m(L)$ on the radio path length L. The lengthy derivations are given in Appendix A. The results are

$$S_\alpha^2(L) = \ln P_0(L) \left\{1 + H(L) \left[\frac{\exp(S_\beta^2)}{P_0(0)} - 1 \right] \right\} \tag{30}$$

and

$$\alpha_m(L) = \beta_m \cdot L \cdot \frac{P_0(0)}{P_0(L)} \cdot \exp\left[\frac{S_\beta^2 - S_\alpha^2}{2} \right], \tag{31}$$

where

$$H(L) = \frac{1}{L^2} \int_0^L \int_0^L \frac{1}{Q(z)} \int_{Q(z)} \int \frac{1}{Q(z')} \int_{Q(z')} \int \psi_u(\mathbf{s}, \mathbf{s}') dv dv', \tag{32}$$

$$\psi_u(\mathbf{s}, \mathbf{s}') = \frac{1}{\sigma_{\beta u}^2} \{E_u[\beta(\mathbf{s}, t) \cdot \beta(\mathbf{s}', t)] - \bar{\beta}_u^2\} \tag{33}$$

is the spatial correlation coefficient[34,35] between $\beta(\mathbf{s}, t)$ and $\beta(\mathbf{s}', t)$, and

$$C_\beta(\mathbf{s}, \mathbf{s}') = \sigma_{\beta u}^2 \cdot \psi_u(\mathbf{s}, \mathbf{s}') \tag{34}$$

is the spatial covariance function[34,35] of $\beta(\mathbf{s}, t)$ and $\beta(\mathbf{s}', t)$, $E_u\{\sim\}$ denotes the unconditional statistical (time) average including both raining time and nonraining time, $\bar{\beta}_u$ and $\sigma_{\beta u}^2$ are the unconditional statistical mean and variance, respectively, of β as defined by eqs. (48) and (49) in Appendix A.

In eq. (32), the integration volume is the entire radio path (Fig. 1) and is a function of both path length L and wavelength λ. However, the spatial correlation coefficient ψ_u of (point) rain attenuation gradient β defined in eq. (33) is not a function of radio path length L.

If the random fluctuations of the βs were "coherent" along the entire radio path, then ψ_u, $H(L)$ and $P_0(L)/P_0(0)$ would be identically unity. Under such conditions, S_α would be identical to S_β and $\alpha_m(L)$ would be equal to $\beta_m \cdot L$ as expected. The complexity of path length dependencies of S_α and α_m in eqs. (30) to (34) is caused by the partially correlated, random fluctuations of βs at various points in the radio path.

We postulate ψ_u to have the functional dependence on distance

$$\psi_u = \frac{G(\lambda, \Delta V)}{[G^2(\lambda, \Delta V) + d^2]^{\frac{1}{2}}} \tag{35}$$

within the range of interest, where

$$d = |\mathbf{s} - \mathbf{s}'|$$

is the distance between the two observation points (s) and (s'), and $G(\lambda, \Delta V)$ is a characteristic distance at which $\psi_u = 1/\sqrt{2}$. The dependence of β on wavelength λ (Fig. 2) and sampling volume ΔV (Section 2.3) indicates that the characteristic distance G may also be a function of λ and ΔV. However, in Table I, the exponents η are all very close to unity for frequencies ranging from 11 to 60 GHz. This means β is approximately *linearly* proportional to R in this frequency range. Therefore, the characteristic distance G of β is approximately equal to that of point rain rate R and will not be very sensitive to frequency in the range from 11 to 60 GHz.

Substituting (35) into (32) and carrying out integrations over ρ' and ϕ' yield

$$H(L) = \frac{4\pi G}{L^2} \cdot \int_0^{\pi} d\theta \int_0^{L} \frac{dz}{Q(z)} \int_0^{h(z)} \rho d\rho \int_0^{L} \frac{F dz'}{Q(z')}, \tag{36}$$

where

$$F = \sqrt{W} - \sqrt{\zeta} + \rho \cos\theta \cdot \ln \frac{\sqrt{W} + h(z') - \rho \cos\theta}{\sqrt{\zeta} - \rho \cos\theta} \tag{37}$$

$$W = G^2 + (z - z')^2 + \rho^2 + h^2(z') - 2\rho h(z') \cos\theta \tag{38}$$

$$\zeta = G^2 + \rho^2 + (z - z')^2 \tag{39}$$

$$\theta = \phi - \phi'. \tag{40}$$

The remaining integrations can be carried out numerically by computer. The calculated $H(L)$ for $G = 0.75$, 1.5, and 3 km, respectively, are shown in Fig. 4.

Notice that the radius $h(z)$ of a "radio beam cross section" is on the order of several meters, whereas the characteristic distance G is on the order of kilometers (see Section 4.1). Therefore,

$$G \gg h(z) \tag{41}$$

for most radio paths at frequencies above 10 GHz. Imposing the condition (41) reduces the complicated integrations in eq. (36) to the simple result

$$H(L) \cong \frac{2G^2}{L^2} \left\{ \frac{L}{G} \ln \left[\frac{L}{G} + \sqrt{1 + \frac{L^2}{G^2}} \right] - \sqrt{1 + \frac{L^2}{G^2}} + 1 \right\}. \tag{42}$$

The differences in numerical values of $H(L)$ calculated by (36) and by approximation (42) are less than 0.1 percent within the range of interest. Notice that $H(L)$ is practically independent of wavelength

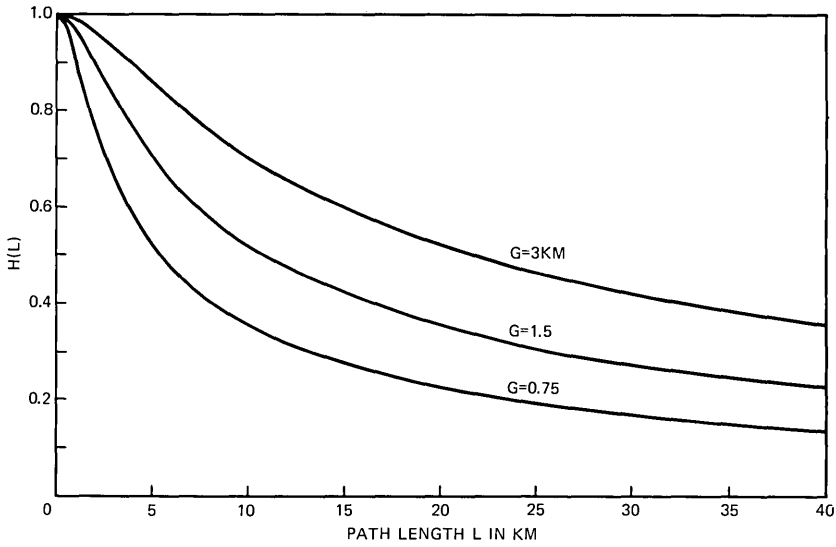


Fig. 4—Dependence of $H(L)$ on path length L and characteristic distance G.

λ because of condition (41). Therefore, we omit the wavelength specification on Fig. 4.

Thus, we have obtained all the necessary relationships. The procedure for calculating rain attenuation distributions from point rain rate distributions is summarized in the next section.

3.2 Procedure for calculating rain attenuation distribution

(*i*) Convert the measured distribution of point rain rate with $T \leqq 2$ min into the distribution of 2-s point rain rates by the conversion factor in Ref. 15.

(*ii*) If $P_0(0)$ with $T \leqq 1$ min is not available at the location of interest, use approximation (19) and the Weather Bureau hourly precipitation data to estimate P_0.

(*iii*) Estimate the lognormal parameters R_m and S_R of the 2-s point rain rate distribution by a least-squares approximation. This step is carried out by a computer iteration process to obtain the (R_m, S_R) pair that minimizes the differences (i.e., the sum of squares of errors) between the data points and the lognormal approximation.

(*iv*) Calculate β_m and S_β by formulas (15) and (16).

(*v*) Calculate $P_0(L)$, $\alpha_m(L)$, and $S_\alpha(L)$ by formulas (20), (30), and (31).

(*vi*) Substitute $P_0(L)$, $\alpha_m(L)$, $S_\alpha(L)$ into eq. (1) to give the attenuation distribution.

IV. COMPARISON OF CALCULATED RESULTS WITH EXPERIMENTAL DATA

The measured rain attenuations in many experiments contain not only the path rain attenuation but also the transmission loss owing to wet radomes. The presently available information is insufficient for *accurate* estimation of wet radome attenuation as a function of rain rate, wind direction, radome shape, size, material, and surface aging effects.

Based upon two measurements of wet radome attenuations discussed in Appendix D, we assume that a flat, vertical radome causes 1.5-dB attenuation during heavy rain. Therefore, 3-dB attenuation, caused by a pair of wet radomes, is added to the calculated path rain attenuation and the result is compared with the measured data utilizing such radomes. In some experiments, the flat radomes are slanted inward to further reduce wetting the radome surfaces. The attenuation caused by a pair of such radomes during rain is assumed* to be less than 3 dB. More detailed discussion of the radome problem is given in Appendix D.

4.1 Determination of characteristic distance G

From the 3-year (1971–1973) distribution of 1-min point rain rates measured at Merrimack Valley, Massachusetts,[36] the conversion factor in Ref. 15, and the Weather Bureau data,[32] we obtain the following approximate lognormal parameters of the distribution of 2-s point rain rate:

$$P_0(0) \simeq 3.3\%, \tag{43}$$

$$R_m \cong 1.23 \text{ mm/h}, \tag{44}$$

and

$$S_R \simeq 1.34. \tag{45}$$

Following the procedure outlined in Section 3.2, we use these parameters to calculate a family of rain attenuation distributions as a function of the distance parameter G. Figure 5 displays the results for an 18-GHz, 4.3-km path subject to Merrimack Valley rain and makes comparison with measured data (1971–1973) at the same location (Ref. 36 and Tables II and III). The radomes on this path are vertical and almost flat. The solid curves on Fig. 5 are calculated path rain attenuations plus assumed 3-dB radome attenuation. Figure 5 indicates that

$$G \cong 1.5 \text{ km} \tag{46}$$

provides good agreement; therefore, a 1.5-km characteristic distance[39,40]

* The slanted radomes may get wet during some heavy rains accompanied by strong wind gusts.

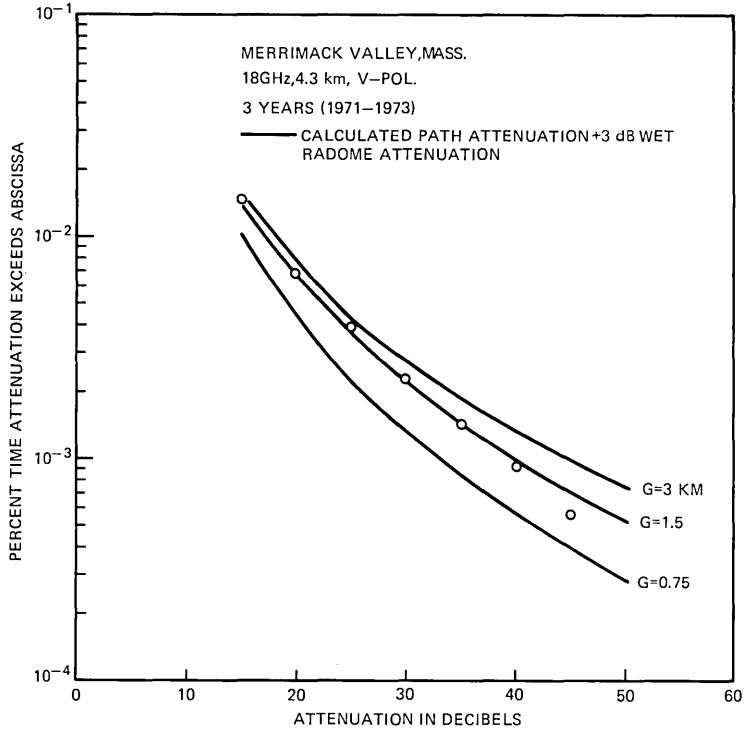


Fig. 5—Determination of characteristic distance G by comparing experimental data with calculated results (solid lines) using rain rate data in Fig. 10.

is used for the calculations and comparisons with other sets of data at other locations in the following sections.

From eqs. (35) and (46), it is easily shown that $d \leqq 15$ km for a spatial correlation coefficient, $\psi_u \geqq 0.1$. In other words, a "rain cell," based upon a definition of $\psi_u \geqq 0.1$ within the cell, has a typical spatial extent of 15 km. Obviously, the cell size depends on its definition.

4.2 Comparisons of calculated results with data in Georgia

Figure 6 compares the calculated result with data from a 5.1-km 17.7-GHz path at Palmetto, Georgia, measured during two 1-year periods (November 1970 through October 1971 and August 1973 through July 1974).* The radomes on this path are flat and canted inward. The calculated result is based upon the rain rate distribution measured by a tipping-bucket rain gauge at Palmetto in the same time

* The data from November 1971 to July 1973 are not used because of intermittent troubles in the rain gauge and the magnetic tape recorder.

Table II — Experimental data on rain attenuation distribution

No.	Authors	Ref No.	Freq (GHz)	Path Length (km)	Polarization	Fig. No.	Path Location	Time Base	Radome Shape Orientation	Rain Rate Data Used in Theoretical Calculation
1	Barnett, Bergmann, Lin, Pursley	30, 37	17.7	5.1	H	6	Rico-Palmetto, Ga.	11/70–10/71 8/73–7/74	Flat, slanted	No. 5 of Table 2
2	Pursley	37	11.6	42.	V	7	Atlanta-Palmetto, Ga.	8/73–7/74	Flat, almost vertical	No. 6 of Table 2
3	Lentz, Kenny	36	18.4	4.3	V	5	Merrimack Valley, Mass.	1971–1973	Flat, vertical	No. 7 of Table 2
4	Semplak	38	18.5	6.4	V	8	Holmdel, N.J.	1968–1969	Flat, slanted*	No. 8 of Table 2

* These radomes are shrouded by wooden rain shields.

Table III — Experimental data on point rain rate distribution

No.	Authors	Ref No.	Fig. No.	Location	Time Base	Rain Gauge Integration Time	Rain Gauge	Estimated Lognormal Parameters		
								Rm mm/hr	S_R	$P_0(0)$
1	Ruthroff, Bodtmann	15	11	Miami, Fla	1966–1970	1 min	Weighing Gauge	2.48	1.54	0.026
2	Jones, Sims	14, 20	11	Miami, Fla	8/57–8/58	1 min	Weighing Gauge	2.48	1.54	0.026
3	Jones, Sims	14, 20	10	Urbana, Ill	5/69–4/72	1 min	Weighing Gauge	1.1	1.47	0.033
4	Ruthroff, Bodtmann, Osborne	15	12	Atlanta, Ga.	1966–1970, 1973	1 min	Weighing Gauge	3.23	1.15	0.026
5	Lin		12	Palmetto, Ga.	11/70–10/71 8/73–7/74	1 min	Tipping Bucket	3.10	1.18	0.031
6	Lin		13	Palmetto, Ga.	8/73–7/74	1 min	Tipping Bucket	3.85	1.11	0.030
7	Lentz	36	10	Merrimack Valley, Mass.	1971–1973	10–90 s	Ruthroff's (Ref 2)	1.23	1.34	0.033
8	*		11	Holmdel, N.J.*	1968–1969	2 s	*	1.53	1.38	0.026
9	Norbury, White	16	10	Slough, England	1970–1971	10 s to 1 h	Special Dropper Gauge	0.42	1.4	0.044
10	Easterbrook, Turner	17	10	Southern England	5/61–5/62 1963	2–60 min	?			

* Point rain rate distribution from 1968 to 1969 at Holmdel, New Jersey, is estimated from rain attenuation data on short paths at Holmdel. See Section 4.3.

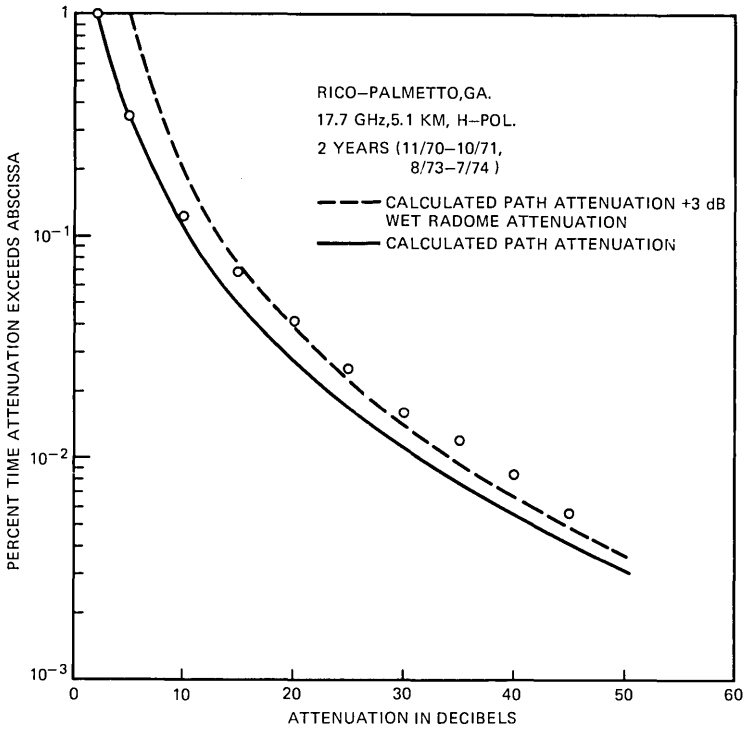


Fig. 6—Comparison of 17.7-GHz rain attenuation data at Palmetto, Georgia, with calculated result (solid line) using rain rate data in Fig. 12.

period. Figure 6 has two calculated attenuation curves, one without radome attenuation and another with 3-dB radome attenuation.

In the 11-GHz band, the path lengths of interest may be as long as 50 km. It is desirable to test the validity of the method for long paths. A preliminary comparison is shown in Fig. 7 for an 11-GHz, 42-km path between Atlanta and Palmetto, Georgia. The radomes on this path are flat and almost vertical. The attenuation and rain rate were observed during a 1-year period (August 1973 through July 1974). The calculated-plus-3-dB-radome-loss result is reasonably close to the data. In Figs. 7 and 13, notice that the measured attenuation and rain rate distributions are both somewhat higher than the lognormal approximations in the probability range from 10^{-2} to 4×10^{-2} percent time. We believe that these deviations are an artifact of the short observation time. A more critical test of this method for long paths awaits longer-term data.

4.3 Comparison of calculated result with data in New Jersey

Rain attenuation experiments on two paths (18 GHz, 6.4 km, and 30.9 GHz, 1.9 km) were carried out simultaneously in 1968 and 1969 at Holmdel, New Jersey.[38] However, local point rain rate distributions were not measured during this period. On the other hand, Bodtmann and Ruthroff[15] have suggested a method for relating point rain rate distribution to path rain attenuation distributions on short paths. Thus, by using the short-path attenuation data from the 30.9-GHz, 1.9-km path and Bodtmann and Ruthroff's method, the 2-year distribution of point rain rate at Holmdel was estimated as shown in Fig. 11. Based on this estimated rain rate distribution, we calculate the 2-year rain attenuation distribution on an 18-GHz, 6.4-km path and compare this with the measured result in Fig. 8. The radomes in these experiments are flat, slanted inward, and shrouded by substantial wooden rain shields. We believe that the wet radome attenuation, with

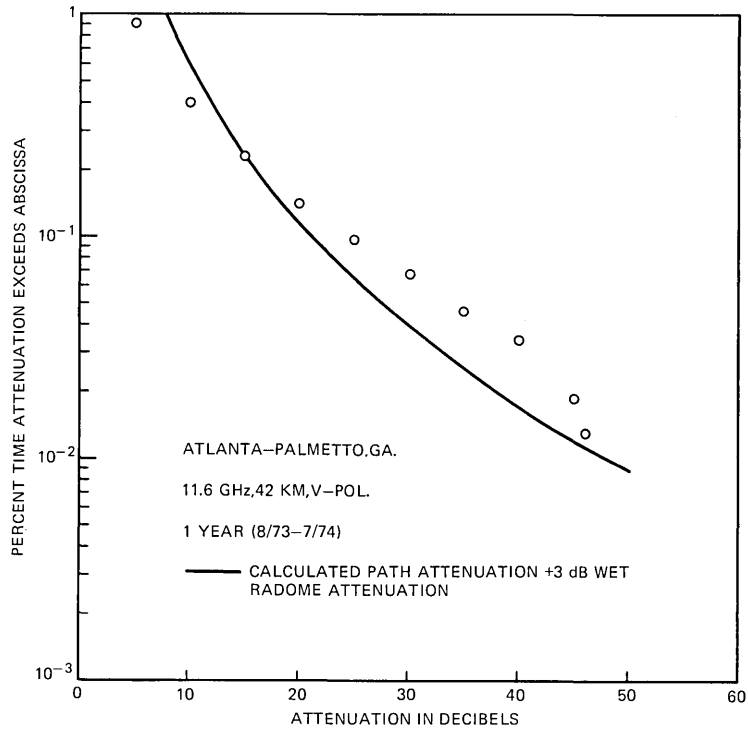


Fig. 7—Preliminary comparison of 1-year rain attenuation data from an 11-GHz, 42-km path in Georgia with calculated results (solid line) using rain rate data in Fig. 13.

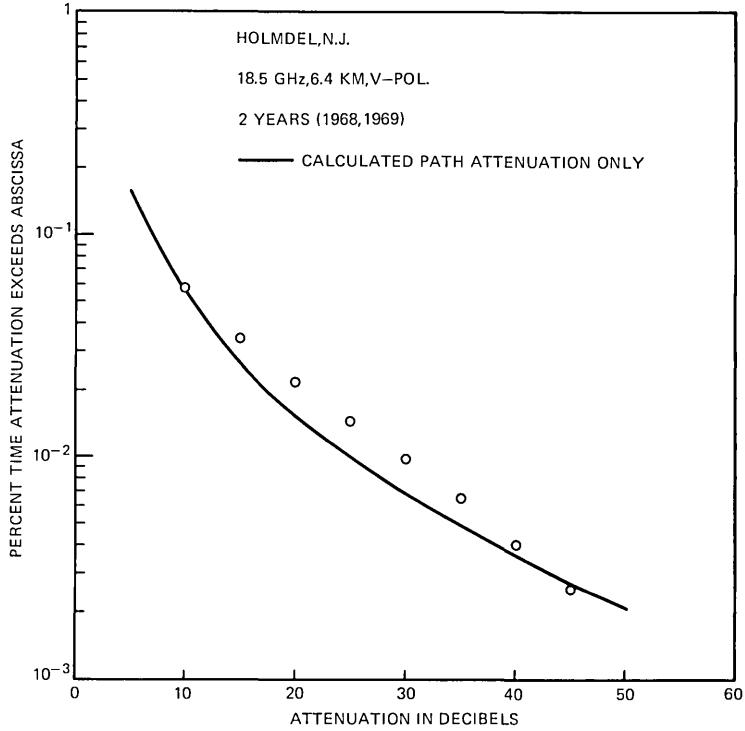


Fig. 8—Comparison of 18.5-GHz rain attenuation data in Holmdel, New Jersey, with calculated result (solid line) using rain rate data in Fig. 11.

such protection, is negligible. Therefore, the calculated curve in Fig. 8 contains only path rain attenuation.

Figures 6, 7, and 8 show that the calculated results agree reasonably well with measured data.

4.4 Excluded rain attenuation data

Many sets of rain attenuation data in the literature are not included in the comparisons in Sections 4.2 and 4.3 because of one or more of the following reasons.

(*i*) Many experiments used cone-shaped or hemispheric-shaped radomes. The transmission loss resulting from rain running on a pair of such radomes may vary from 0 to 14 dB depending on frequency, rain rate, and radome surface aging. The uncertainty in estimating such radome attenuation is too large for a meaningful comparison between the data and the calculated path rain attenuation distribution.

(*ii*) The antennas in the experiment are not covered by any radomes, exposing the antenna feeds and reflecting surfaces to rain, snow, and ice. No information is available for estimating the possible transmission loss from the wetting of these elements.

(*iii*) The published information does not specify whether the antennas are covered by radomes or not. The configuration of the radome, if used, is also unknown.

(*iv*) The polarization of the transmitted signal is unstated.

(*v*) The time base of the experiment is too short to yield long-term representative statistics.

(*vi*) No rain rate data with $T \leqq 2$ min is available at or near the location of the rain attenuation experiment.

V. OUTAGE ESTIMATION FOR 11- AND 18-GHz RADIO LINKS

For a constant transmitter output power, the dependence of fade margin $F_0(L)$ in dB on the path length L is

$$F_0(L) = F_0(L_0) - 20 \log_{10}\left(\frac{L}{L_0}\right) \text{dB}, \tag{47}$$

where L_0 is a reference repeater spacing and $F_0(L_0)$ is the corresponding reference fade margin. For 11- and 18-GHz radio, reasonable clear-day reference fade margins are

$$F_0 = 40 \text{ dB for 18 GHz at } L_0 = 4 \text{ km}$$

and

$$F_0 = 40 \text{ dB for 11 GHz at } L_0 = 40 \text{ km}.$$

A radio outage occurs when the path rain attenuation plus the wet radome attenuation exceeds the clear-day fade margin F_0. By substituting the fade margin (47) into the attenuation distribution (1), we can calculate the probability of radio outage per hop as a function of repeater spacing L. As an example, Fig. 9 shows the outage probabilities* for 11- and 18-GHz radio links in Atlanta, Georgia.

The wet radome attenuation A_R is assumed to be 3 dB in these calculations of outage probabilities.

VI. SOME QUALIFICATIONS

This section discusses some limitations, approximations, and assumptions in the theoretical calculation procedure, the data employed, and the calculated results.

* Multipath interference fading can also cause outages. An empirical formula for estimating the multipath-caused outage probability can be found in Ref. 41.

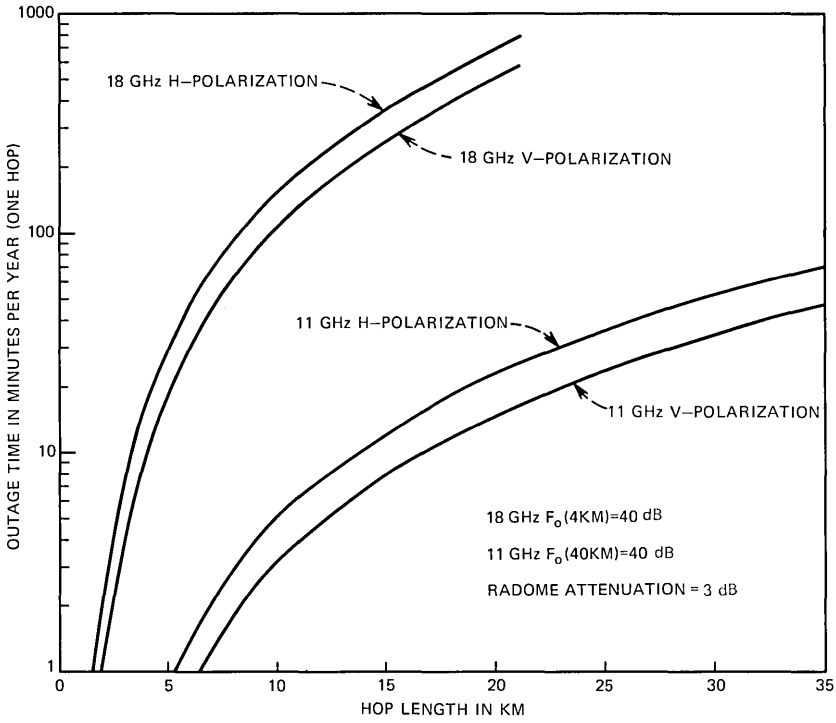


Fig. 9—Expected outage times of 11- and 18-GHz radio links as a function of hop length in Atlanta, Georgia.

6.1 Uncertainty in estimation of lognormal parameters P_0, R_m, and S_R

Some point rain rate measurements report only the heavy rain (e.g., $\geqq 30$ mm/h) portion of the distribution, neglecting the light rain statistics completely. Table III indicates that the median rain rates R_m at many locations are less than 4 mm/h. In other words, the major portion ($\cong$98 percent) of the distribution is missing, and accurate estimation of the statistical parameters R_m and S_R from the tail region ($\leqq 2$ percent) is difficult.

Furthermore, high rain rates (e.g., >140 mm/h) require a long observation time to yield representative, long-term statistics. The time bases of most available data may not be sufficient to yield stable statistics for these extreme rain rates. For example, at Newark, New Jersey, the 1-min point rain rate exceeded 180 mm/h *only once* in the 5-year period processed by Bodtmann and Ruthroff. To obtain reasonably stable statistics, we need a sample size much larger than 1. The omission of light-rain statistics together with the inherent instability of the extreme rain rate statistics causes considerable un-

certainty in the estimation of P_0, R_m, and S_R. This uncertainty can be reduced significantly if the light rain portion of the distribution is also measured and reported.

6.2 Path length dependence of P_0

The empirical formula (20) for the dependence of P_0 on path length L is obtained from the rain-gauge network data in Florida. The test of the applicability of this empirical formula to other locations and the improvement of this approximation will require further multiple rain gauge experiments at other locations.

6.3 Dependence of point rain rate distribution on rain gauge integration time

The dependence of the point rain rate distribution on rain-gauge integration time T has been obtained by Bodtmann and Ruthroff[15] from a 2-year experiment at Holmdel, New Jersey. Since a 2-year time base may not be sufficient to yield stable statistics for high rain rates, a longer time base may be needed to improve this empirical conversion factor. The applicability of this result to other geographic locations also remains to be verified.

6.4 Radome problem

In the comparisons of calculated and measured rain attenuation distributions, the wet radome attenuations are assumed values. To improve this approximation, a more systematic experimental study on the dependence of radome attenuation on rain rate is needed.

6.5 Anisotropic spatial correlation ψ

At some geographic locations, the squall lines of heavy rain may have a predominant orientation related to the predominant orientation of weather fronts.[42–44] This means the spatial correlation ψ may depend not only on the spacing but also on the orientation. However, the presently available information is not sufficient for a *quantitative* description of such an anisotropic correlation. Therefore, we use the isotropic correlation coefficient (35) throughout our theoretical calculations. Some of the difference between calculated and measured attenuation distributions may be caused by neglecting the anisotropy of the spatial correlation function.

VII. CONCLUSION

By using lognormal approximations, we have described a method for calculating rain attenuation distributions on microwave paths. The calculated results agree reasonably well with experimental data in Massachusetts, New Jersey, and Georgia. This procedure may

prove useful for the design of radio paths using frequencies above 10 GHz. To demonstrate the application, Fig. 9 shows the calculated outage probability as a function of repeater spacing for 11- and 18-GHz radio links in Georgia.

VIII. ACKNOWLEDGMENTS

I wish to express appreciation to K. A. Jarett who has helped in developing a computer program for the numerical calculations in this paper; to N. Levine, E. E. Muller, W. T. Barnett, D. C. Hogg, A. A. M. Saleh, and T. S. Chu for valuable comments and suggestions that greatly improved the consistency and precision of this analysis; to D. C. Hogg, R. A. Semplak, R. A. Desmond, A. E. Freeny, and J. D. Gabbe for rain rate and attenuation data at Holmdel; to W. T. Barnett, M. V. Pursley, and H. J. Bergmann for attenuation data in Georgia; to G. H. Lentz and J. J. Kenny for rain rate and attenuation data in Massachusetts; and to C. L. Ruthroff, W. F. Bodtmann, and A. L. Sims for rain rate data at many locations.

APPENDIX A

Derivation of Formulas Relating Rain Attenuation Distribution to Path Length

Since the random fluctuations of the attenuation gradient $\beta(s)$ at various positions in the radio path are partially correlated, we require the spatial covariance function of β to relate the variance of β to the variance of the path attenuation α. However, raining intervals at separated observation points are not always coincident. Hence, definition of the spatial covariance function for β requires a time base for all βs common to all observations. A natural common time base fulfilling this requirement is the total time, including both raining and nonraining intervals. This means the unconditional* statistical means and variances of β and α are also needed in this formulation. We therefore define

$$\overline{\beta_u} = E_u\{\beta(t)\} \tag{48}$$

and

$$\sigma_{\beta u}^2 = E_u\{\beta^2(t)\} - \bar{\beta}_u^2 \tag{49}$$

as the unconditional mean and variance, respectively, of β, where $E_u\{\sim\}$ denotes the (unconditional) statistical average including both raining and nonraining time. We assume that, on a long-term basis, $\bar{\beta}_u$ and $\sigma_{\beta u}^2$ are independent of position in or near the radio path of

* Fading caused by other atmospheric effects, such as multipath interference and "earth-bulge," is not treated in this paper. Therefore, the path rain attenuation α and the attenuation gradient β are taken to be identically zero during nonraining time.

interest; therefore, we omit the position argument(s) in eqs. (48) and (49). Similarly,

$$\bar{\alpha}_u = E_u\{\alpha(t)\} \tag{50}$$

$$\sigma_{\alpha u}^2 = E_u\{\alpha^2(t)\} - \bar{\alpha}_u^2 \tag{51}$$

are the unconditional mean and variance, respectively, of α.

Based upon the definitions (48) to (51) and the relationships (22) to (29), it can be shown that conditional and unconditional means and variances are related by

$$\bar{\beta} = \bar{\beta}_u / P_0(0) \tag{52}$$

$$\bar{\alpha} = \bar{\alpha}_u / P_0(L) \tag{53}$$

$$S_\beta^2 = \ln\left\{P_0(0)\left[1 + \frac{\sigma_{\beta u}^2}{\bar{\beta}_u^2}\right]\right\} \tag{54}$$

and

$$S_\alpha^2 = \ln\left\{P_0(L)\left[1 + \frac{\sigma_{\alpha u}^2}{\bar{\alpha}_u^2}\right]\right\}. \tag{55}$$

Obtaining the unconditional statistical averages of both sides of eq. (6) yields

$$\bar{\alpha}_u = \bar{\beta}_u \cdot L. \tag{56}$$

Substituting eqs. (6) and (56) into definition (51) yields

$$\begin{aligned}\sigma_{\alpha u}^2 &= E_u\left\{\int_0^L \frac{1}{Q(z)} \int_{Q(z)} \int\int_0^L \frac{1}{Q(z')} \right. \\ &\qquad \left. \times \int_{Q(z')} \int \beta(\mathbf{s}, t)\cdot\beta(\mathbf{s}', t)\cdot dv dv' \right\} - \bar{\beta}_u^2 \cdot L^2 \\ &= \int_0^L \frac{1}{Q(z)} \int_{Q(z)} \int\int_0^L \frac{1}{Q(z')} \\ &\qquad \times \int_{Q(z')} \int \{E_u[\beta(\mathbf{s}, t)\cdot\beta(\mathbf{s}', t)] - \bar{\beta}_u^2\}\cdot dv dv'. \end{aligned} \tag{57}$$

Let us define a spatial covariance function[34,35] $C_\beta(\mathbf{s}, \mathbf{s}')$ for $\beta(\mathbf{s}, t)$ and $\beta(\mathbf{s}', t)$ such that

$$C_\beta(\mathbf{s}, \mathbf{s}') = E_u\{\beta(\mathbf{s}, t)\cdot\beta(\mathbf{s}', t)\} - \bar{\beta}_u^2. \tag{58}$$

In other words,

$$\psi_u(\mathbf{s}, \mathbf{s}') = \frac{C_\beta(\mathbf{s}, \mathbf{s}')}{\sigma_{\beta u}^2} \tag{59}$$

is the (spatial) correlation coefficient[34,35] between $\beta(\mathbf{s}, t)$ and $\beta(\mathbf{s}', t)$. Substituting definitions (58) and (59) into (57) yields

$$\sigma_{\alpha u}^2 = \sigma_{\beta u}^2 \cdot L^2 \cdot H(L), \tag{60}$$

where

$$H(L) = \frac{1}{L^2}\int_0^L \frac{1}{Q(z)}\int_{Q(z)}\int\int_0^L \frac{1}{Q(z')}\int_{Q(z')}\int \psi_u(\mathrm{s}, \mathrm{s}')dvdv'. \quad (61)$$

Substituting (56) and (60) into (55) gives

$$S_\alpha^2(L) = \ln P_0(L)\cdot\left[1 + H(L)\cdot\frac{\sigma_{\beta u}^2}{\bar{\beta}_u^2}\right]. \quad (62)$$

Combining (54) and (62) gives

$$S_\alpha^2(L) = \ln P_0(L)\cdot\left\{1 + H(L)\left[\frac{\exp(S_\beta^2)}{P_0(0)} - 1\right]\right\}. \quad (63)$$

Combining eqs. (24), (25), (52), (53), and (56) gives

$$\alpha_m(L) = \beta_m\cdot L\cdot\frac{P_0(0)}{P_0(L)}\cdot\exp\left[\frac{S_\beta^2 - S_\alpha^2}{2}\right]. \quad (64)$$

This completes the derivation for $S_\alpha^2(L)$ and $\alpha_m(L)$.

APPENDIX B

Lognormal Distribution of Point Rain Rate

Figures 10 to 13 display the distributions of 2-s point rain rate observed in Miami, Florida; Urbana, Illinois; Atlanta and Palmetto, Georgia; Merrimack Valley, Massachusetts; Holmdel, New Jersey; and Southern England. The time bases range from 1 to 6 years. It can be seen that these distributions of 2-s point rain rate are very close to the lognormal approximation in the range below 100 mm/h. The rain rates beyond 100 mm/h are generally separated by more than 3 sigma from the median, and constitute the tail of the lognormal distribution. A very long observation time (e.g., more than 20 years) is necessary to obtain stable statistics of extreme rain rates beyond 100 mm/h.[45–48] Since the time bases of the data in Figs. 10 to 13 are much less than 20 years, the deviations of the data from the lognormal distributions in the tails are not unexpected.

