

THE JULY–AUGUST 1981 VOL. 60, NO. 6, PART 1

BELL SYSTEM TECHNICAL JOURNAL



| | | |
|--|---|------------|
| A. Vigants | Microwave Radio Obstruction Fading | 785 |
| J. A. Schiavone | Prediction of Positive Refractivity Gradients for Line-of-Sight Microwave Radio Paths | 803 |
| N. H. E. Weste | MULGA—An Interactive Symbolic Layout System for the Design of Integrated Circuits | 823 |
| W. B. Gardner, S. R. Nagel, M. I. Schwartz, F. V. DiMarcello, C. R. Lovelace, D. L. Brownlow, M. R. Santana, and E. V. Sigety | The Effect of Optical Fiber Core and Cladding Diameter on the Loss Added by Packaging and Thermal Cycling | 859 |
| A. S. Acampora | Analysis of Maximum-Likelihood Sequence Estimation Performance for Quadrature Amplitude Modulation | 865 |
| X. M. Chang, D. Z. Du, and F. I. Hwang | Characterizations for Series-Parallel Channel Graphs | 887 |
| L. R. Rabiner, J. G. Wilpon, and J. G. Ackenhusen | On the Effects of Varying Analysis Parameters on an LPC-Based Isolated Work Recognizer | 893 |
| P. E. Butzien | Three-Dimensional Radiation Characteristics of a Pyramidal Horn-Reflector Antenna | 913 |
| K. Ogawa | Noise Caused by GaAsMESFETs in Optical Receivers | 923 |

T CARRIER CHARACTERIZATION

| | | |
|--|---|------------|
| J. R. Davis and A. K. Reilly | T Carrier Characterization Program—Overview | 929 |
| J. R. Davis, L. E. Forman, and L. T. Nguyen | The Metropolitan Digital Trunk Plant | 933 |

(Contents continued on back cover)

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

D. E. PROCKNOW, *President*, *Western Electric Company*
I. M. ROSS, *President*, *Bell Telephone Laboratories, Incorporated*
W. M. ELLINGHAUS, *President*, *American Telephone and Telegraph Company*

EDITORIAL COMMITTEE

A. A. PENZIAS, *Chairman*
A. G. CHYNOWETH L. SCHENKER
R. P. CLAGETT W. B. SMITH
T. H. CROWLEY G. SPIRO
I. DORROS J. W. TIMKO
R. A. KELLEY S. HORING

EDITORIAL STAFF

B. G. KING, *Editor*
PIERCE WHEELER, *Associate Editor*
HEDWIG A. DEUSCHLE, *Assistant Editor*
H. M. PURVIANCE, *Art Editor*
B. G. GRUBER, *Circulation*

THE BELL SYSTEM TECHNICAL JOURNAL is published monthly, except for the May-June and July-August combined issues, by the American Telephone and Telegraph Company, C. L. Brown, Chairman and Chief Executive Officer; W. M. Ellinghaus, President; V. A. Dwyer, Vice President and Treasurer; F. A. Hutson, Jr., Secretary. Editorial inquiries should be addressed to the Editor, The Bell System Technical Journal, Bell Laboratories, Room WB 1L-331, Crawfords Corner Road, Holmdel, N.J. 07733. Checks for subscriptions should be made payable to The Bell System Technical Journal and should be addressed to Bell Laboratories, Circulation Group, Whippany Road, Whippany, N.J. 07981. Subscriptions \$20.00 per year; single copies \$2.00 each. Foreign postage \$1.00 per year; 15 cents per copy. Printed in U.S.A. Second-class postage paid at New Providence, New Jersey 07974 and additional mailing offices.

© 1981 American Telephone and Telegraph Company. ISSN0005-8580

Single copies of material from this issue of The Bell System Technical Journal may be reproduced for personal, noncommercial use. Permission to make multiple copies must be obtained from the editor.

Comments on the technical content of any article or brief are welcome. These and other editorial inquiries should be addressed to the Editor, The Bell System Technical Journal, Bell Laboratories, Room WB 1L-331, Crawfords Corner Road, Holmdel, N.J. 07733. Comments and inquiries, whether or not published, shall not be regarded as confidential or otherwise restricted in use and will become the property of the American Telephone and Telegraph Company. Comments selected for publication may be edited for brevity, subject to author approval.

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 60

July-August 1981

Number 6

Copyright © 1981 American Telephone and Telegraph Company. Printed in U.S.A.

Microwave Radio Obstruction Fading

By A. VIGANTS

(Manuscript received February 20, 1981)

The occurrence of obstruction fading on line-of-sight microwave radio paths can require tower heights that are substantially larger than those needed for transmission in a normal atmosphere. In this paper, we show that the occurrence probability of obstruction fading can be estimated from the probability distribution of positive vertical refractivity gradients. The occurrence of positive refractivity gradients is related to general meteorological variables in a companion paper. The work summarized is part of a recent effort that has resulted in a new tower-height design method where tower-height requirements are quantitatively determined from transmission performance requirements.

I. INTRODUCTION

The design and the transmission performance of terrestrial line-of-sight microwave radio paths is greatly influenced by two clear-air propagation phenomena, obstruction and multipath fading. Both types of fading are caused by anomalous stratification of the atmosphere. In the case of obstruction fading, atmospheric stratification temporarily changes the transmission path to such an extent that it becomes blocked by the terrain. In the case of multipath fading, a different type of atmospheric stratification temporarily creates multiple transmission paths that cause destructive interference of a number of waves at the receiving antenna.

Satisfactory transmission performance is attained when the height of microwave radio towers is sufficient to reduce the occurrence of obstruction fading to a tolerable value, and when diversity protection is used to eliminate transmission impairments caused by multipath

fading. Methods for the estimation of multipath fading and the associated transmission performance have been developed, but a similar body of knowledge does not exist for obstruction fading.¹⁻⁶ One reason is that the variation of the occurrence of obstruction fading is much larger than that of multipath fading, which implies that the cost for establishing a data base for obstruction fading by direct measurement would be prohibitive.

However, obstruction fading can be related to meteorological variables for which a data base already exists. Recent work on this approach has resulted in a method that permits estimation of obstruction fading for any location in the United States. One part of this work, described in a companion paper, is the determination of probability distributions of positive vertical gradients of atmospheric microwave refractivity.⁷ In this paper, we show that the probability of obstruction fading can be determined from probability distributions of the positive refractivity gradients. A mathematical model of obstruction fading is formulated to accomplish this. Obstruction-fading measurements are used to determine an obstruction loss expression for deep obstruction fading. A height interval for the definition of a positive refractivity gradient probability is also determined.

Tower-height requirements follow from transmission performance requirements when we use the mathematical model of obstruction fading. A new tower-height design method incorporating this approach supersedes the historical clearance rules (summarized in Ref. 2) previously used in the Bell System that did not relate performance to height in a quantitative manner. Additional advantages of the new method are geographic resolution and performance-based comparison of designs. We discuss transmission performance requirements applicable to tower-height determination in Section IX.