The rain gauge integration time T in the original data range from 1.5 s to 2 min, depending upon the source. As discussed in Section 2.3, the appropriate integration time T, corresponding to 1-m^3 sampling volume in our formulation, is about 2 s. From 2-year experimental data at Holmdel, New Jersey, Bodtmann and Ruthroff[15] have obtained an empirical relationship for the dependence of point rain rate distributions on rain gauge integration t time in the range

$$1.5\ \mathrm{s} \leqq T \leqq 120\ \mathrm{s}. \quad (65)$$

This empirical result enables us to convert the original data into the 2-s point rain rate distributions shown in Figs. 10 to 13.

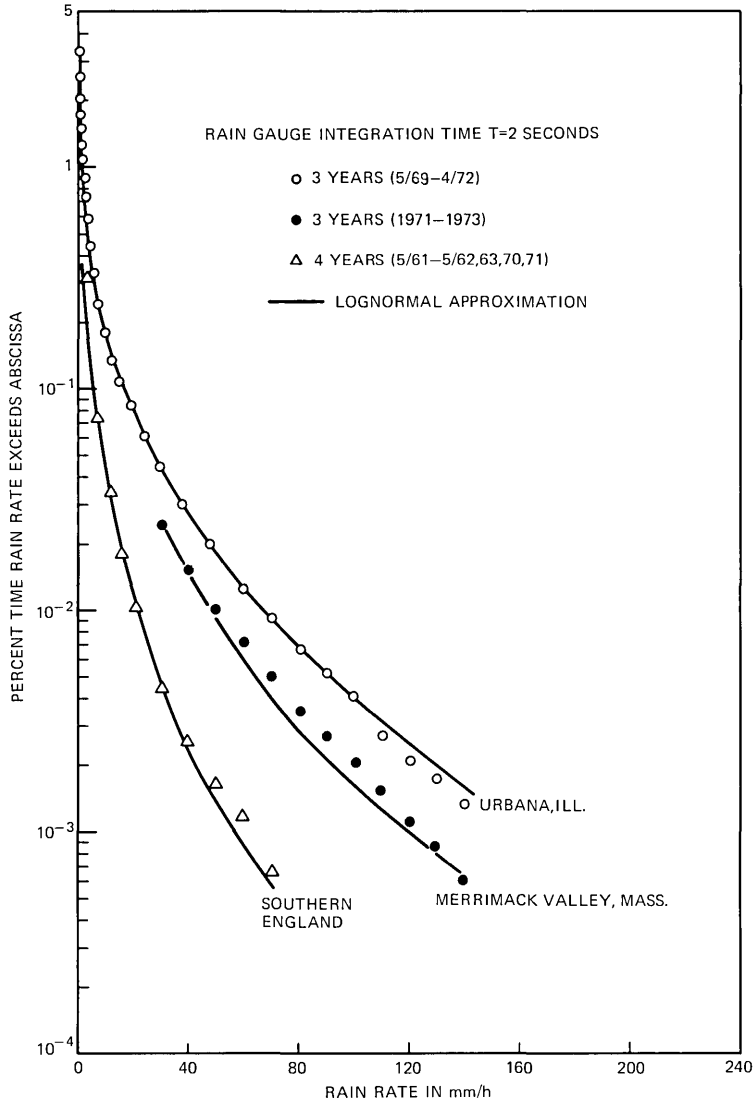


Fig. 10—Lognormal distribution of 2-s point rain rate at Urbana, Illinois; Merrimack Valley, Massachusetts; and Southern England.

APPENDIX C

Derivation for Path Length Dependence of $P_0(L)$

Let $P_0(L_1)$ and $P_0(L_1 + \Delta L)$ be the probabilities that rain falls on the radio path with length L_1 and an extended path of length $L_1 + \Delta L$, where ΔL is a small incremental length. The relation between $P_0(L_1)$

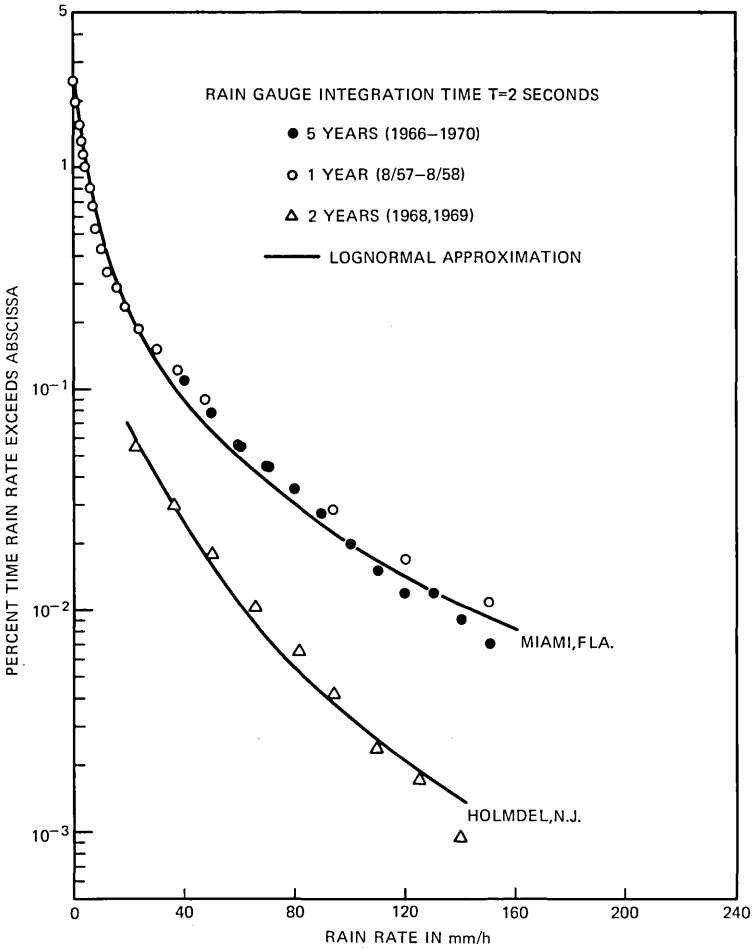


Fig. 11—Lognormal distribution of 2-s point rain rate at Miami, Florida and Holmdel, New Jersey.

and $P_0(L_1 + \Delta L)$ can be written as

$$P_0(L_1 + \Delta L) = P_0(L_1) + \Delta P_0(\Delta L), \tag{66}$$

where $\Delta P_0(\Delta L)$ is the incremental probability of rainfall associated with the incremental length ΔL. This incremental probability can be written as

$$\Delta P_0(\Delta L) = \Delta P_0(\Delta L | \text{no rain for } 0 \leqq L \leqq L_1) \cdot P(\text{no rain for } 0 \leqq L \leqq L_1), \tag{67}$$

where $\Delta P_0(\Delta L | \text{no rain for } 0 \leqq L \leqq L_1)$ is the (incremental) probability that rain falls on the incremental length ΔL under the condi-

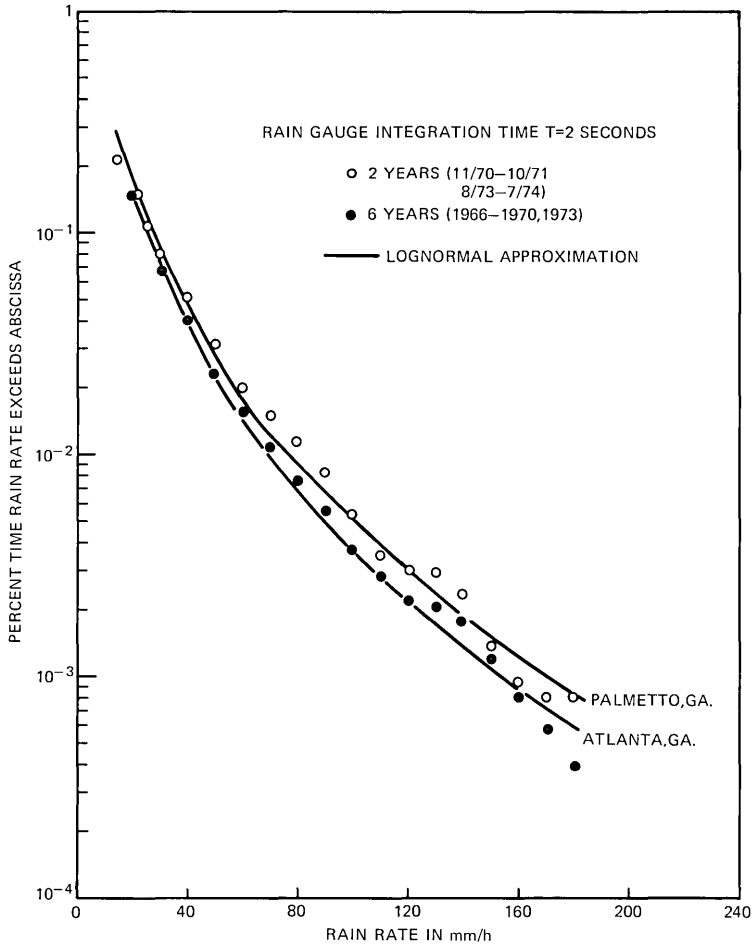


Fig. 12—Lognormal distribution of 2-s point rain rate at Palmetto and Atlanta, Georgia.

tion that rain is not falling on the path L_1. This condition is required because rainfall on L_1 and ΔL are partially correlated.

We assume that

$$\Delta P_0(\Delta L | \text{no rain for } 0 \leqq L \leqq L_1) \propto \Delta L. \tag{68}$$

The justification for this assumption is

(*i*) $P_0(L)$ is expected to be a smooth, continuous function of L, i.e.,

$$\lim_{\Delta L \to 0} P_0(L_1 + \Delta L) = P_0(L_1), \tag{69}$$

$$\lim_{\Delta L \to 0} \Delta P_0(\Delta L | \text{no rain for } 0 \leqq L \leqq L_1) = 0. \tag{70}$$

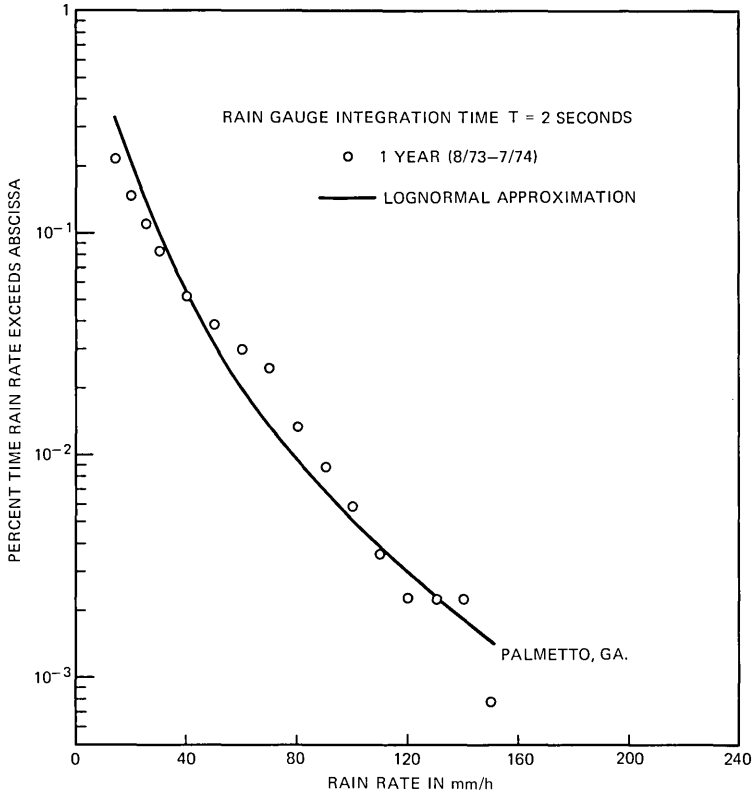


Fig. 13—Lognormal distribution of 2-s point rain rate at Palmetto, Georgia.

(*ii*) The rain-gauge network data (Fig. 3) indicate that the slope of $P_0(L)$ is not zero, i.e.,

$$\frac{\Delta P_0}{\Delta L} \neq 0, \qquad \text{for } L \neq 0.$$

Let b be the proportional parameter in assumption (68). Then

$$\Delta P_0(\Delta L | \text{no rain for } 0 \leqq L \leqq L_1) = b(L) \cdot \Delta L. \tag{71}$$

The unknown proportional parameter $b(L)$ will be determined from rain-gauge network data.

By definition,

$$P(\text{no rain for } 0 \leqq L \leqq L_1) = 1 - P_0(L_1). \tag{72}$$

Combining eqs. (67), (71), and (72) yields

$$\Delta P_0(\Delta L) = b(L) \cdot \Delta L \cdot [1 - P_0(L_1)], \tag{73}$$

from which

$$\frac{dP_0(L)/dL}{1 - P_0(L)} = b(L). \tag{74}$$

Integrating (74) yields

$$P_0(L) = 1 - c \exp\left\{-\int_0^L b(z)dz\right\}, \tag{75}$$

where c is an unknown constant to be determined by the condition

$$\lim_{L \to 0} P_0(L) = P_0(0). \tag{76}$$

Applying condition (76) to (75) gives

$$P_0(L) = 1 - [1 - P_0(0)] \exp\left\{-\int_0^L b(z)dz\right\}. \tag{77}$$

Since the rain-gauge network data yield $P_0(L)$ at only a few discrete distances, we need the quantized version

$$\frac{\Delta P_0(L_i)/\Delta L}{1 - P_0(L_i)} \cong b(L_i), \qquad i = 1, 2, 3, \cdots, n, \tag{78}$$

of eq. (74) for estimating $b(L)$. By using eq. (78) and the rain-gauge network data in Fig. 3, we can calculate $b(L_i)$ at several discrete points. From these results, we find that $b(L)$ can be approximately described by the empirical formula

$$b(L) \cong \frac{0.028L}{21.5 + L^2} \text{ (km}^{-1}\text{)}. \tag{79}$$

Substituting (79) into (77) and carrying out the integration yields

$$P_0(L) \cong 1 - \frac{1 - P_0(0)}{\left[1 + \dfrac{L^2}{21.5}\right]^{0.014}}, \tag{80}$$

which is the same as (20).

Figure 3 shows that the empirical result (80) is reasonably close to all the data points measured by the Florida rain-gauge network. Admittedly, eqs. (79) and (80) are empirical. Further theoretical work and multiple rain-gauge experiments are needed to improve these approximations.

APPENDIX D

Transmission Loss Due to Wet Radomes

A 20-GHz experiment by Anderson[49] on a section of a 90-ft diameter radome, pertaining to an earth satellite radio link, indicated a transmission loss of 2 to 3 dB at 10 mm/h rain rate when the radome

was new. However, after 6 months of weathering, the transmission loss increased to 8 dB at 10 mm/h rain rate. A 4-GHz experiment[50] on an earth-satellite radio link indicated a 3-dB transmission loss resulting from the wet radome. These experimental data agree reasonably well with theoretical calculations[51–54] for rain rates $\leqq 10$ mm/h, assuming laminar water flow on hemispherical radome surface pointing towards the zenith.

For typical 11- and 18-GHz terrestrial radio paths, the transmission losses from wet radomes are expected to be smaller than those of earth-satellite radio links because of the smaller radome size, different radome shape, and orientation. However, the terrestrial radio passes through a pair of radomes on each link; therefore, the contribution of wet radome loss to the total path attenuation may not be negligible. Theoretical calculation of wet radome attenuation pertaining to terrestrial radio links is not available at the present time because of the difficulty in calculating the nonuniform thickness of the water film. A semiquantitative experiment[55] was carried out on the 12.2-GHz radio link between Murray Hill and Crawford Hill, New Jersey (22 miles). The 10-ft dish antenna was covered by a cone-shaped radome that was made of resin-coated fiberglass and, at 10 years of age, was well weathered. Water was sprayed on the radome-covered antenna by a manually controlled sprinkler. The results indicated that a uniform light sprinkle caused approximately 2.5-dB attenuation, whereas a very heavy spray (maximum stream of water) caused between 4- and 7-dB attenuation. After the spray was turned off, 2 to 3 minutes elapsed before the signal recovered to within 1 dB of its nonfaded level. The residual wet radome attenuation is estimated to be 0.5 dB.

On an 18-GHz, 4.3-km path at Merrimack Valley, Massachusetts, Kenny[36] has also observed a residual wet radome attenuation of 0.75 dB (i.e., 1.5 dB for two radomes).

APPENDIX E

List of Symbols and Their Definitions

A_g	Collecting aperture of a rain gauge.
A_R	Attenuation by two wet radomes on a radio link.
$C_\beta(\mathbf{s}, \mathbf{s}')$	Spatial covariance function of $\beta(\mathbf{s}, t)$ and $\beta(\mathbf{s}', t)$ as defined by eq. (58).
dv	$= \rho d\rho d\phi dz$.
$E_L\{\sim\}$	Conditional statistical (time) average under the condition that the point rain rates, in the radio path of length L (see Fig. 1), are not all zero.

$E_0\{\sim\}$	Conditional statistical (time) average under the condition that the point rain rate (defined in Section 2.3) is not zero at the position of interest.
$E_u\{\sim\}$	Unconditional statistical (time) average including both raining and nonraining time.
erfc($\sim$)	Complementary error function.
F_0	Fade margin of radio links.
G	Characteristic distance defined in eq. (35). See also Section 4.1.
$h(z)$	Radius of the circular cross section of radio beam at a distance z from the transmitter. See Fig. 1 and eq. (3).
$H(L)$	Defined by eqs. (32) and (36).
H-Pol	Horizontal polarization.
L	The path length of a radio link; see Fig. 1.
L_0	A reference repeater spacing defined by eq. (47).
ln ($\sim$)	Natural logarithm.
$P(\alpha \geqq A)$	Probability that rain attenuation α exceeds A.
$P(R \geqq r)$	Probability that rain rate R exceeds r.
$P(\beta \geqq B)$	Probability that attenuation gradient β exceeds B.
$P_0(L)$	Probability that rain is falling on a radio link of length L.
$P_0(0)$	$= \lim_{L \to 0} P_0(L)$; the probability that rain is falling at the position of interest.
$Q(z)$	$= \pi h^2(z)$; area of the circular cross section of radio beam at a distance z from the transmitter. See eqs. (3) and (4).
$R(\mathbf{s}, t)$	Point rain rate measured by a 1-m^3 sampling volume located at $\mathbf{s}$.
R_m	Median value of the point rain rate R during raining time.
$\mathbf{s}$	A vector to denote the position (ρ, ϕ, z).
S_α	Standard deviation of ln α during raining time.
S_R	Standard deviation of ln R during raining time.
S_β	Standard deviation of ln β during raining time.
T	Integration time of rain gauge.
t	Time.
$T_{\Delta V}$	Defined in eq. (10).
V-Pol	Vertical polarization.
V_R	Average falling velocity of raindrops.
ΔV	Incremental sampling volume for measurement of point rain rate. See Section 2.3.
z	Distance from the radio transmitter (Fig. 1).
α	Rain attenuation in decibels.
α_m	Median value of α during raining time.
$\bar{\alpha}$	Mean value of α during raining time.

$\bar{\alpha}_u$	Unconditional statistical mean of α as defined by eq. (50).
β	Point rain attenuation gradient measured in dB/km by a 1-m^3 sampling volume as discussed in Section 2.3.
β_m	Median value of β during raining time.
$\bar{\beta}$	Mean value of β during raining time.
$\beta_q(z, t)$	Average of $\beta(\mathbf{s}, t)$ over the circular cross section Q of the radio beam. See eqs. (5) and (7).
$\bar{\beta}_u$	Unconditional statistical mean of β as defined by eq. (48).
θ	$= \phi - \phi'$.
γ	A parameter defined by eq. (12) relating point rain attenuation gradient β and point rain rate R.
λ	Radio wavelength.
η	A parameter defined by eq. (12) relating point rain attenuation gradient β and point rain rate R.
σ_α	Standard deviation of α during raining time.
$\sigma_{\alpha u}$	Unconditional standard deviation of α as defined by eq. (51).
σ_β	Standard deviation of β during raining time.
$\sigma_{\beta u}$	Unconditional standard deviation of β as defined by eq. (49).
ρ	Radial distance from the z-axis in the cylindrical coordinate system in Fig. 1.
ϕ	Angle in the cylindrical coordinate system in Fig. 1.
$\psi_u(\mathbf{s}, \mathbf{s}')$	Correlation coefficient between $\beta(\mathbf{s}, t)$ and $\beta(\mathbf{s}', t)$ as defined by eqs. (33) and (59).

REFERENCES

1. A. E. Freeny and J. D. Gabbe, "A Statistical Description of Intense Rainfall," B.S.T.J., *48*, No. 6 (July–August 1969), pp. 1789–1851.
2. A. E. Freeny and J. D. Gabbe, private communication.
3. D. C. Hogg, "Statistics on Attenuation of Microwaves by Intense Rain," B.S.T.J., *48*, No. 9 (November 1969), pp. 2949–2962.
4. C. R. Stracca, "Propagation Tests of 11 GHz and 18 GHz on Two Paths of Difference Length," Alta Frequenza, Italy, *38*, 1969, pp. 345–360.
5. D. C. Hogg, "Path Diversity in Propagation of Millimeter Waves Through Rain," IEEE Trans. Ant. Prop., *AP-15*, No. 3 (May 1967), pp. 410–415.
6. S. H. Lin, "Statistical Behavior of Rain Attenuation," B.S.T.J., *52*, No. 4 (April 1973), pp. 557–581.
7. J. A. Morrison, M. J. Cross, and T. S. Chu, "Rain-Induced Differential Attenuation and Differential Phase Shift at Microwave Frequencies," B.S.T.J., *52*, No. 4 (April 1973), pp. 599–604.
8. T. Oguchi, "Attenuation and Phase Rotation of Radio Waves due to Rain: Calculation at 19.3 and 34.8 GHz," Radio Science, *8*, No. 1 (January 1973), pp. 31–38.
9. R. G. Medhurst, "Rainfall Attenuation of Centimeter Waves: Comparison of Theory and Measurement," IEEE Trans. Ant. Prop., *AP-13*, No. 4 (July 1965), pp. 550–564.
10. D. E. Setzer, "Computed Transmission Through Rain at Microwave and Visible Frequencies," B.S.T.J., *49*, No. 8 (October 1970), pp. 1873–1892.

11. J. W. Ryde, "Echo Intensity and Attenuation due to Clouds, Rain, Hail; Sand and Dust Storms at Centimeter Wavelength," Report 7831, General Electric Company Research Laboratories, Wembly, England, October 1941.
12. D. Ryde and J. W. Ryde, "Attenuation of Centimeter Waves by Rain, Hail and Clouds," Report 8516, General Electric Company Research Laboratories, Wembly, England, August 1944.
13. D. Ryde and J. W. Ryde, "Attenuation of Centimeter and Millimeter Waves by Rain, Hail, Fogs and Clouds," Report 8670, General Electric Company Research Laboratories, Wembly, England, May 1945.
14. D. M. A. Jones and A. L. Sims, "Climatology of Instantaneous Precipitation Rates," Illinois State Water Survey at the University of Illinois, Urbana, Illinois. Project No. 8624, Final Report, December 1971.
15. W. F. Bodtmann and C. L. Ruthroff, "Rain Attenuation on Short Radio Paths: Theory, Experiment, and Design," B.S.T.J., *53*, No. 7 (September 1974), pp. 1329–1349.
16. J. R. Norbury and W. J. K. White, "Point Rainfall Rate Measurements at Slogh, U.K.," Conference on Propagation of Radio Waves at Frequencies above 10 GHz, April 10–13, 1973, London, England, Conference Records, pp. 190–196 and IEE (London) Conference Publication Number 98.
17. B. J. Easterbrook and D. Turner, "Predication of Attenuation by Rainfall in the 10.7–11 GHz Communication Band," Proc IEE (London), *114*, No. 5, (May 1967), pp. 557–565.
18. K. Funakawa and J. Kato, "Experimental Studies of Propagational Characteristics of 8.6-m Wave on the 24-km Path," J. Radio Research Labs (Japan), *9*, No. 45 (September 1962), pp. 351–367.
19. K. Moritta and I. Higuti, "Statistical Studies on Electromagnetic Wave Attenuation Due to Rain," Review of the Electrical Communication Laboratories (Japan), *19*, No. 7–8 (July–August 1971), pp. 798–892.
20. A. L. Sims and D. M. A. Jones, "Climatology of Instantaneous Precipitation Rates," Illinois State Water Survey at the University of Illinois, Urbana, Illinois. Project No. 8624, March 1973.
21. C. L. Ruthroff, "Rain Attenuation and Radio Path Design," B.S.T.J., *49*, No. 1 (January 1970), pp. 121–135.
22. E. A. Mueller and A. L. Sims, "Investigation on the Quantitative Determination of Point and Areal Precipitation by Radar Echo Measurement," Technical Report ECOM-00032-F, Illinois State Water Survey at the University of Illinois, Urbana, Illinois, December 1966.
23. D. C. Hogg, "How the Rain Falls," unpublished work.
24. R. A. Semplak and R. H. Turrin, "Some Measurements of Attenuation by Rainfall at 18.5 GHz," B.S.T.J., *48*, No. 6 (July–August 1969), pp. 1767–1787 and Figure 11.
25. E. A. Mueller and A. L. Sims, "The Influence of Sampling Volume on Raindrop Size Spectra," Proc. Twelfth Conference on Radar Meterology, American Meteorological Society, October 1966, pp. 135–141.
26. J. O. Laws and D. A. Parsons, "The Relation of Rain Drop Size to Intensity," Trans. Am. Geophysical Union, *25*, 1943, pp. 452–460.
27. J. A. Morrison and M. J. Cross, "Scattering of a Plane Electromagnetic Wave by Axisymmetric Raindrops," B.S.T.J., *53*, No. 6 (July–August 1974), pp. 955–1020.
28. T. S. Chu, private communications.
29. T. S. Chu, "Rain Induced Cross Polarization at Centimeter and Millimeter Wavelength," B.S.T.J., *53*, No. 8 (October 1974), pp. 1557–1579.
30. W. T. Barnett, "Some Experimental Results on 18 GHz Propagation," The 1972 National Telecommunications Conference, December 4–6, 1972, IEEE, Houston, Texas, Conference Record, pp. 10E-1–10E-4 and IEEE Publication 72 CHO 601-5-NTC.
31. R. A. Semplak, "Effect of Oblate Raindrops on Attenuation at 30.9 GHz," Radio Science, *5*, No. 3 (March 1970), pp. 559–564.
32. Local Climatological Data, U. S. Department of Commerce, Environmental Science Services Administration, National Weather Records Center, Ashville, North Carolina, 28801. Also available from Superintendent of Documents, U. S. Government Printing Office, Washington, D. C. 20402.
33. J. Aitchison and J. A. C. Brown, *The Lognormal Distribution*, London: Cambridge University Press, 1957.

34. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York: McGraw-Hill, 1965, pp. 210 and 282.
35. P. Beckmann, *Probability in Communication Engineering*, New York: Harcourt, Brace and World, 1967, pp. 88 and 204.
36. G. H. Lentz and J. J. Kenny, private communication.
37. W. T. Barnett, H. J. Bergmann, and M. V. Pursley, private communication.
38. R. A. Semplak, "The Influence of Heavy Rainfall on Attenuation at 18.5 GHz and 30.9 GHz," IEEE Trans. Ant. Prop., *AP-18*, No. 4 (July 1970), pp. 507–511.
39. B. N. Harden, J. R. Norbury, and W. J. K. White, "Model of Intense Convective Rain Cells For Estimating Attenuation on Terrestrial Millimetric Radio Links," Elec. Letters, *10*, No. 23 (November 1974), pp. 483–484.
40. D. C. Hogg, "Intensity and Extent of Rain on Earth-Space Paths," Nature, *243*, June 8, 1973, pp. 337–338.
41. W. T. Barnett, "Multipath Propagation at 4, 6, and 11 GHz," B.S.T.J., *51*, No. 2 (February 1972), pp. 321–361.
42. R. R. Braham, Jr., and H. R. Byers, *The Thunderstorms*, Report of the Thunderstorm Project, Government Printing Office, Washington, D. C., 1949, Chapter VIII Squall Lines.
43. R. J. Boucher and R. Wexler, "The Motion and Predictability of Precipitation Lines," J. Meteorology, *18*, No. 2 (April 1961), pp. 160–171.
44. D. A. Gray, "Earth-Space Path Diversity Dependence on Base Line Orientation," 1973 International IEEE/G-AP Symposium and URSI Meeting, Boulder, Colorado, August 1973, Symposium Record pp. 366–369.
45. "Rainfall Intensity-Duration-Frequency Curves for Selected Stations in the United States, Alaska, Hawaiian Islands, and Puerto Rico," Weather Bureau, U. S. Department of Commerce, Tech. Paper 25, December 1955.
46. L. H. Seamon and G. S. Bartlett, "Climatological Extremes," Weatherwise, *9*, 1956, pp. 194–213.
47. A. E. Cole, R. J. Donaldson, R. Dyer, A. J. Kantor and R. A. Skrivanek, "Precipitation and Clouds, a Revision of Chapter 5, Handbook of Geophysics and Space Environments," AFCRL-69-0487, Air Force Surveys in Geophysics, No. 212, Office of Aerospace Research, U.S. Air Force Cambridge Research Laboratories, Bedford, Massachusetts, November 1969.
48. W. Y. S. Chen, private communication.
49. I. Anderson, "Measurements of 20 GHz Transmission Through a Wet Radome," 1973 International IEEE/G-AP Symposium and URSI Meeting, Boulder, Colorado, August 1973, Symposium Record, pp. 239–240.
50. A. J. Giger, "4-gc Transmission Degradation Due to Rain at Andover, Maine, Satellite Station," B.S.T.J., *44*, No. 7 (September 1965), pp. 1528–1533.
51. A. Cohen and A. Smolski, "The Effect of Rain on Satellite Communications Earth Terminal Rigid Radomes," Microwave J., *9*, No. 9 (September 1966), pp. 111–121.
52. D. Gibble, "Effects of Rain on Transmission Performance of a Satellite Communication System," IEEE International Convention Record, Part VI, March 1964, p. 52.
53. B. Blevis, "Losses Due to Rain on Radomes and Antenna Reflecting Surfaces," IEEE Trans. Ant. Prop., *AP-13*, No. 1 (January 1965), pp. 175–176.
54. B. Blevis, "Rain Effects on Radomes and Antenna Reflectors," Proc. IEE Conference on Large Steerable Aerials, London, 1966, pp. 148–152.
55. H. J. Bergmann and S. H. Lin, "Measurements of 12 GHz Transmission Loss Through a Wet Radome," unpublished work.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

A Geometric Derivation of Forney's Upper Bound

By J. E. MAZO

(Manuscript received January 17, 1975)

Effective analyses of performance for detection schemes that optimally decode digital data in the presence of intersymbol interference have been slow in coming. Recently, however, Forney has given an upper bound on the bit error probability for maximum-likelihood sequence estimation. Starting from a standard geometrical framework, we give a much simplified derivation of this upper bound. Our derivation places the validity of this important bound more in evidence in that the concepts of whitened matched filter and error event are not introduced.

Let a_j, $j = 1, 2, \cdots, N$, be independent, equilikely binary random variables taking values ± 1. Data transmission usually involves estimating the a_i from a pulse sequence of the form

$$\sum_{j=1}^{N} a_j h(t - jT), \quad -\infty < t < \infty, \tag{1}$$

which is observed in white gaussian noise of (two-sided) spectral density $N_0/2$. In (1), the minimum assumption put on the pulse waveform $h(t)$ is that it be L_2. One possible detection procedure is to decide, on the basis of the received noisy signal, which one of the 2^N equilikely signals given in (1) was "most likely" (maximum-likelihood sequence detection) and use the sequence $\{a_n\}$ which is associated with that sequence as the detected symbols. As N grows large, the probability of deciding incorrectly on the sequence approaches unity; however, the real question revolves about the *bit*-error probability for maximum-likelihood *sequence* detection. Important work on this problem was done recently by Forney,[1,2] who showed that, under certain conditions on $h(t)$, the bit-error probability for large signal-to-noise ratios goes

exponentially to zero as

$$P_e \approx \text{(coeff.)} \exp\left(-\frac{d_{\min}^2}{4N_0}\right), \quad N_0 \to 0, \tag{2}$$

where $d_{\min}^2$ is the minimum distance between the signal sequences in (1); i.e., if we add a superscript to distinguish among sequences thus

$$s^{(i)}(t) = \sum_{j=1}^{N} a_j^{(i)} h(t - jT), \quad -\infty < t < \infty \quad i = 1, 2, \cdots, 2^N, \tag{3}$$

then

$$d_{\min}^2 = \lim_{N \to \infty} \min_{\substack{i,k \\ i \neq k}} \int_{-\infty}^{\infty} |s^{(i)}(t) - s^{(k)}(t)|^2 dt. \tag{4}$$

Forney's demonstration consists of two steps. First, a lower bound on P_e of the form (2) is established, valid for any $h(t)$. Second, if $H(\omega)$ denotes the Fourier transform of $h(t)$ and

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 e^{ik\omega} d\omega = 0 \text{ for integer } k, \ |k| > \nu, \ \nu \text{ integer}, \tag{5}$$

then an upper bound for P_e can be given which is convergent for large signal-to-noise ratios and which, furthermore, also has an asymptotic form given by (2).

In our opinion, Forney's discussion of the upper bound is sufficiently complicated that some question remains as to how firmly the result is established. We shall give a much simpler derivation, but first let us review the situation when $\nu = 0$, i.e., when there is no intersymbol interference. Using the reduction to the standard geometrical picture,[3] the signal points (sequences) received in the absence of noise are as shown in Fig. 1. These signal points are to be regarded as being perturbed by spherically symmetric, N-dimensional, zero mean gaussian noise; the variance of each component of the noise is $N_0/2$. Thus any point in the N-dimensional space may be received and the maximum-likelihood decoder chooses the unperturbed sequence nearest to the received point as the transmitted one. The decoding regions are shown in Fig. 2 by dashed lines, and are labeled $R_{i,j}$ in an obvious way.

Now assume we transmit (1, 1) and ask for the probability that the first bit is in error. This is the same as the probability that the received signal point is in $R_{-1,1} \cup R_{-1,-1}$, or equivalently that the received signal point is to the left of the line labeled S in Fig. 1. In N dimensions it would be the probability that the received signal point is on the opposite side of an $(N - 1)$ dimensional hyperplane. This is clearly a simple one-dimensional gaussian problem having the well-known Q-

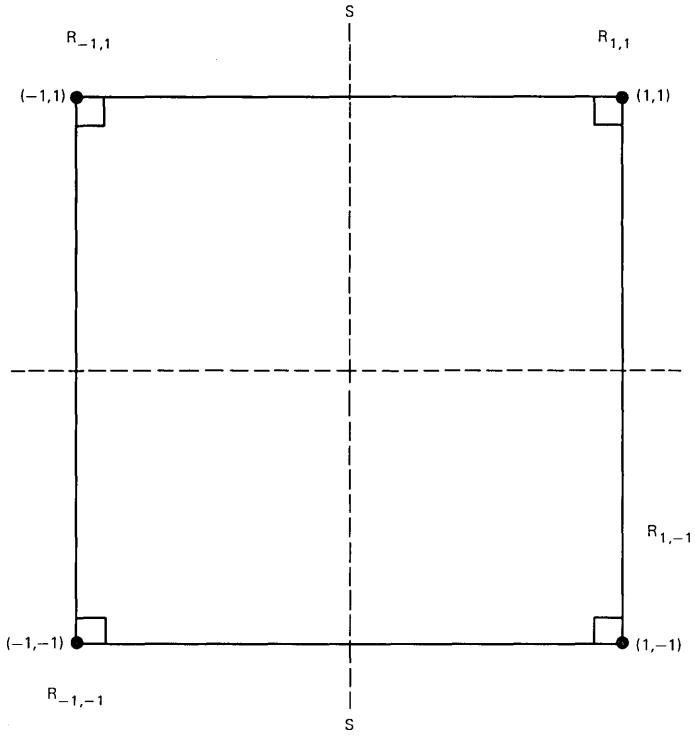


Fig. 1—Four signal points corresponding to sequences for $N = 2$ in the absence of intersymbol interference.

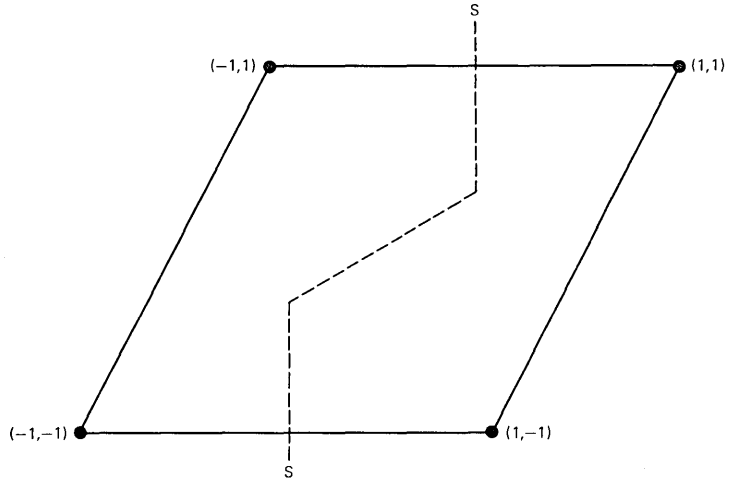


Fig. 2—Four signal points with intersymbol interference.

function for an answer, independent (because of the simple geometry) of the dimension or of which bit in the sequence is transmitted.

When intersymbol interference is present, the error probability for the kth bit may well depend on k and N. In addition, the surface S which separates the sequences that have $a_k = +1$ from those which have $a_k = -1$ is no longer a hyperplane, although it is made up of segments which are hyperplanes. Finally, and perhaps a bit vaguely, the "shape" of the surface may depend on k. An example for $N = 2$ is given in Fig. 2 showing the separating surface for the first bit.

Our goal is to derive Forney's upper bound by geometrical arguments about as simple as those used in the discussion of Fig. 1.