II. DAYTIME PROPAGATION

Microwave radio propagation on a line-of-sight path in a normal daytime atmosphere can be characterized by the trajectory of a ray that passes from the transmitting antenna to the receiving antenna. The trajectory is curved because the microwave index of refraction of the atmosphere decreases with height. The index of refraction (n) is usually described in terms of refractivity,⁸

$$N = (n - 1)10^6. \quad (1)$$

As an example, a refractivity of 320 N -units corresponds to a value of 1.000320 for the index of refraction. The vertical change of the refractivity can be considered to be a linear decrease with height in the first few hundred feet above ground.^{9,10} A single parameter of the atmosphere, the refractivity gradient (N'), determines the trajectory of the

ray under such conditions. An alternate parameter, denoted by K , is frequently used to describe the refractivity gradient.¹¹⁻¹³ The origin of K is a geometric transformation to an equivalent earth without atmosphere and a radius which is K times that of the actual earth. The conversion between the two parameters is

$$N' = -157(1 - K^{-1}), \quad (2)$$

where K is dimensionless, and N' is expressed in N -units per kilometer.^{12,13} The daytime "standard" value of K is $4/3$, which corresponds to a refractivity gradient of -39.25 N -units/km. Expressed as a function of K , the equation for the trajectory of the ray is

$$H_d = (D_1 D_2 / 1.5)(1 - K^{-1}), \quad (3)$$

where H_d is the vertical deviation of the trajectory from a straight line between the antennas, and D_1 and D_2 are the distances from a point of interest to the ends of the path (Fig. 1). The units of H_d are feet when D_1 and D_2 are in miles.

The nearness of the trajectory of the ray to the terrain is usually described by a normalized vertical distance, which is the vertical distance between a terrain point of interest and the trajectory, divided by the radius of the first Fresnel zone,^{11,12}

$$F_1 = 72.1(D_1 D_2 / fD)^{1/2}. \quad (4)$$

The units of F_1 are feet when the frequency f is in GHz and the path length D is in miles ($D = D_1 + D_2$). The normalization to F_1 introduces

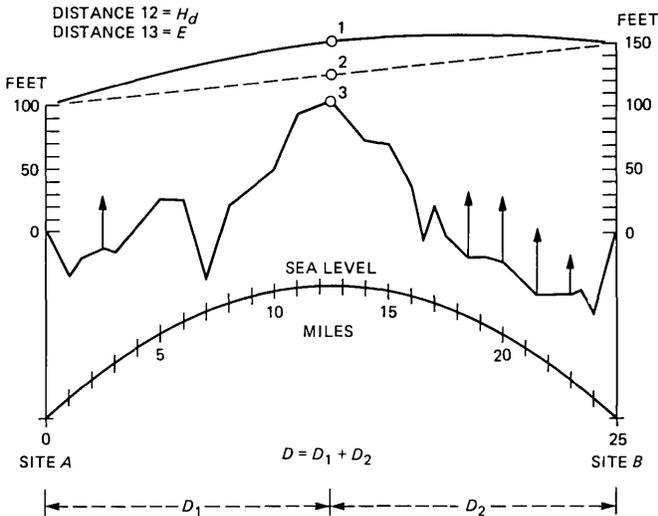


Fig. 1—Example of ray trajectory between antennas when $K = 4/3$ (solid line). In the absence of the atmosphere, or in the case of an atmosphere in which the vertical refractivity gradient is zero, the trajectory would be a straight line (dashed line). Arrows on the path profile denote tree cover.

scaling for frequency and distance that is convenient when signal attenuation due to the proximity of the terrain is considered. The attenuation is controlled by the terrain point with the smallest normalized vertical distance to the trajectory of the ray. This point on the terrain is referred to as the controlling obstruction. Tree cover and buildings must be included in the description of the terrain when the controlling obstruction is determined. The distance from the controlling obstruction to the trajectory of the ray is called clearance (distance E in Fig. 1).

Microwave radio paths are usually designed to avoid daytime attenuation caused by the proximity of the terrain to the trajectory of the ray. This is accomplished by using antenna heights such that the normalized clearance E/F_1 is unity or larger when K is $4/3$.^{2,12} Normalized clearances on many paths exceed unity at $K = 4/3$ because of antenna heights required to avoid excessive obstruction fading.

III. OBSTRUCTED PROPAGATION

Humid air can become positioned above drier surface-based air as a result of atmospheric stratification (layering).⁷ Such stratification, when it occurs, develops gradually during the night and persists until broken up by air turbulence, caused either by a change in wind pattern or by a change in atmospheric conditions associated with sunrise. The resulting increase of refractivity with height causes the trajectory of the ray between the antennas to deviate downward from its normal daytime location. When the refractivity increase becomes sufficiently steep, the trajectory is blocked by the terrain, and the receiving antenna is placed in a shadow zone. The associated reduction in the received signal power is referred to as obstruction fading.

During obstruction fading, the lowest layer of the atmosphere controls propagation. The increases of refractivity with height in this layer are not necessarily linear, but it can be argued that the deviation of the trajectory is determined by a space-averaged gradient. Characterization of obstructed propagation by a single atmospheric parameter, a positive refractivity gradient, is therefore appropriate. This leads to the concept of a virtual trajectory (Fig. 2), which is a trajectory between the antennas that would exist for a particular positive refractivity gradient in the absence of the obstruction. The resulting geometrical description of blockage is similar to that of clearance, except that vertical distances from the terrain to the virtual trajectory are negative to signify blockage. The controlling obstruction in this case is that for which the normalized vertical distance has the largest negative value.

Quantitative description of obstruction fading in terms of the above geometry requires consideration of two topics. First, a relationship

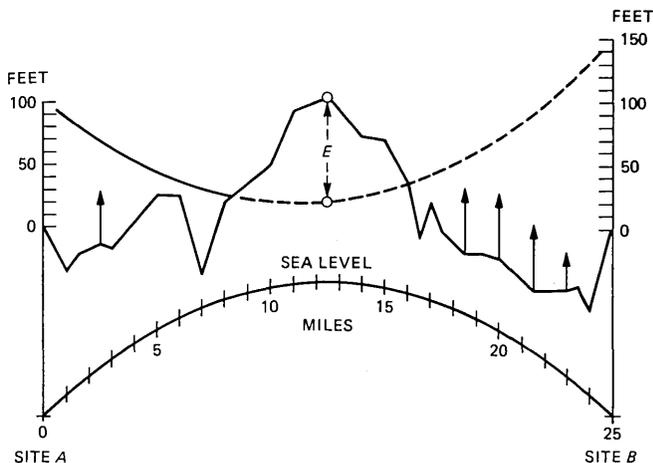


Fig. 2—Example of virtual trajectory between antennas when $K = 1/2$.

between obstruction loss and the normalized blockage E/F_1 at the controlling obstruction must be obtained. Second, probabilities of positive refractivity gradients must be translated into probabilities of blockage. We use the obstruction-fading measurements summarized in Section IV to accomplish this.

IV. MEASURED OBSTRUCTION FADING

We obtained quantitative information on obstruction fading from fading measurements carried out for one year (February 1974–January 1975) at 4.13 GHz on a 24.3-mi path in northeastern Florida where obstruction fading is prevalent. The measurement path was flat with substantial tree cover. At the transmitting site, referred to as station T , the ground elevation was 40 ft relative to the mean sea level. The centerline of the transmitting antenna (10-ft horn reflector) was 270 ft above ground. The controlling obstruction, located at 11.9 mi from the receiving site, consisted of 50-ft tree cover on ground that was 15 ft above the mean sea level. At the receiving site, referred to as station R , the ground elevation was 15 ft above the mean sea level. The centerline of the receiving antenna (10-ft horn reflector) was 220 ft above ground. Reception was also monitored on an antenna (10-ft dish) with a centerline at 185 ft above ground. We adjusted instrumentation gains to simulate operation with equal antenna gains and equal waveguide losses. A third receiving antenna (6-ft dish), added on November 1, 1974, had a centerline height of 73 ft above ground. During substantial fading experienced in November, the presence of the third receiving antenna permitted further verification of earlier conclusions that the amount of fading at station R was inversely

related to antenna height in a manner that indicated the presence of obstruction fading.

We used Portable Propagation Recorders (PPRs) to record the experimental results. These devices sampled the received signals ten times per second and accumulated the time during which the strength of the signals was below a set of levels. The levels for the top antenna were, in dB relative to daytime normal, -10 and then -20 to -40 in 5-dB steps. Limitations of recording capacity forced us to assign a smaller set of levels to the two lower antennas. We transmitted the accumulated data to Bell Laboratories at Holmdel, New Jersey, once every working day. The accumulated data were also automatically punched out on paper tape at station *R* at noon every day. This permitted the signal statistics to be separated into noon-to-noon periods. The interpretation of the statistics was aided by strip charts (3 in/h) that accompanied each 24-h block.

Obstruction fading was present on four nights during 24-h periods that ended at noon on March 8, June 20, November 3, and November 17 in the 1-yr measurement period. We determined the presence of obstruction fading on the basis of three criteria: depressed levels of received signals on a time scale of tens of minutes, increase of fading with decrease of antenna height, and high degree of simultaneous fading on the two upper antennas. The last criterion separates nights with obstruction fading from those with heavy multipath fading.

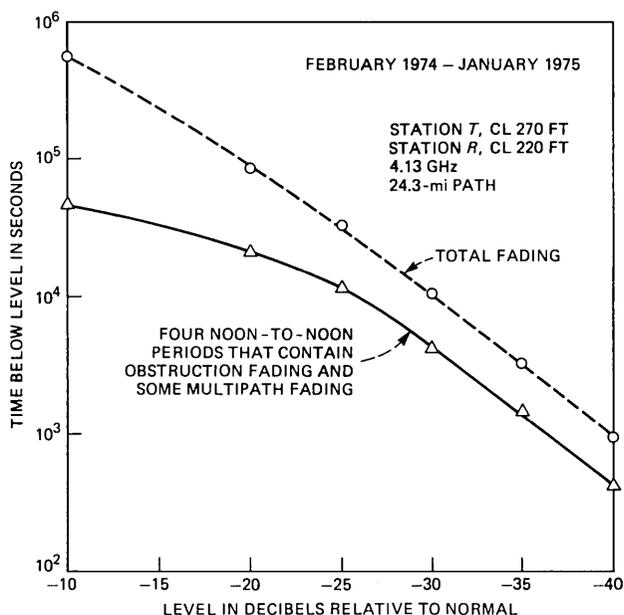


Fig. 3—Measured obstruction fading.

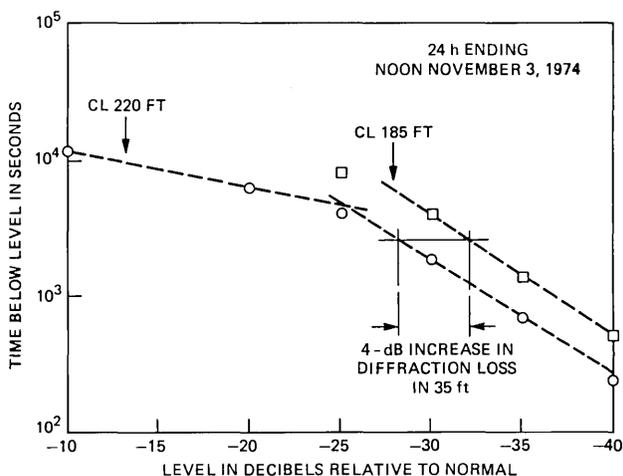


Fig. 4—Variation of obstruction fading with reception height at station *R*.

The measured time below -35 dB on the top antenna during nights with obstruction fading was 1425 s (solid curve in Fig. 3). The corresponding simultaneous time below -35 dB on the two upper antennas was 562 s. The space-diversity improvement is only 2.5 (1425/562), which demonstrates the incapacitation of diversity protection in the presence of obstruction fading.

The measured time below -35 dB on the top antenna caused by all fading was 3217 s (dashed curve in Fig. 3). Subtracting the obstruction-fading dominated time from this yields the amount of multipath fading (1792 s below -35 dB). The corresponding simultaneous time below -35 dB caused by multipath fading on the two upper antennas was 16 s. The space-diversity improvement of 112 (1792/16) is typical of multipath fading.

V. EMPIRICAL OBSTRUCTION LOSS EXPRESSION

We obtain a relationship between obstruction loss and blockage when measurements made at station *R* are related to diffraction theory. The measurements imply a linear relation between large obstruction loss in dB and normalized blockage, with a slope of 20 dB per unit of normalized blockage. This follows from the observation that (Fig. 4), for a given amount of loss larger than about 25 dB on the top antenna at station *R*, the loss on the antenna 35 feet below it is 4 dB larger. From the geometry of the path (specified in Section IV), the vertical offset corresponding to 35 ft at station *R* is 17.9 ft ($35 \times 12.4/24.3$) at the controlling obstruction 11.9 mi from station *R* (12.4 mi from station *T*). The radius of the first Fresnel zone at this point on the path is 87.4 ft. Therefore, the vertical offset is 0.20 (17.9/87.4) in

terms of a change of normalized blockage. This corresponds to a change of 20 dB per unit of normalized blockage (4/0.20).

Large values of the theoretical obstruction loss for extended obstructions, such as large spheres, vary approximately linearly with normalized blockage.¹⁴ The theoretical results suggest that straight-line approximations for large losses intersect zero loss at 0.5 units of normalized clearance. Given this, and the experimentally inferred slope of 20 dB, the loss expression applicable to deep obstruction fading becomes

$$M = -10 + 20(E/F_1), \quad M < -20, \quad (5)$$

where M is the fade level in dB relative to free space, and the value of the vertical distance E is negative to signify blockage. The term "free space" refers to a signal level that would be attained in the absence of atmospheric effects and without the proximity of the terrain.