As in (3), we consider signal points identified by their respective data sequences $\{a_i\}^N$ and label them with a superscript. We focus on the kth bit being in error, and define sets A and B:

$$A = \{\mathbf{a}^{(i)} \mid a_k^{(i)} = +1\}, \quad B = \{\mathbf{a}^{(i)} \mid a_k^{(i)} = -1\} \equiv \{\mathbf{b}^{(i)}\}. \tag{6}$$

We are mainly interested in the chance that the maximum-likelihood decoder selects a point $\mathbf{b} \in B$, given that a particular sequence $\mathbf{a}^{(1)}$ (say), $\mathbf{a}^{(1)} \in A$, was transmitted. One upper bound on this is the union bound

$$P_e(k) \leqq \sum_{\mathbf{b} \in B} Q\left(\frac{d(\mathbf{a}^{(1)}, \mathbf{b})}{\sqrt{2N_0}}\right), \tag{7}$$

where $d(\mathbf{a}^{(1)}, \mathbf{b})$ is the euclidean distance between $\mathbf{a}^{(1)}$ and $\mathbf{b}$; i.e., if l labels the particular $\mathbf{b}$ sequence,

$$d^2(\mathbf{a}^{(1)}, \mathbf{b}) = \int_{-\infty}^{\infty} |s^{(1)}(t) - s^{(l)}(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 \left| \sum_{j=1}^{N} (a_j^{(1)} - b_j^{(l)}) e^{ij\omega} \right|^2 d\omega. \tag{8}$$

On writing (8), and henceforth, we set $T = 1$. Equation (7) is a bad bound because it includes too many terms on the right-hand side. This is easily seen by applying it to the N-dimensional hypercube (no intersymbol interference), for which we obtain (ignoring unessential coefficients)

$$P_e(k) \leqq 2^N e^{-(d^2/4N_0)}, \tag{9}$$

where d^2 is the length of a hypercube edge. Thus, the bound, for any fixed N_0, approaches ∞ as the length of the sequence N increases.

Our next step will be to make some simple observations about the geometry of the received signal points [when (5) is true] that will allow us to delete most of the terms on the right-hand side of (7). Following

Forney, we are motivated to define another set B_1 of signal points which is a subset of B. A vector $\mathbf{b} \in B_1$ if $\{\mathbf{b}_j - \mathbf{a}_j^{(1)}\}_{j=1}^{N}$ (after deleting all zeros which begin the sequence and end the sequence of coefficients $\{\mathbf{b}_j - \mathbf{a}_j^{(1)}\}$) does not contain ν or more consecutive zeros either to the right of the kth position or to the left of the kth position. Forney's upper bound then reads

$$P_e(k) \leqq \sum_{\mathbf{b} \in B_1} Q\left(\frac{d(\mathbf{a}^{(1)}, \mathbf{b})}{\sqrt{2N_0}}\right). \tag{10}$$

To see why no further terms need be included in (10), select an arbitrary signal point b^* not included in the sum in (10); i.e., $b^* \in B - B_1 \equiv B^*$. From the way things have been defined, we may write

$$\mathbf{a}^{(1)} = (\boldsymbol{\alpha}^{\mathrm{L}}, 1, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3) \tag{11}$$

$$\mathbf{b}^* = (\boldsymbol{\beta}^{\mathrm{L}}, -1, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_3), \tag{12}$$

where $\boldsymbol{\alpha}_1 - \boldsymbol{\beta}_1$ does not contain ν consecutive zeros in its coefficient sequence and $\boldsymbol{\alpha}_3 - \boldsymbol{\beta}_3 \neq 0$. The 1 in (11) and the (-1) in (12) occur in the kth position, and $\boldsymbol{\alpha}_2$ is at least of dimension ν, corresponding to the ν (or more) positions where $\mathbf{a}^1$ and $\mathbf{b}^*$ are to agree. Now $\boldsymbol{\alpha}_{\mathrm{L}} - \boldsymbol{\beta}_{\mathrm{L}}$ may or may not contain ν consecutive zeros,† and we distinguish these two cases in our discussion. First assume that $\boldsymbol{\alpha}_{\mathrm{L}} - \boldsymbol{\beta}_{\mathrm{L}}$ does not contain ν consecutive zeros. Then note that

$$\mathbf{b}^{(1)} \equiv (\boldsymbol{\beta}_{\mathrm{L}}, -1, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3) \in B_1 \tag{13}$$

$$\mathbf{a}^{(2)} \equiv (\boldsymbol{\alpha}_{\mathrm{L}}, 1, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_3) \in A. \tag{14}$$

Statement (13) is true by the absence of ν consecutive zeros in $\boldsymbol{\beta}_{\mathrm{L}} - \boldsymbol{\alpha}_{\mathrm{L}}$ and also in $\boldsymbol{\beta}_1 - \boldsymbol{\alpha}_1$; (14) is true because $\mathbf{a}_k^{(2)} = +1$. We may at this point imagine the four signal points $\mathbf{a}^{(1)}$, $\mathbf{b}^*$, $\mathbf{b}^{(1)}$, $\mathbf{a}^{(2)}$ in general position as in Fig. 3. Now focus attention on the triangle $(a^{(1)}, a^{(2)}, b^*)$.‡ We have (letting 0_i denote a string of at least i zeros)

$$\begin{aligned} d^2(a^{(1)}, b^*) &= \|a^{(1)} - b^*\|^2 \\ &= \|\alpha^{\mathrm{L}} - \beta^{\mathrm{L}}, 2, \alpha_1 - \beta_1, 0_\nu, \alpha_3 - \beta_3\|^2 \\ &= \|\alpha^{\mathrm{L}} - \beta^{\mathrm{L}}, 2, \alpha_1 - \beta_1\|^2 + \|\alpha_3 - \beta_3\|^2. \end{aligned} \tag{15}$$

The last step in (15) follows from the Fourier integral expression (8) for the distance and from (5), the requirement that intersymbol interference not extend beyond ν. The right member of (15) can readily be seen from (11, 12) and (13, 14) to be $d^2(a^{(2)}, b^*) + d^2(a^{(1)}, a^{(2)})$.

† Recall that a string of zeros in the beginning does not count.

‡ We drop bold-face notation for vectors here, and also allow ourselves the freedom of writing the subscript L as a superscript.

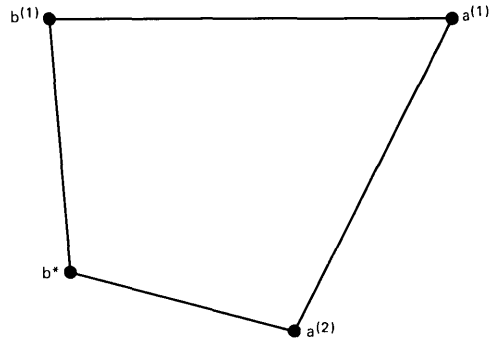


Fig. 3—The four signal points defined in the test illustrated in general position (not necessarily planar).

Hence,

$$d^2(a^{(1)}, b^*) = d^2(a^{(2)}, b^*) + d^2(a^{(1)}, a^{(2)}) \tag{16}$$

and $(a^{(1)}, a^{(2)}, b^*)$ forms a right triangle. In an entirely similar manner, one verifies

$$d^2(a^{(2)}, b^{(1)}) = d^2(a^{(1)}, b^{(1)}) + d^2(a^{(1)}, a^{(2)}), \tag{17}$$

implying that $(a^{(1)}, a^{(2)}, b^{(1)})$ is a right triangle. Since $b^{(1)} - b^* = a^{(1)} - a^{(2)}$, we have

$$d^2(a^{(1)}, a^{(2)}) = d^2(b^{(1)}, b^*) \tag{18}$$

and the additional fact that the four points lie in the same plane. Equations (16), (17), and (18), and planarity, imply that $(a^{(1)}, a^{(2)}, b^{(1)}, b^*)$ form a rectangle as shown in Fig. 4. This demonstration assumes that there are ν consecutive zeros to the right of the kth position and not to the left. If we interchange the words "right" and "left", the same type of demonstration will apply. There remains the case when there are ν consecutive zeros both to the right and to the left of the kth position. In this case we write

$$a^{(1)} = (\alpha_3^L, \alpha_2^L, \alpha_1^L, 1, \alpha_1, \alpha_2, \alpha_3) \tag{19}$$

$$b^* = (\beta_3^L, \alpha_2^L, \beta_1^L, -1, \beta_1, \alpha_2, \beta_3), \tag{20}$$

where $\alpha_3 - \beta_3 \neq 0 \neq \alpha_3^L - \beta_3^L$ and neither $\alpha_1 - \beta_1$ nor $\alpha_1^L - \beta_1^L$ contain ν consecutive zeros in their coefficients. Further, we are to assume α_2 and α_2^L each have dimension at least ν. If we define

$$b^{(1)} = (\alpha_3^L, \alpha_2^L, \beta_1^L, -1, \beta_1, \alpha_2, \alpha_3) \in B_1 \tag{21}$$

$$a^{(2)} = (\beta_3^L, \alpha_2^L, \alpha_1^L, 1, \alpha_1, \alpha_2, \beta_3) \in A, \tag{22}$$

the proof that $(a^{(1)}, a^{(2)}, b^{(1)}, b^*)$ is a rectangle can be carried out using the same techniques as earlier.

Figure 4 makes it clear why, if $a^{(1)}$ is transmitted, terms like b^* do not have to be included in the right-hand side of (10). The term

$$Q\left(\frac{d(a^1, b^1)}{\sqrt{2N_0}}\right)$$

in (10) is the probability that, if $a^{(1)}$ is transmitted, the received signal will be on the "wrong side" of the hyperplane H which perpendicularly bisects the line (b^1a^1). Now b^* only needs to be included in (10) if its associated decoding region contributes some set of points of positive measure not accounted for in some other way. Thus, from Fig. 4, this is only the case if these new points were on the same side of H as $a^{(1)}$. But any point on that side is closer to $a^{(2)}$ than b^* and, hence, would never be decoded into b^*. Hence the gaussian measure of the decoding region for b^* is already included in the term

$$Q\left(\frac{d(b^1, a^1)}{\sqrt{2N_0}}\right).$$

If we calculate further upper bounds for (10) by letting $N = \infty$, we obtain a bound independent of k. Averaging this over the possible transmitted symbols gives precisely Forney's upper bound. The fact that the resulting upper bound converges for N_0 small enough (for

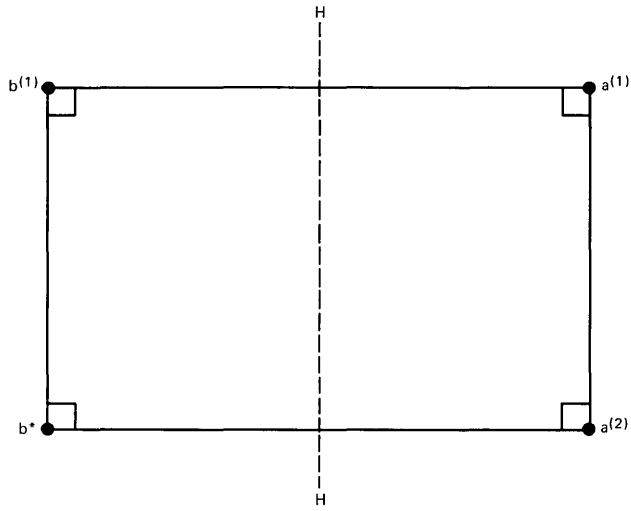


Fig. 4—The actual relationship of the four signal points defined in the text.

$N = \infty$) has been recently discussed by Foschini.[4] This last step is an important one in a full proof, and was overlooked in the initial work.[2]

REFERENCES

1. G. David Forney, "Lower Bounds on Error Probability in the Presence of Large Intersymbol Interference," IEEE Trans. Commun., *COM-20* (February 1972), pp. 76–77.
2. G. David Forney, "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. Inform. Theory, *IT-18* (May 1972), pp. 363–378.
3. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering.* New York: John Wiley, 1965, Ch. 4.
4. G. J. Foschini, "Performance Bound for Maximum Likelihood Reception of Digital Data," IEEE Trans. Inform. Theory, *IT-21*, No. 1 (January 1975), pp. 47–50.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

Power Spectrum of a Digital, Frequency-Modulation Signal

By H. E. ROWE and V. K. PRABHU

(Manuscript received December 31, 1974)

We present the power spectrum of a sinusoidal carrier, frequency modulated by a random baseband pulse train in which the signaling-pulse duration is finite and the signal pulses may overlap and have different shapes. Symbols transmitted during different time slots are assumed to be statistically independent and identically distributed. The spectral density appears as a Hermitian form suitable for numerical computation by a digital computer. Simple conditions in terms of the modulation parameters are given under which discrete spectral lines are present in the spectrum. Several examples are given to illustrate the method.

I. INTRODUCTION

In recent years, digital-frequency and phase-modulation techniques have been increasingly important in radio, waveguide, and optical communication systems.

An important parameter in the statistical description of a signal is its spectral density, which defines the average power density of the signal as a function of frequency. In addition to furnishing an estimate of bandwidth requirements, the knowledge of the spectral density is also essential in the evaluation of mutual interference between channels.

In this paper, we extend the techniques developed in Ref. 1 for digital PSK to the case of digital FSK with phase-continuous transitions, such as may be obtained at the output of a voltage-controlled oscillator driven by a digital baseband wave. We assume that the sinusoidal carrier is frequency modulated by a random, baseband pulse train in which the signaling pulse duration is finite and the signal pulses may overlap and have different shapes. It is generally assumed that the symbols transmitted during different time slots are statistically independent and identically distributed.

We express the spectral density of such an ensemble of continuous-phase, constant-envelope, digital, FM waves as a compact Hermitian

form that provides an appropriate division between analysis and machine computation. The present work permits simpler numerical computation of digital FSK spectra than earlier studies,[2–5] and contains a simpler statement of the conditions, in terms of the modulation parameters, that determine whether discrete spectral lines are present in the spectrum or not.

Examples give the spectra of binary and quaternary FSK waves with overlapping baseband modulation pulses of several shapes.

II. M-ARY FREQUENCY-MODULATED SIGNALS

We seek the spectrum of the digital frequency-modulated wave:

$$x(t) = \cos [2\pi f_c t + \phi(t)], \quad f_c > 0. \tag{1}$$

$$\phi(t) = \int^t f_d(\mu) d\mu. \tag{2}$$

$$f_d(t) = \sum_{k=-\infty}^{\infty} h_{s_k}(t - kT), \qquad s_k = 1, 2, \cdots, M. \tag{3}$$

The symbol ϕ is the phase and f_d the frequency deviation of the carrier at frequency f_c. The signaling alphabet consists of M waveforms h_1, h_2, $\cdots$, h_M, that may have different shapes; one of these is transmitted for each signaling interval of duration T. The different signaling waveforms in (3) may overlap, but are statistically independent in most of the present work; i.e., s_k is statistically independent of s_l for $k \neq l$.

Define for convenience

$$v(t) \equiv e^{j\phi(t)}; \tag{4}$$

then

$$x(t) = \text{Re}\ \{e^{j2\pi f_c t} v(t)\}. \tag{5}$$

The spectral density of $v(t)$ is

$$P_v(f) = \int_{-\infty}^{\infty} \Phi_v(\tau) e^{-j2\pi f\tau} d\tau, \tag{6}$$

where

$$\Phi_v(\tau) = \overline{\Phi_v(t, \tau)} \equiv \lim_{A \to \infty} \frac{1}{2A} \int_{-A}^{A} \Phi_v(t, \tau) dt, \tag{7}$$[§]

$$\Phi_v(t, \tau) = \langle v(t + \tau) v^*(t) \rangle = \langle e^{j[\phi(t+\tau) - \phi(t)]} \rangle$$

$$= \left\langle \exp \left(j \int_t^{t+\tau} f_d(\mu) d\mu \right) \right\rangle. \tag{8}$$

§ The symbol $\overline{\quad}$ denotes average on t throughout.

An arbitrary constant of integration is implicit in the notation of (2), which defines the phase. This is of little consequence in the remainder of the present paper, since most of our study is directed toward the spectrum of the complex wave $v(t)$; as seen by the final line of (8), the absolute phase is irrelevant in determining $P_v(f)$. The absolute phase must be rendered explicit only for the following three purposes in the present work:

(*i*) Relating $P_x(f)$ to $P_v(f)$ (the spectra of the real FSK wave $x(t)$ of (1) and of the complex FSK wave of (4), respectively).
(*ii*) Separating possible line-frequency components from the FSK wave.
(*iii*) Specializing the present FSK treatment to the prior PSK results.[1]

To make the absolute phase explicit, we write

$$\phi(t) = \int_0^t f_d(\mu)d\mu + \phi(0); \quad v(0) = e^{j\phi(0)}. \tag{9}$$

The term $\phi(0)$ is the phase of the FSK wave at $t = 0$. We consider three representative assumptions for $\phi(0)$:

(*i*) $\phi(0)$ deterministic, e.g.:

$$\phi(0) = 0. \tag{10}$$

(*ii*) $\phi(0)$ random, uniform, and independent of the modulation s_k:

$$\Pr[\phi < \phi(0) \leqq \phi + d\phi] = \frac{d\phi}{2\pi}, \quad 0 \leqq \phi < 2\pi. \tag{11}$$

(*iii*) $\phi(0)$ dependent only on the modulation parameters (or signaling pulses) that contribute to $f_d(0+)$§ in (3):

$$\phi(0) = \sum_{k:h_{s_k}(-kT+)\neq 0} \int_{-\infty}^{-kT} h_{s_k}(\mu)d\mu. \tag{12}$$§

The signal-pulse duration is assumed finite; consequently, in (12), the $\sum$ has a finite number of terms, and the lower limit on the $\int$ becomes finite. Specific examples of (12) appear in Section IV below. In other words, the net phase contributed by all past signaling pulses that are over by $t = 0$, i.e., for which $h_{s_k}(-kT+) = 0$,§ is normalized to 0.

---

§ The $+$'s indicate that f_d and h_{s_k} are to be evaluated an infinitesimal time later than 0 and $-kT$, respectively, if any of the signal pulses have discontinuities at their upper limits at a time-slot boundary (see Section IV). If the signal pulses and, hence, the instantaneous frequency deviation $f_d(t)$ are continuous, these two $+$'s may be dropped.

In the appendix we show that, except for special modulations with low carrier frequencies f_c that take on special values, the following simple relation gives the spectrum of the real wave $x(t)$ of (1):

$$P_x(f) = \tfrac{1}{4} P_v(f - f_c) + \tfrac{1}{4} P_v(-f - f_c). \tag{13}$$§

The first term of (13) is the spectrum of the complex baseband wave $v(t)$ shifted to the carrier frequency $+f_c$; the second term is the spectrum of $v^*(t)$ shifted to $-f_c$. More specifically, (13) is valid if any of the following are true:

(*i*) Eq. (11) holds, independently of any other considerations.
(*ii*) Eq. (10) or (12) holds, and $P_v(f)$ [and hence $P_x(f)$] has no line components.
(*iii*) Eq. (10) or (12) holds, $P_v(f)$ [and hence $P_x(f)$] has (equally spaced) line components, and f_c is low enough so that the two terms of (13) overlap, but f_c is such that the line components of the two terms of (13) do not coincide. The discrete values forbidden to f_c under these conditions are given by (146) in the appendix.
(*iv*) The carrier frequency is so large that the two terms of (13) do not significantly overlap, independently of any other considerations.

Consequently, we study only $P_v(f)$ throughout the remainder of this paper. This suffices for all cases except that of a low carrier frequency f_c that takes on special discrete values related to the baud rate $1/T$ and the modulation, for very special modulations that result in line-spectral components (PSK is one such case, but there are others).

The condition expressed by (12) is assumed in much of what follows. This results in no significant loss in generality in treating the spectrum of the real wave $x(t)$, as discussed above. It permits economy of notation in the general case, a convenient treatment of line components present in special cases, and simple specialization of the present FSK results to prior PSK results.[1]

III. NOTATION AND STATISTICAL MODEL

We introduce the vector-matrix notation of the prior PSK study.[1] Since that study allowed correlated modulation parameters s_k while the present work is largely restricted to independent s_k, only a portion of Section III of Ref. 1 need be summarized here.

---

§ Compare (5) of Ref. 1.

We write (3) as

$$f_d(t) = \sum_{k=-\infty}^{\infty} \underbrace{\mathbf{a}_k} \cdot \mathbf{h}(t - kT)\rfloor, \tag{14}$$§

where

$$\underbrace{\mathbf{a}_k} = \mathbf{a}_k\rfloor' \equiv [a_k^{(1)}\ a_k^{(2)} \cdots a_k^{(M)}\rfloor, \tag{15}$$

$$\underbrace{\mathbf{h}(t)} = \mathbf{h}(t)\rfloor' \equiv [h_1(t)\ h_2(t) \cdots h_M(t)\rfloor. \tag{16}$$

For a given k (i.e., for a given time slot) one of the a_k's is unity and the rest are zero:

$$a_k^{(s_k)} = 1; \quad a_k^{(i)} = 0, \quad i \neq s_k. \tag{17}$$

Thus, $\mathbf{a}_k$ is a unit basis vector, i.e., $\mathbf{a}_k$ has one component unity and all other components zero.

The modulation process s_k is assumed stationary, as in the prior PSK study,[1] but here we make the stronger assumption of independence in most of what follows. Define the first-order probability

$$w_i \equiv \Pr\{s_k = i\} \tag{18}$$

as the probability that the ith signaling waveform is transmitted in the kth time slot. w_i is independent of k by stationarity. By independence,

$$\Pr\{s_k = i,\ s_l = j\} = w_i w_j, \quad k \neq l. \tag{19}$$

Then

$$w_i = \Pr\left\{\mathbf{a}_k = [\underset{1}{0}\ \underset{2}{0}\ \cdots\ \underset{i-1}{0}\ \underset{i}{1}\ \underset{i+1}{0}\ \cdots\ \underset{M}{0}\rfloor\right\}. \tag{20}$$

Normalization of the total probability requires

$$\sum_{i=1}^{M} w_i = 1. \tag{21}$$

We use the following vector notation for the probabilities:

$$\underbrace{\mathbf{w}} = \mathbf{w}\rfloor' \equiv [w_1\ \ w_2 \cdots w_M\rfloor. \tag{22}$$

---

§ The following notational conventions are adopted throughout:

(*i*) Boldface quantities denote matrices.
(*ii*) Row and column vectors are distinguished by the additional notation $\llcorner$ and $\rfloor$, respectively.
(*iii*) Ordinary matrix multiplication is indicated by $\cdot$, Kronecker matrix products by $\times$ (see second footnote, page 908, Ref. 1, for properties of Kronecker products used throughout the present paper).
(*iv*) The transpose of a matrix is indicated by $'$.
(*v*) The Hermitian transpose of a matrix is indicated by $\dagger$.
(*vi*) Multiple Kronecker products are indicated by $\prod\times$ (see footnote, page 911, Ref. 1) and the Kronecker power is indicated by an integer exponent enclosed in square brackets.

Further define a vector (here of dimension M) with all elements unity as

$$\underline{1} = 1]' \equiv [1 \quad 1 \cdots 1]. \tag{23}$$

Then (21) may be written

$$\underline{1}\cdot\mathbf{w}] = \underline{w}\cdot 1] = 1. \tag{24}$$

Finally, for convenience later, define the diagonal matrix

$$\mathbf{w}_d \equiv \begin{bmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_M \end{bmatrix}. \tag{25}$$

Then, from (22) and (23),

$$\mathbf{w}_d\cdot 1] = \mathbf{w}], \quad \underline{1}\cdot\mathbf{w}_d = \underline{w}, \tag{26}$$

Note that

$$\langle\underline{\mathbf{a}_k}\rangle = \underline{w}, \tag{27}$$

$$\langle\mathbf{a}_k]\cdot\underline{\mathbf{a}_k}\rangle = \mathbf{w}_d. \tag{28}$$

IV. FSK AS A BASEBAND PULSE TRAIN

We seek an expression of $v(t)$, given by (4), (9), (12), and (14), of the form

$$v(t) = \sum_{k=-\infty}^{\infty} \underline{c_k}\cdot\mathbf{r}(t - kT)] \tag{29}$$

for signaling pulses of finite duration. Assume the $h_i(t)$ of (3) are strictly time-limited to an interval KT, as follows:

$$\mathbf{h}(t)] = \mathbf{0}], \quad t \leqq L_K, \quad t > U_K. \tag{30}$$[§]

$$L_K \equiv \begin{cases} -\dfrac{K-1}{2}\,T, & K \text{ odd.} \\ -\dfrac{K}{2}\,T, & K \text{ even.} \end{cases}$$

$$U_K \equiv \begin{cases} \dfrac{K+1}{2}\,T, & K \text{ odd.} \\ \dfrac{K}{2}\,T, & K \text{ even.} \end{cases} \tag{31}$$

L_K and U_K are respectively the lower and upper limits of the pulses. Figure 1 shows portions of $f_d(t)$ for four different maximum signal-pulse durations; the terms $k = -1, 0, 1, 2$ of (14) are shown, and for

[§] $\underline{0} = \mathbf{0}]'$ is a vector of appropriate dimension (here M) with all elements zero.

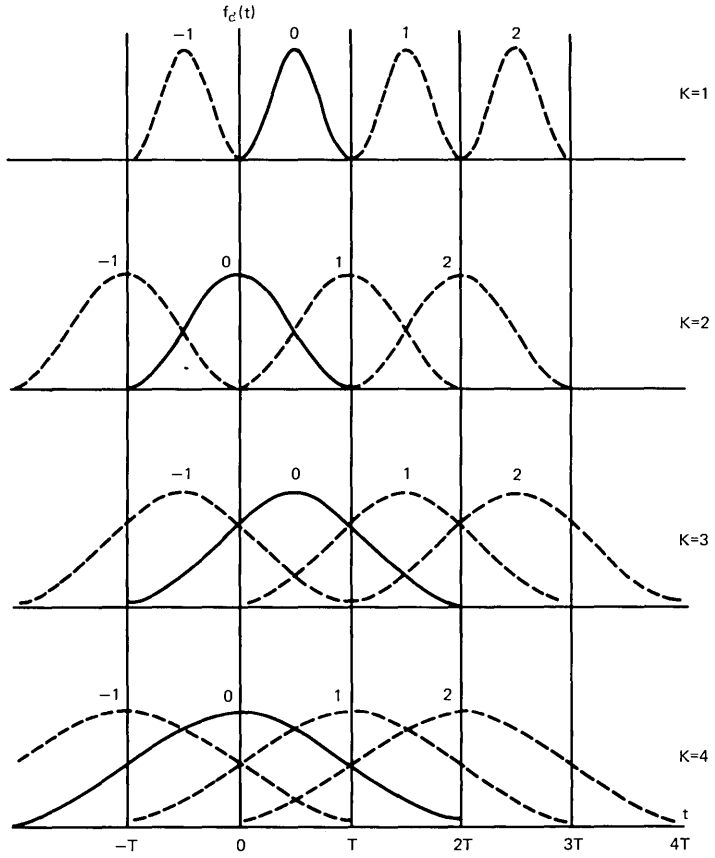


Fig. 1—Frequency modulation for different signal pulse durations. Index k is shown near peak of each pulse. Also, for simplicity, same signal pulse is shown for each k. T = time slot duration or signaling period. KT = maximum signal-pulse duration. Note different pulse center location for odd and even K.

convenience a_k has been taken the same for each of these time slots. The pulses are positioned along the time slots such that the limits of each signal pulse lie on the boundary between adjacent time slots (i.e., t = integer$\cdot T$); this results in different definitions for L_K, U_K for K even and odd. Since symmetric pulses have been chosen for illustration in Fig. 1, their maxima are centered in the time slots for K odd, and lie on the time-slot boundaries for K even. Discontinuities are permitted at the pulse edges (and elsewhere), but are not present in the example of Fig. 1 (and would not normally be present in a mathematical model of a physical system); discontinuities at the time-slot boundaries are restricted by the inequalities at the right of (30). Examine the $(0, T]$

time slot in Fig. 1 as typical; then the number of pulses contributing to $f_d(t)$ at every instant equals K.§

It remains for us to express the pulse shapes $\mathbf{r}(t)$ and coefficients $\mathbf{c}_k$ of (29) in terms of the signal pulses $\mathbf{h}(t)$ and coefficients $\mathbf{a}_k$ of (14). We give separate treatments for the cases $K = 1$ and $K = 2$, and extend these results to general K. The treatment is an extension of that for the PSK case, given in Section IV of Reference 1.

4.1 Nonoverlapping pulses: $K = 1$

The top portion of Fig. 1 shows digital frequency modulation for which the signal pulses in different time slots never overlap; in this case from (30) to (31),

$$\mathbf{h}(t)\rceil = 0\rceil, \quad t \leqq 0, \quad t > T. \tag{32}$$

Define

$$\underline{\mathbf{q}(t)} \equiv \begin{cases} \left[\exp\left(j\int_0^t h_1(\mu)d\mu\right) \exp\left(j\int_0^t h_2(\mu)d\mu\right) \right. \\ \qquad \left. \cdots \exp\left(j\int_0^t h_M(\mu)d\mu\right)\right], & 0 < t \leqq T. \\ \underline{0}, & t \leqq 0, \quad t > T. \end{cases} \tag{33}$$

Equations (4), (9), (12), (14), and (33) yield

$$\phi(0) = 0; \quad v(0) = 1. \tag{34}$$

$$v(t) = \sum_{k=-\infty}^{\infty} S_k \underline{\mathbf{a}_k} \cdot \mathbf{q}(t - kT)\rceil, \tag{35}$$

where

$$S_k = \begin{cases} \prod_{i=k+1}^{0} \underline{\mathbf{a}_{i-1}} \cdot \mathbf{q}^*(T)\rceil, & k < 0. \\ 1, & k = 0. \\ \prod_{i=1}^{k} \underline{\mathbf{a}_{i-1}} \cdot \mathbf{q}(T)\rceil, & k > 0. \end{cases} \tag{36}$$

Comparing (35) and (29), the parameters of the latter are given as follows for nonoverlapping signal pulses:

$$\left.\begin{array}{l} \underline{\mathbf{c}_k} = S_k \underline{\mathbf{b}_k}, \quad \underline{\mathbf{b}_k} \equiv \underline{\mathbf{a}_k}; \\ \mathbf{r}(t)\rceil = \mathbf{q}(t)\rceil; \end{array}\right\} \quad K = 1. \tag{37}$$

S_k in (37) is, of course, given by (36).

§ The same conventions were used in Fig. 1, Ref. 1, for the baseband modulation pulses in digital PSK.

4.2 Overlapping pulses: $K = 2$

This case is illustrated in the second portion of Fig. 1. In the $(0, T]$ time slot, the $k = 0, 1$ pulses contribute. We have from (30) to (31)

$$\mathbf{h}(t)] = \mathbf{0}], \quad t \leqq -T, \quad t > T. \tag{38}$$

Define

$$\underline{\mathbf{q}(t)} \equiv \begin{cases} \left[\exp\left(j\int_{-T}^{t} h_1(\mu)d\mu\right) \exp\left(j\int_{-T}^{t} h_2(\mu)d\mu\right) \right. \\ \qquad \left. \cdots \exp\left(j\int_{-T}^{t} h_M(\mu)d\mu\right)\right], & -T < t \leqq T. \\ \underline{\mathbf{0}}, & t \leqq -T, \quad t > T. \end{cases} \tag{39}$$

§

Equations (4), (9), (12), (14), and (39) yield

$$\phi(0) = \int_{-T}^{0} \dot{h}_{s_0}(\mu)d\mu. \tag{40}$$

$$v(t) = \sum_{k=-\infty}^{\infty} S_k\{\underline{\mathbf{a}_k}\cdot\mathbf{q}(t - kT)]\}\{\underline{\mathbf{a}_{k+1}}\cdot\mathbf{q}(t - (k+1)T)]\}, \tag{41}$$

where S_k remains as given in (36), the same as for the prior $K = 1$ case. Proceeding exactly as in (51) of Ref. 1, (41) above yields (29) with the following parameters, when no more than two signal pulses overlap:

$$\begin{aligned} &\underline{\mathbf{c}_k} = S_k\underline{\mathbf{b}_k}, \quad \underline{\mathbf{b}_k} \equiv \underline{\mathbf{a}_k} \times \underline{\mathbf{a}_{k+1}}; \\ &\mathbf{r}(t)] = \mathbf{q}(t)] \times \mathbf{q}(t - T)]; \end{aligned} \qquad K = 2. \tag{42}$$

S_k is given by (36), and $\times$ denotes the Kronecker product [see footnote to (14)]. The term $\mathbf{b}_k$, like $\mathbf{a}_k$, is a unit basis vector, i.e., it has one element unity and the remaining $M^2 - 1$ elements zero.[1] Note from (42) and (39) that

$$\mathbf{r}(t)] = \mathbf{0}], \quad t \leqq 0, \quad t > T. \tag{43}$$

Binary FSK offers a simple example of these results, governed by the same relations between $\mathbf{a}_k$ and $\mathbf{b}_k$ as for binary PSK given in (57) of Ref. 1.

4.3 Overlapping pulses: general K

The general case follows by straightforward extension of the above; see (58) to (62) of Ref. 1. Figure 1 illustrates the frequency modulation for $K = 3, 4$. The modulation pulse restrictions are given by (30) and

§ Comparing (39) with (33), note that the definition of $\mathbf{q}(t)$ is different for different K; $\mathbf{q}(t) \neq 0$ over the same interval in which $\mathbf{h}(t)$ may be nonzero.

(31). Define

$$\underline{\mathbf{q}(t)} \equiv \begin{cases} \left[\exp\left(j\int_{L_K}^{t} h_1(\mu)d\mu\right)\ \exp\left(j\int_{L_K}^{t} h_2(\mu)d\mu\right)\right. \\ \qquad \left.\cdots \exp\left(j\int_{L_K}^{t} h_M(\mu)d\mu\right)\right], \quad L_K < t \leqq U_K. \\ \underline{\mathbf{0}}, \qquad\qquad t \leqq L_K, \quad t > U_K. \end{cases} \tag{44}$$

$\phi(0)$ of (12) is

$$\phi(0) = \begin{cases} \sum_{k=-(K-1)/2}^{(K-3)/2} \int_{L_K}^{-kT} h_{s_k}(\mu)d\mu, & K \text{ odd}, \quad K > 1. \\ \sum_{k=-(K-2)/2}^{(K-2)/2} \int_{L_K}^{-kT} h_{s_k}(\mu)d\mu, & K \text{ even}, \quad K > 0. \end{cases} \tag{45}$$

The parameters of (29) are

$$\underline{\mathbf{c}_k} = S_k\underline{\mathbf{b}_k}; \quad \underline{\mathbf{b}_k} = \begin{cases} \prod_{i=-(K-1)/2}^{(K-1)/2} \times \underline{\mathbf{a}_{k+i}}, & K \text{ odd}. \\ \prod_{i=-(K-2)/2}^{K/2} \times \underline{\mathbf{a}_{k+i}}, & K \text{ even}. \end{cases} \tag{46}$$

$$\mathbf{r}(t)] = \begin{cases} \prod_{i=-(K-1)/2}^{(K-1)/2} \times \mathbf{q}(t - iT)], & K \text{ odd}. \\ \prod_{i=-(K-2)/2}^{K/2} \times \mathbf{q}(t - iT)], & K \text{ even}. \end{cases} \tag{47}$$

$\prod\times$ denotes a multiple Kronecker product [see footnote to (14)]. S_k is given by

$$S_k = \begin{cases} \prod_{i=k+1}^{0} \underline{\mathbf{a}_{i-(K+1)/2}} \cdot \mathbf{q}^*(U_K)], & k < 0; \\ 1, & k = 0; \quad K \text{ odd}. \\ \prod_{i=1}^{k} \underline{\mathbf{a}_{i-(K+1)/2}} \cdot \mathbf{q}(U_K)], & k > 0. \end{cases} \tag{48}$$

$$S_k = \begin{cases} \prod_{i=k+1}^{0} \underline{\mathbf{a}_{i-(K/2)}} \cdot \mathbf{q}^*(U_K)], & k < 0; \\ 1, & k = 0; \quad K \text{ even}. \\ \prod_{i=1}^{k} \underline{\mathbf{a}_{i-(K/2)}} \cdot \mathbf{q}(U_K)], & k > 0. \end{cases} \tag{49}$$

Note that

$$\mathbf{r}(t)\rceil = \mathbf{0}\rceil, \quad t \leqq 0, \quad t > T. \tag{50}$$

4.4 Discussion

It is instructive to obtain the prior PSK results[1] by specializing the present FSK results. Define the phase shift produced by each signaling pulse as

$$g_i(t) \equiv \int_{L_K}^{t} h_i(\mu)d\mu, \quad i = 1, 2, \cdots, M \tag{51}$$

or in vector notation

$$\mathbf{g}(t)\rceil \equiv \int_{L_K}^{t} \mathbf{h}(\mu)\rceil d\mu, \tag{52}$$

and substitute into (44)§ (or into (33) or (39) for special cases $K = 1$, 2). Then the present results of (29), (31), and Section 4.3 are similar to the former PSK results of (43) and Section 4.3 of Ref. 1, except for the factor S_k; in particular, the equations for $\mathbf{q}(t)$, $\mathbf{b}_k$, and $\mathbf{r}(t)$ have an identical form. The additional factor S_k present in the FSK case accounts for the total phase shift introduced by each of the signaling pulses.