The above empirical obstruction loss expression describes diffraction by "large" obstructions in the presence of atmospheric stratification. Paths with obstructions of "smaller" horizontal extent, such as a mountain ridge, may have obstruction losses smaller than those described by the empirical expression. This is suggested by a comparison with the theoretical limiting case of knife-edge diffraction (Fig. 5).^{13,14} Such "smaller" obstructions usually occur in mountainous terrain where obstruction fading is not a problem. Based on general scaling

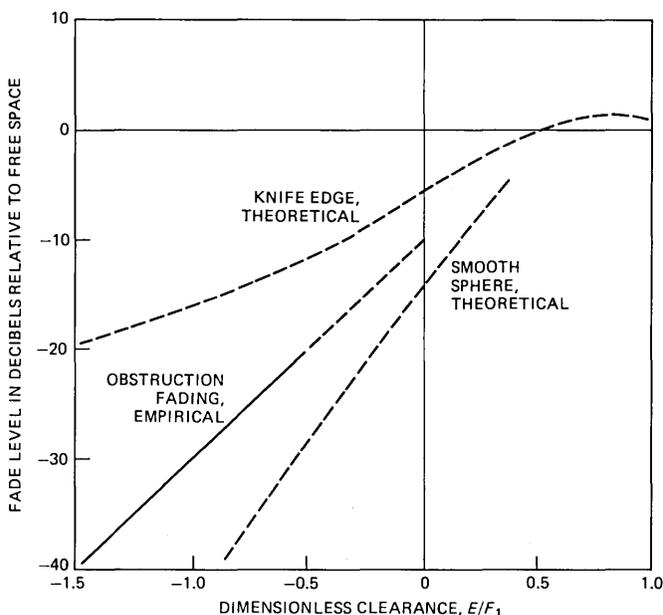


Fig. 5—Obstruction loss for deep obstruction fading on line-of-sight microwave radio paths.

experience in path engineering, the empirical loss expression can be applied at frequencies from 2 to 11 GHz on most practical paths between 20 and 30 mi in length.

VI. PROBABILITY OF OBSTRUCTION FADING

The practical problem in microwave radio path engineering is estimating the amount of time during which obstruction fading causes the received signal to be below a particular fade level. This time can be obtained from the probability of obstruction fading, which can be estimated from the probability distribution of refractivity gradients.

The steps in the procedure for making such an estimation are outlined in a simplified flow chart in Fig. 6. A fade level of interest corresponds to a normalized blockage E/F_1 as specified by the obstruction loss expression in eq. (5). The value of F_1 can be calculated when the radio frequency is specified and the location of the controlling obstruction is determined. This can be an iterative procedure, since the location of the controlling obstruction can change with antenna height and refractivity gradient. When F_1 has been calculated, the vertical distance E can be determined from a rearranged obstruction loss expression

$$E = (F_1/20)(M + 10). \quad (6)$$

Next, given the tentative heights of the transmitting and receiving antennas and the profile of the path, eq. (3) for the trajectory of the ray can be applied to determine a positive refractivity gradient (S) required to produce the value of E determined above.

The annual amount of time during which obstruction fading causes a received signal to be below a particular fade level can then be expressed, in seconds per year, as

$$T = T_0 P(N' > S), \quad (7)$$

where T_0 is the number of seconds in a year and $P(N' > S)$ is the annual probability that the refractivity gradient N' exceeds S .

The annual probability that the refractivity gradient exceeds a particular positive value S is composed of seasonal parts,⁷

$$P(N' > S) = (1/4) \sum_i P_i(N' > S), \quad (8)$$

where the summation extends over the four seasons of the year. Each seasonal probability consists of two components

$$P_i(N' > S) = 0.8P_{m,i} + 0.2P_{s,i}, \quad (9)$$

where the subscript m refers to the mixed (daytime) atmosphere, and the subscript s refers to a stratified (nighttime) atmosphere. Both

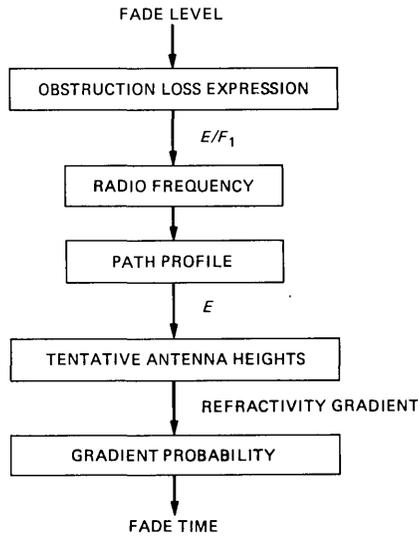


Fig. 6—Obstruction fading calculation.

components are Gaussian probability functions specified by appropriate values of their means and standard deviations.⁷ The means vary with geographic locations, but in a given season they are the same for the mixed and stratified regimes in the above mathematical formulation. The standard deviation for the mixed regime is 15 N -units/km for all locations and all seasons of the year in the United States.⁷ The amount of obstruction fading is governed largely by the standard deviations for the stratified regime, which exhibit considerable geographic and seasonal variation.⁷

VII. GRADIENT SCALE

Probability distributions of refractivity gradients usually describe the occurrence of increments of refractivity associated with an increment in height.^{7,15} A frequently used height increment is 100 m. Analyzing data for other height increments shows that the standard deviation of positive refractivity increments in the stratified regime is inversely proportional to the fourth root of the height increment.⁷ This arises because the absolute increase in refractivity that the atmosphere can support is limited. This behavior of the gradient statistics necessitates use of obstruction-fading measurements to determine the height increment (gradient scale) appropriate for microwave obstruction fading.

Simultaneous fading on the upper two receiving antennas at station R is the experimental measurement best suited for the determination of the gradient scale (centerlines of 220 and 185 ft above ground).

Simultaneous fading below the -35 dB fade level was measured on these antennas continuously throughout the experiment. The total simultaneous time below -35 dB was 562 s in the four 24-h periods when obstruction fading was present. This time represents “pure” obstruction fading below -35 dB at the 220-ft height, since the constraint of simultaneity removes multipath effects.

The calculated time below the -35 dB fade level can be expressed as a function of the height increment, or more specifically from eq. (9) in Ref. 7, as a function of the coefficient in the expression that relates the standard deviation of positive refractivity increments to meteorological parameters. This coefficient is inversely proportional to the fourth root of the height increment. The value of the coefficient is 440 when the height increment is 100 m.⁷ The corresponding standard deviations for the area where the path is located are, in N -units/km, 71 for winter, 56 for spring, 58 for summer, and 96 for fall.¹⁶ The coefficient that corresponds to the measured 562 s is 540 (Fig. 7). Therefore, the height interval (gradient scale) appropriate for calculation of obstruction fading becomes, as determined from Fig. 4 in Ref.