To specialize the present FSK results to the PSK case, we require the total area of each of the present modulation pulses to be zero. Thus, (51) and (52) become

$$g_i(U_K) \equiv \int_{L_K}^{U_K} h_i(t)dt = 0, \quad i = 1, 2, \cdots, M \tag{53}$$§§

or

$$\mathbf{g}(U_K)\rceil \equiv \int_{L_K}^{U_K} \mathbf{h}(t)\rceil dt = 0\rceil. \tag{54}$$§§

Substituting in (44),

$$\mathbf{q}(U_K)\rceil = 1\rceil. \tag{55}$$

Therefore, (17) yields

$$\underline{\mathbf{a}_i}\cdot\mathbf{q}(U_K)\rceil = 1. \tag{56}$$

Substituting in (48) to (49),

$$S_k = 1, \quad \text{all } k, K, \tag{57}$$

completing the specialization of the present FSK results to the PSK case.[1]

In the general FSK case, the results of Sections 4.1 to 4.3 above reduce the FSK problem to determining the spectrum of (29). This is ac-

§ Only the range $L_k < t \leqq U_k$ is relevant in (51), (52), since $\mathbf{q}(t) \equiv \mathbf{0}$ outside this range.

§§ Equations (30) to (31) and (53), (54) render (51) and (52) zero for $t \leqq L_K$, $t \geqq U_K$, satisfying (58) of Ref. 1.

complished by deleting (149), (153), and (163) to (171) of Ref. 1, Appendix B.§ Then the spectral density of (29) above is given by setting $\mathbf{b} \rightarrow \mathbf{c}$ in (150) to (152) and (160) to (162) of Ref. 1 as follows:

$$P_v(f) = \frac{1}{T}\underline{\mathbf{R}(f)}\cdot\tilde{\mathbf{P}}_c(fT)\cdot\mathbf{R}^*(f)]. \tag{58}$$

$$\tilde{\mathbf{P}}_c(f) = \sum_{n=-\infty}^{\infty} e^{-j2\pi fn}\tilde{\mathbf{\Phi}}_c(n) \tag{59}$$§§

$$\tilde{\mathbf{\Phi}}_c(n) = \langle \mathbf{c}_{k+n}]\cdot\underline{\mathbf{c}_k^*}\rangle \tag{60}$$

$$\mathbf{R}(f)] = \underline{\mathbf{R}(f)'} = \int_{-\infty}^{\infty} e^{-j2\pi ft}\mathbf{r}(t)]dt. \tag{61}$$

In these relations $\mathbf{R}(f)$ is the Fourier transform of $\mathbf{r}(t)$ of (37), (42), or (47), depending on K (i.e., the amount of pulse overlap). $\tilde{\mathbf{\Phi}}_c(n)$ is determined from (36) and (37), (36) and (42), or (46) and (48) or (49) in Section VI for $K = 1$, 2, and general K, respectively.

V. FSK WITH LINE COMPONENTS

From (21), (22), and (44),

$$|\underline{\mathbf{w}}\cdot\mathbf{q}(U_K)]| \leqq 1. \tag{62}$$§§§

We show in the present section that equality in (62) corresponds to the presence of line components in the FSK spectrum. Conversely, if the inequality of (62) is satisfied, we see in Section VI that the FSK spectrum contains no line components.

Assume throughout the remainder of the present section that

$$|\underline{\mathbf{w}}\cdot\mathbf{q}(U_K)]| = 1. \tag{63}$$

This yields

$$\int_{L_K}^{U_K} h_i(t)dt = 2\pi f_l + \text{integer}\cdot 2\pi, \quad i = 1, 2, \cdots, M, \tag{64}$$

where $2\pi f_l$ is defined as the common area (mod 2π) of all of the FSK signaling pulses in the line case. For definiteness we take

$$-\tfrac{1}{2} < f_l \leqq \tfrac{1}{2}. \tag{65}$$

Equivalently,

$$\mathbf{q}(U_K)] = e^{j2\pi f_l}\,\mathbf{1}], \quad -\tfrac{1}{2} < f_l \leqq \tfrac{1}{2}. \tag{66}$$

---

§ These deleted portions were relevant to the study of line spectral components of PSK.[1] The line spectral components of FSK are treated separately in Section V.

§§ See footnote, page 905, Ref. 1.

§§§ This follows because $|q_i(t)| = 1$.

Now substitute (66) into (48) and (49); we have

$$S_k = e^{jk2\pi f_l}, \quad \text{all } k, K. \tag{67}$$

Thus, when (62) is satisfied

$$v(t) = \sum_{k=-\infty}^{\infty} e^{jk2\pi f_l} \underline{\mathbf{b}_k} \cdot \mathbf{r}(t - kT)]. \tag{68}$$

The relation (68) is the same as the PSK result of (43) of Ref. 1 except for the factor $e^{jk2\pi f_l}$; as noted following (52), $\mathbf{b}_k$ and $\mathbf{r}(t)$ have the same form in FSK and PSK. Consequently, when (63) is satisfied, the FSK spectrum may be obtained by simple transformation of the PSK results of Ref. 1. One way this may be done is to rewrite (68) as

$$v(t)e^{-j2\pi(f_l/T)t} = \sum_{k=-\infty}^{\infty} \underline{\mathbf{b}_k} \cdot (e^{-j2\pi(f_l/T)(t-kT)}\mathbf{r}(t - kT)]). \tag{69}$$

Comparing with (43) of Ref. 1, all the PSK results of Ref. 1 apply directly to the present case by making the substitutions

$$\begin{aligned} v(t) &\rightarrow v(t)e^{-j2\pi(f_l/T)t}, \\ \mathbf{r}(t)] &\rightarrow e^{-j2\pi(f_l/T)t}\,\mathbf{r}(t)]. \end{aligned} \tag{70}$$

Therefore, when (68) is satisfied, and consequently (68) holds, the FSK wave has line spectra. We separate the line and continuous components as

$$v(t) = v_l(t) + v_c(t). \tag{71}$$

The line component is given from (70), and (114)§ of Ref. 1, as

$$\begin{aligned} v_l(t) &= \frac{1}{T}\underline{\mathbf{w}^{[K]}} \cdot \sum_{n=-\infty}^{\infty} \mathbf{R}\left(\frac{n + f_l}{T}\right) e^{j2\pi((n+f_l)/T)t}, \\ P_{v_l}(f) &= \frac{1}{T^2}\,|\mathbf{w}^{[K]} \cdot \mathbf{R}(f)]|^2 \sum_{\mathbf{n}=-\infty}^{\infty} \delta\left(\mathbf{f} - \frac{n + f_l}{T}\right). \end{aligned} \tag{72}$$§§

The spectral density of the continuous component is found from (70), and (69),§§§ (96), or (116) of Ref. 1 as follows:

$$P_{v_c}(f) = \frac{1}{T}\underline{\mathbf{R}(f)} \cdot \{\quad\} \cdot \mathbf{R}^*(f)]. \tag{73}$$

---

§ Or from (66) and (95) of Ref. 1 for the special cases $k = 1, 2$.

§§ See footnote to (14); an exponent enclosed in square brackets denotes the Kronecker power, i.e., the Kronecker product of a matrix (or vector) with itself the indicated number of times.

§§§ This is equation (69) of Ref. 1 specialized to independent signal pulses, i.e., with (32) used for all $n \neq 0$, yielding (72) or (73), of Ref. 1. If this restriction is not imposed, and (70) above is used in the general form of (69) of Ref. 1, we obtain the spectrum of digital FSK with line spectra, with nonoverlapping signal pulses having arbitrary correlation (rather than being independent, as in the remainder of the present paper). This is the only correlated case that can be readily treated by the present methods.

$$\{\ \} = \mathbf{w}_d - \mathbf{w}\rangle\cdot\underline{\mathbf{w}}, \quad K = 1. \tag{74}$$

$$\begin{aligned}\{\ \} = \mathbf{w}_d^{[2]} - \mathbf{w}\rangle^{[2]}\cdot\underline{\mathbf{w}}^{[2]} \\ + (\underline{\mathbf{w}} \times \mathbf{w}_d \times \mathbf{w}\rangle - \mathbf{w}\rangle^{[2]}\cdot\underline{\mathbf{w}}^{[2]})e^{-j2\pi(fT-f_l)} \\ + (\mathbf{w}\rangle \times \mathbf{w}_d \times \underline{\mathbf{w}} - \mathbf{w}\rangle^{[2]}\cdot\underline{\mathbf{w}}^{[2]})e^{+j2\pi(fT-f_l)}, \quad K = 2.\end{aligned} \tag{75}$$

§

$$\{\ \} = \begin{array}{l}\text{expression in } \{\ \} \text{ in (116) of} \\ \text{Ref. 1 with } fT \to fT - f_l,\end{array} \qquad \text{general } K. \tag{76}$$

The condition of (63), that has here been shown sufficient, is shown in Section VI also to be necessary for line components to be present in FSK spectra. This condition has a simple physical interpretation. Every signal pulse must introduce the same total phase change in the modulated carrier in order to have line components. This phase change has been denoted as $2\pi f_l$; the line components appear at frequencies

$$\pm\left(f_c + \frac{f_l + n}{T}\right), \quad n = \cdots, -1, 0, 1, \cdots. \tag{77}$$

When $f_l = 0$, (73) to (76) show that the FSK spectra are identical to the prior PSK results.[1] However the wave in this case is not necessarily a PSK wave. The stronger condition of (53) or (54) is required to have a PSK wave; this condition demands the net phase shift introduced by every signaling pulse to be zero. A wave with $f_l = 0$, but one or more signal pulses with net phase change equal, for example, to $\pm 2\pi$, will have a spectrum given by the PSK formula, but will not be a PSK wave.

VI. FSK WITH NO LINE COMPONENTS

Assume throughout this section that the inequality of (62) is satisfied:

$$|\underline{\mathbf{w}}\cdot\mathbf{q}(U_K)\rangle| < 1. \tag{78}$$

We demonstrate that under these conditions the FSK spectrum contains no lines. The properties of Kronecker products given in Ref. 1 in the second footnote, p. 908, and in the first footnote, p. 915, are used throughout without further comment.

6.1 FSK spectrum with no line components: nonoverlapping pulses, $K = 1$

From (37) and (60),

$$\tilde{\mathbf{\Phi}}_c(n) = \langle (S_{k+n}S_k^*)(\mathbf{a}_{k+n}\rangle\cdot\underline{\mathbf{a}_k})\rangle. \tag{79}$$

From (36),

$$S_{k+n}S_k^* = \sum_{i=0}^{n-1} \underline{\mathbf{a}_{k+i}}\cdot\mathbf{q}(T)\rangle, \quad n > 0. \tag{80}$$

---

§ See footnote to eq. (72).

Let us first consider $\tilde{\mathbf{\Phi}}_c(0)$. Since $S_k S_k^* = |S_k|^2 = 1$, for all k, we have from (79) and (28)

$$\tilde{\mathbf{\Phi}}_c(0) = \mathbf{w}_d. \tag{81}$$

Next, (79) and (80) yield

$$\tilde{\mathbf{\Phi}}_c(n) = \left\{ \prod_{i=1}^{n-1} \underline{a_{k+i}} \cdot \mathbf{q}(T) \right\} \{\underline{\mathbf{q}(T)} \cdot \mathbf{a}_k]\} \{\mathbf{a}_{k+n}] \cdot \underline{\mathbf{a}_k}\}, \quad n \geqq 1, \tag{82}$$

where we have split off the first factor of (80). For $n = 1$, the first $\{\ \}$, containing the $\prod$, is dropped. The first $\{\ \}$ is independent of the remainder of the expression by (19); using the fact that $\underline{\mathbf{q}(T)} \cdot \mathbf{a}_k]$ is a complex number or equivalently a 1×1 complex matrix, and using (27) and (28),

$$\begin{aligned} \tilde{\mathbf{\Phi}}_c(n) &= \{\underline{\mathbf{w}} \cdot \mathbf{q}(T)]\}^{n-1} \langle \mathbf{a}_{k+n}] \cdot \underline{\mathbf{q}(T)} \cdot \mathbf{a}_k] \cdot \underline{\mathbf{a}_k} \rangle \\ &= \{\underline{\mathbf{w}} \cdot \mathbf{q}(T)]\}^{n-1} \{\mathbf{w}] \cdot \underline{\mathbf{q}(T)} \cdot \mathbf{w}_d\}, \quad n \geqq 1, \end{aligned} \tag{83}$$

the last step following from the independence of $\mathbf{a}_{k+n}$ and $\mathbf{a}_k$.

Finally, taking the Hermitian transpose[§] of (79),

$$\tilde{\mathbf{\Phi}}_c^\dagger(n) = \langle (S_{k+n}^* S_k)(\mathbf{a}_k] \cdot \underline{\mathbf{a}_{k+n}}) \rangle. \tag{84}$$

Alternatively, setting $n \to -n$ in (79),

$$\tilde{\mathbf{\Phi}}_c(-n) = \langle (S_k^* S_{k-n})(\mathbf{a}_{k-n}] \cdot \underline{\mathbf{a}_k}) \rangle. \tag{85}$$

Since (79) has been shown independent of k, we may set $k \to k + n$ in (85), to yield upon comparison with (84)

$$\tilde{\mathbf{\Phi}}_c(-n) = \tilde{\mathbf{\Phi}}_c^\dagger(n). \tag{86}$$

From (58) to (61), (81), (83), and (86),

$$\tilde{\mathbf{P}}_c(fT) = \mathbf{A} + \mathbf{A}^\dagger, \tag{87}$$

$$\mathbf{A} = \tfrac{1}{2}\mathbf{w}_d + e^{-j2\pi fT}\mathbf{w}] \cdot \underline{\mathbf{q}(T)} \cdot \mathbf{w}_d \sum_{n=1}^{\infty} \{e^{-j2\pi fT} \underline{\mathbf{w}} \cdot \mathbf{q}(T)]\}^{n-1}. \tag{88}$$

Because of (78) the geometric series in (88) converges, and

$$\mathbf{A} = \frac{1}{2}\mathbf{w}_d + \frac{e^{-j2\pi fT}\,\mathbf{w}] \cdot \underline{\mathbf{q}(T)} \cdot \mathbf{w}_d}{1 - e^{-j2\pi fT} \underline{\mathbf{w}} \cdot \mathbf{q}(T)]}. \tag{89}$$

Finally, the spectrum of the complex FSK wave of (4) is

$$P_v(f) = \frac{1}{T} \underline{\mathbf{R}(f)} \cdot (\mathbf{A} + \mathbf{A}^\dagger) \cdot \mathbf{R}^*(f)], \tag{90}$$

with $\mathbf{A}$ given by (89) for nonoverlapping pulses, $K = 1$.

---

§ See footnote to (14).

It is clear that the restriction of (78) renders $\mathbf{A}$ and, hence, $P_v(f)$ finite for all f; consequently, there can be no line components in the FSK spectrum. If we take the limit as $|\underline{\mathbf{w}}\cdot\mathbf{q}(T)]| \to 1$, and substitute (66) into (88), (90) yields directly the appropriate results for the line case, i.e., (72) and (73) to (74) of Section V.

6.2 FSK spectrum with no line components: overlapping pulses, $K = 2$

$$\tilde{\mathbf{\Phi}}_c(n) = \langle (S_{k+n}S_k^*)(\mathbf{b}_k]\cdot\underline{\mathbf{b}_k}) \rangle \tag{91}$$

with

$$\underline{\mathbf{b}_k} \equiv \underline{\mathbf{a}_k} \times \underline{\mathbf{a}_{k+1}}. \tag{92}$$

We have noted in Section 4.4. that the present $\mathbf{b}_k$ for FSK has identical form to the prior $\mathbf{b}_k$ of Ref. 1 for PSK. In Section 4.2, we saw that S_k is identical for $K = 1$ and $K = 2$; therefore, (80) applies to the present case as well. Comparing the PSK analysis of Section VII of Ref. 1, we evaluate (91) above by inserting (80) above inside the $\langle\ \rangle$ in (87) of Ref. 1.

For $n = 0$, (87) and (90) of Ref. 1 yield

$$\tilde{\mathbf{\Phi}}_c(0) = \mathbf{w}_d^{[2]}. \tag{93}$$

For $n = 1$, (87) and (91) of Ref. 1 yield

$$\begin{aligned}\tilde{\mathbf{\Phi}}_c(1) &= \langle \{\underline{\mathbf{q}(T)}\cdot\mathbf{a}_k]\} \times \underline{\mathbf{a}_k} \times \{\mathbf{a}_{k+1}] \times \underline{\mathbf{a}_{k+1}}\} \times \mathbf{a}_{k+2}] \rangle \\ &= \langle \{\underline{\mathbf{q}(T)}\cdot\mathbf{a}_k]\cdot\underline{\mathbf{a}_k}\} \times \{\mathbf{a}_{k+1}]\cdot\underline{\mathbf{a}_{k+1}}\} \times \mathbf{a}_{k+2}] \rangle. \end{aligned} \tag{94}$$

Since $\mathbf{a}_k$, $\mathbf{a}_{k+1}$, and $\mathbf{a}_{k+2}$ are independent, (27) and (28) yield

$$\tilde{\mathbf{\Phi}}_c(1) = \{\underline{\mathbf{q}(T)}\cdot\mathbf{w}_d\} \times \mathbf{w}_d \times \mathbf{w}]. \tag{95}$$

Finally, substituting (80) inside the $\langle\ \rangle$ of the third line of (87) of Ref. 1,

$$\begin{aligned}\tilde{\mathbf{\Phi}}_c(n) = \Bigg\langle \Bigg\{ \prod_{i=2}^{n-1} \underline{a_{k+i}}\cdot\mathbf{q}(T)] \Bigg\} \{\underline{\mathbf{q}(T)}\cdot\mathbf{a}_k]\}\{\underline{\mathbf{q}(T)}\cdot\mathbf{a}_{k+i}]\}\{\mathbf{a}_{k+n}]\cdot\underline{\mathbf{a}_k}\} \\ \times \{\mathbf{a}_{k+n+i}]\cdot\underline{\mathbf{a}_{k+i}}\} \Bigg\rangle, \quad n \geqq 2, \end{aligned} \tag{96}$$

where we have split off the first two factors of (80). For $n = 2$, the first $\{\ \}$, containing the $\prod$, is dropped. Regarding the factors $\underline{\mathbf{q}(T)}\cdot\mathbf{a}_k]$ and $\underline{\mathbf{q}(T)}\cdot\mathbf{a}_{k+1}]$ as complex numbers or alternatively as 1×1 complex matrices, and using the independence of the different $\mathbf{a}_k$, (96) yields

$$\begin{aligned}\tilde{\mathbf{\Phi}}_c(n) &= \{\underline{\mathbf{w}}\cdot\mathbf{q}(T)]\}^{n-2}\langle \{\mathbf{a}_{k+n}]\cdot\underline{\mathbf{q}(T)}\cdot\mathbf{a}_k]\cdot\underline{\mathbf{a}_k}\} \\ &\qquad \times \{\mathbf{a}_{k+n+i}]\cdot\underline{\mathbf{q}(T)}\cdot\mathbf{a}_{k+i}]\cdot\underline{\mathbf{a}_{k+i}}\} \rangle \\ &= \{\underline{\mathbf{w}}\cdot\mathbf{q}(T)]\}^{n-2}\{\mathbf{w}]\cdot\underline{\mathbf{q}(T)}\cdot\mathbf{w}_d\}^{[2]}, \quad n \geqq 2. \end{aligned} \tag{97}$$

The remainder of the analysis proceeds as in Section 6.1. The FSK spectrum for overlapping pulses with $K = 2$ is given by

$$P_v(f) = \frac{1}{T}\underline{\mathbf{R}(f)}\cdot(\mathbf{A} + \mathbf{A}^\dagger)\cdot\mathbf{R}^*(f)], \tag{98}$$

where

$$\mathbf{A} = \frac{1}{2}\mathbf{w}_d^{[2]} + e^{-j2\pi fT}\{\underline{\mathbf{q}(T)}\cdot\mathbf{w}_d\} \times \mathbf{w}_d \times \mathbf{w}] + \frac{\{e^{-j2\pi fT}\mathbf{w}]\cdot\underline{\mathbf{q}(T)}\cdot\mathbf{w}_d\}^{[2]}}{1 - e^{-j2\pi fT}\underline{\mathbf{w}}\cdot\mathbf{q}(T)]}. \tag{99}$$

$P_v(f)$ again contains no spectral lines by (78); the limit $\underline{\mathbf{w}}\cdot\mathbf{q}(T)] \to 1$ again yields the appropriate results of Section V for the line case.

6.3 FSK spectrum with no line components: overlapping pulses, general K

Similar analysis yields the generalization to any overlapping signal pulses. From (46) and (60),

$$\tilde{\mathbf{\Phi}}_c(n) = \langle(S_{k+n}S_k^*)(\mathbf{b}_k]\cdot\underline{\mathbf{b}_k})\rangle \tag{100}$$

with $\mathbf{b}_k$ given by (46) and S_k by (48) to (49). Thus,

$$S_{k+n}S_k^* = \begin{cases} \prod_{j=1}^{n} \underline{\mathbf{a}_{k-((K+1)/2)+j}}\cdot\mathbf{q}\left(\dfrac{K+1}{2}T\right)\Big], & k \text{ odd}, \\ \prod_{j=1}^{n} \underline{\mathbf{a}_{k-(K/2)+j}}\cdot\mathbf{q}\left(\dfrac{K}{2}T\right)\Big], & k \text{ even}, \end{cases} \qquad n > 0. \tag{101}$$

Since the present $\mathbf{b}_k$ have identical form to $\mathbf{b}_k$ of Ref. 1, (100) is evaluated by substituting (101) inside the $\langle\ \rangle$ of (106) of Ref. 1, and making corresponding changes in the remainder of Section VII, Ref. 1. The factor

$$\prod_{j=1}^{n}\underline{\mathbf{a}_j}\cdot\mathbf{q}(U_K)] \tag{102}$$

is inserted inside the $\langle\ \rangle$ of (107) of Ref. 1. In the following we recall that the factors of (102), $\underline{\mathbf{a}}\cdot\mathbf{q}] = \underline{\mathbf{q}}\cdot\mathbf{a}]$, may be regarded alternatively as complex numbers or as 1×1 complex matrices.

In (109) of Ref. 1, the first factor of the second line is modified as

$$\langle\underline{\mathbf{a}_i}\rangle \to \underline{\mathbf{q}(U_K)}\cdot\langle\mathbf{a}_i]\cdot\underline{\mathbf{a}_i} = \underline{\mathbf{q}(U_K)}\cdot\mathbf{w}_d; \tag{103}$$

consequently,

$$\tilde{\mathbf{\Phi}}_c(1) = \{\underline{\mathbf{q}(U_K)}\cdot\mathbf{w}_d\} \times \mathbf{w}_d^{[K-1]} \times \mathbf{w}]. \tag{104}$$

In (110), Ref. 1, the first two factors of the fourth line are modified

§ U_K is given by eq. (31).

as in (103), yielding

$$\tilde{\mathbf{\Phi}}_c(2) = \{\underline{\mathbf{q}(U_K)} \cdot \mathbf{w}_d\}^{[2]} \times \mathbf{w}_d^{[K-2]} \times \mathbf{w}]^{[2]}. \tag{105}$$

By induction,

$$\tilde{\mathbf{\Phi}}_c(n) = \{\underline{\mathbf{q}(U_K)} \cdot \mathbf{w}_d\}^{[n]} \times \mathbf{w}_d^{[K-n]} \times \mathbf{w}]^{[n]}, \quad n \leqq K. \tag{106}$$

Next, for $n \geqq K$, inserting (102) inside the $\langle\ \rangle$ of the second line of (107) of Ref. 1, associating the factors of (102) with corresponding factors of the second $\prod\times$ of (107), Ref. 1, with $n - K$ factors of (102) left over, and then noting that all factors have different indices and, hence, are independent,

$$\begin{aligned}\tilde{\mathbf{\Phi}}_c(n) &= \mathbf{w}]^{[K]} \cdot \{\underline{\mathbf{q}(U_K)} \cdot \mathbf{w}_d\}^{[K]} \{\underline{\mathbf{q}(U_K)} \cdot \mathbf{w}]\}^{n-K} \\ &= \{\underline{\mathbf{w}} \cdot \mathbf{q}(U_K)]\}^{n-K} \{\mathbf{w}] \cdot \underline{\mathbf{q}(U_K)} \cdot \mathbf{w}_d\}^{[K]}, \quad n \geqq K. \end{aligned} \tag{107}$$

Finally, from (108) of Ref. 1 and (100) above,

$$\tilde{\mathbf{\Phi}}_c(0) = \mathbf{w}_d^{[K]}. \tag{108}$$

The FSK spectrum for overlapping pulses with general K is given by

$$P_v(f) = \frac{1}{T} \underline{\mathbf{R}(f)} \cdot (\mathbf{A} + \mathbf{A}^\dagger) \cdot \mathbf{R}^*(f)], \tag{109}$$

where

$$\begin{aligned}\mathbf{A} = \frac{1}{2} \mathbf{w}_d^{[K]} + \sum_{n=1}^{K-1} e^{-jn2\pi fT} \underline{\mathbf{q}(U_K)} \cdot \mathbf{w}_d^{[n]} \times \mathbf{w}_d^{[K-n]} \times \mathbf{w}]^{[n]} \\ + \frac{\{e^{-j2\pi fT}\mathbf{w}] \cdot \underline{\mathbf{q}(U_K)} \cdot \mathbf{w}_d\}^{[K]}}{1 - e^{-j2\pi fT} \underline{\mathbf{w}} \cdot \mathbf{q}(U_K)]}. \end{aligned} \tag{110}$$

The condition (78) again guarantees no spectral lines; as equality is approached in (78) the present results approach those of Section V for the line case. The condition (63) is, therefore, necessary and sufficient for $P_v(f)$ to have line components.

VII. ILLUSTRATIVE EXAMPLES

The computation of the digital FM spectral density from the above methods is straightforward. For a given set of baseband signaling pulse shapes and symbol probability distribution, we determine K, the overlap parameter, and the probability row vector $\underline{\mathbf{w}}$. We then evaluate $\mathbf{q}(U_K)$ and $\underline{\mathbf{w}} \cdot \mathbf{q}(U_K)]$.

If $|\underline{\mathbf{w}} \cdot \mathbf{q}(U_K)| < 1$, we know that there are no line components in $P_v(f)$. The continuous part of the spectrum is evaluated from the appropriate Hermitian form given in Section VI.

If $|\underline{\mathbf{w}} \cdot \mathbf{q}(U_K)]| = 1$, we know that there are line components in the spectrum, and also that

$$\mathbf{q}(U_K)] = e^{j2\pi f_l} \cdot \mathbf{1}], \quad |f_l| \leqq \tfrac{1}{2}. \tag{111}$$

We determine f_l, and then $P_{v_l}(f)$ and $P_{v_c}(f)$ from the methods given in Section V.

In the following examples, the digital computer is programmed to work directly with the Hermitian forms (both ordinary and Kronecker matrix multiplications are performed by the computer). In this way, complicated cases involving multilevel signal pulses overlapping several time slots may be simply treated.

The case of rectangular-pulse FSK modulation is treated in Ref. 3, and consequently will not be considered here.§

We make the following assumptions for convenience; none are essential.

(*i*) The number of frequency levels is a power of 2,

$$M = 2^N, \quad N \text{ an integer.} \tag{112}$$

(*ii*) The M baseband signaling pulses have a common shape;

$$g_k(t) = C_k g(t). \tag{113}$$

(*iii*) All signal pulses are equally likely;

$$\mathbf{w}] = \frac{1}{M}\,\mathbf{1}], \quad \mathbf{w}_d = \frac{1}{M}\,\mathbf{I}_M, \tag{114}$$

where $\mathbf{I}_M$ is the identity matrix of order M.

7.1 Raised-cosine nonoverlapping signal pulses: $K = 1$

If the pulses have a common raised-cosine shape,

$$\mathbf{g}(t)] = \begin{cases} \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_M \end{bmatrix} \pi f_d \left[1 - \cos\frac{2\pi t}{T}\right], & 0 < t \leqq T, \quad f_d > 0, \\ \mathbf{0}], & \text{otherwise,} \end{cases} \tag{115}$$

where $\Delta_1, \Delta_2, \cdots, \Delta_M$ are the peak-frequency-deviation parameters of the FSK signals. We assume that $\Delta_1 = 1$, $\Delta_2 = -1$, $\Delta_3 = 3$, $\Delta_4 = -3$, $\cdots$, $\Delta_M = -(M-1)$.

Since $K = 1$,

$$R_k(f) = T \sum_{n=-\infty}^{\infty} J_n\left(\frac{f_d T \Delta_k}{2}\right) e^{-j\pi[f-(f_d\Delta_k/2)]T} \cdot \frac{\sin \pi\left(fT - \dfrac{f_d\Delta_k T}{2} + n\right)}{\pi\left(fT - \dfrac{f_d\Delta_k T}{2} + n\right)}, \tag{116}$$

where $J_n(x)$ is the Bessel function of the first kind and of order n.

§ This is the only discrete-frequency-modulation spectrum given in Ref. 3.

7.1.1 Raised-cosine signaling with no line spectrum: $K = 1$

Since

$$\underline{\mathbf{q}(U_1)} = \underbrace{e^{j\Delta_1 \pi f_d T}\, e^{j\Delta_2 \pi f_d T} \cdots e^{j\Delta_M \pi f_d T}}, \tag{117}$$

line spectra are absent if and only if $f_d T$ is not an integer. In this case,

$$P_v(f) = P_{v_c}(f) = \frac{1}{T}\underline{\mathbf{R}(f)} \cdot (\mathbf{A} + \mathbf{A}^\dagger) \cdot \underline{\mathbf{R}(f)}^\dagger, \tag{118}$$

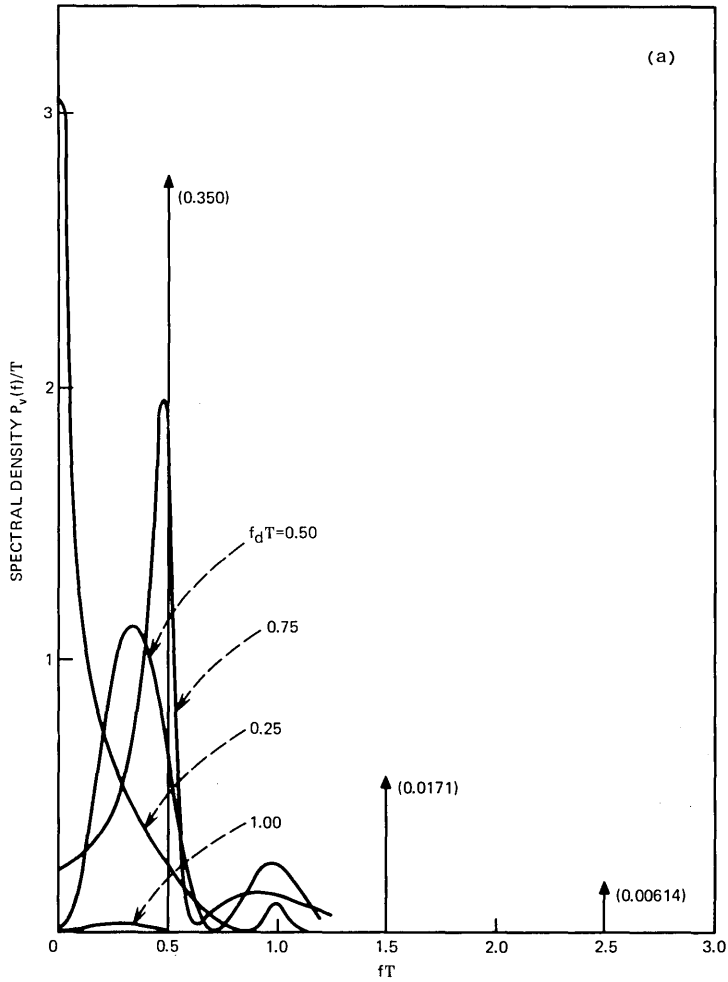


Fig. 2—Spectral density of binary FSK system with raised-cosine signaling and pulse duration T. $K = 1$. $2f_d$ is the spacing between two adjacent *a priori* chosen frequencies.

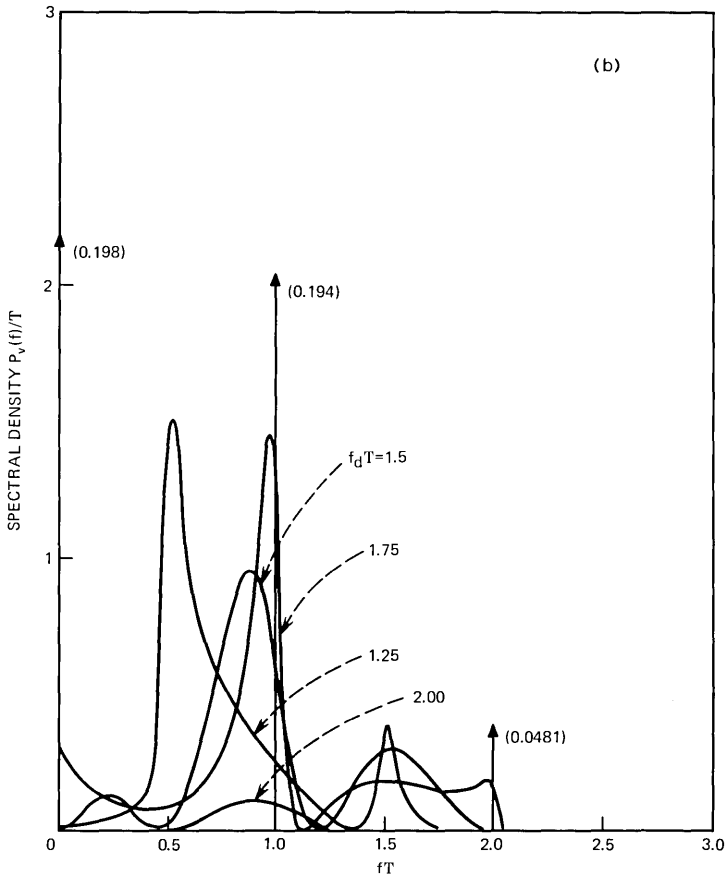


Fig. 2 (continued).

where $\mathbf{R}(f)]$ is given by (116) and

$$\mathbf{A} = \frac{1}{2M}\mathbf{I}_M + \frac{1}{M^2}\frac{e^{-j2\pi fT}\,\mathbf{1}]\cdot\underline{\mathbf{q}(U_1)}}{1 - e^{-j2\pi fT}\,\frac{1}{M}\,\frac{\sin M\pi f_dT}{\sin \pi f_dT}}. \tag{119}$$

For $M = 2$, 4, and 8 and various values of $k = f_dT$, $P_v(f)$ is plotted in Figs. 2, 3, and 4.

7.1.2 Raised-cosine signaling with line spectrum: K = 1

The FM spectral density $P_v(f)$ contains lines if and only if $|\underline{\mathbf{w}}\cdot\mathbf{q}(U_1)]| = 1$; that is, if and only if

$$f_dT = 1, 2, 3, \cdots. \tag{120}$$

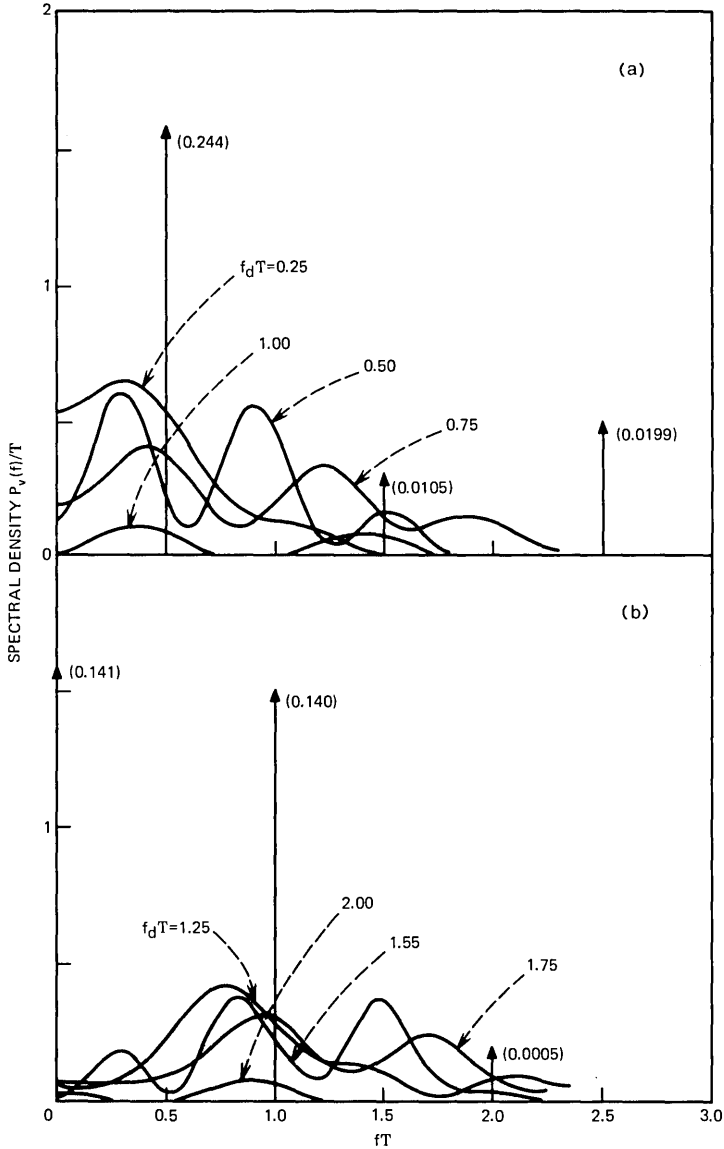


Fig. 3—Spectral density of quaternary FSK system with raised-cosine signaling and pulse duration T. $K = 1.2f_d$ is the spacing between two adjacent *a priori* chosen frequencies.

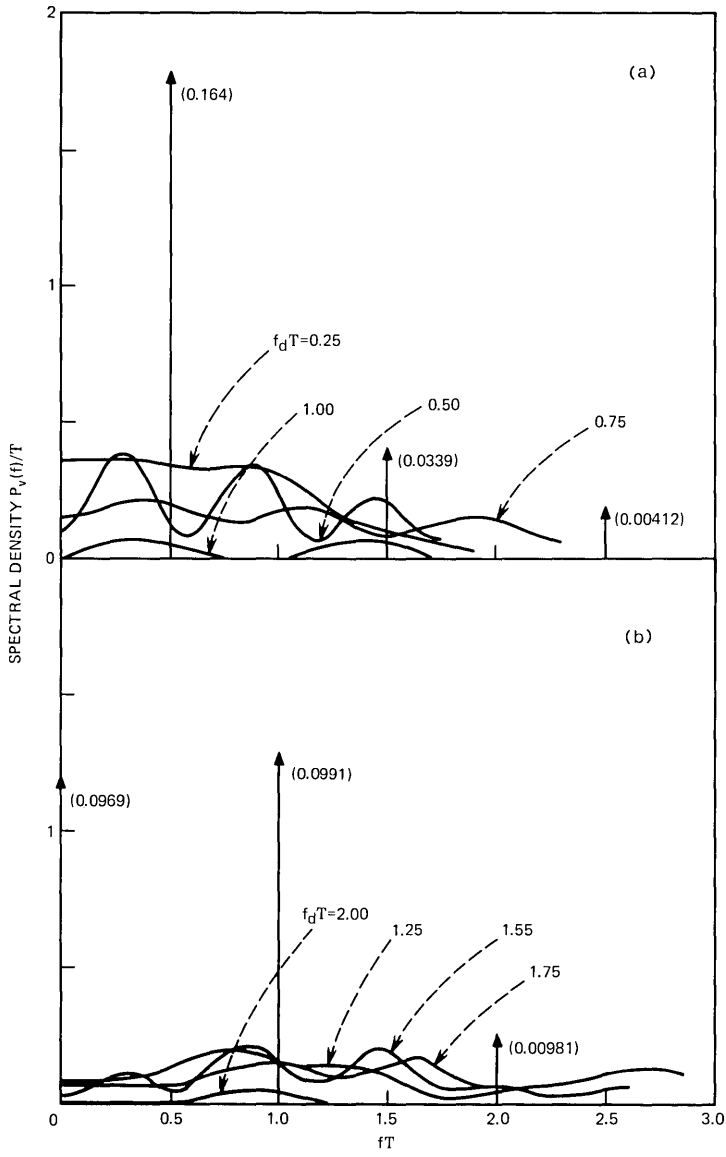


Fig. 4—Spectral density of octonary FSK system with raised-cosine signaling and pulse duration T. $K = 1.2f_d$ is the spacing between two adjacent *a priori* chosen frequencies.

In this case,

$$f_l = \tfrac{1}{2}, \quad f_d T = 1, 3, 5, \cdots, \tag{121}$$

$$f_l = 0, \quad f_d T = 2, 4, 6, \cdots. \tag{122}$$

From Section V,

$$P_v(f) = \frac{1}{T^2}\frac{1}{M^2}\,|\underline{1}\cdot\mathbf{R}(f)]|^2 \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n + f_l}{T}\right) + \frac{1}{T}\underline{\mathbf{R}(f)}\cdot(\mathbf{B} + \mathbf{B}^\dagger)\cdot\underline{\mathbf{R}(f)}^\dagger, \tag{123}$$

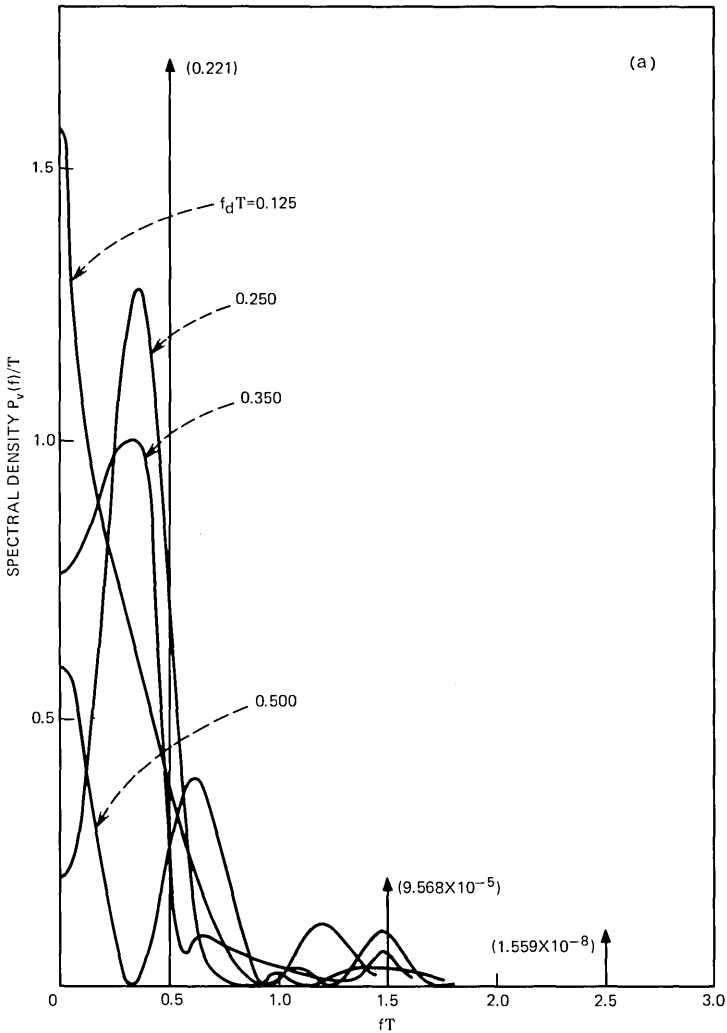


Fig. 5—Spectral density of binary FSK system with raised-cosine signaling and pulse duration $2T$. $K = 2.2f_d$ is the spacing between two adjacent *a priori* chosen frequencies.

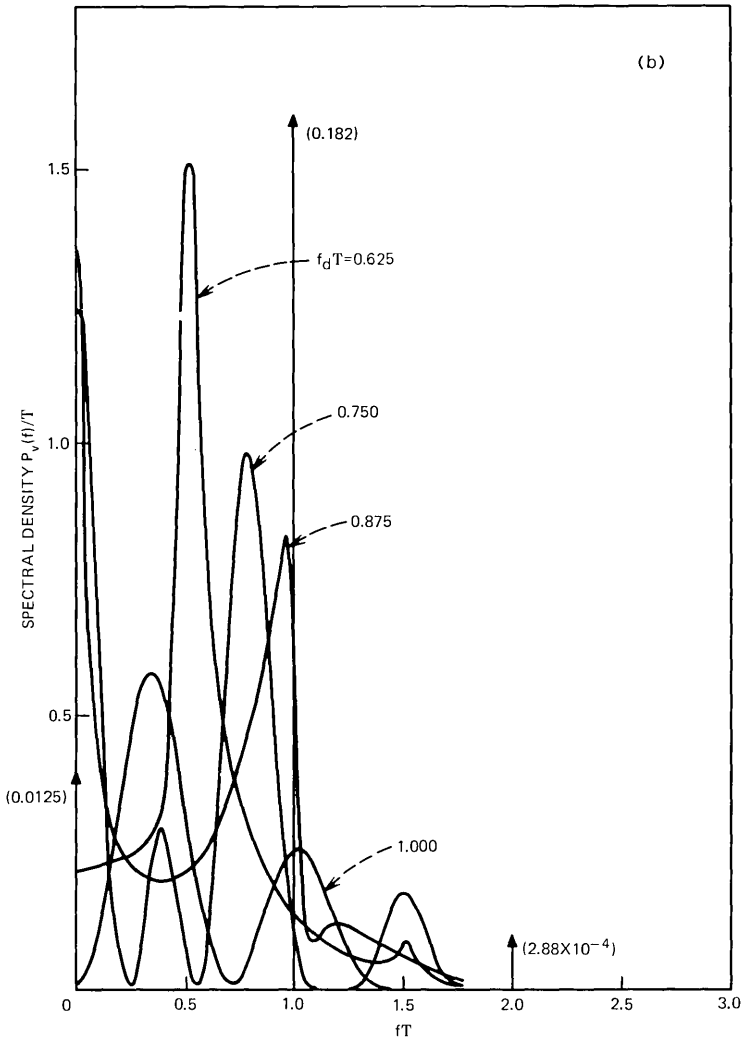


Fig. 5 (continued).

where§

$$\mathbf{B} + \mathbf{B}^\dagger = \frac{1}{M}\mathbf{I}_M - \frac{1}{M^2}\mathbf{1}\rceil\cdot\underline{1}. \tag{124}$$

For $M = 2$, 4, and 8 and for some integral values of f_dT, $P_v(f)$ is also plotted in Figs. 2, 3, and 4.

§ Note that $\mathbf{B} + \mathbf{B}^\dagger$ for $K = 1$ is given in (74).

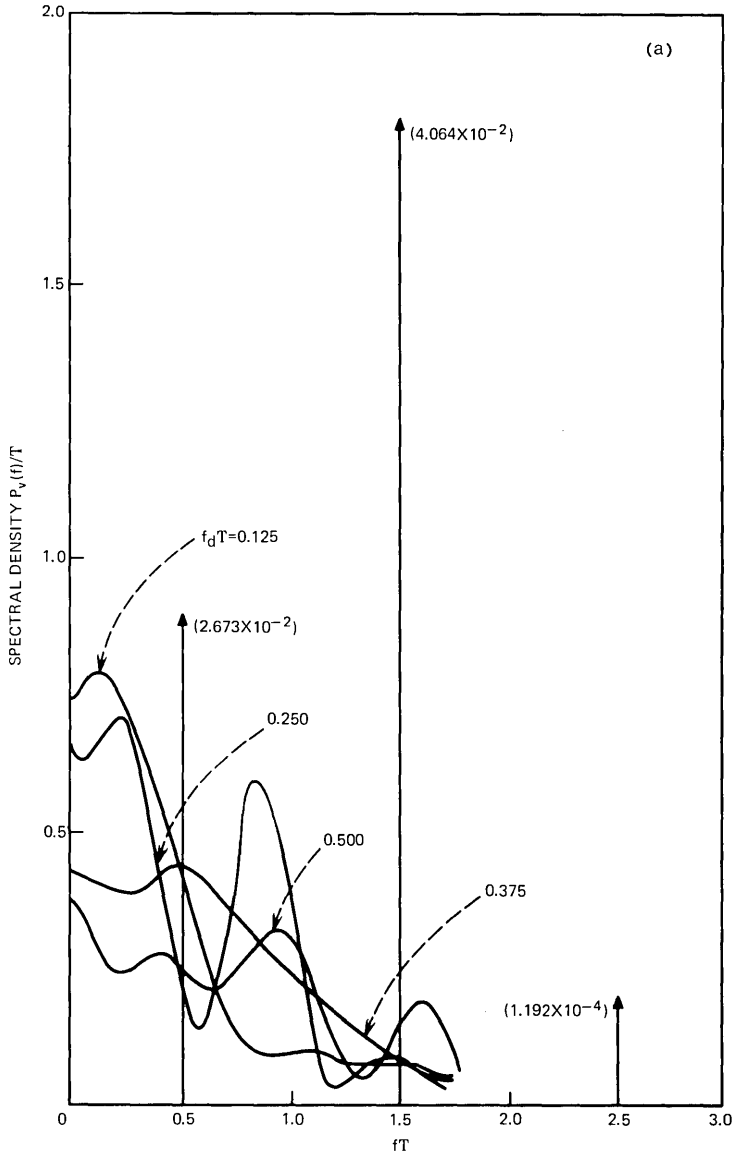


Fig. 6—Spectral density of quaternary FSK system with raised-cosine signaling and pulse duration $2T$. $K = 2.2f_d$ is the spacing between two adjacent *a priori* chosen frequencies.

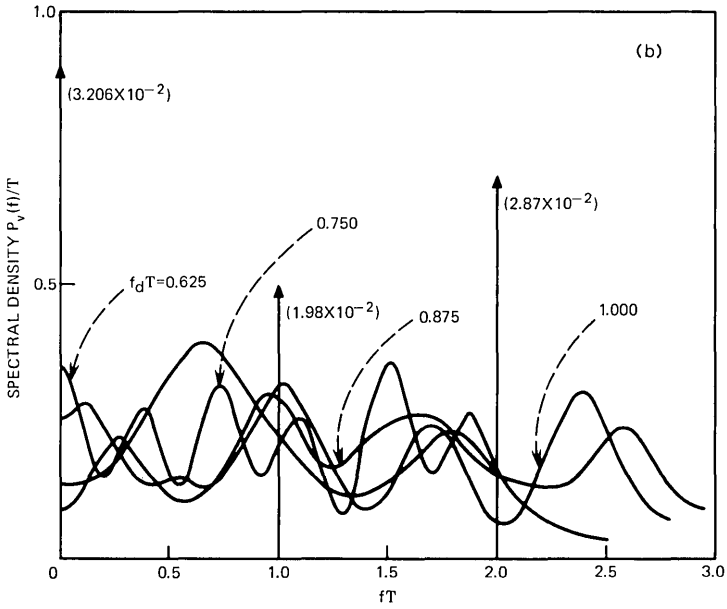


Fig. 6 (continued).

7.2 Raised cosine overlapping signal pulses: K = 2

If a raised-cosine signal pulse just fills up two time slots,

$$\mathbf{g}(t)] = \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_M \end{bmatrix} \pi f_d \left(1 + \cos \frac{\pi t}{T}\right), \quad -T < t \leqq T, \quad f_d > 0$$

$$0], \qquad \text{otherwise,} \tag{125}$$

$K = 2$, and the spectral density may be calculated from Sections 4.2, V, or 6.2 according to whether

$$|\underline{\mathbf{w}} \cdot \mathbf{q}(U_2)]| = 1 \tag{126}$$

or

$$|\underline{\mathbf{w}} \cdot \mathbf{q}(U_2)| < 1. \tag{127}$$

If $\Delta_1 = 1$, $\Delta_2 = -1$, $\Delta_3 = 3$, $\Delta_4 = -3$, $\cdots$, $\Delta_{M-1} = (M-1)$, $\Delta_M = -(M-1)$, (116) can be shown to be satisfied if and only if $2f_dT$ is an integer. Further,

$$f_l = 0, \quad \text{if } f_dT = 1, 2, 3, \cdots, \tag{128}$$

and

$$f_l = \tfrac{1}{2}, \quad \text{if } f_dT = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots. \tag{129}$$

In this case, note that

$$\underline{\mathbf{w}}\cdot\mathbf{q}(T)] = \frac{1}{M}\frac{\sin(2M\pi f_d T)}{\sin(2\pi f_d T)}. \tag{130}$$

In this case, the FSK spectrum contains line components, and the continuous and line spectra for $M = 2$ and 4 are shown in Figs. 5 and 6.

If $2f_dT$ is not an integer, the FSK spectrum does not contain any lines, and the continuous spectrum given by (97) to (98) for $M = 2$ and 4 is also plotted in Figs. 5 and 6.

Several observations can be made from Figs. 2 to 6. For both $K = 1$ and $K = 2$, discrete spectral lines appear as f_dT approaches the limiting value and the power in the lines is substantial for $K = 1$ and $M = 2$. Also note that power in the line components with $K = 2$ is smaller than the power in the lines with $K = 1$.

For the same value of f_dT and K, the principal portion of the spectrum of binary FSK is narrower than that of quaternary FSK. For $K = 1$, quaternary FSK spectrum is narrower than that of octonary FSK spectrum for the same value of f_dT.

Since lines can appear in the spectrum for a set of values of f_dT, the FSK spectrum is quite different from the PSK spectrum even when lines are present in the FSK spectrum.

VIII. SUMMARY AND CONCLUSIONS

Matrix methods are given to express the spectral density of a carrier, frequency-modulated by a random baseband pulse train, in a concise and computable form. Arbitrary pulse shapes may be used for M-ary digital signaling, and they may overlap over a finite number of signal intervals.

The spectral density is expressed as a compact Hermitian form suitable for numerical computation by a digital computer. The computer is readily programmed to perform matrix operations directly, rather than expanding the Hermitian form and evaluating the individual terms. In this way, quite complicated cases involving multilevel signal pulses overlapping several time slots may be treated simply.

Simple conditions in terms of the modulation parameters are given under which discrete spectral lines are present in the spectrum. A method is given to evaluate the power in the discrete spectral lines. The utility of the method is illustrated by giving several examples.

The present results are restricted to independent signal pulses, except for the very special case of nonoverlapping ($K = 1$) signal pulses with line spectra present [see footnote following (72)]. In Ref. 1 we saw that for overlapping signal pulses with $K = 2$ (i.e., signal pulses that occupy no more than two time slots), PSK with correlated

signal pulses required the fourth-order statistics of the digital modulation process. FSK is more complicated; the correlated $K = 2$ case requires statistics of *all* orders for the modulation process, except in the line spectrum case. However, certain special cases with correlated signal pulses have been handled by extension of the present methods.

APPENDIX

Spectra of Complex and Real FSK Waves

From (130), (131) of Ref. 1, Appendix A, we see that (13) of the present paper holds true if

$$\overline{e^{j4\pi f_c t}\Phi_{vv^*}(t, \tau)} = 0, \tag{131}$$

where

$$\Phi_{vv^*}(t, \tau) = \langle v(t + \tau)v(t)\rangle = \langle e^{j[\phi(t+\tau)+\phi(t)]}\rangle. \tag{132}$$

From (3) and (9),

$$\Phi_{vv^*}(t, \tau) = \left\langle e^{j2\phi(0)} \exp\left(j \sum_{k=-\infty}^{\infty} \left[\int_{-kT}^{t+\tau-kT} + \int_{-kT}^{t-kT}\right] h_{s_k}(\mu)d\mu\right)\right\rangle. \tag{133}$$

If $\phi(0)$ is independent of the modulation parameters s_k, (133) becomes

$$\Phi_{vv^*}(t, \tau) = \langle e^{j2\phi(0)}\rangle \left\langle \exp\left(j \sum_{k=-\infty}^{\infty} \left[\int_{-kT}^{t+\tau-kT} + \int_{-kT}^{t-kT}\right] h_{s_k}(\mu)d\mu\right)\right\rangle. \tag{134}$$

Thus, if (11) holds, the first factor in (134) $= 0$, and (131) and, hence, (13) follow immediately.

If, instead, (10) holds, and imposing in addition the independence of the s_k, (134) becomes

$$\Phi_{vv^*}(t, \tau) = \prod_{k=-\infty}^{\infty} \left\langle \exp\left(j\left[\int_{-kT}^{t+\tau-kT} + \int_{-kT}^{t-kT}\right] h_{s_k}(\mu)d\mu\right)\right\rangle. \tag{135}$$

For large enough $|t|$ and finite pulse length, the integrals in the exponent fall into three classes, depending on k:

(*i*) Both limits on both integrals lie to one side of the signal pulses. The integrals equal 0 and the corresponding factor in the infinite product equals 1 and, hence, may be ignored.

(*ii*) The limits on integrals straddle the pulse, and the limits may consequently be replaced by $\int_{L_K}^{U_K}$, where the maximum pulse length is KT and L_K, U_K are given in (31).

(*iii*) The common lower limit, or one or both upper limits, lie within the pulse.

Therefore, for large enough $|t|$ (depending on the pulse length and on τ)

$$\Phi_{vv^*}(t + T, \tau) = \Phi_{vv^*}(t, \tau) \left\langle \exp\left(j2 \int_{L_K}^{U_K} h_{s_k}(\mu) d\mu\right)\right\rangle, \quad t \text{ sufficiently positive.} \tag{136}$$

$$\Phi_{vv^*}(t - T, \tau) = \Phi_{vv^*}(t, \tau) \left\langle \exp\left(-j2 \int_{L_K}^{U_K} h_{s_k}(\mu) d\mu\right)\right\rangle, \quad t \text{ sufficiently negative.} \tag{137}$$

Condition (12) may be substituted for condition (10) with only minor changes in the above discussion, and identical results (136) and (137). From (44),

$$\mathfrak{q}(U_K)] = \begin{bmatrix} \exp\left(j \int_{L_K}^{U_K} h_1(\mu) d\mu\right) \\ \exp\left(j \int_{L_K}^{U_K} h_2(\mu) d\mu\right) \\ \vdots \\ \exp\left(j \int_{L_K}^{U_K} h_M(\mu) d\mu\right) \end{bmatrix}; \tag{138}$$

the integrals in the exponents are the areas of the signal pulses. We consider the two cases of Sections V and VI:

(*i*) All pulse areas are identical (mod 2π); line components are present in the spectrum.
(*ii*) The contrary; no line components are present.

In case (*i*), from (6)

$$\mathfrak{q}(U_K)] = e^{j2\pi f_l \cdot \mathbf{1}}], \quad -\tfrac{1}{2} < f_l \leqq \tfrac{1}{2}; \tag{139}$$

$2\pi f_l$ is the common area (mod 2π) of all the signal pulses and $\mathbf{1}]$ is the unit column vector of (23). It therefore follows that the final factors of (136) and (137) satisfy the following:

$$\left\langle \exp\left(\pm j2 \int_{L_K}^{U_K} h_{s_k}(\mu) d\mu\right)\right\rangle = e^{\pm j4\pi f_l}, \quad \text{line spectrum present.} \tag{140}$$

$$\left|\left\langle \exp\left(\pm j2 \int_{L_K}^{U_K} h_{s_k}(\mu) d\mu\right)\right\rangle\right| < 1, \quad \text{line spectrum absent.} \tag{141}$$

With line spectra absent, (141), (136), and (137) show immediately that (131) holds for all f_c, and, hence, condition (10) or (12) guarantees the result of (13) for independent s_k without further restriction.

With line components present, (140), (136), and (137) show that $\Phi_{vv^*}(t, \tau)$ regarded as a function of t contains a periodic (line) com-

ponent. Therefore, we write

$$\Phi_{vv^*}(t, \tau) = \Phi_{vv^*}(t, \tau)_l + \Phi_{vv^*}(t, \tau)_c, \tag{142}$$

where

$$\Phi_{vv^*}(t, \tau)_c = 0, \quad |t| \text{ sufficiently large}, \tag{143}$$

and

$$\Phi_{vv^*}(t, \tau)_l = e^{j4\pi f_l(t/T)} \sum_{n=-\infty}^{\infty} \varphi_n(\tau) e^{jn2\pi(t/T)}. \tag{144}$$§

Only the line component (144) can possibly contribute to the average on the left-hand side of (131):

$$\begin{aligned}\overline{e^{j4\pi f_c t}\Phi_{vv^*}(t, \tau)} &= \sum_{n=-\infty}^{\infty} \varphi_n(\tau) \lim_{A\to\infty} \frac{1}{2A} \int_{-A}^{A} e^{j2\pi[2f_c+(2f_l+n)/T]t} dt \\ &= \sum_{n=-\infty}^{\infty} \varphi_n(\tau) \lim_{A\to\infty} \frac{\sin 2\pi[2f_c + (2f_l + n)/T]A}{2\pi[2f_c + (2f_l + n)/T]A}. \end{aligned} \tag{145}$$

The $\lim_{A\to\infty} \to 0$ if $2f_c + (2f_l + n)/T \neq 0$ for every integer n, thus satisfying (131). Hence, condition (10) or (12) and

$$2\left(f_c + \frac{f_l}{T}\right) \neq \frac{n}{T}, \quad n = 0, 1, 2, 3, \cdots, \tag{146}$$

guarantees the result of (13) for independent s_k. Since, by (72), the line spectral components of $v(t)$ occur at $(f_l + n)/T$, (146) is equivalent to requiring that the line components of $v(t)e^{j2\pi f_c t}$ and of $v^*(t)e^{-j2\pi f_c t}$ never coincide. Note that condition (5) of Ref. 1 is the special case of (146) above for $f_l = 0$.

Finally, (135) to (140) of Ref. 1, Appendix A apply also in the present case, and establish that (13) is a good approximation if f_c is high enough so that there is no significant overlap between the two terms of (13).

REFERENCES

1. V. K. Prabhu and H. E. Rowe, "Spectra of Digital PM by Matrix Methods," B.S.T.J., *53*, No. 5 (May–June 1974), pp. 899–935.
2. W. R. Bennett and S. O. Rice, "Spectral Density and Autocorrelation Functions Associated with Binary Frequency Shift Keying," B.S.T.J., *42*, No. 5 (September 1963), pp. 2355–2385.
3. R. R. Anderson and J. Salz, "Spectra of Digital FM," B.S.T.J., *44*, No. 6 (July–August 1965), pp. 1165–1189.
4. J. E. Mazo and J. Salz, "Spectra of Frequency Modulation with Random Waveforms," Information and Control, *9*, No. 4 (August 1966), pp. 414–422.
5. O. Shimbo, "General Formula for Power Spectra of Digital FM Signals," Proc. IEE (GB), *113*, No. 11 (November 1966), pp. 1783–1789.
6. F. A. Graybill, *Introduction to Matrices with Applications in Statistics*, Belmont, Calif.: Wadsworth, 1969, Section 8.8.
7. D. E. Cartier, "The Power Spectrum of PCM/FSK-AM/PM," IEEE Trans. on Comm., *COM-21*, No. 7 (July 1973), pp. 847–850.

§ These φ_n are generalizations of those in (133) of Ref. 1, Appendix A.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

Analysis of Trunk Groups Containing Short-Holding-Time Trunks

By L. J. FORYS and E. J. MESSERLI

(Manuscript received July 1, 1974)

This paper presents models for the behavior of trunk groups containing short-holding-time faulty trunks. The models, referred to as ordered selection, two-sided selection, random selection, or queuing selection, are applicable to selection procedures used by a number of switching systems. Each model is analyzed to obtain the fraction of group attempts carried on a faulty trunk in the group, and the corresponding fraction of group attempts that find all trunks busy (blocking or overflow). Numerical results for the basic models are also presented. The results indicate that factors such as the trunk selection procedure or the type of group (high usage or final) can lead to significant differences in the performance of a group containing a faulty trunk.

I. INTRODUCTION

A message trunk, the basic connecting link in the switched telephone network, provides the supervising, signaling, and ringing capabilities essential to call set-up, as well as the communication path. When a condition that prevents proper functioning of a message trunk occurs, and causes a call failure, the trunk is normally released by a switching system on customer abandonment of the failed attempt and is then available to fail another call. As a result, a single undetected faulty trunk can fail a disproportionate fraction of the offered attempts to a group. Figure 9 shows an illustrative case where one trunk, subsequently verified to have been faulty, carried 35 out of 77 attempts offered to a trunk group during one hour. Because of their potential service impact, such trunks are of major concern throughout the Bell System and the object of many preventive maintenance and trouble detection programs.

This paper presents models and analyses for the behavior of trunk groups containing short-holding-time trunks. This terminology is used to emphasize that the faulty trunks of interest are accessible by the customer, as opposed to faulty trunks that inhibit seizure by a switching system (false busys) or that result in an automatic retrial on

another trunk when an abnormal condition is detected. The models were developed to provide tools for quantifying the potential impact of short-holding-time trunks, to help determine more effective ways to use the trouble detection resources in the network. Although the main emphasis here is on the analytical development of the models, some numerical results are also presented. The reader primarily interested in these results can refer directly to Section III. (Necessary terminology and the relationship of the idealized models to switching systems are summarized in Section 2.1 and Table I, respectively.) A more detailed outline follows.

The basic models studied, referred to as ordered selection, two-sided selection, random selection, and queuing selection, and the switching systems to which they primarily apply, are summarized in Table I. Justification for the choice of the models is given in Ref. 1. Basic assumptions are Poisson input, exponential holding times for both normal- and short-holding-time trunks,* and no retrials because of blocked or ineffective attempts.† Within this framework, the impact of a short-holding-time trunk depends on its position in the group and on the way idle trunks are selected. Impact is quantified in terms of the fraction ineffective (fraction of offered attempts carried on the short-holding-time trunk) and the blocking (fraction of offered attempts that find all trunks busy in the group). These measures enable total ineffectives for both high usage groups (fraction ineffective) and final groups (fraction ineffective plus blocking) to be determined, as well as permit the distortion in standard traffic measurements (group usage and counts of offered and overflow attempts) owing to a short-holding-time trunk to be assessed. It should be noted that blocking, used here in a conventional traffic engineering sense, is used by some to refer to call failures resulting from any cause.

In Section II, prior work is briefly discussed and analysis for the various models is developed. Terminology common to the models is summarized in Section 2.1. In Section 2.2, ordered selection is considered. An exact solution for the fraction ineffective is obtained. The computations for the special case of one short-holding-time trunk are summarized in (14) to (18). Blocking is treated in an approximate way, using the equivalent random method.[2]

For two-sided selection (Section 2.3), the fraction ineffective is approximated by using results for ordered selection. The computa-

* Although our main interest is short-holding-time trunks, some results hold for a short or long holding time on one trunk. In these cases, the trunk is referred to as abnormal.

† The models, however, can be used with retrials, if these are also treated as Poisson (this approximation was found to be reasonable). However, the inclusion of retrials (which are an important factor in considering volume changes) does not substantially change the service impact of a short-holding-time trunk.

Table I — Summary of models for machine trunk selection

Trunk Selection Model	Description	Switching System*
Ordered selection	Fixed-order hunt for an idle trunk.	Step-by-step, Panel, No. 4 Crossbar
Two-sided selection	Traffic split into two parts. Each part uses a fixed-order hunt for an idle trunk, with the two orders reversed.	No. 1 Crossbar, Crossbar Tandem, No. 4 Crossbar (two-way)
Random selection	Equally likely choice of an idle trunk.	No. 5 Crossbar
Queuing selection	Trunk that has been idle the longest is chosen.	No. 1 Electronic switching system

* Groups are assumed one-way (trunks are selected from only one end of the group), except for No. 4 Crossbar (two-way), where a two-way group between No. 4 Crossbar systems is assumed. The applications are not always precise; trunk assignments to frames, gradings, and certain subgrouping arrangements can affect the selection procedure of some systems.

tions summarized in (25) to (37) give good results for the fraction ineffective. Blocking is also treated in an approximate way.

Random and queuing selection are considered in Section 2.4. It is shown that, for fraction ineffective and blocking, these are equivalent. Simple exact solutions involving the Erlang B formula are obtained. Equation (60) gives the fraction ineffective, and eq. (61) gives the blocking.

Numerical results are given in Section III. For reasonable values of normal- to short-holding-time ratios as suggested by field data (about 5 to 30, depending on the type of fault and type of traffic), the results confirm that a single short-holding-time trunk can have a severe impact on service. For example, Fig. 6 shows the impact of one short-holding-time trunk in a group with random/queuing selection. It is apparent that the short-holding-time trunk has a significant impact over a wide range of conditions.

For any group, assuming that a short-holding-time trunk is equally likely to be in any position,* the various disciplines can be compared. We find that (see Fig. 8)

$$\begin{bmatrix}\text{Average}\\ \text{fraction}\\ \text{ineffective}\\ \text{for}\\ \text{random or}\\ \text{queuing}\\ \text{selection}\end{bmatrix} \leqq \begin{bmatrix}\text{Average}\\ \text{fraction}\\ \text{ineffective}\\ \text{for}\\ \text{two-sided}\\ \text{selection}\end{bmatrix} \leqq \begin{bmatrix}\text{Average}\\ \text{fraction}\\ \text{ineffective}\\ \text{for}\\ \text{ordered}\\ \text{selection}\end{bmatrix},$$

* For random/queuing selection, the position is irrelevant.

i.e., the random/queuing selection tends to minimize the impact of a short-holding-time trunk. Of course, for a fixed position for a short-holding-time trunk, either ordered or two-sided selection can have a significantly lower fraction ineffective than random/queuing (e.g., last trunk for ordered selection). These results suggest that, if expected service improvement is used as a criterion, then, all other things being equal, there is a higher service payoff in eliminating faults from trunk groups in older switching systems than in more modern systems. Conversely, the more modern systems provide better service for a given level of trunk faults.

Section 3.5 briefly discusses practical limitations to the model assumptions and gives further consideration to Fig. 9.

II. ANALYSIS MODELS FOR GROUPS CONTAINING SHORT-HOLDING-TIME TRUNKS

In this section, the equilibrium behavior of groups containing short-holding-time trunks is considered. Prior work in this area is limited. Klimontowicz[3] appears to be the first to give the problem attention, considering random selection of idle trunks, ordered selection, and cyclic random (sequential with initial starting point chosen randomly). He develops some analytic results for special cases such as zero holding time on the abnormal trunk, but relies on simulation for most of his results. In support of work to detect faulty trunks from operator trouble reports,[4] Forys[5] has considered ordered selection for a trunk group with mean holding time dependent on trunk position. In this general case, the fraction of attempts carried on any trunk can be determined by applying renewal theory, but the results are numerically inconvenient. (For completeness, these results are included in Section 2.1.) However, ordered selection with a single abnormal trunk permits application of a known recursion formula from Ref. 6, to give the computationally convenient solution developed in Section 2.1.

The random selection model has also been considered in the context of fault detection from trouble reports.[7] Except for limiting cases, analytic solutions were not obtained in Ref. 7; numerical solutions of the state equations to determine equilibrium occupancy probabilities were obtained. For a single abnormal trunk in the group, however, a simple exact solution to the state equations is possible. This is given in eq. (62). Subsequently, Kaufman[8] has shown that this solution can be generalized to an arbitrary number of abnormal trunks. However, the fraction ineffective and the blocking can no longer be expressed in terms of the Erlang B formula, which is the case for one abnormal trunk.

2.1 Terminology

- N Number of trunks in group, considered to be numbered from 1, 2, $\cdots$, N. (For ordered selection, search for an idle trunk is from 1, 2, $\cdots$, N. For two-sided selection, one traffic parcel searches from 1, 2, $\cdots$, N, and the other from N, $N-1$, $\cdots$, 1.)
- K Number of the short-holding-time trunk.
- λ Arrival rate for the Poisson traffic offered to the group.
- μ Hang-up rate for normal trunks, i.e., the holding-time distribution is $F(x) = 1 - e^{-\mu x}$, $x \geqq 0$.
- $\tilde{\mu}$ Hang-up rate for short-holding-time trunk.
- r Normal- to short-holding-time ratio $(\tilde{\mu}/\mu)$.
- a Normal offered load to the group (λ/μ). (For two-sided selection, $a = a_1 + a_2$, where a_1 corresponds to the load offered in direction 1, $\cdots$, N and a_2 the load offered in direction N, $N-1$, $\cdots$, 1).
- P Fraction ineffective (probability that an offered attempt is carried on the short-holding-time trunk).
- B Blocking* (probability that an offered attempt finds all trunks busy in the group).

2.2 Analysis for ordered selection†

In this section, we derive results used to determine the effects of an abnormal trunk in a trunk group with an ordered selection of idle trunks. We first derive results for the case where each trunk has a different holding time. We then specialize our results to the case of one abnormal trunk because convenient computational algorithms are available for this case, and because it is the case of most interest.

Of main interest is the proportion of offered calls serviced by each trunk (in equilibrium). Denote these by P_K, $K = 1, 2, \cdots, N$. The P_Ks may be obtained by first calculating the stationary occupany probabilities. If p_K is the stationary occupancy probability of the Kth trunk, then, using (for example) Little's Law,[9] one can show that

$$P_K = \mu_K p_K / \lambda, \tag{1}$$

where μ_K is the hang-up rate for trunk K.

The method used to compute the p_Ks is conceptually straightforward. The interarrival time distribution function of the traffic presented to the Kth trunk is determined, and known results for a single trunk with renewal inputs are used to find p_K.

---

* The term *blocking* is used without reference to whether a group is a high usage or final. Blocked attempts for finals are also ineffective attempts.

† This section reproduces and extends some of the results in Ref. 5.

Thus, if the interarrival distribution function of the traffic into the Kth trunk is $A_K(t)$, and

$$\alpha_K(s) = \int_0^\infty e^{-st} dA_K(t), \tag{2}$$

$$a_K = \int_0^\infty t dA_K(t), \tag{3}$$

$$\rho_K = \frac{1}{a_K \mu_K}; \tag{4}$$

then, using the results in Ref. 10, p. 93,

$$p_K = \rho_K[1 - \alpha_K(\mu_K)]. \tag{5}$$

Hence, the proportion of calls carried by the Kth trunk is

$$P_K = [1 - \alpha_K(\mu_K)]/a_K\lambda. \tag{6}$$

Following the same argument as in Ref. 10, p. 37, we can obtain

$$A_K(t) = \int_0^t \{\exp(-\mu_{K-1}x) + [1 - \exp(\mu_{K-1}x)]A_K(t - x)\} dA_{K-1}(x). \tag{7}$$

The idea behind the argument is to consider that an overflow from the $(K - 1)$st trunk occurred at time 0. In order for the next overflow to occur in less than t, we have two cases: an overflow occurs from the $(K - 2)$nd trunk at time x, $0 < x < t$ and the $(K - 1)$st trunk is busy, or the $(K - 1)$st trunk was free at time x, and so the next overflow from the $(K - 1)$st trunk must come in less than $t - x$ units of time.

Since the input to the entire trunk group is Poisson,

$$A_1(t) = 1 - e^{-\lambda t} \tag{8}$$

and

$$\alpha_1(s) = \frac{\lambda}{s + \lambda}. \tag{9}$$

Taking the Laplace-Stieltjes transform of (7), we obtain

$$\alpha_K(s) = \frac{\alpha_{K-1}(s + \mu_{K-1})}{1 - \alpha_{K-1}(s) + \alpha_{K-1}(s + \mu_{K-1})}. \tag{10}$$

Equation (10) can be used to obtain $\alpha_K(\mu_K)$ for computing P_K via (6). Unfortunately, this computation is quite cumbersome if K is large.