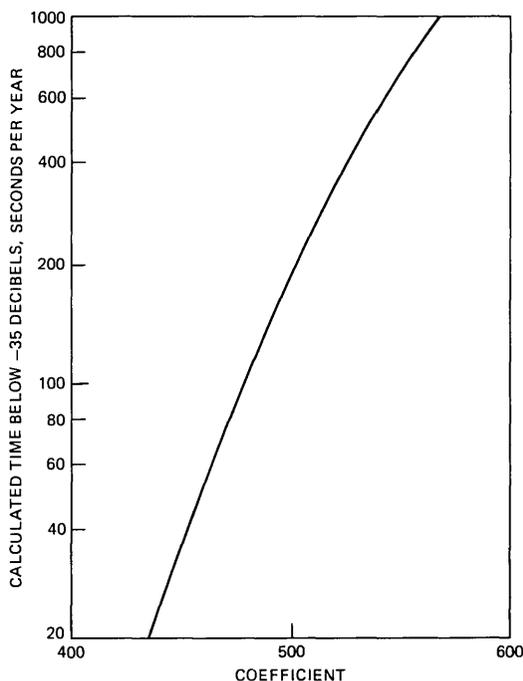


Fig. 7—Calculated obstruction fading as a function of the coefficient in the expression for the standard deviation of refractivity gradient.⁷ Measured obstruction fading of 562 s and meteorological parameters for the propagation path between stations T and R determine 540 as the value of the coefficient.¹⁶

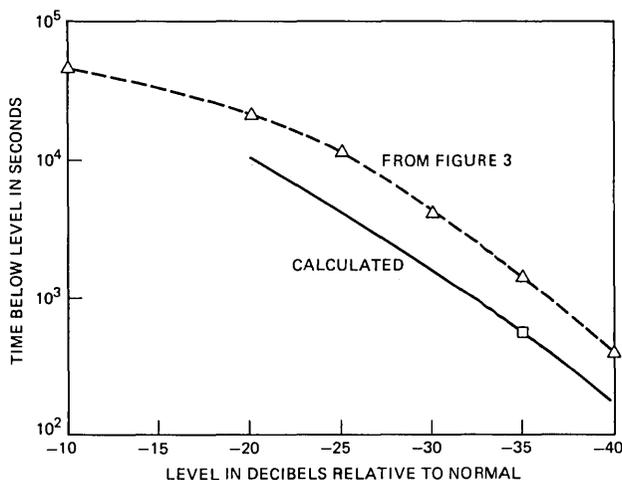


Fig. 8—Calculated obstruction fading at station *R*. Measured fading on nights when obstruction fading was dominant is shown by dashed curve (from Fig. 3), which contains some multipath fading.

7, $(1400/540)^4$ or 45 m. The corresponding standard deviations of the positive refractivity increments are, in *N*-units/km, 87 for winter, 69 for spring, 71 for summer, and 118 for fall.

The mean values of the refractivity gradients do not require scaling, since they describe the daytime reference condition of the atmosphere. The values of the means used in the above calculations are, in *N*-units/km, -46 for winter, -53 for spring, -58 for summer, and -53 for fall.¹⁶

VIII. CALCULATED OBSTRUCTION FADING

A comparison of calculated and measured obstruction fading at station *R* demonstrates the validity of the mathematical model. The slope of the calculated obstruction fading curve matches the measured results (Fig. 8). The slope is determined by the mathematical model, which has not been forced to conform in any way to slopes of measured fading curves. The calculated curve passes through the simultaneous fading point at -35 dB, which was used to determine the gradient scale and which describes "pure" obstruction fading. The dashed curve (from Fig. 3) describes measured obstruction fading accompanied by some multipath fading. Extension of the calculated curve to fade levels significantly shallower than -20 dB is not appropriate, since the obstruction loss curve incorporated in the mathematical model applies only to deep obstruction fades.

Additional calculated results for station *R* (Figs. 9 and 10) illustrate the variation of obstruction fading with antenna height and radio

frequency. The time below -35 dB varies by more than two orders of magnitude as the receiving height at station R changes 200 ft. When the radio frequency is varied from 2 to 11 GHz, the fade level at 10 s/yr (Fig. 10) varies over a range of almost 20 dB. This variation is a consequence of the frequency dependence of the obstruction loss in eq. (5). The obstruction loss increases with frequency because, as illustrated in Fig. 11, the blockage of Fresnel zones increases with frequency.

IX. DESIGN SPECIFICATIONS

Tower-height design specifications can be derived from Bell System's annual all-cause two-way transmission unavailability design limit of 0.02 percent, which applies to a 4000-mi long-haul circuit or a 250-mi short-haul circuit.^{1,2,6} For route engineering purposes, half of this is allocated to multipath fading in the 2-GHz, 4-GHz, and 6-GHz bands where clear-air propagation phenomena are the dominant causes of propagation impairments.^{1,2} A new allocation to obstruction fading is 0.005 percent, which is both two-way and one-way, assuming that both directions of transmission use radio channels on the same terrestrial radio route. The other 0.005 percent is allocated to equipment impair-

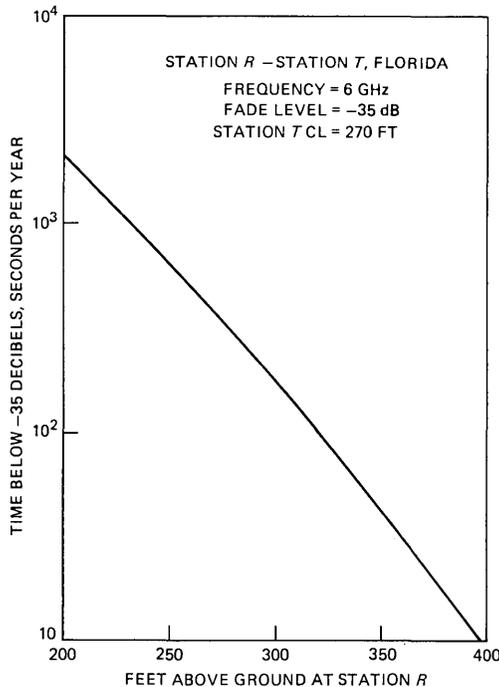


Fig. 9—Variation of obstruction fading with antenna height.

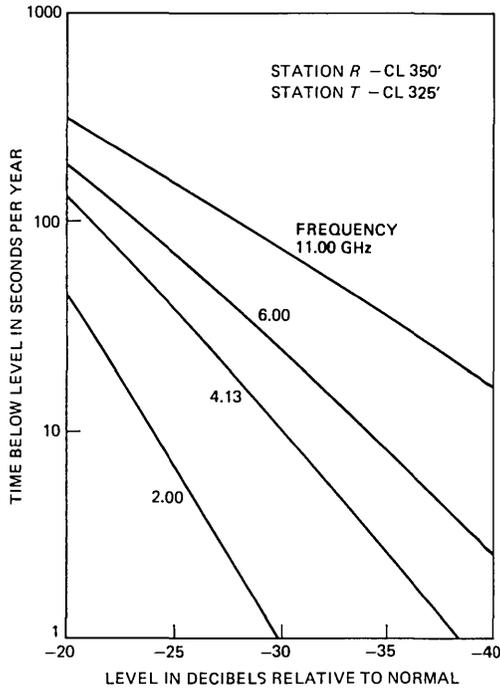


Fig. 10—Calculated obstruction fading for station R - station T.

ments and other causes. These design allocations emphasize that, in modern solid state radio systems, transmission performance is governed by propagation impairments. In the case of short-haul radio, tradeoffs between allocations are frequently invoked, particularly so in the 11-GHz band where transmission unavailability caused by rain can be dominant.