To evaluate a_K, we make use of the fact that

$$a_K = -\frac{d}{ds}\alpha_K(s)\Big|_{s=0}. \tag{11}$$

After some algebra,

$$a_K = \frac{1}{\alpha_{K-1}(\mu_{K-1})}\, a_{K-1} \tag{12}$$

and, hence,

$$a_K = \frac{1}{\lambda\alpha_1(\mu_1)\alpha_2(\mu_2)\cdots\alpha_{K-1}(\mu_{K-1})}, \tag{13}$$

where $\alpha_j(\mu_j)$ can be obtained from (10). Equations (6) and (13) combine to give the obvious result,

$$P_K = \begin{cases} 1 - \alpha_1(\mu_1) & K = 1 \\ \alpha_1(\mu_1)\cdots\alpha_{K-1}(\mu_{K-1})(1 - \alpha_K(\mu_K)), & K > 1, \end{cases} \tag{14}$$

i.e., the probability that an offered attempt is carried on trunk K is the probability it is blocked on the first $K - 1$ trunks times the probability it is carried on trunk K, given that it is offered to trunk K. The probability that a call is blocked on the first $K - 1$ trunk is simply the product of the individual call congestions, i.e., $B_K = B_{K-1}\alpha_K(\mu_K)$, where B_K is the blocking probability on K trunks, with $B_0 = 1$.

In the special case of $\mu_j = \mu$, $j = 1, 2, \cdots, K - 1$ and $\mu_K = \tilde{\mu}$, we obtain from (10) and (13)

$$\alpha_K(s) = \frac{1 + \sum_{j=1}^{K-1}\binom{K-1}{j}\lambda^{-j}s(s+\mu)\cdots(s+(j-1)\mu)}{1 + \sum_{j=1}^{K}\binom{K}{j}\lambda^{-j}s(s+\mu)\cdots(s+(j-1)\mu)}$$

$$a_K = 1/\lambda B(K-1, a). \tag{15}$$

Here $B(n, a)$ is the Erlang B formula for n trunks offered load a, $(a = \lambda/\mu)$.

In this case, a simple recursion exists for $\alpha_K(s)$. From Ref. 6 we can obtain

$$\alpha_j^{-1}(s) = \frac{s + \lambda + (j-1)\mu}{\lambda} - (j-1)\frac{\mu}{\lambda}\alpha_{j-1}(s) \qquad \text{for} \quad j \leqq K. \tag{16}$$

There is also a simple recursion for $B(j, a)$:

$$B^{-1}(j, a) = 1 + \frac{j}{aB(j-1, a)}, \tag{17}$$

with

$$B(0, a) = 1. \tag{18}$$

Use of recursion (16) with $s = \tilde{\mu}$ together with (17) makes the calculation of the fraction ineffective $P_K = B(K - 1, a)[1 - \alpha_K(\tilde{\mu})]$ straightforward.

The blocking B (the probability of a call finding all N trunks occupied) can be obtained by calculating the load overflowing the last trunk. Thus,

$$B = \alpha_N(\mu_N)/\lambda a_N. \tag{19}$$

Even in the case where there is only one abnormal trunk in the group, this calculation of B can be quite tedious. This is especially true in the case where a large number of trunks follow the abnormal trunk in the ordering. Instead of (19), we shall approximately calculate B for the special case $\mu_j = \mu$, $j \neq K$.

We consider offering the overflow from trunk K to a hypothetical infinite trunk group with normal holding times. The mean m and variance v of the number of occupied trunks in the infinite group can be calculated from (this follows from Ch. 3 of Ref. 11, or p. 36 of Ref. 10)

$$m = 1/\mu a_{K+1} \tag{20}$$

$$v = m\left[\frac{1}{1 - \alpha_{K+1}(\mu)} - m\right]. \tag{21}$$

Using (13) and (10), we obtain

$$m = \alpha_K(\tilde{\mu})/a_K\mu = aB(K - 1, a)\alpha_K(\tilde{\mu}) \tag{22}$$

$$v = m\left[1 - m + \frac{\alpha_K(\mu + \tilde{\mu})}{1 - \alpha_K(\mu)}\right]. \tag{23}$$

Recursions (16) (used with $s = \mu, \tilde{\mu}, \mu + \tilde{\mu}$) and (17) again make the calculation of m, v straightforward.

We now apply the equivalent random method.[2] That is, we approximate the overflow from trunk K by the overflow from an "equivalent" trunk group having normal trunk holding times, and which produces the same m, v on a hypothetical infinite trunk group. To determine blocking, we proceed along the same lines as in the equivalent random method, giving

$$B \approx \lambda_e B\left(N - K + N_e, \frac{\lambda_e}{\mu}\right)\Big/\lambda, \tag{24}$$

where N_e is the number of trunks in the equivalent group and λ_e is the attempt rate for the traffic offered to the equivalent group. (This approximation is exact where one trunk follows trunk K. In other cases, small errors can be expected.)

Although the results in this section have assumed that the input stream is Poisson, it is a simple matter to extend them to handle

peaked* (overflow) traffic. In particular, one can use the equivalent random method to model the peaked traffic as the overflow from a single trunk group offered Poisson traffic, append this "primary" group to the trunk group of interest, and assume an ordered hunt selection for the extended trunk group. The hunting is done first over the primary group. Thus, if the Kth trunk were faulty in the trunk group of interest and there were N_e trunks in the primary group modeling the peaked traffic, the faulty trunk will now be in the $(N_e + K)$th position.† The proportion of calls carried by the faulty trunk is determined by solving the extended trunk group problem and scaling the resulting proportion up by the ratio of the mean traffic intensity of the input to the extended group to the mean traffic intensity to the original trunk group.

This completes our treatment of ordered selection. Numerical results are given in Section III.

2.3 Analysis for two-sided selection

In this section, two-sided selection is considered. We derive approximations for the fraction ineffective P and for the blocking B.

First, consider the extreme case with trunk K having zero holding time. The fraction ineffective seen by each traffic is $P_1 = B(K - 1, a_1)$, $P_2 = B(N - K, a_2)$, where $B(\cdot,\cdot)$ is the Erlang B formula. This case motivates the approximation to be described which, in its simplest form, ignores interaction between the separated subgroups of good trunks, and in more refined form accounts for some interaction.‡ This approximation, which is quite accurate, was used because a computationally feasible solution to the state equations proved difficult.

The notation to be used in this section is indicated in Fig. 1: m_i, v_i are the mean and variance of the traffic offered to trunk K, and c_{ij}, v_{ij} are the mean and variance of the "crossover traffic." To use the convenient recursions developed in the preceding section, only the mean of the crossover traffic is used. This has a negligible effect on the computation for P, but can have a somewhat larger effect on the computation for B, because the peaked crossover traffics would experience a higher blocking than the Poisson traffics. However, their mean is normally small compared to the Poisson traffic, which reduces the effect of their peakedness somewhat. In particular, for $a_1 = a_2$

* The peakedness factor $z(\mu)$ of a traffic stream is the equilibrium variance-to-mean ratio of busy servers when this traffic is offered to an infinite group of exponential servers with service rate μ. The peakedness factor is larger than one for overflow traffic.

† Since nonintegral N_e can occur in the equivalent random method, interpolation may be necessary.

‡ The only real difference is that the refined form requires iteration.

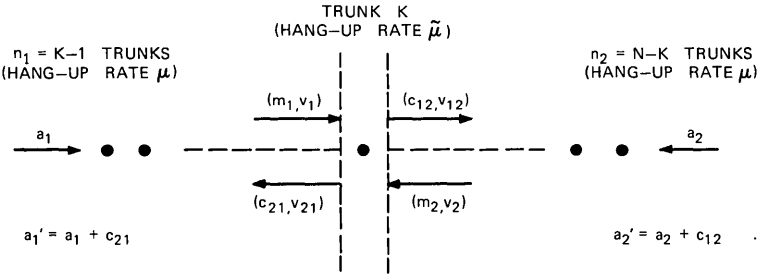


Fig. 1—Terminology for two-sided selection.

(the condition for which the approximation is intended) we usually have $c_{12} \ll a_2$, $c_{21} \ll a_1$. When this is not the case, such as a large trunk group with the short-holding-time trunk in the first position, so that c_{12} may be of the same magnitude as a_2, then c_{12} has a low peakedness.

To describe the computations, assume that the c_{ij} are given and independent of the a_i. Thus, consider a load $a_i' = a_i + c_{ji}$ (treated as Poisson) offered to a group of n_i trunks with hang-up rate μ, overflowing to a trunk with hang-up rate $\tilde{\mu}$. From the results in Section 2.1, the mean and variance of the overflow (in terms of hang-up rate $\tilde{\mu}$) can be determined. First, we define $\lambda_i' = a_i'/\mu$ and solve the recursion [see (16)]

$$\alpha_j^{-1} = \frac{\tilde{\mu} + \lambda_i' + (j-1)\mu}{\lambda_i'} - (j-1)\frac{\mu}{\lambda_i'}\alpha_{j-1}, \tag{25}$$

given

$$\alpha_1 = \frac{\lambda_i'}{\tilde{\mu} + \lambda_i'} \tag{26}$$

for $j = 1, 2, \cdots, n_i + 1$. The parameter $\tilde{\alpha}_i = \alpha_{n_i+1}$ is the call congestion on a (fictitious) trunk with hang-up rate $\tilde{\mu}$. The mean and variance of the offered load to this trunk is

$$m_i' = \frac{1}{r} a_i' B(n_i, a_i') \tag{27}$$

$$v_i = m_i'\left(\frac{1}{1-\tilde{\alpha}_i} - m_i'\right). \tag{28}$$

We must now get the mean and variance of the overflow traffic corresponding to a_i, i.e., for the traffic offered to trunk K. Since c_{ji} is treated as Poisson, clearly

$$m_i = \frac{1}{r} a_i B(n_i, a_i'). \tag{29}$$

To get the variance, we use a result from Ref. 12 on the peakedness of split overflow, which gives

$$\frac{v_i}{m_i} = 1 + \frac{a_i}{a_i'}\left(\frac{1}{1-\tilde{\alpha}_i} - m_i' - 1\right). \tag{30}$$

Unless $a_i = a_i'$, this gives a value less than v_i'/m_i'. The total mean and variance of the offered traffic to trunk K is then

$$m = m_1 + m_2 \tag{31}$$

$$v = v_1 + v_2. \tag{32}$$

To compute the call congestion for trunk K, we assume (as in the equivalent random approach) that m, v are the mean and variance of a renewal input to trunk K. If $\alpha(\cdot)$ represents the Laplace-Stieljes transform of the interarrival distribution for this renewal process, then the call congestion is given by $\tilde{\alpha} = \alpha(\tilde{\mu})$. But for the renewal model,

$$v = m\left(\frac{1}{1-\tilde{\alpha}} - m\right) \tag{33}$$

and, hence,

$$\tilde{\alpha} = \frac{m + z - 1}{m + z}, \quad (z = v/m). \tag{34}$$

This results in a fraction ineffective for the total offered traffic to the group given by

$$P = \frac{rm}{a_1 + a_2} \cdot \frac{1}{m + z} \tag{35}$$

and an approximation for blocking,

$$B = 1 - P - \frac{a_1'[1 - B(n_1, a_1')] + a_2'[1 - B(n_2, a_2')]}{a_1 + a_2}. \tag{36}$$

Finally, we require that the crossover traffics assumed be consistent with the mean call congestion on trunk K,

$$c_{ij} = a_i B(n_i, a_i') \frac{m + z - 1}{m + z}. \tag{37}$$

This condition can be met by a few simple iteration steps, beginning with $c_{ij} = 0$, and using (37) to update the c_{ij}.

When condition (37) is satisfied, eqs. (35) and (36) give the basic results for the two-sided selection procedure. It should be noted that, in addition to the treatment of each crossover traffic as Poisson and independent of the Poisson traffic offered to the same subgroup of good trunks, a second (minor) approximation is implicit in the computations. This is the use of the average call congestion in (37) to update

each crossover traffic. It is a straightforward matter to determine a separate call congestion for each traffic offered to trunk K, but the approximation is sufficiently accurate without this refinement. Numerical results using the approximations and comparisons with exact solutions (for small problems) are presented in Section III.

2.4 Analysis for queuing and random selection

To derive results for the case where the idle trunks form a queue, we define a two-dimensional state that represents the number of idle trunks in the queue together with the position of the abnormal trunk in the queue. We then derive equations for the equilibrium probabilities of each state. One can show that these probabilities exist and are unique.* The equilibrium probability that the abnormal trunk is busy, denoted by $\tilde{p}$, is then simply a sum for states in which the abnormal trunk is busy. As indicated in Section 2.2, it is easy to show that the mean number of requests serviced by the abnormal trunk in a unit of time is given by

$$\tilde{\mu}\tilde{p} \tag{38}$$

and, hence, the proportion of requests serviced by the abnormal trunk is

$$P = \tilde{\mu}\tilde{p}/\lambda = r\tilde{p}/a. \tag{39}$$

In the course of deriving our results, we show that it is equally likely that the abnormal trunk occupies any position in the queue of idle trunks. This, however, is equivalent to a random selection of idle trunks as far as the fraction ineffective and blocking are concerned. Hence, the results derived for P, B also apply to the random selection model. Strictly speaking, the two selection procedures are not equivalent, of course. For example, with queuing selection, an abnormal trunk would not serve two successive attempts if other trunks were idle, but this is a possibility with random selection. Under other assumptions, e.g., dependent retrials or non-Poisson input, these differences become important.

To proceed with the derivation, define

E_{ij} = the event that there are j idle trunks in the queue and that the abnormal trunk is in the ith position in the queue.

P_{ij} = Prob $\{E_{ij}\}$.

We let $i = 0$ denote that the abnormal trunk is occupied. Thus,

$$\tilde{p} = \sum_{j=0}^{N-1} P_{0j}. \tag{40}$$

* We have a continuous-time Markov process, with a well-behaved embedded Markov chain (see Ref. 13).

The following equations can be readily derived relating the P_{ij}'s:

$$[\tilde{\mu} + (N-1)\mu]P_{00} = \lambda P_{01} + \lambda P_{11}, \tag{41}$$

$$[\tilde{\mu} + (N-j-1)\mu + \lambda]P_{0j} = \lambda P_{0,j+1} + (N-j)\mu P_{0,j-1} + \lambda P_{1,j+1} \quad \text{for} \quad 1 \leqq j < N-1, \tag{42}$$

$$[\tilde{\mu} + \lambda]P_{0,N-1} = \lambda P_{1N} + \mu P_{0,N-2}, \tag{43}$$

$$[(N-j)\mu + \lambda]P_{ij} = (N-j+1)\mu P_{i,j-1} + \lambda P_{i+1,j+1} \quad \text{for} \quad 0 < i < j \leqq N, \tag{44}$$

$$[(N-j)\mu + \lambda]P_{jj} = \tilde{\mu} P_{0,j-1} + \lambda P_{j+1,j+1} \quad \text{for} \quad 0 < j \leqq N, \tag{45}$$

and, trivially,

$$P_{ij} = 0 \quad \text{for} \quad i > j, \quad \text{or} \quad j > N;$$
$$P_{0j} = 0 \quad \text{for} \quad j \geqq N.$$

These equations appear quite formidable because of the apparent lack of simple structure. A brute force algebraic approach seems unfruitful, as does a generating function approach.

Instead, we make the conjecture that

$$P_{1j} = P_{2j} = \cdots = P_{jj} \qquad \text{for all } j \geqq 1. \tag{46}$$

We should note that (46) is equivalent to assuming that the idle trunk is selected at random. This is true because P_{ij} represents the probability of seeing the state (i, j) at a random instant in equilibrium and Poisson arrivals see the same distribution.

We justify our conjecture by using (46) in (41) to (45) together with the fact that the probabilities must add to 1 and showing that the resulting equations have a simple solution. This solution can be substituted into the original equations to show that we indeed have the correct solution. However, since there exists a unique equilibrium solution to these equations and we have, in effect, solved them by adding additional constraints, the proof is complete without this step.

Proceeding in the manner described, we first use (46) in (41) to (45) and obtain the following equations:

$$[\tilde{\mu} + (N-1)\mu]P_{00} = \lambda P_{01} + \lambda P_{11}, \tag{47}$$

$$[\tilde{\mu} + (N-j-1)\mu + \lambda]P_{0j} = \lambda P_{0,j+1} + (N-j)\mu P_{0,j-1} + \lambda P_{j+1,j+1} \quad \text{for} \quad 1 \leqq j < N-1, \tag{48}$$

$$[\tilde{\mu} + \lambda]P_{0,N-1} = \lambda P_{NN} + \mu P_{0,N-2}, \tag{49}$$

$$[(N-j)\mu + \lambda]P_{jj} = (N-j+1)\mu P_{j-1,j-1} + \lambda P_{j+1,j+1} \quad \text{for} \quad 0 < j \leqq N, \tag{50}$$

and

$$[(N - j)\mu + \lambda]P_{jj} = \tilde{\mu}P_{0,j-1} + \lambda P_{j+1,j+1} \quad \text{for} \quad 0 < j \leqq N. \tag{51}$$

Putting $j = N$ in eq. (50), recalling that $P_{N+1,N+1} = 0$, we get

$$P_{N-1,N-1} = \frac{\lambda}{\mu} P_{NN} = aP_{NN}. \tag{52}$$

Repeatedly reducing j by 1 and solving for $P_{j-1,j-1}$ in terms of P_{NN}, we get

$$P_{jj} = \frac{a^{N-j}}{(N - j)!} P_{NN}, \qquad j = 1, \cdots, N. \tag{53}$$

Using (53) in (51) results in

$$P_{0j} = \frac{\lambda}{\tilde{\mu}} \frac{a^{N-j-1}}{(N - j - 1)!} P_{NN}. \tag{54}$$

Finally, from (47) we get

$$P_{00} = \frac{\lambda}{\tilde{\mu}} \frac{a^{N-1}}{(N - 1)!} P_{NN}. \tag{55}$$

One can check that expressions (53), (54), and (55) satisfy eqs. (48) and (49), which proves the conjecture.

We compute P_{NN} from the fact that the sum of the probabilities must be 1. This yields

$$P_{NN} = \frac{1}{N}\left[\frac{a^N}{N!}\frac{\mu}{\tilde{\mu}} + \cdots + \frac{a^j}{j!}\left(\frac{N - j}{N} + \frac{j}{N}\frac{\mu}{\tilde{\mu}}\right) + \cdots + 1\right]. \tag{56}$$

Thus,

$$\tilde{p} = \sum_{j=0}^{N-1} P_{0j} = \frac{\mu a \sum_{j=0}^{N-1} \frac{a^j}{j!}}{\tilde{\mu}N \sum_{j=0}^{N} \frac{a^j}{j!}\left(\frac{N - 1}{N} + \frac{j\mu}{\tilde{\mu}}\right)} \tag{57}$$

and

$$P = \frac{r\tilde{p}}{a} = \frac{\sum_{j=0}^{N-1} \frac{a^j}{j!}}{N \sum_{j=0}^{N} \frac{a^j}{j!}\left(\frac{N - j}{N} + \frac{j}{rN}\right)}, \tag{58}$$

recalling that $r = \tilde{\mu}/\mu$.

Equation (58) can be rewritten in terms of Erlang's loss function as

$$P = \frac{\sum_{j=0}^{N-1} \frac{a^j}{j!}}{N \sum_{j=0}^{N} \frac{a^j}{j!} + \left(\frac{1}{r} - 1\right) \sum_{j=1}^{N} \frac{a^j}{(j-1)!}}$$

$$= \frac{\sum_{j=0}^{N-1} \frac{a^j}{j!}}{N \sum_{j=0}^{N-1} \frac{a^j}{j!} + \frac{Na^N}{N!} + \left(\frac{1}{r} - 1\right) a \sum_{j=0}^{N-1} \frac{a^j}{j!}}.$$

Thus,

$$P = \frac{r}{a + Nr - ar[1 - B(N-1, a)]}. \tag{59}$$

An alternative form of (59) can be obtained via the standard recursion for Erlang B [see (17)], giving

$$P = \frac{r[1 - B(N, a)]}{Nr - (r-1)a[1 - B(N, a)]}. \tag{60}$$

To get an expression for the blocking B, notice that

$$B = P_{00} = \frac{\lambda}{\tilde{\mu}} \frac{a^{N-1}}{(N-1)!} P_{NN} = \frac{(\lambda/\tilde{\mu})(a^{N-1}/(N-1)!)}{N\left[\sum_{j=0}^{N} \frac{a^j}{j!}\left(\frac{N-j}{N} + \frac{j}{rN}\right)\right]}$$

or

$$B = \frac{NB(N, a)}{Nr - (r-1)a[1 - B(N, a)]}. \tag{61}$$

Although (61) and (62) also apply to the random model, it should be noted that the equivalence between random and queuing selection extends to the equilibrium distribution for *redefined* states (i, j), where i is the number of good trunks occupied and j is 1 if the abnormal trunk is occupied, 0 otherwise. In fact, it is straightforward to show from (53) to (55) that the redefined probabilities must satisfy

$$\left.\begin{aligned} P_{i0} &= \frac{a^i(N-i)}{i!\,N} P_{00} \\ P_{i1} &= \frac{a^{i+1}}{ri!\,N} P_{00} \end{aligned}\right\} i = 0, 1, \cdots, N-1, \tag{62}$$

which solve the somewhat simpler equations for the random model (these equations are given in Ref. 4).

III. NUMERICAL RESULTS

This section presents selected numerical results obtained with the solution procedures of Section II.

3.1 Ordered selection (Figs. 2 to 5)

Figure 2 gives the fraction ineffective for $r = 5$ and the blocking* for $r=5$, $N=20$. Even for this relatively small value of r, the short-holding-time trunk has a significant impact, which decreases as the position in the order of selection increases. For large values of offered load, the dependence of fraction ineffective on K (position of the short-holding-time trunk) is decreased. In fact, it follows from Little's Law[9] that, for all K (and all selection procedures), $P \leqq r/a$, with the bound approached as the time congestion on the short-holding-time trunk approaches 1.0; i.e., for very large a. For $a = 25$, $K = 1$, we note that $P = 0.17$, $r/a = 0.2$.

The results for fraction ineffective do not depend on the total number of trunks $N \geqq K$. For blocking, N is important. For $N = 20$, $r = 5$, blocking is decreased from Erlang B blocking by about $\frac{1}{3}$ in the (design) range for 20 to 30 percent overflow, with the decrease relatively insensitive to K. Blocking results are approximate, with errors resulting only from the application of the equivalent random method. In cases where exact solutions were compared to the approximate, the agreement was good. The decreased overflow resulting from a short-holding-time trunk can contribute substantial errors in estimation procedures for which overflow enters. Usage measurements are also affected by a short-holding-time trunk, with mean carried usage given by $aP/r + a(1 - P - B)$.

Figure 3, for $r = 15$, is similar to Fig. 2, but with larger impacts and more spread between the curves. The ratio $r = 15$ is typical for a short-holding-time trunk and shows that $P > 0.4$ can easily occur, even for relatively large loads. The limiting ratio $r = \infty$ (zero-holding-time trunk) would result in $P^{\infty} = B(K - 1, a)$, and $B^{\infty} = 0$, which are substantially different from the $r = 15$ results. For example, $P^{\infty} = 1$ for $K = 1$.

3.2 Two-sided selection (Figs. 4 and 5)

Figure 4 gives the behavior for $N = 5$, $r = 15$, including a comparison with the exact solution obtained by solving the state equations for the system. The approximation for P displays essentially no error for low blocking. This is to be expected for conditions under which the

* Recall that blocking refers to the fraction of attempts that find all trunks busy in the group, whether or not it is a final or high-usage group.

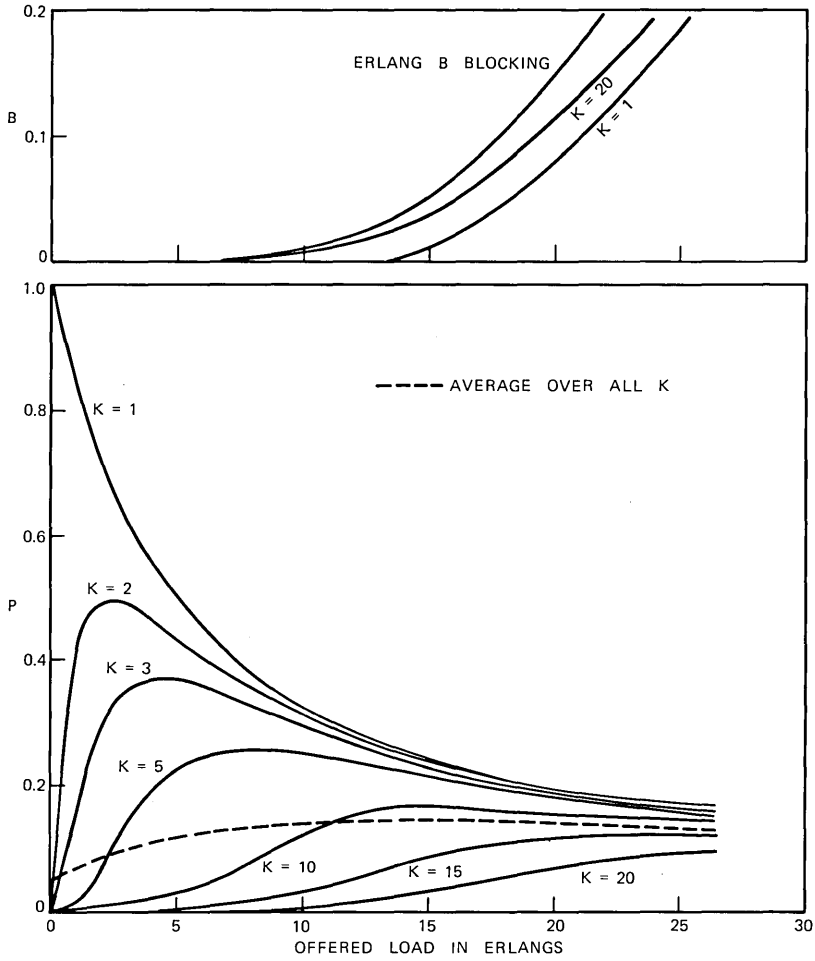


Fig. 2—Behavior of ordered selection for $r = 5$, $N = 20$.

offered load to trunk K comes from only one side (e.g., $K = 1$, low offered load), since the approximation for P becomes exact. For other low blocking conditions (such as $K = 3$, low offered load), no discernible error indicates that modeling the traffics offered to trunk K as a renewal stream introduces essentially no error in the computed call congestion for trunk K. As the crossover traffics increase, and, hence, there is interaction between the separated subgroups of good trunks, the approximation shows some error. The error is well-behaved, and increases to only about 2 to 3 percent. For larger groups, the error is generally less and is not shown on the figures.

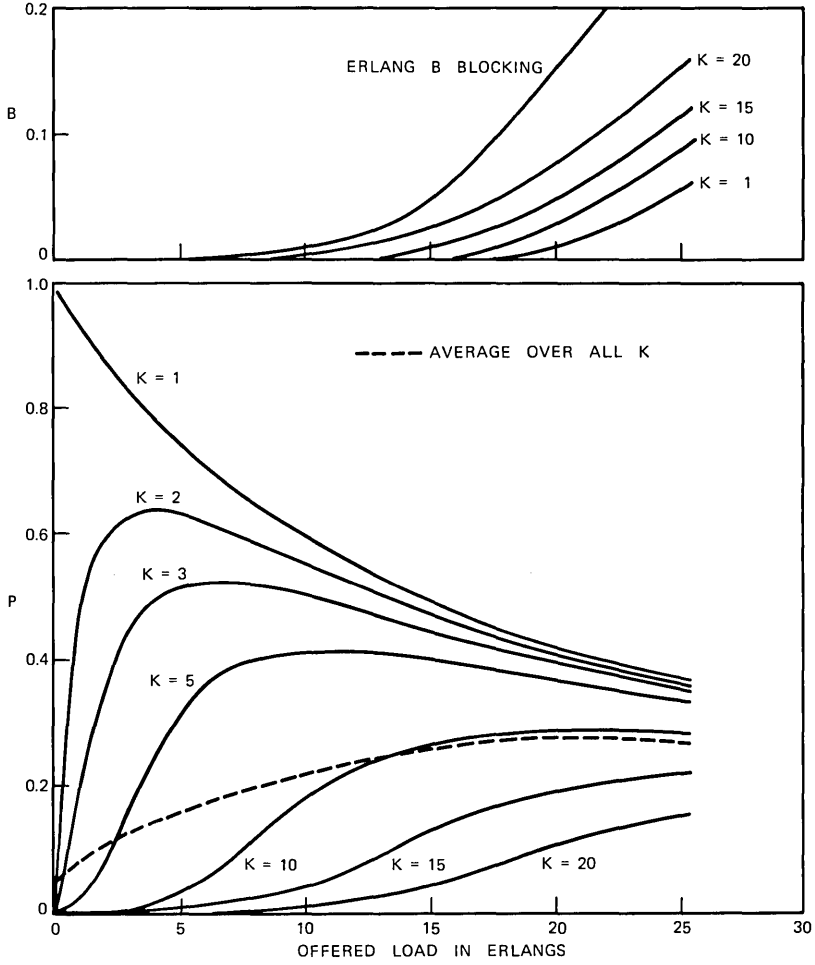


Fig. 3—Behavior of ordered selection for $r = 15$, $N = 20$.

Errors in the blocking are larger. Since peakedness is ignored in determinining the crossover traffics and the carried load on the good trunks, this is to be expected. However, the absolute error does not increase significantly with offered load. As the change from Erlang B blocking is the important factor in applications, this error behavior is acceptable. As for ordered selection, the blocking is relatively insensitive to K.

The interesting behavior for $K = 1$ is apparently due to the size of the group. For $K = 1$, $N \geqq 2$, qualitatively, one would expect that the fraction ineffective seen by load a_1 monotonically decreases from

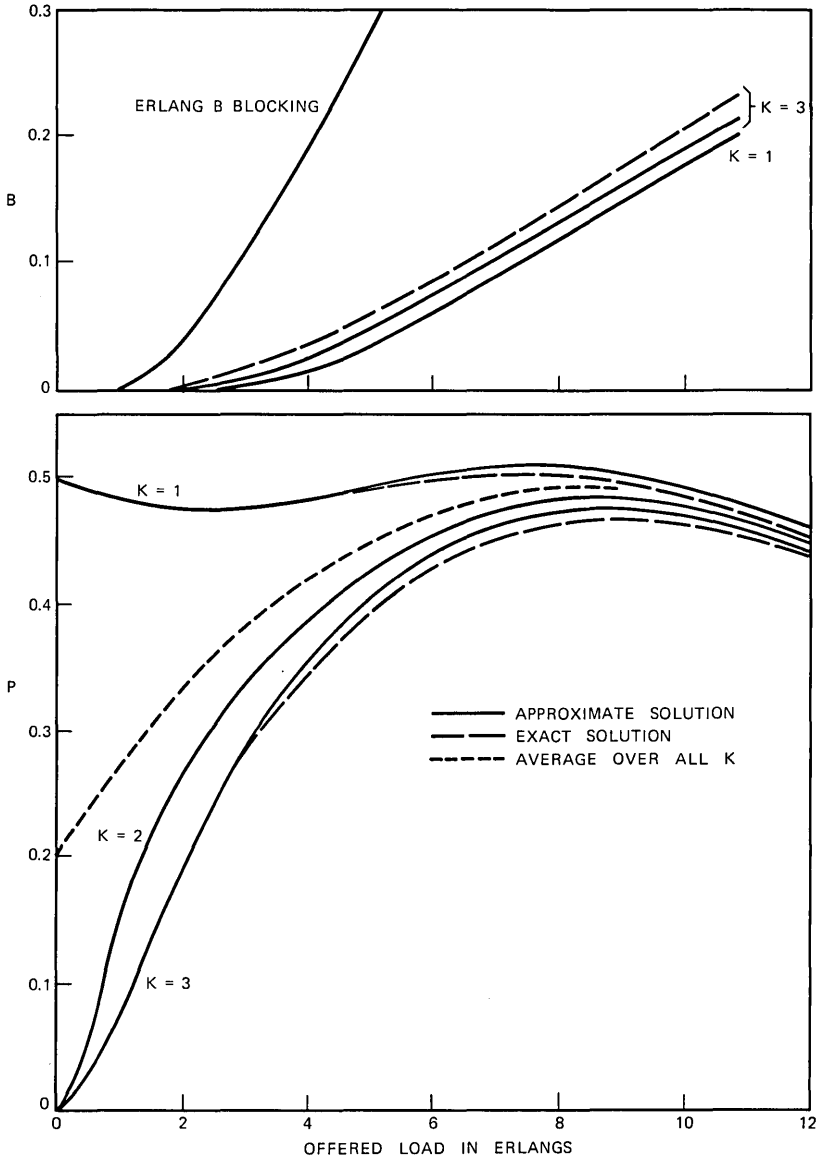


Fig. 4—Behavior of two-sided selection for $N = 5$, $r = 15$.

1.0, while for a_2 it initially increases from 0, as the offered load increases. The exact solution for $N = 5$ confirmed this behavior. However, the fraction ineffective seen by load a_2 increases fast enough to create an increase in P, before its ultimate decrease. As a result,

$P \approx 0.5$ in the relatively wide range 0 to 10 erlangs. For larger values of N, P was found to be monotonically decreasing for $K = 1$. On the other hand, for $N = 2$, P can easily be shown to initially increase. Since $a_1 = a_2$, for $N = 2$, two-sided and random selection are equivalent, and (59) applies. Thus, the conditions that produce the qualitative behavior of Fig. 4 are difficult to predict.

Figure 5 illustrates the behavior for $r = 15$, $N = 20$. Compared to ordered selection, the impact is less, and the dependence on K less pronounced, especially for blocking. This is attributable to the diversity introduced into the selection procedure by two-sided selection. However, a substantial overall impact is still evident. For $N = 20$, even for the most favorable case $K = 10$, for a design overflow of 20 percent, $P \approx 0.2$, and overflow is reduced to 5 to 6 percent.

The limiting case for $r = \infty$ gives

$$P^{\infty} = \frac{B(K - 1, a/2) + B(N - K, a/2)}{2}.$$

For $N = 20$, $K = 1$, this gives a value of 0.5 over the range 0 to 25 erlangs, compared to the value of $P = 0.29$ for $a = 25$, $r = 15$. For $N = 20$, $K = 10$, $P^{\infty} = 0.35$ at $a = 25$, or about 50 percent higher than the value for $r = 15$.

3.3 Random/queuing selection (Figs. 6 and 7)

In each case in Figs. 6 and 7, P begins at $1/N$, then increases over the range of offered loads shown. The $1/N$ initial behavior is intuitively obvious for random selection. Queuing selection, on the other hand, guarantees that, when a trunk becomes idle, trunks already idle must serve calls before it can be picked for service. For very low offered loads, a short-holding-time trunk essentially serves every Nth call to give the $1/N$ initial behavior.

Analysis of eq. (59) indicates that P can have at most one extremum. For $r > 1$, this occurs at the unique root of the equation

$$\frac{\partial}{\partial a}\{a[1 - B(N - 1, a)]\} = \frac{1}{r}.$$

Thus, random/queuing selection never displays the complex behavior observed in Fig. 4 for two-sided selection. Because the extremum occurs for relatively large values of a, a short-holding-time trunk has a larger impact in a high usage group than in a final group of the *same* size. For example, for ten trunks, 4.5 erlangs is a typical offered load for a final, and 9.5 erlangs for a high usage group. From Fig. 6, $P = 0.17$ at $a = 4.5$ erlangs, whereas $P = 0.28$ at $a = 9.5$ erlangs, i.e., the *relative* impact is larger. The time congestion aP/r for the short-holding-

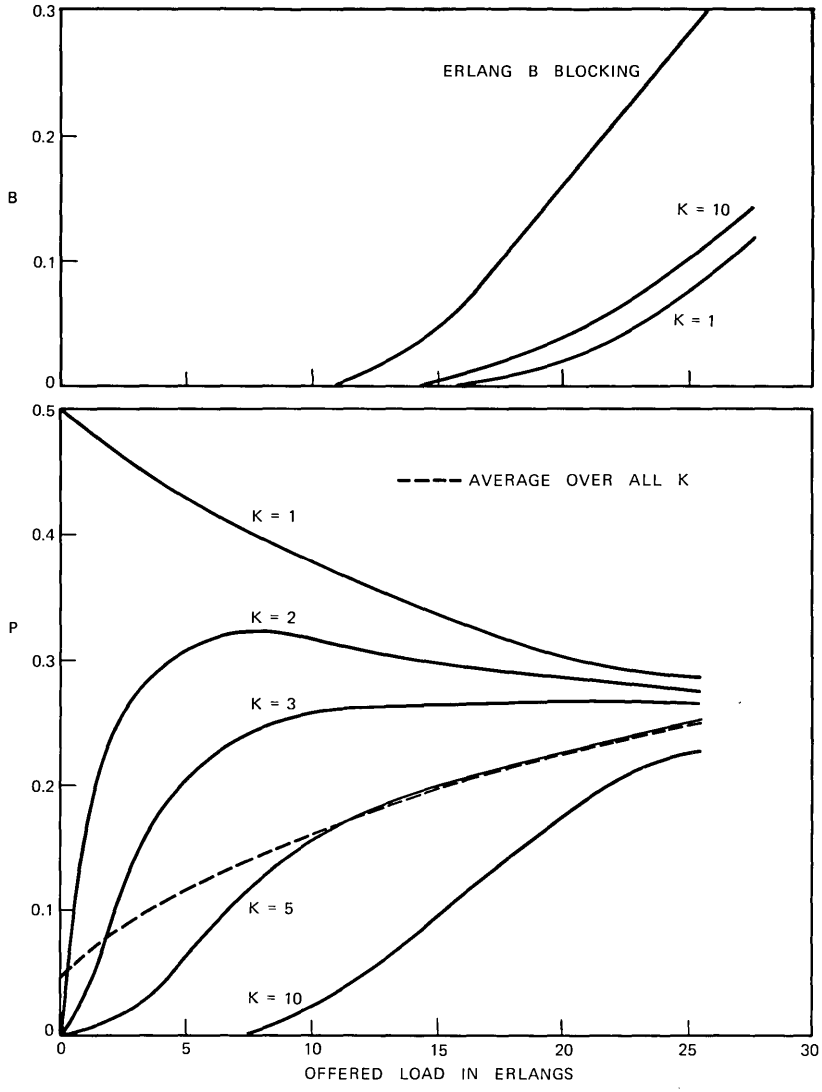


Fig. 5—Behavior of two-sided selection for $N = 20$, $r = 15$.

time trunk is 0.052 at $a = 4.5$ erlangs, and 0.175 at $a = 9.5$ erlangs, i.e., the *absolute* impact of the short-holding-time trunk is over three times as much for the high-usage application. In terms of the expected time congestion with a short-holding-time trunk equally likely in any position, this general behavior also holds for two-sided and ordered selection. This follows from the increase in average fraction ineffective as the offered load increases.