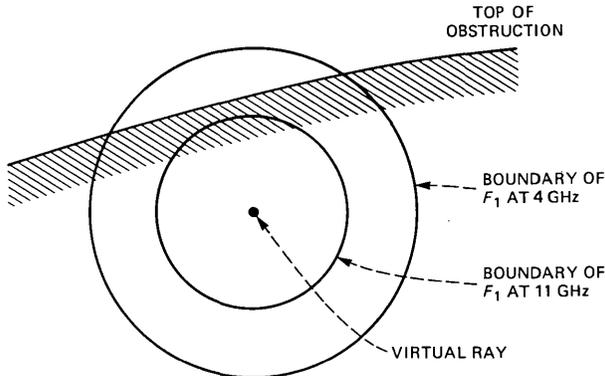


Fig. 11—Illustration of blockage.

The allocations are prorated to distance. For long-haul service, this implies that antenna centerline heights (tower heights) on a 25-mi path have to be such that the annual time below a specified fade level, caused by obstruction fading, is 10 s or less. A fade level of -35 dB is recommended as the obstruction fading outage threshold in tower-height design procedures. The choice of -35 dB as a design value is related to general characteristics of radio systems, interference considerations, and possible joint occurrence of obstruction fading and multipath fading. An advantage of an equipment-independent fade-level specification is design stability, since radio towers are "permanent" parts of the plant that can be expected to be used for a number of future varieties of radio equipment and service. Combined use of the allocation and the -35 dB value is necessary to achieve a balanced design where terrain clearance is adequate and where impairment of diversity protection is minimal when obstruction fading and multipath fading occur jointly.

When used as design criteria, the allocations are prorated to the length of a frequency-diversity switching section or to a route segment between terminals or drop and add points. For example, the allocation to a switching section containing three 25-mi paths (hops) is 30 s. The hops are permitted to contribute unequally to the 30-s total. A design where two hops contribute 5 s each and the third hop contributes 20 s is satisfactory. Design allocations such as these are related to measured performance in an average sense, since fading can vary from year to year, and because interactions of fading, actual fade margins, and protection system operation can be complicated.

Tower heights on radio hops in short-haul service frequently have been determined by clearance rules that are the same as those for hops in long-haul service. This permits flexibility in the routing of communications traffic with differing reliability requirements. Therefore, tower-height requirements for short-haul routes can be determined from the obstruction-fading allocation of 10 s/yr prorated to a 25-mi hop. An option for the determination of short-haul tower heights is an obstruction fading allocation of 160 s/yr prorated to a 25-mi hop. This retains the standard one-to-sixteen ratio of per-mile performance between long-haul and short-haul service. The 160-s option can be used only when reliability requirements are expected to remain constant over the life of the towers. When used as design criteria, these allocations are prorated to the length of a switching section or route segment.

In the case of paths equipped for space-diversity reception, the obstruction fading allocation for the lower antenna prorated to a 25-mi hop is 50 s/yr for long-haul service and for those hops in short-haul service where the height of the upper antenna is determined by the

10-s allocation. We apply the 50-s allocation on a per-hop basis to avoid impairment of space-diversity protection. A separate allocation for lower antennas is not made when we use the 160-s option to design short-haul paths, since this option represents a design limit.

The design criteria outlined above indicate that tower heights larger than previously required can be beneficial in regions where obstruction fading is severe. For example, a previous clearance rule has been grazing (zero) clearance at $K = 1/2$ in difficult propagation areas,² which requires approximately 180-ft towers on a 20-mi flat path covered by 50-ft trees. The new criteria indicate that towers on such a path may need to be as high as 250 ft for long-haul service in an area with a high occurrence probability of obstruction fading.

X. CONCLUSION

A new tower-height design method outlined in this paper permits more efficient use of construction funds, since tower heights are quantitatively related to transmission performance. High towers can be constructed where needed, whereas towers of minimal height, determined by daytime clearance requirements, can be used in the cooler regions of the United States.

XI. ACKNOWLEDGMENTS

The work on obstruction fading benefited considerably from interactions with W. T. Barnett. We had many helpful discussions with J. A. Schiavone about the meteorological aspects of radio propagation. Portable Propagation Recorders (PPRS) designed by G. A. Zimmerman made acquisition of the propagation data possible. A number of colleagues at Bell Laboratories, AT&T Long Lines, and Southern Bell provided assistance needed to carry out the propagation measurements.

REFERENCES

1. A. Vigants and M. V. Pursley, "Transmission Unavailability of Frequency-Diversity Protected Microwave FM Radio Systems Caused by Multipath Fading," B.S.T.J., 58, No. 8 (October 1979), pp. 1779-96.
2. A. Vigants, "Space-Diversity Engineering," B.S.T.J., 54, No. 1 (January 1975), pp. 103-42.
3. W. T. Barnett, "Microwave Line-of-Sight Propagation With and Without Frequency Diversity," B.S.T.J., 49, No. 8 (October 1970), pp. 1827-71.
4. W. T. Barnett, "Multipath Propagation at 4, 6, and 11 GHz," B.S.T.J. 51, No. 2 (February 1972), pp. 321-61.
5. S. H. Lin, "Statistical Behavior of Deep Fades of Diversity Signals," IEEE Trans. Commun., COM-20, No. 6 (December 1972), pp. 1100-7.
6. W. Y. S. Chen, "Estimated Outage in Long-Haul Radio Relay Systems with Protection Switching," B.S.T.J., 50, No. 4 (April 1971), pp. 1455-85.
7. J. A. Schiavone, "Prediction of Positive Refractivity Gradients for Line-of-Sight Microwave Radio Paths," B.S.T.J., this issue.
8. B. R. Bean and E. J. Dutton, *Radio Meteorology*, National Bureau of Standards, U. S. Government Printing Office, 1966.

9. C. L. Pekeris, "Accuracy of the Earth-Flattening Approximation in the Theory of Microwave Propagation," *Phys. Rev.* 70, Nos. 7 and 8 (October 1 and 15, 1946), pp. 518-22.
10. D. E. Kerr, *Propagation of Short Radio Waves*, New York: McGraw-Hill, 1951.
11. D. C. Livingston, *The Physics of Microwave Propagation*, Englewood Cliffs, New Jersey: Prentice-Hall, 1970.
12. R. F. White, *Engineering Considerations for Microwave Communications Systems*, San Carlos, California: Lenkurt Electric Co., 1970.
13. M. P. M. Hall, "Effects of the Troposphere on Radio Communications," London: Peter Peregrinus Ltd. on behalf of the Institution of Electrical Engineers, 1979.
14. K. Bullington, "Radio Propagation Fundamentals," *B.S.T.J.*, 36 (May 1957), pp. 593-626.
15. C. A. Samson, "Refractivity Gradients in the Northern Hemisphere," OT Report 75-59, U. S. Department of Commerce, Office of Telecommunications, U. S. Government Printing Office, April 1975.
16. J. A. Schiavone, private communication, 1979.