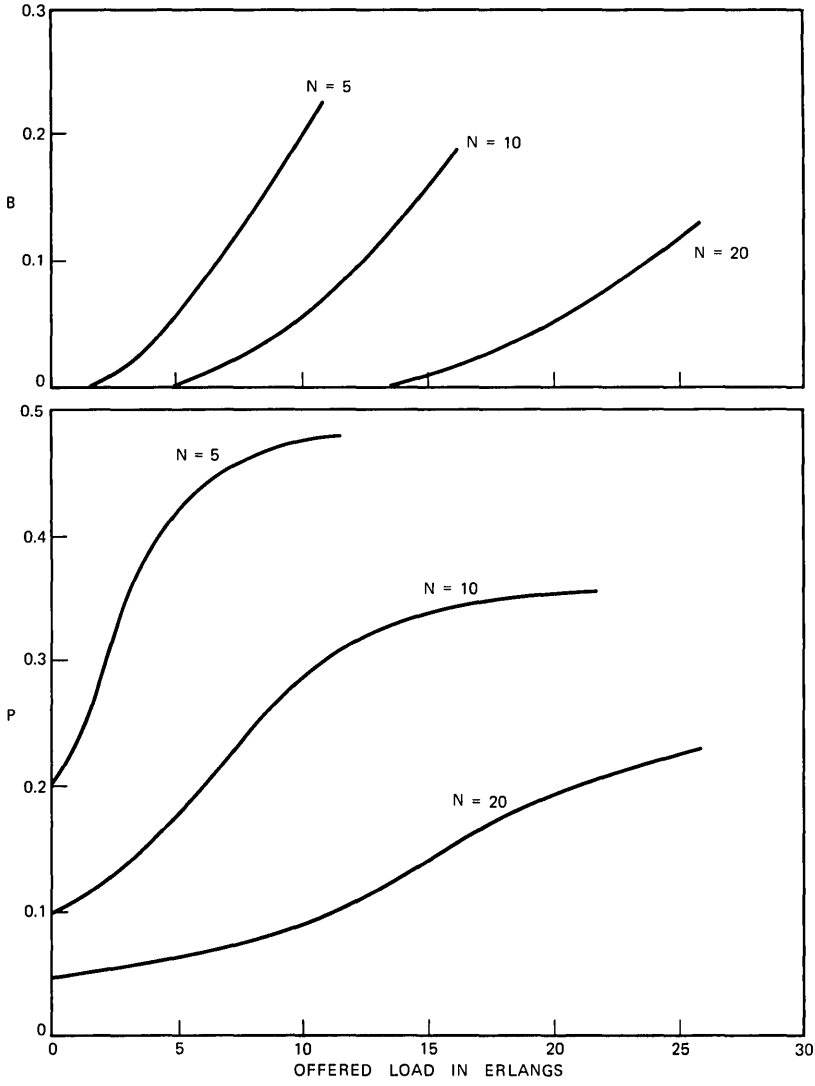


Fig. 6—Behavior of random selection for $r = 15$, $N = 5, 10, 20$.

Figure 7 displays the behavior of random/queuing selection as r varies. In all cases, P is close to its maximum over a fairly wide range. The limiting case $r = \infty$ gives [see (59)]

$$P^{\infty} = \frac{1}{N - a[1 - B(N - 1, a)]}.$$

For $N = 20$, $a = 25$ erlangs, $P^{\infty} = 0.36$, compared to the $r = 20$

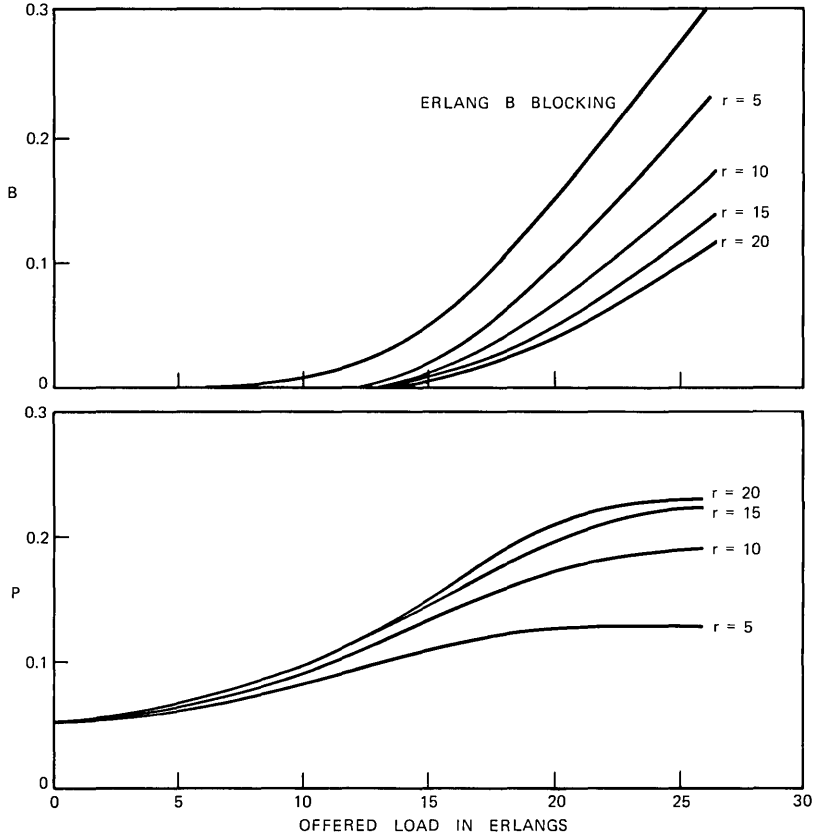


Fig. 7—Behavior of random selection for $N = 20$, $r = 5, 10, 15, 20$.

value from Fig. 7 of $P = 0.23$. Since P is no more difficult to compute than P^{∞}, the upper bound P^{∞} is of limited use in this case.

3.4 Comparison of selection procedures

To compare selection procedures, the measure taken is the expected value of P, denoted by $\bar{P}$, given that a short-holding-time trunk is equally likely in any position. Computationally, it was found that

$$\bar{P}\begin{bmatrix}\text{Random}\\ \text{or}\\ \text{queuing}\\ \text{selection}\end{bmatrix} \leqq \bar{P}\begin{bmatrix}\text{Two-sided}\\ \text{selection}\end{bmatrix} \leqq \bar{P}\begin{bmatrix}\text{Ordered}\\ \text{selection}\end{bmatrix},$$

with equality occurring only at $a = 0$. Figure 8 for $N = 20$, $r = 15$ displays typical behavior. The differences are substantial. It is likely that

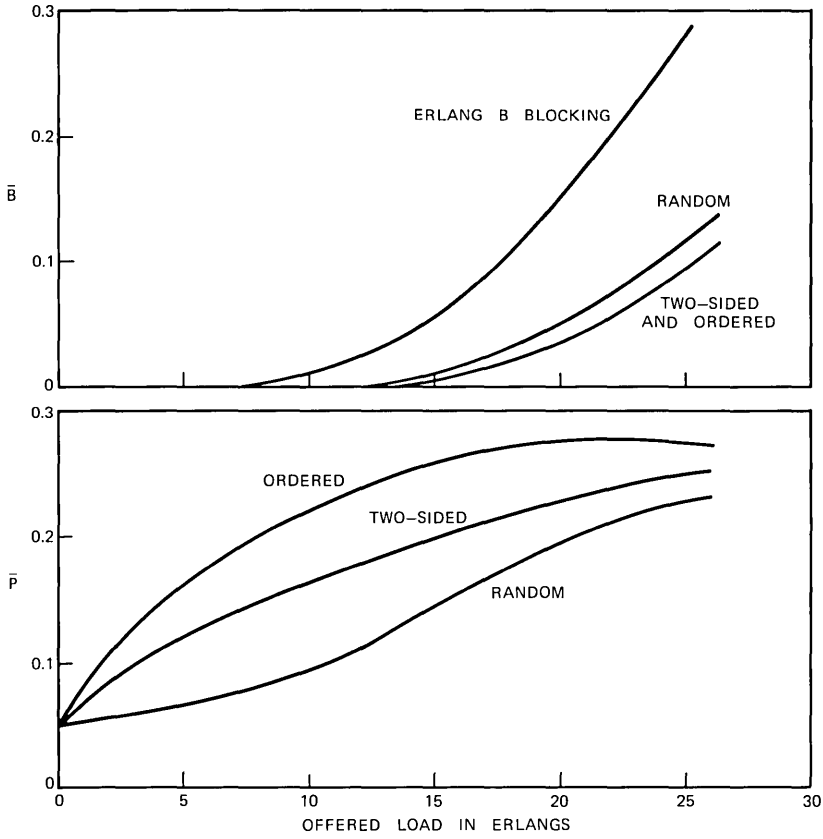


Fig. 8—Comparison of selection procedures for short-holding-time trunk equally likely in any position ($N = 20$, $r = 15$).

any improvement over the random/queuing behavior would require a selection procedure based on more information. For example, choosing the idle trunk whose last holding time was longest would further reduce the impact of a short-holding-time trunk.

The expected blocking does not show much difference from one selection procedure to another, but all disciplines show a substantial reduction from the Erlang B results.

3.5 Validation remarks

The basic traffic assumptions of the model concern holding times and the arrival process. For normal trunk holding times (which result from a mix of conversations, busys, don't answers), the exponential assumption is reasonable. For short-holding-time trunks, this assumption may be less valid. For example, trunks resulting in immediate

reorder returned to the customer would not likely display the exponential behavior. However, since the Erlang B formula holds for arbitrary holding-time distributions, it is reasonable to expect that the results should not be too sensitive to the form of the distribution.

The Poisson assumption for the arrival process can be invalidated due to time variations, to peakedness (for groups receiving overflow traffic), or to retrial behavior. For ordered selection with $K = 1$, peakedness clearly reduces the impact of a short-holding-time trunk. This behavior may be reversed for higher values of K since peakedness increases the mean attempt rate to trunk K. For example, for $r = \infty$, $K \geqq 2$, P is increased. However, several peaked traffic computations for $r = 15$ give values of the fraction ineffective that are very close to the Poisson value.

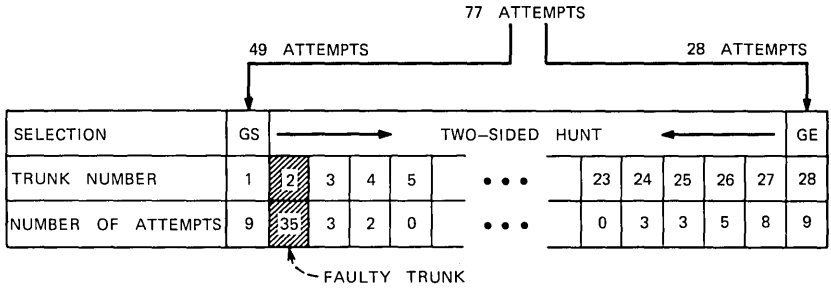
Even if the Poisson assumption is valid with all trunks good, it can be violated in the presence of a short-holding-time trunk. The short time between a retrial and the failure that initiated it gives the retrial a high probability of meeting the same conditions as the initial failure. For two-sided, ordered, and random selection, this could increase the impact of the short-holding trunk over what would be expected from a Poisson assumption.

The model is also susceptible to variations from the idealized selection procedures. This holds particularly for applications to older systems where grading and multipling arrangements can lead to many different selection procedures. It also holds for applications to 5XB systems, where any irregularity in the distribution of trunks on the frames can distort the selection procedure from the idealized random selection model. In fact, individual circuit usage results would seem to indicate that significant discrepancies from (cyclic) random selection may be quite common in 5XB. Thus, for any specific application, suitability of the model would have to be determined.

Figure 9 shows data obtained by special measurements from a Crossbar Tandem group containing a short-holding-time trunk. The peg count and usage measurements during the study hour indicate that trunk 2 had a short holding time relative to other trunks in the group, and a failure condition was subsequently verified. Thus, all 35 attempts on the trunk are assumed to have failed. The large imbalance in the offered attempts from either side of the group could be statistical, or it could be due to irregularities in incoming trunk assignment to frames (which determines the order of selection). It is more likely due to retrials. Thus, the reasonable assumption that $a_1 = a_2$ can be violated in some situations.

Since the measured hour was not the busy hour, the left side of the group can be treated like ordered selection. The estimate for a mean

STUDY RESULTS–TANDEM COMPLETING FINAL TRUNK GROUP
JUNE 21 1972, 3–4 PM



HOLDING TIME ESTIMATES FROM INDIVIDUAL CIRCUIT USAGE AND PEG COUNT DATA

AVERAGE NORMAL HOLDING TIME = 246 s
AVERAGE HOLDING TIME FOR TRUNK 2 = 19 s

Fig. 9—Example of a short-holding-time trunk.

attempt rate for a Poisson offered traffic is $\hat{\lambda} = 49$ attempts/hour. For the offered load a in erlangs, a normal holding time is associated with each attempt, to give $\hat{a} = 3.3$ erlangs. The ratio of holding times is $r = 13$. For these values, the ordered selection model with $K = 2$ predicts $P = 0.62$. Thus, the realized value of ineffectives (35) is close to the average value (30.5) predicted from the same data. The discrepancy could be statistical, or to modeling the stream as Poisson, which smoothes out the retrial stream.

The preceding very limited comparison is not a validation. Because of the various factors noted, a comprehensive validation of the models is difficult. Since the models are not intended for design purposes, but only to estimate impact of short-holding-time trunks, such validation has not been attempted.

IV. CONCLUSIONS

This paper has considered analytical models for groups containing short-holding-time trunks. The models confirm that these trunks can have a substantial impact on customer service. Although traffic and system characteristics can differ from those of the model, it is felt that the models are adequate for estimation purposes.

The numerical results indicate that the type of system and type of group (high usage or final) lead to significant differences in performance in the presence of a short-holding-time trunk. These models and results are directly useful in devising optimum strategies for deploying maintenance resources to minimize the service impact of short-holding-time trunks.

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge comments by S. Horing, discussions with M. F. Morse on trunk selection, and the programming assistance of Mary Zeitler and Anne Vottero. Data for the Fig. 9 example were supplied by J. L. Laude.

REFERENCES

1. L. J. Forys and E. J. Messerli, "Analysis of Trunk Groups with a Short Holding Time Trunk," unpublished work.
2. R. I. Wilkinson, "Theories for Toll Traffic Engineering in the U. S. A.," B.S.T.J., *35*, No. 2 (March 1956), pp. 421–514.
3. Andrezej Klimontowicz, "Grade of Service for Full Available Trunk Groups with Faulty Trunks," Proceedings of the 6th International Teletraffic Conference, Munich, 1970, pp. 212/1—212/6.
4. P. J. Bogert, W. S. Gifford, and I. J. Shapiro, "Directed Maintenance: Troubles Encountered in Placing DDD Calls," unpublished work.
5. L. J. Forys, "Occupancy Probabilities for an Ordered Hunt Group with Unequal Service Times," unpublished work.
6. A. Descloux, "On Overflow Processes of Trunk Groups with Poisson Input and Exponential Service Time," B.S.T.J., *42*, No. 2 (March 1963), pp. 388–397.
7. W. S. Gifford and J. Shapiro, "Effect of the Change in Mean Holding Time Associated with an Equipment Irregularity on Network Trouble Detection and Customer Service," IEEE Trans. Communications, *COM-21*, No. 1 (January 1973), pp. 1–6.
8. J. S. Kaufman, "Distribution of Busy Trunks in a Random Hunt Group Containing Killer Trunks," unpublished work.
9. J. D. C. Little, "A Proof of the Queueing Formula: $L = \lambda W$," Operations Research, *9*, No. 3 (May–June 1961), pp. 383–87.
10. J. Riordan, *Stochastic Service Systems*, New York: John Wiley, 1962.
11. L. Takács, *Introduction to the Theory of Queues*, New York: Oxford University Press, 1962.
12. A. Descloux, "On the Components of Overflow Traffic," unpublished work.
13. J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Princeton, N. J.: Van Nostrand, 1960.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

Entropy Measurements for Nonadaptive and Adaptive, Frame-to-Frame, Linear-Predictive Coding of Video-telephone Signals

By B. G. HASKELL

(Manuscript received January 7, 1975)

Linear predictive coding is an efficient method for transmitting the amplitudes of moving-area picture elements (pels) in a conditional replenishment coder for video-telephone signals. It has been conjectured that if the linear predictor can dynamically adapt to the speed and direction of motion in the scene, then greatly improved performance should result. To test this conjecture and to get a first-order estimate of the possible saving, computer simulations were carried out using pairs of video-telephone frames stored on digital discs. Using this data, picture quality could not be studied. However, differential signal entropies could be estimated, and this was done for several nonadaptive and adaptive linear predictors. Entropies (in bits per moving-area pel) for adaptive linear predictors were significantly lower than for nonadaptive predictors, indicating that substantial bit-rate savings should be possible. However, simpler implementations will have to be devised before adaptive prediction becomes practicable.

I. INTRODUCTION

In coding television pictures for transmission over a digital channel, it is well known that the required bit rate can be significantly reduced by removing various redundancies that exist in the signal, and in recent years methods for removing frame-to-frame redundancy have been investigated.[1] In a conditional replenishment[2] system, only the picture elements (pels) that have changed significantly since the previous frame are transmitted. Their amplitudes as well as their locations must be sent; however, most of the transmission capacity is used in sending the amplitudes. During periods of rapid motion, only every other moving-area pel need be transmitted, i.e., the moving area of the picture can be subsampled[3,4] at half-rate with the unsampled pels being replaced by the average of their neighbors.

Linear predictive coding is an efficient method of transmitting these amplitudes. Channel rates of 1 bit per pel and below have been obtained.[4–7] With this technique, a prediction is formed of each pel to be sent by computing a linear combination of previously transmitted pels. The difference between the actual value and prediction is then quantized and coded for transmission. Since the differential signal is small usually and large only occasionally, variable word-length coding can be used to good advantage in reducing the overall bit rate.

The entropy of the quantized moving-area differential signal provides an estimate of the average number of bits required to transmit a pel. Thus, it is a good yardstick for comparing the performance of various frame-to-frame predictive coders. The entropy will depend on the amount of detail (frequency and amplitude of brightness variations) in the moving area of a frame as well as on the speed of movement in the scene. The overall bit rate, however, is strongly dependent on the number of pels in the moving area which, in turn, is determined by the type of picture to be transmitted. See Ref. 8 for statistics on the number of moving-area pels per frame in typical video-telephone signals.

In Refs. 9 and 10, and by simple extension of the techniques of Ref. 11, it is suggested that if the predictor can dynamically adapt to the speed and direction of motion in the scene, then greatly improved performance should result. For example, if an object is moving left to right at a speed of about 1 pel per frame period (PEF) then for each moving-area pel of the present frame a very good prediction should be obtainable by going back to the previous frame and looking 1 pel to the left. Other types of adaptive linear prediction are described in Refs. 12 to 14. They suggest that the weighting coefficients in the linear predictor be varied adaptively to make the differential signal smaller.

II. COMPUTER SIMULATION

To get some comparison between nonadaptive and adaptive, frame-to-frame, linear predictors, computer simulations were carried out using about three dozen video-telephone picture sequences stored as 8-bit PCM* on digital disc (two successive frames per sequence). With only two frames available per sequence, picture quality could not be studied. However, moving-area differential signal entropies could be estimated, and this was done for several nonadaptive and adaptive predictors.

* Characteristics: 30-Hz frame-rate, 271 lines, 2:1 interlace, 3 dB down at 1 MHz, 2-MHz sampling-rate, 8 bits/sample, 210 visible samples/line.

Frame-to-frame noise in the pictures was small—in most cases, less than 1.5 percent of black-to-white signal amplitude. Thus, detecting the moving-area pels was not difficult. This was done as follows:

(*i*) Frame-to-frame differences larger in magnitude than 4 out of a possible 255 were detected.

(*ii*) If a significant change had two insignificant changes directly to the left and two insignificant changes directly to the right, or if it had two insignificant changes directly above and two insignificant changes directly below, it was deemed to be insignificant, i.e., caused by noise rather than movement.

(*iii*) Finally, horizontal gaps of six pels or less between significant changes were deemed to be also in the moving area.

This procedure defines the moving-area very well. See Ref. 4, Figs. 1, 4, and 7 for pictorial examples.

Figure 1 shows some of the pictures used. Figures 1a to 1c are scenes containing a mannequin's head, which could be moved horizontally at various speeds. The smaller the head size, the more detail there is in the moving area. Thus, with these scenes results could be obtained for various speeds and for various amounts of moving-area detail.

About half of the picture sequences were of live subjects engaged in typical video-telephone conversations, such as shown in Figs. 1d and 1e. These scenes were important in comparing different linear predictors because they were more representative of what would normally be encountered in practice. The speed was not constant over the whole moving area as it was with the mannequin head, and there were more variations in lighting and picture detail. For these scenes the speed of movement could only be estimated to the nearest PEF (pels per frame period) by observing the frame-to-frame displacement of the edge of the moving area.

In video-telephone scenes, speeds range from slow (0.5 PEF) to very fast (4 PEF). Very rapid movement is rare, however, and in such instances the viewer is less critical of picture quality since he is already accustomed to seeing blurred moving-areas in cinema and television pictures.

III. NONADAPTIVE LINEAR PREDICTIVE CODING

Figure 2 shows two successive frames with interlacing fields (two interlaced fields per frame). Suppose Z is a moving-area pel we wish to transmit. Pels A, B, C, G, and H are in the field presently being scanned; pels D, E, F, R, S, and T are in the previous field; and the remaining pels are one frame period back from the present field. Pel M is the previous frame value of Z. The general linear prediction of

Fig. 1—Typical pictures used in the simulations. (a) Small head. (b) Medium head. (c) Large head. (d) and (e) Live subjects.

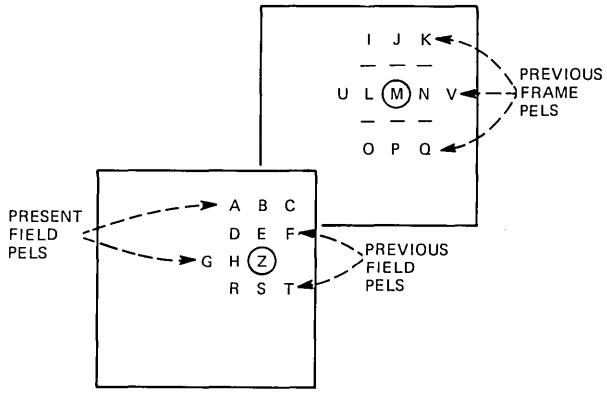


Fig. 2—Two successive television frames with interlacing assumed (two interlaced fields per frame). Pels Z and M are exactly one frame period apart.

Z based on the previously transmitted pels, which are nearby both spatially and temporally, is given by

$$P_Z = \alpha_1 A + \alpha_2 B + \alpha_3 C + \cdots + \alpha_{22} V, \tag{1}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_{22}$ are the weighting coefficients.

Since Z is 8-bit PCM and P_Z is clipped to 8-bit PCM, the differential signal Z-P_Z can assume any of 511 levels. The entropy of the 511-level signal is a measure of the relative performance of different predictors. However, in practice, a more coarsely quantized signal (consistent with acceptable picture quality) would probably be transmitted to reduce the overall bit rate.

Picture quality could not be observed in the simulations. However, to get a rough estimate of bit rate for linear predictors with coarser quantization, a compromise 35-level* quantization scale was chosen that is slightly coarser than the quantizer of Ref. 4 for frame differences and somewhat finer than the one used in Ref. 7 during periods of movement for element differences. Using this quantization scale, slope overload rarely occurs, and the predominant picture degradation is granular noise in the moving areas. This has been verified in recent preliminary simulations using thirty consecutive frames stored on digital disc. Using the nonadaptive and adaptive linear predictors mentioned in this paper, there is little difference in picture quality in the limited picture sample studied. However, much work remains to be done in this area.

* On a scale of 0 to 255, the levels are: 0, ±5, ±14, ±22, ±30, ±40, ±50, ±60, ±70, ±82, ±94, ±106, ±118, ±130, ±142, ±154, ±166, and ±178.

The effects of the predictive coder feedback loop are ignored in the entropy measurements. Feedback can affect results to a significant degree if quantization is extremely coarse. Even so, the entropy measurements reported here are in close agreement with those of Refs. 4 and 7.

For various nonadaptive linear predictors, Figs. 3 to 8 show the entropy (bits per moving-area pel) of the differential signal as a function of the speed of movement (pels per frame period or PEFs). Results are summarized in Table I. For these cases, more moving-area detail (smaller head) resulted in somewhat higher entropies, as would be expected, but moving-head results were still fairly close to each other. The live-subject entropies were close to the moving-head entropies for the most part, even though there was considerable variation in moving-area picture detail.

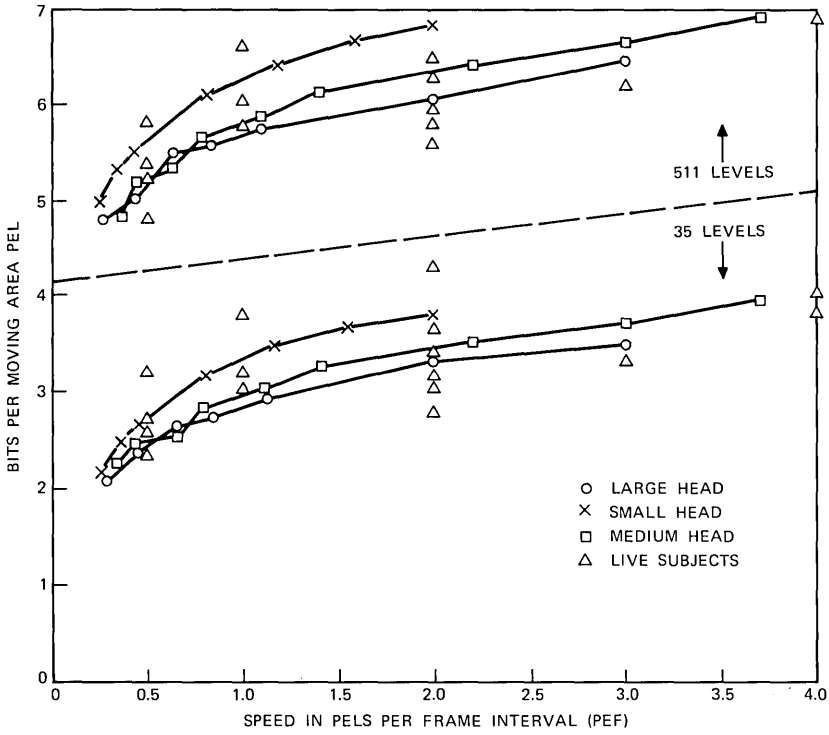


Fig. 3—Entropy of the frame difference signal in the moving area versus speed. $P_Z = M$. Starting with 8-bit PCM as was done in these simulations the differential signal could assume any of 511 levels. Results are also shown for coarser quantization to 35 levels which still gives very good picture quality. Solid curves are for the mannequin head at various distances from the camera. Unconnected points are for live subjects, such as in Figs. 1d and 1e.

Table I — Entropies of some nonadaptive linear predictors

Transmitted Signal $Z - P_Z$	P_Z (See Fig. 2)	Entropies in Bits Per Moving-Area Pel (35-Level Quantization)
Frame difference	M	≈2.1–3.9
Element difference	H	≈2.0–3.7
Element difference of frame difference	$M + H - L$	≈1.8–3.1
Line difference of frame difference	$M + B - J$	≈1.5–3.5
Field difference	$(E + S)/2$	≈1.8–3.2
Element difference of field difference	$H + (E + S)/2 - (D + R)/2$	≈1.5–2.5

The frame-difference[4] entropy (Fig. 3) increases with speed as expected. However, the element-difference entropy (Fig. 4) decreases as the speed increases, because of blurring introduced by the camera.[7] It drops below the frame-difference entropy at a speed of about 1

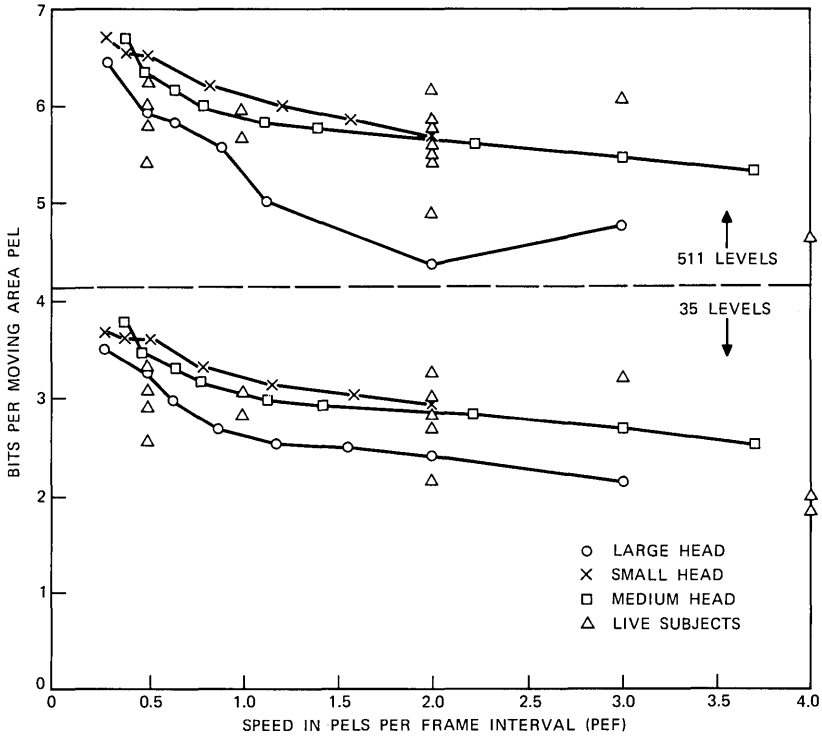


Fig. 4—Entropy of the element difference signal in the moving area versus speed. $P_Z = H$.

PEF.[7] Subsampling at half-rate, however, causes it to rise above the levels shown in Fig. 4.

The element difference of frame-difference entropy[15] and line difference of frame-difference entropy[16] (Figs. 5 and 6) are very close to each other even though with interlace (see Fig. 2) the previous line is further away from pel Z than is the previous element. D. J. Connor has pointed out that this occurs because movement in the scenes is mostly in the horizontal direction. However, the line-difference of frame-difference signal has the advantage of being unaffected by subsampling along the line.

The field-difference entropy[17] (Fig. 7) is lower than the frame-difference entropy, except at very slow speeds, because of the spatial and temporal closeness of the previous field pels. It compares well even with the double derivative signal of Figs. 5 and 6. The element difference of field-difference entropy (Fig. 8) is smaller than any of the others and, other factors ignored, would be the logical choice in a non-adaptive, frame-to-frame, linear predictive coder. However, sub-

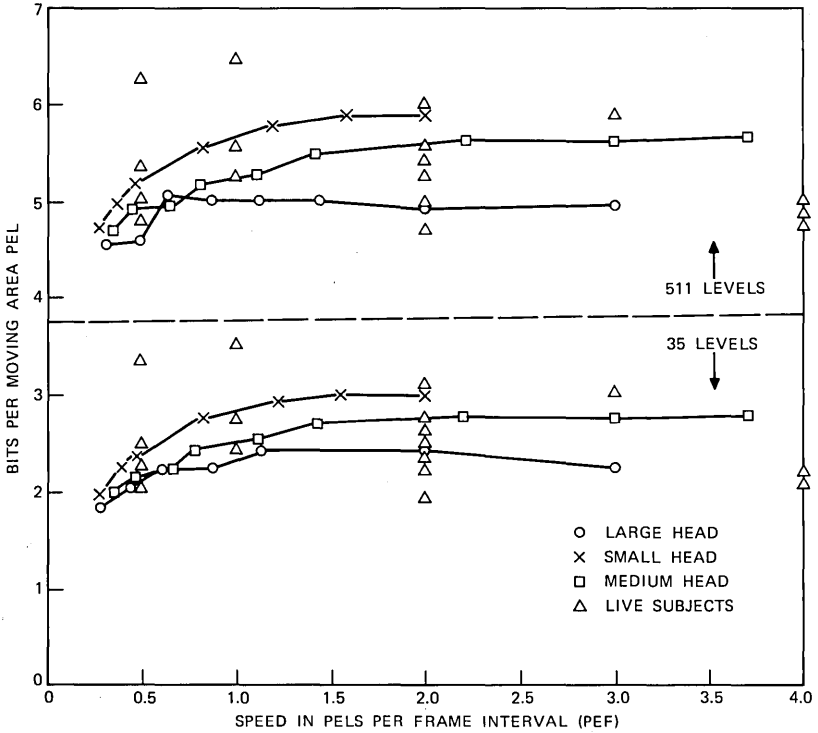


Fig. 5—Entropy of the element difference of frame-difference signal in the moving area versus speed. $P_Z = M + H - L$.

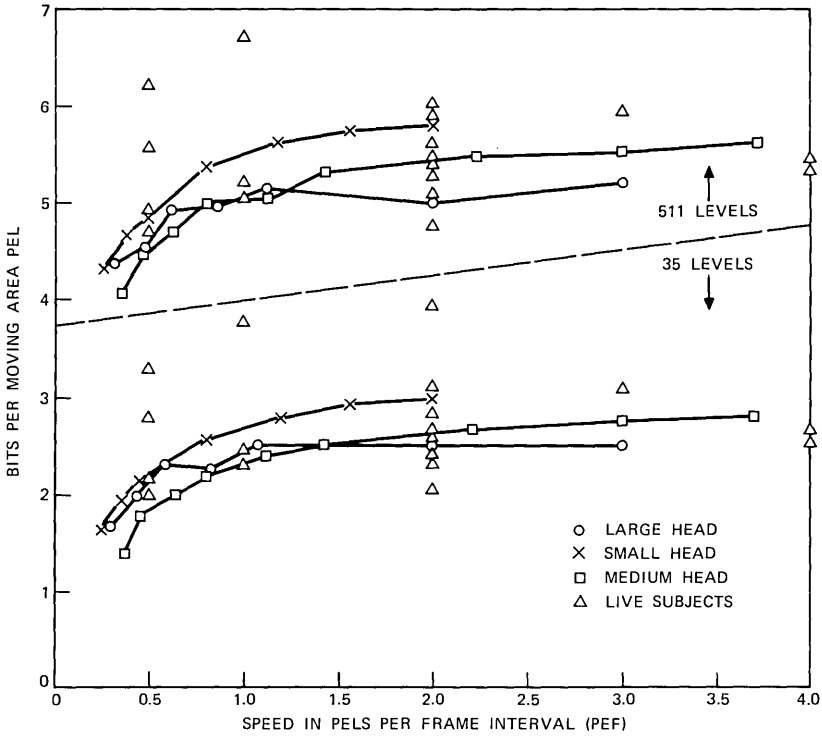


Fig. 6—Entropy of the line difference of frame-difference signal in the moving area versus speed. $P_Z = M + B - J$.

sampling along the line leads to a considerable increase in entropy, putting it above the line difference of frame-difference entropy. Furthermore, conditional vertical subsampling[18,6] makes impracticable the use of any predictive coder that uses previous field pels.

IV. ADAPTIVE FRAME-DIFFERENTIAL CODING WITH MOVEMENT COMPENSATION

In Refs. 9 and 10, and by simple extension of the techniques of Ref. 11, frame-differential coding is adapted to the speed and direction of movement in the scene. Thus, if in Fig. 2 the speed of the moving object is about 1 PEF left to right, then pel L is a much better prediction of Z than M is. Similarly, if movement is 1 PEF right to left, then pel N is a better prediction. Such a coder must first estimate the velocity (speed and direction) of the moving object and then transmit this estimate to the receiver before sending the quantized moving-area differential signal.

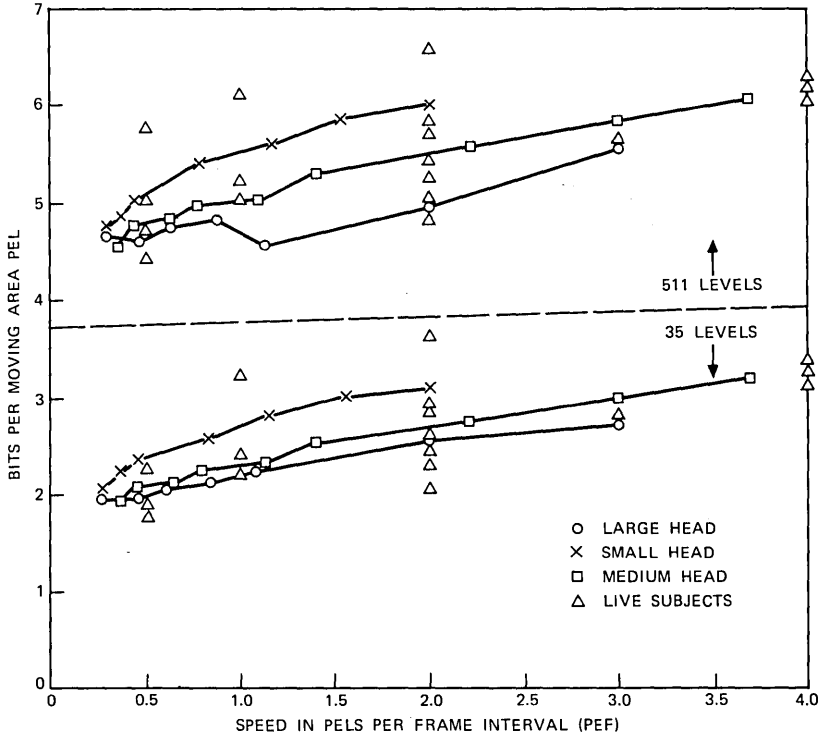


Fig. 7—Entropy of the field-difference signal in the moving area versus speed. $P_Z = (E + S)/2$.