Prediction of Positive Refractivity Gradients for Line-of-Sight Microwave Radio Paths

By J. A. SCHIAVONE

(Manuscript received January 15, 1981)

Estimating obstruction fading on line-of-sight microwave radio paths requires statistics on positive refractivity gradients in the lowest 100 meters of the atmosphere. In this paper, we describe a semiempirical climatological model that predicts quantitatively the occurrence frequency of positive refractivity gradients and the geographical variation of this frequency within the contiguous United States. The occurrence frequency is parameterized by six factors, for which numerical values are obtained from available climatological and physiographical data. By modeling the two diurnal atmospheric boundary layer regimes with separate probability density functions, we produce cumulative probability distributions for refractivity gradients. The model is normalized using measured refractivity gradient distributions available for 17 sites within the United States. The model can be used for line-of-sight microwave radio path engineering; a companion paper presents a method for determining antenna heights.

I. INTRODUCTION

Line-of-sight microwave radio transmission performance is susceptible to atmospheric conditions, in particular to spatial variations in the microwave refractive index. Nearly always during the day and frequently at night, sufficient vertical mixing occurs of the low-level air layers, through which the microwave beam propagates, so that vertical refractive index gradients encountered are small. However, horizontal layers containing strong vertical gradients in air temperature and water vapor pressure, which determine the refractive index, sometimes develop at night. The presence of these layers can cause the received signal to fade (decrease in strength).

This paper addresses the meteorology of obstruction fading, which

is caused by a large positive gradient in the refractive index, which is, in turn, caused by a large positive water vapor pressure gradient. The ray path between the transmitting and the receiving antennas is lower than normal in the presence of a positive refractive index gradient, and the terrain can block the ray path when the positive gradient is sufficiently large. Although infrequent, the consequences of obstruction fading are severe: transmitted signal is steadily blocked for as long as an hour at a time. Obstruction fading can be avoided by locating microwave antennas sufficiently high above the ground.

Prolonged fades can also be caused by strong negative gradients that refract the transmitted microwave beam downward to such an extent that it ceases to illuminate the receiving antenna. Climatic conditions and path geometries in the United States are such that this is usually a secondary phenomenon compared to obstruction fading.

The presence of negative gradients of more modest proportions is associated with multiple propagation paths through the atmosphere, which cause phase interference fading (multipath fading). Phase interference fading is a major factor in microwave system design. The meteorology of multipath fading is a separate topic, and is not addressed in this paper.

Probability distributions for the occurrence of refractive index gradients have been compiled by Samson,¹ Bean et al.,² and Dougherty³ from measurements of vertical profiles of meteorological parameters. The distributions are intended for use in engineering microwave paths. However, these sets of compilations are for isolated locations, and they have not been generalized to account for the varied climates to which different microwave routes are exposed.

The purpose of the present work is to devise a climatological model to predict quantitatively the geographical variation of the occurrence frequency of positive refractive index gradients within the contiguous United States. The results can be used to estimate obstruction fading occurrence on microwave radio paths to help determine optimum microwave antenna heights. The result of the model is a predicted cumulative probability distribution for positive refractive index gradients calculated for an arbitrary location within the contiguous United States for arbitrarily large gradients. The present model represents the first attempt to predict refractivity gradient probability distributions from climatological data. Ikegami et al.⁴ and Akiyama and Sasaki⁵ have performed fits to measured refractive index gradient distributions but without meteorological parameterization.

This model is designed to provide predictions for the 48 contiguous United States. Two submodels must be used to account for the varied climates experienced by this large geographical entity: a coastal region submodel and an interior region submodel. The interior region sub-

model is based on a relatively small experimental data base and the reliability of the overall model for interior regions is possibly lower than for coastal regions. However, since obstruction fading occurs mostly in coastal regions, the model is most reliable where it is most useful.

The predictions of this model are compared to measured distributions¹ and, in general, show good agreement. We attribute the imperfect agreement that exists in a few cases to local effects that cannot be accounted for without introducing additional complexity with the attendant risk of making the model unwieldy to implement. That is, we use a geographical resolution for the model, which has been judged optimum by weighing resolution improvements against simplicity of model use.

II. MODEL FORMULATION

The model is formulated in two stages. First, we choose parameters (Section 2.1) based on meteorological principles, to relate the occurrence frequency and magnitude of refractive index gradients to available climatological and physiographical data. Second, we develop a method for producing a refractive index gradient probability distribution from these parameters (Section 2.2). The model is quantified by comparison with measured distributions.¹ We make final minor adjustments to the model to provide consistency with information on the geographical distribution of obstruction fading gleaned from transmission performance experience in the Bell System.⁶

2.1 Meteorology

The refractive index, n , for microwaves can be expressed as refractivity,⁷

$$N = 10^6(n - 1) = \frac{77.6}{T} \left(P + 4810 \frac{e}{T} \right), \quad (1)$$

where N is the refractivity in N -units, T is the air temperature in kelvin, P is the air pressure in millibars, and e is the water vapor pressure in millibars. Water vapor pressure is the governing contributor to producing positive refractivity gradients (N') at night. Therefore, a positive refractivity gradient model must predict climatologically the frequency of occurrence of positive e gradients.

During obstruction fading, the important part of the microwave beam generally lies within the lowest 100 m of the atmosphere and available radio meteorological data have been obtained, therefore, for 100 m. However, since probability distributions are functions of height increments, we present a method (Section 2.2) for using a height increment appropriate to the particular application.

Several mechanisms can produce inverted (positive gradient) water vapor vertical profiles:

(i) Condensation of water vapor at the bottom of a surface-based nocturnal temperature inversion, e.g., dew or ground fog, causes absolute humidity to decrease there.⁸ However, the amount of water vapor condensed is insufficient to produce large gradients across a layer as thick as 100 m.

(ii) Strong convection processes in relatively dry air can produce water vapor inversions within the daytime boundary layer.⁹ Development of a nocturnal temperature inversion could cause the water vapor inversion to persist throughout some of the night. The magnitude of these gradients is insufficient to cause obstruction fading, however.

(iii) Uplands generally receive more rainfall than surrounding lowlands, provided the altitude difference is large enough. When situated within relatively dry lowland regions, daytime evaporation from the uplands can produce elevated moist air layers. Such an effect has been observed.¹⁰ Nocturnal radiative temperature inversions could cause the water vapor inversion to persist during the night. Such a mechanism is greatly limited geographically, however.