Some computer simulations were carried out to test such schemes. First, each field was divided into 64 smaller blocks of 27 pels × 16 lines each, in an attempt to accommodate velocity variations within the field. Within a block the velocity was assumed constant. Then, for pel Z in the moving-area of a block, the 17 differences (see Fig. 2) Z-D, Z-E, $\cdots$, Z-Q were computed between Z and the six previous field pels and the 11 previous frame pels located relative to Z, as shown in Fig. 2. The magnitudes of these 17 differences were each summed over the moving area of the block. Following this, the 17 accumulated sums were examined to determine the smallest one. Then, for each moving-area pel Z, the pel in the relative position that yielded the smallest accumulated sum was used as a prediction within that block, and statistics of Z-P_Z were measured.

This technique always gave a prediction which corresponded to the correct direction of motion, and was fairly accurate with regard to speed for speeds $\leqq 2$ PEF. However, for a given block, the accumulated sums were all quite close to each other, reflecting the high pel-to-pel correla-

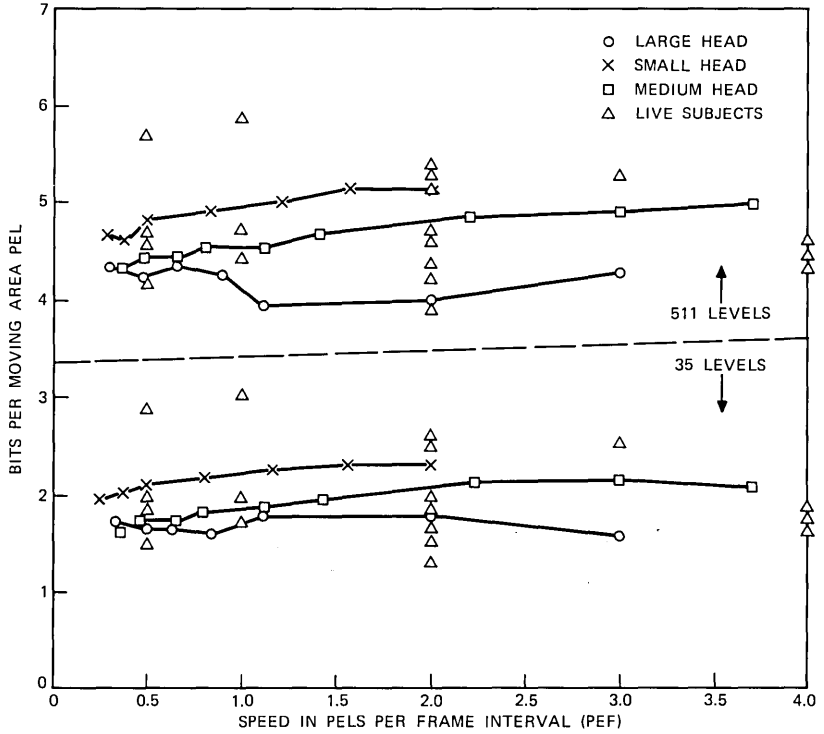


Fig. 8—Entropy of the element difference of field-difference signal in the moving area versus speed. $P_Z = (E + S)/2 + H - (D + R)/2$.

tion present in most television pictures. In low-detail moving areas, of course, tracking was less accurate; but in these regions choice of the pel to be used as a prediction was also less important since they all had approximately the same value anyway.

Entropies of the differential signal Z-P_Z are shown in Fig. 9, again with no additional quantization and with 35-level quantization. Results are comparable to those for the element difference of field difference, but they are much less regular. In addition, there is a sharp dip around 1 PEF where, for the moving head at least, pel L (see Fig. 2) should be a near perfect predictor of pel Z. There is another less sharp dip near 2 PEF. These results are encouraging, but they indicate that movement at a nonintegral number of PEF should be handled in some other manner.

V. MINIMUM MEAN-SQUARE-ERROR LINEAR PREDICTION

In the general case of adaptive linear predictive coding, the weighting coefficients of eq. (1) are changed periodically to obtain a small

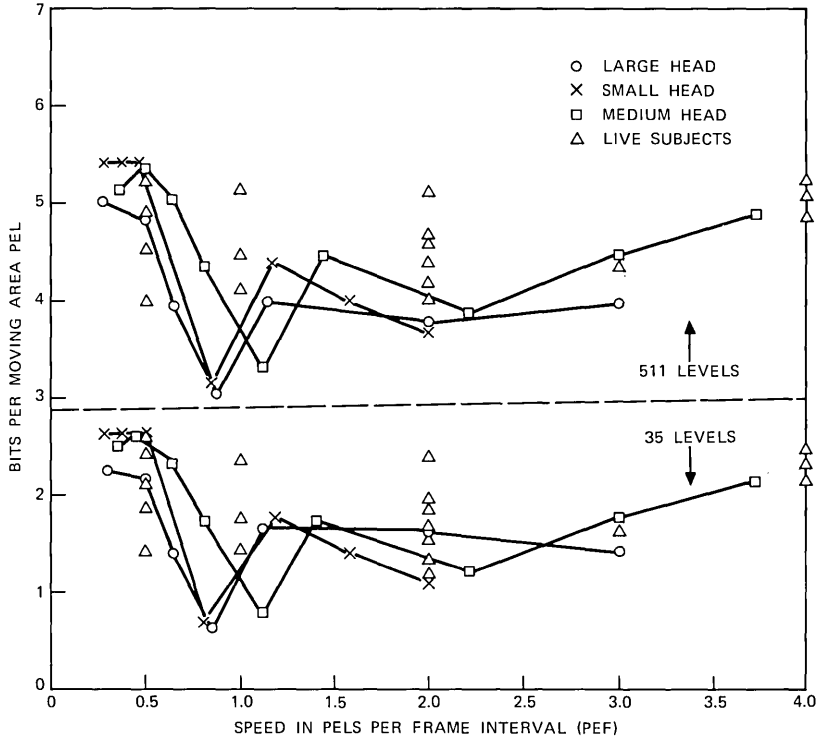


Fig. 9—Entropy of the differential signal in the moving area when movement compensation techniques are carried out. In this case, P_Z equals the pel in the previous field or frame that gives the smallest $|Z - P_Z|$ averaged over the moving area. A sharp dip in the entropy occurs near 1 PEF.

entropy for the differential signal Z-P_Z. If a block of pels is scanned before coding to derive an appropriate set of α's, then these would have to be sent to the receiver, followed by the moving-area differential signals Z-P_Z.

In general, it is not known how to find the α's that minimize the entropy of Z-P_Z. It is possible, however, to determine the α's that minimize the mean squared value of Z-P_Z. As will be seen later, the resulting entropy turns out to be smaller than with any of the previously discussed methods. In fact, using search techniques, it is not possible to find α's yielding entropies very much smaller than those obtained using minimum, mean-square-error (MSE), linear prediction.

Let $\mathfrak{M}$ be the set of moving-area pels in the block of pels presently being scanned, and let Z_i be the ith member of $\mathfrak{M}$. Let $Y_{i1} = A_i$, $Y_{i2} = B_i$, $Y_{i3} = C_i$, $\cdots$, $Y_{i22} = V_i$ be the pels located relative to Z_i, as shown in Fig. 2. Then, to minimize the mean squared prediction

error over $\mathfrak{M}$ it is sufficient to minimize

$$Q = \sum_i (Z_i - P_{Z_i})^2 = \sum_i (Z_i - \sum_{j=1}^{22} \alpha_j Y_{ij})^2. \tag{2}$$

Set the partial derivatives with respect to $\alpha_1, \alpha_2, \cdots, \alpha_{22}$ all equal to zero, i.e., for $k = 1, 2, \cdots, 22$

$$\frac{\partial Q}{\partial \alpha_k} = -\sum_i 2(Z_i - \sum_{j=1}^{22} \alpha_j Y_{ij}) Y_{ik} = 0. \tag{3}$$

In matrix notation, these simultaneous equations can be written

$$\mathfrak{A}\boldsymbol{\alpha} = \mathfrak{B}, \tag{4}$$

where the jkth element of the square matrix $\mathfrak{A}$ is

$$\sum_i Y_{ij} Y_{ik}, \tag{5}$$

the kth element of the column matrix $\mathfrak{B}$ is

$$\sum_i Z_i Y_{ik}, \tag{6}$$

and the kth element of the column matrix $\boldsymbol{\alpha}$ is α_k.

Matrix $\mathfrak{A}$ is symmetric, and if it has an inverse, eq. (4) has a unique solution. Otherwise, many solutions exist. Furthermore, any solution to (4) will minimize the Q of (2) since Q is a convex downward function of α.

This procedure should take advantage of many types of redundancy in the picture. By tracking movement in the scene, frame-to-frame redundancy is removed. By using pels in the present and previous field, intraframe redundancy is also removed.

VI. ADAPTIVE CODING USING ONE SET OF α's PER FIELD

In the simulations to be described next, a set of α's was chosen to minimize the squared differential signal averaged over the moving area of the entire field. That is, in the preceding equations, summations with respect to i were over the entire moving area of each field. In this case the α's can be sent using negligible channel capacity.

Entropy results are shown in Fig. 10. Values are significantly less than all of the previously described results. The live subject results are not as good as those of the moving head, however, mainly because of the velocity variation within the moving area of a field. At speeds above about 2 PEF, results are surprisingly good given the pel configuration of Fig. 2, which would, at first, be expected to encompass

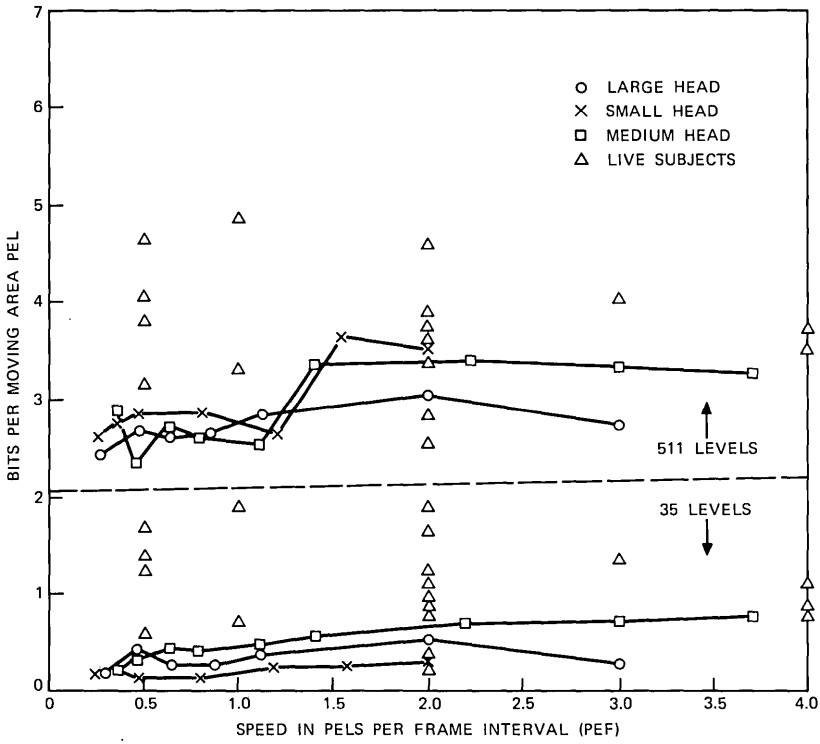


Fig. 10—Entropy of the differential signal in the moving area when minimum MSE linear prediction is carried out. In each field P_Z is the linear combination of the 22 pels shown in Fig. 2, which minimizes $(Z - P_Z)^2$ averaged over the moving area. Entropies are significantly smaller than any of the previous results; however, live subject results are, in general, above the moving area results.

only speeds less than 2 PEF. This may happen because at the higher speeds, significant blurring is introduced by the camera making interpolative predictions more efficient.

Figure 11 shows a representative set of derived weighting coefficients (multiplied by 100) arranged in the configuration of Fig. 2. The speed of the moving head is 0.63 PEF, left to right. Thus, the point on the moving object which is now at pel Z is in the previous frame 0.63 pel to the left of M. The linear predictor places very little weight on pels K, N, Q, and V, since motion is from left to right. It interpolates between pels L and M as it should for this speed, but numerous other differentiations are also present. For example, a form of element-difference of field-difference prediction is attempted in the present frame. Some weight is also placed on previous line pels.

A question which arises at this point is: how much of the removed redundancy is frame-to-frame, and how much of it is intraframe?

Thus, some simulations were carried out where the predictor was only allowed to use pels from the previous frame in its prediction. This is done by deleting in eq. (4) the rows and columns of matrix $\mathfrak{A}$ and the elements of column vectors α and $\mathfrak{B}$ that correspond to the unused pels. Entropy results are shown in Fig. 12.

The results are practically the same as in Fig. 10 up to a speed of about 1.3 PEF which is nearly the range of speeds for which one should expect good performance, given the configuration of Fig. 2. Thus, most of the redundancy removed is frame-to-frame redundancy. Intraframe redundancy becomes important only when the speed is outside the tracking range of the algorithm.

Some simulations were also carried out to test the effect of subsampling at half rate. In this case the predictor is only allowed to use pels *B*, *D*, *F*, *G*, *J*, *M*, *P*, *R*, *T*, *U*, and *V* in the prediction. Entropy results are shown in Fig. 13 for the speed range of 1 to 4 PEF, which is where subsampling would normally be carried out.

Subsampling increased the entropy as one would expect, but the amount of the increase depended on how much detail there was in the moving area of the scene. With 35-level quantization, subsampling increased the entropies by factors ranging from 1.5 to 2.3 for the large head, 1.7 to 2.9 for the medium head, and 3.2 to 4.6 for the small head. As the size of the head decreases, the amount of detail increases and, thus, so does the detrimental effect of subsampling. Of course, if the entropy is more than doubled by subsampling, then subsampling at half rate does not pay.

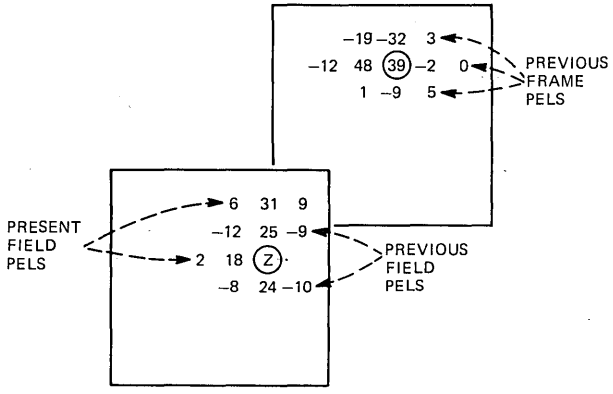


Fig. 11—A representative set of 22 minimum MSE weighting coefficients multiplied by 100 (see Fig. 2). The mannequin head was moving at 0.63 PEF. Thus, a moving object point which is at *Z* in the present frame was 0.63 pels to the left of *M* in the previous frame.

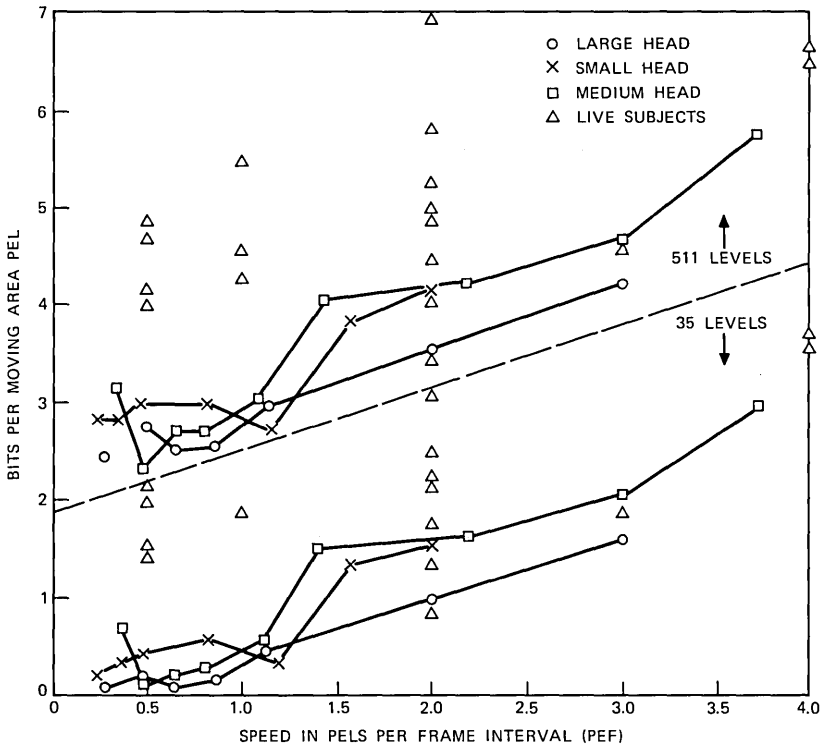


Fig. 12—Entropy of the differential signal in the moving area with minimum MSE linear prediction; P_Z is a linear combination only of the 11 pels in the previous frame. Results are about the same as those in Fig. 10 up to a speed of 1.3 PEF. Thus, in this range, the coder is removing mostly frame-to-frame redundancy.

VII. SUBDIVISION OF THE FIELD INTO SMALLER BLOCKS

In Fig. 10, where one set of weighting coefficients is used for the entire moving area of the field, the entropies corresponding to the live subjects are, in general, larger than those of the moving head because the head moves with pure translation, whereas different parts of live subjects move with different velocities. Velocity variations within a field can be accommodated by first dividing the field into smaller blocks, and then using a different set of weighting coefficients within each block. The α's which give the minimum mean square prediction error over the moving area of a given block can be computed from eq. (4) exactly as before.

Some simulations were carried out to determine the effect of using smaller blocks. A slightly different set of pictures was used for these computations than was used for the previous results. The mannequin head was slightly larger than in Fig. 1b, but the live subjects were

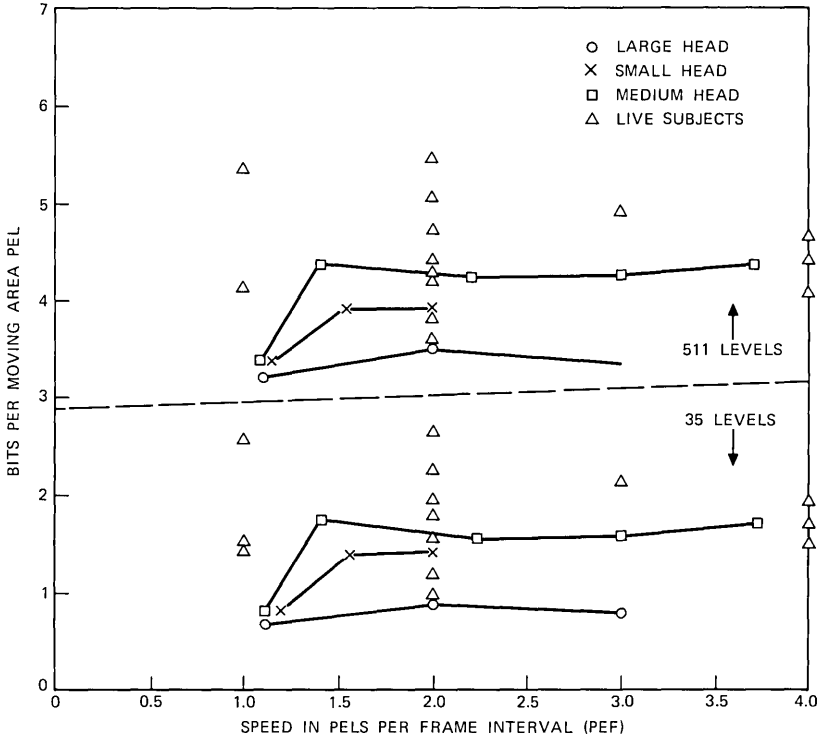


Fig. 13—Entropy of the differential signal in the moving area with minimum MSE linear prediction with subsampling, i.e., P_Z is a linear combination only of pels B, D, F, G, J, M, P, R, T, U, and V.

about the same as in Figs. 1d and 1e. It was found in the simulations that the determinant of matrix $\mathcal{A}$ in eq. (4) was frequently zero, indicating nonunique solutions, i.e., one or more of the α's could be chosen arbitrarily. In this case, the arbitrary α's were set equal to zero and, whenever possible, they were chosen to correspond to pels in the present frame.

Figure 14 shows entropies for 511-level differential signals. The uppermost curve and its associated live-subject points are obtained with no subdivision, as in Fig. 10. The single points are all above the moving-head curve, as before.

The middle curve and points are the entropies which result from subdividing the field into 120 blocks (10 horizontal $\times$ 12 vertical) and using a different set of weighting coefficients for each block. Each block is 21 pels $\times$ 10 lines. Not only are the single points closer to the moving-head curve, but all of the points are considerably below those obtained using no subdivision. Using smaller blocks, therefore, not only ac-

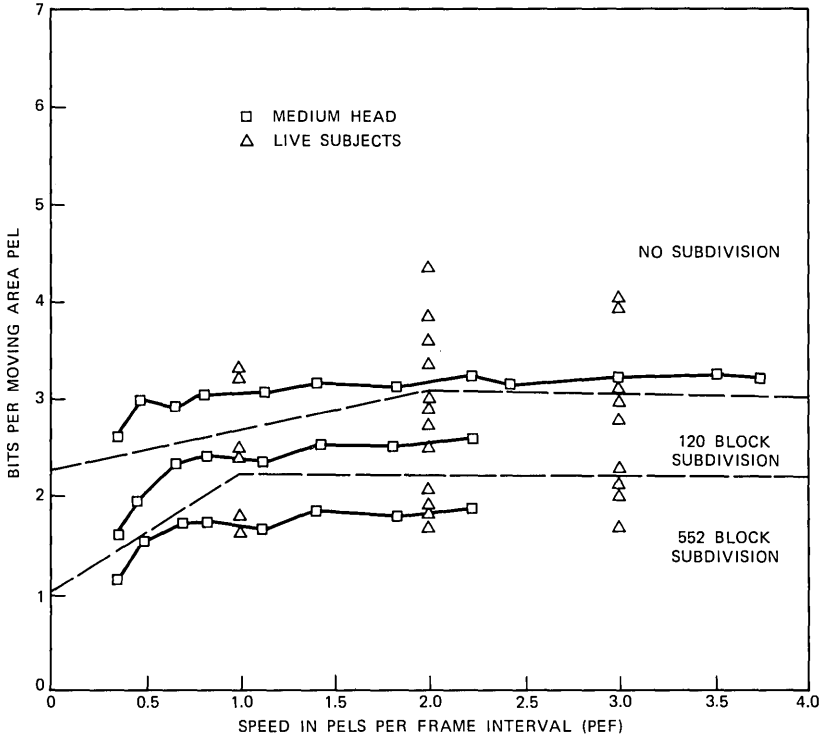


Fig. 14—Entropy of the differential signal in the moving area with minimum MSE linear prediction and prior subdivision of each field into smaller blocks. A different set of weighting coefficients is used in each block. As the block size is reduced, the live-subject results come closer to the moving-head results because variations in velocity within a field are accommodated. In addition, all entropy figures are reduced because more intraframe redundancy is removed.

commodates variations in velocity within the field, but it also removes some in-frame redundancy that could not be removed previously.

There is a penalty to pay, however. The α's of each block in which there is movement must be transmitted to the receiver so that it can correctly decode the received differential signal. If the α's are sent via 8-bit PCM, and movement occurs in 50 percent of the blocks, then $22 \times 8 \times 60 = 10{,}560$ bits (or about 0.84 bit per moving-area pel) must be transmitted just to send the α's. Clearly, some alternate method of transmitting the α's should be used. For example, they might be estimated from the minimum MSE α's of some nearby previously transmitted block, or the vector $\boldsymbol{\alpha}$ might be chosen from some predefined finite set of vectors known both to the coder and decoder.

With even smaller blocks the entropies are further reduced. The bottom curve and points of Fig. 14 result from subdividing the field

into 552 blocks (23 horizontal × 24 vertical). Each block is 9 pels × 5 lines. Coarse quantization of the differential signal in this case gives zero nearly always. Thus, nearly all of the information about the pels is carried in the weighting coefficients which, with 50 percent of the area moving, require 48,576 bits for transmission using 8-bit PCM. So if the blocks are made too small, the algorithm is inefficient because of the bits required to send the α's.

VIII. CONCLUSION

In a conditional replenishment system for transmitting video-telephone pictures, the moving-area picture elements can be coded for transmission in many ways. Sending frame-to-frame differences and subsampling at half rate during active movement is simple, reasonably efficient, and does not fail catastrophically with an occasional transmission error. Sending line-to-line differences of frame-differences and subsampling during active motion is even more efficient and is reasonably simple, but it is somewhat less impervious to transmission errors.

Adaptive linear predictive coding, which takes into account the speed and direction of movement in the scene, results in much greater coding efficiency, indicating that substantial bit-rate savings should be possible. However, system complexity and vulnerability to transmission errors are increased accordingly. This means that some kind of error detection-correction scheme is absolutely necessary if adaptive coding is to be used.

Adaptive coding, which handles speeds of a nonintegral number of PEFs, performs better than if an integral number of PEFs is assumed. Relatively good results are obtained if for each block in the picture the weighting coefficients are chosen to minimize the mean squared moving-area differential signal. This technique requires transmission in some way of the weighting coefficients so that the decoder can correctly decode the received differential signal. Thus, if the block size is too small, then the system is inefficient because of the number of bits required to send the coefficients. However, even using only one set of weighting coefficients per field (block size = entire field), significant reductions in differential signal entropies are measured compared with nonadaptive coding.

Much work remains to be done before these techniques can be used in a practical coder. For example, the subjective aspects of interframe coders which track moving objects are as important as the statistical aspects which are discussed in this paper. Also, many short cuts in implementation must be developed before these methods are economical. Nevertheless, these results provide a valuable yardstick with which

simpler and more practical adaptive frame-to-frame coders can be compared.

IX. ACKNOWLEDGMENTS

Much use was made of the picture-processing computer facility constructed by R. C. Brainard and J. D. Beyer. Without it these results would not have been possible. Also, many fruitful discussions were held with D. J. Connor, J. O. Limb, R. F. W. Pease, and F. W. Mounts.

REFERENCES

1. B. G. Haskell, F. W. Mounts, and J. C. Candy, "Interframe Coding of Videotelephone Pictures," Proceedings of the IEEE, *60*, No. 7 (July 1972), pp. 792–800.
2. F. W. Mounts, "Video Encoding with Conditional Picture Element Replenishment," B.S.T.J., *48*, No. 7 (September 1969), pp. 2545–2554.
3. R. F. W. Pease and J. O. Limb, "Exchange of Spatial and Temporal Resolution in Television Coding," B.S.T.J., *50*, No. 1 (January 1971), pp. 191–200.
4. J. C. Candy et al., "Transmitting Television as Clusters of Frame-to-Frame Differences," B.S.T.J., *50*, No. 6 (July–August 1971), pp. 1889–1917.
5. J. B. Millard, Y. C. Ching, and D. M. Henderson, private communication.
6. D. J. Connor, B. G. Haskell, and F. W. Mounts, "A Frame-to-Frame *Picturephone*® Coder for Signals Containing Differential Quantizing Noise," B.S.T.J., *52*, No. 1 (January 1973), pp. 35–51.
7. J. O. Limb, R. F. W. Pease, and K. A. Walsh, "Combining Intraframe and Frame-to-Frame Coding for Television," B.S.T.J., *53*, No. 6 (July–August 1974), pp. 1137–1173.
8. B. G. Haskell, "Buffer and Channel Sharing by Several Interframe *Picturephone*® Coders," B.S.T.J., *51*, No. 1 (January 1972), pp. 261–289.
9. F. Rocca, "TV Bandwidth Compression Utilizing Frame-to-Frame Correlation and Movement Compensation," Symposium on Picture Bandwidth Compression (M.I.T.), Gordon and Breach, 1972.
10. B. G. Haskell and J. O. Limb, "Predictive Video Encoding Using Measured Subject Velocity," U. S. Patent No. 3,632,865, January 1972.
11. F. K. Manasse, "Directional Correlation—A Technique to Reduce Bandwidth in PCM Television Transmissions," IEEE Transactions on Communication Technology, *COM-15*, No. 2 (April 1967), pp. 204–208.
12. A. V. Balakrishnan, "An Adaptive Nonlinear Data Predictor," Proc. 1962 National Telemetry Conference, 2, Session 6-5.
13. L. D. Davisson, "Theory of Adaptive Data Compression," in *Advances in Communication Systems, 2*, New York: Academic Press, 1966.
14. R. W. Lucky, "Adaptive Redundancy Removal in Data Transmission," B.S.T.J., *47*, No. 4 (April 1968), pp. 549–573.
15. F. W. Mounts, private communication.
16. D. J. Connor, private communication.
17. R. F. W. Pease, private communication.
18. R. F. W. Pease, "Conditional Vertical Subsampling—A Technique to Assist in the Coding of Television Signals," B.S.T.J., *51*, No. 4 (April 1972), pp. 787–802.

Contributors to This Issue

James E. Bennett, B.Sc. (Met. E.), 1959, Carnegie Institute of Technology; M.Sc. (Metallurgy), 1961, and Ph.D. (Metallurgy), 1965, Case Institute of Technology; General Electric Refractory Metals Laboratory, 1964–1966; Battelle Memorial Institute, 1966–1968; Bell Laboratories, 1968—. At the Columbus Laboratories, Mr. Bennett has been conducting fundamental studies on the interaction of liquid mercury with contact metals, interdiffusion in connector materials, phase transformations in and processing behavior of Fe/Co/2–3% V alloys, and contact materials development. Member, AIME, IMS, Alpha Sigma Mu, Sigma Xi.

Dan Bisbee, B.S., 1965, Monmouth College; Bell Laboratories, 1955—. Mr. Bisbee has worked on the design and measurement of millimeter waveguide components. He has also been involved with the study and measurement of transmission losses in bulk glass and optical-fiber waveguides. He is presently engaged in developing techniques for splicing optical fibers and cables.

Edwin L. Chinnock, Stevens Institute of Technology; Bell Laboratories, 1939—. Mr. Chinnock has worked on microwave components, microwave radio relay, and helix waveguide fabrication. He is presently working on optical waveguide components.

Leonard J. Forys, B.S.E.E., 1963, University of Notre Dame; M.S. and E.E., 1965, Massachusetts Institute of Technology; Ph.D., 1968, University of California at Berkeley; Acting Assistant Professor of Electrical Engineering, University of California at Berkeley, 1967–1968; Bell Laboratories, 1968—. Upon joining Bell Laboratories, Mr. Forys was engaged in research and consulting on communication and control theory problems. He later was involved in studies of the analysis and control of an air traffic system and worked on the real-time prediction of demands for a domestic satellite system. He is currently supervisor of the Traffic Studies Group doing capacity studies of electronic switching machines.

D. Gloge, Dipl. Ing., 1961, Dr. Ing., 1964, Technical University of Braunschweig, Germany; Bell Laboratories, 1965—. Mr. Gloge's work has included the design and field testing of various optical trans-

mission media and the application of ultra-fast measuring techniques to optical component studies. He is presently engaged in transmission research related to optical-fiber communication systems.

Barry G. Haskell, B.S., 1964, M.S., 1965, and Ph.D. (Electrical Engineering), 1968, University of California, Berkeley; Research Assistant, University of California, 1965–1968; Bell Laboratories, 1968—. Mr. Haskell is a member of the Visual Communication Research Department. His primary interest is the efficient coding of pictures for transmission at reduced bit rate. Member, Phi Beta Kappa, Sigma Xi, IEEE.

Sing-Hsiung Lin, B.S.E.E., 1963, National Taiwan University; M.S.E.E., 1966, and Ph.D., 1969, University of California, Berkeley; Bell Laboratories, 1969—. At the Electronics Research Laboratory, University of California at Berkeley, Mr. Lin was engaged in research on antennas in plasma media and numerical solutions of antenna problems. Mr. Lin is presently working on wave propagation problems on terrestrial radio systems and earth-satellite radio systems. Member, IEEE, Sigma Xi, AIAA.

Dietrich Marcuse, Diplom Vorpruefung, 1952, Dipl. Phys., 1954, Berlin Free University; D.E.E., 1962, Technische Hochschule, Karlsruhe, Germany; Siemens and Halske (Germany), 1954–1957; Bell Laboratories, 1957—. At Siemens and Halske, Mr. Marcuse was engaged in transmission research and studying coaxial cable and circular waveguide transmission. At Bell Laboratories, he has been engaged in studies of circular electric waveguides and work on gaseous masers. He spent one year (1966–1967) on leave of absence from Bell Laboratories at the University of Utah. He is presently working on the transmission aspect of a light communications system. Mr. Marcuse is the author of three books. Fellow, IEEE; member, Optical Society of America.

J. E. Mazo, B.S. (Physics), 1958, Massachusetts Institute of Technology; M.S. (Physics), 1960, and Ph.D. (Physics), 1963, Syracuse University; Research Associate, Department of Physics, University of Indiana, 1963–1964; Bell Laboratories, 1964—. At the University of Indiana, Mr. Mazo worked on studies of scattering theory. At Bell Laboratories, he has been concerned with problems in data transmission and is now working in the Mathematical Research Center. Member, American Physical Society, IEEE.

E. J. Messerli, B.A.Sc. (E.E.) 1965, University of British Columbia; M.S. (E.E.) 1966, University of California at Berkeley; Ph.D. (E.E. and Comp. Sc.) 1968, University of California at Berkeley; Acting Assistant Professor of Electrical Engineering, University of California at Berkeley, 1968–1969; Bell Laboratories, 1969—. Mr. Messerli has been involved in studies on the analysis and control of an air traffic system and on the demand assignment of capacity for a domestic satellite system. He has also worked on the development of optimization algorithms. Since 1971, his work has been concerned with telephone traffic analysis and modeling. Currently, his work is concerned with the effect of measurement and forecasting errors on the trunk provisioning process.

M. Robert Pinnel, B.Sc. (E.E.), 1966, M.Sc. (Met.E.), 1968, and Ph.D. (Materials Sci.), 1970, Drexel University; Bell Laboratories, 1970—. Mr. Pinnel has been engaged in research on the characterization of the physical and mechanical behavior of numerous copper-based alloys used as spring materials, interdiffusion in electrical connector materials, the detailed characterization of the Fe/Co/2–3% V alloy system, and the interaction of liquid mercury with numerous metallic elements. His current interests are in the areas of magnetic materials and solid-state diffusion. Member, ASM, AIME, Tau Beta Pi, Phi Kappa Phi, Alpha Sigma Mu, Eta Kappa Nu.

Vasant K. Prabhu, B.E. (Dist.), 1962, Indian Institute of Science, Bangalore, India; S.M., 1963, Sc.D., 1966, Massachusetts Institute of Technology; Bell Laboratories, 1966—. Mr. Prabhu has been concerned with various theoretical problems in solid-state microwave devices, noise, and optical communication systems. Member, IEEE, Eta Kappa Nu, Sigma Xi, Tau Beta Pi, and Commission 6 of URSI.

Harrison E. Rowe, B.S., 1948, M.S., 1950, and Sc.D., 1952, Massachusetts Institute of Technology; Bell Laboratories, 1952—. Mr. Rowe's fields of interest have included parametric amplifier theory, noise and communication theory, modulation theory, propagation in random media, and related problems in waveguide, radio, and optical systems. Fellow, IEEE; member, Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and Commission 6 of URSI.

Adel A. M. Saleh, B.Sc. (Electrical Engineering), 1963, University of Alexandria, Alexandria, Egypt; S.M., 1967, Ph.D. (Electrical

Engineering), 1970, Massachusetts Institute of Technology; University of Alexandria, 1963–1965, Instructor and Demonstrator; Bell Laboratories, 1970—. Mr. Saleh is engaged in studies of millimeter-wave components, devices, and propagation. Member, Sigma Xi, IEEE.

Peter W. Smith, B.Sc., Mathematics and Physics, 1958, and M.Sc. and Ph.D., Physics, 1961 and 1964, McGill University; Visiting Mackay Lecturer in Electrical Engineering, 1970, University of California, Berkeley; Bell Laboratories, 1964—. Mr. Smith has investigated a number of systems for obtaining single-frequency laser operation and is currently investigating the use of waveguide techniques for producing miniature gas lasers. Member, American Physical Society, Optical Society of America, IEEE.

Richard H. Turrin, B.S.E.E., 1956, Newark College of Engineering; M.S.E.E., 1960, New York University; Bell Laboratories, 1956—. Mr. Turrin has been concerned with propagation and antenna work at micro- and millimeter wavelengths. He participated in the design of the *Telstar*® satellite ground-station antennas and is presently engaged in studies of satellite antennas. Member, IEEE, Eta Kappa Nu, Tau Beta Pi.

THE BELL SYSTEM TECHNICAL JOURNAL is abstracted or indexed by *Abstract Journal in Earthquake Engineering, Applied Mechanics Review, Applied Science & Technology Index, Chemical Abstracts, Computer Abstracts, Computer & Control Abstracts, Current Papers in Electrical & Electronic Engineering, Current Papers on Computers & Control, Electrical & Electronic Abstracts, Electronics & Communications Abstracts Journal, The Engineering Index, International Aerospace Abstracts, Journal of Current Laser Abstracts, Language and Language Behavior Abstracts, Mathematical Reviews, Metals Abstracts, Science Abstracts,* and *Solid State Abstracts Journal.* Reproductions of the Journal by years are available in microform from University Microfilms, 300 N. Zeeb Road, Ann Arbor, Michigan 48106.

Bell System