(iv) Nocturnal advection of moist air can produce inverted water vapor profiles. Nocturnal land breezes at water body-land interfaces, are a possible mechanism. However, the effects of land breezes usually do not penetrate more than a few tens of kilometers inland and, thus, cannot explain severe obstruction fading observed farther inland.

(v) We propose large-scale advection of moist air as the mechanism for producing most of the obstruction fading on line-of-sight microwave radio routes. This mechanism is supported by various characteristics of the limited information on observed obstruction fading in the Bell System:⁶ (1) Often large areas are affected by a single episode. (2) The length of microwave paths requires vertical gradients to be horizontally homogeneous over approximately 40-km distances to produce fading. (3) Relatively flat areas are especially susceptible; flat areas are conducive to large-scale, low-level advection at night. (4) Regions within 300 km of coastlines are especially susceptible; 300 km is the distance air could travel at typical speeds at approximately 100-m altitude during 12 night hours. (5) Finally, a study of a severe fading episode in the Bell System showed that the onset of obstruction fading occurred later for greater distances from the coast.

An understanding of the details of the proposed large-scale moisture advection mechanism requires knowledge of atmospheric motion on various distance scales. We will now discuss atmospheric motion within the boundary layer, across the continent, and within air masses.

Obstruction fading usually occurs during fine weather conditions, when anticyclonic air masses envelop the affected region. Thus, the

characteristics of the boundary layer and their diurnal variation only in fine weather conditions need be considered.

During the day, solar radiation causes convective mixing which produces a relatively deep (≥ 1 km) boundary layer which is vertically well mixed throughout. Thus, vertical gradients of water vapor and horizontal momentum are relatively small across the lowest 200 m where the microwave beam lies. Small vertical gradients of horizontal momentum produces small wind speed and direction gradients (Fig. 1a).

At night, radiative cooling of the earth's surface produces a relatively thin (as thin as ~ 100 m) density-stable boundary layer. The high stability suppresses turbulence and very little vertical mixing occurs; water vapor inversions would be maintained once produced. Relatively large wind speed and direction gradients could be maintained across the boundary layer¹¹ (Fig. 1b) because of the limited vertical mixing of horizontal momentum. Calm winds often occur at the surface.

Large-scale air motion is determined by the midtropospheric (~ 6 km) winds. These currents steer air masses across the North American continent, in general, from west to east. Air masses that affect the

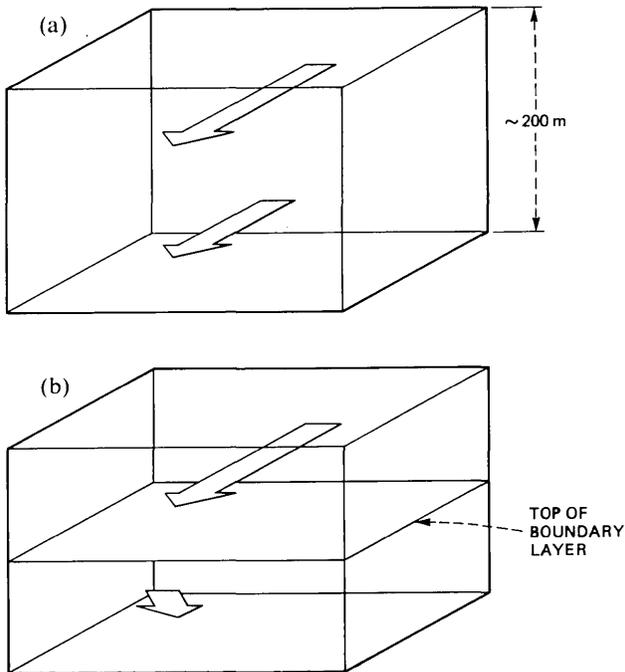


Fig. 1—Air flow in lowest 200 m. Arrows indicate relative wind speed and directions at two heights, illustrating the enhancement of differential flow at night which contributes to moisture inversions in coastal regions.

United States usually originate over northwest Canada or the eastern Pacific Ocean.¹² Pacific air masses crossing the continental divide lose much of their moisture through lifting-induced precipitation over the western mountains. Therefore, except for the west coast, air masses are relatively dry until they reach the Great Lakes, Atlantic Ocean, or Gulf of Mexico. Once they reach these large water bodies, evaporation increases their moisture content.

Air flow at moderately low levels (≤ 2 km) within anticyclonic air masses is essentially circular (for the purpose of this discussion) with respect to the air mass core, and in a clockwise direction. Thus, moderately low level winds at a given location change direction as an air mass traverses the region. Furthermore, winds near the center of the air mass are very light because of small horizontal pressure gradients; therefore, highly stratified air layers can readily form there.

The mechanism for producing water vapor inversions by the large-scale moisture advection that we are proposing is as follows: dry air masses approach the east (south) coast of the United States from the west (northwest). Before the air masses reach the coasts, moderately low level (≤ 2 km) wind flow is mainly offshore in coastal regions, thus, maintaining dry air there. As the air mass core comes sufficiently close to the coast, winds to the south and west of the air mass core shift to an onshore direction, advecting moist air inland. If the offshore-onshore wind shift occurs near sunset, the characteristics of the nighttime boundary layer discussed earlier, wind speed and direction differences across it, cause moist air to be advected over the top of surface-based dry air. The differential wind speed advects moist air inland faster at approximately 200 m than at the surface. The differential wind direction, which is normally a clockwise shift with height, also contributes to maintaining faster inland penetration of moist air at approximately 200 m relative to the surface.

Production of positive e gradients at a particular site requires the site to be sufficiently close to a moisture source so that moist air can reach the site during the nocturnal stratified air regime. Measured refractivity gradient distributions¹⁻³ show that elevated moist layers occur most frequently near large bodies of water such as oceans, gulfs, and large lakes. However, measurements show also that elevated moist layers occur at sites too far from these large water bodies to have had them for their source of moisture. Thus, mechanisms other than large-scale advection probably also contribute to positive refractivity gradients. Each mechanism cannot be modeled specifically, as for large-scale moisture advection. Therefore, a large-scale moisture-advection model is used near large water bodies (coastal regions), while a mechanism-independent model is used for other areas (interior regions).

2.1.1 Coastal region

The coastal region model is parameterized by six factors relating the occurrence of positive e gradients to parameters obtainable from seasonal climatological and physiographical maps. The factors are frequency of temperature inversion occurrence, f_i ; air homogeneity, H ; water body proximity, D ; frequency of wind direction occurrence, f_w ; air moisture capacity, V ; and surface moisture availability, M . The proportionality among the six factors and their individual normalization is obtained empirically from measured N' probability distributions (Section 2.2).¹ The factors f_i and H specify the occurrence of stratified air, D and f_w specify the production of positive e gradients, and V and M specify the magnitude of the positive e gradients.

The occurrence of stratified air in the lowest 100 m coincides with the existence of temperature inversions (temperature increase with height). Hosler has compiled maps of the occurrence frequency of surface-based temperature inversions for four seasons within the United States.¹³ The frequency of temperature inversion occurrence, f_i , whose units represent the fraction of total time that inversions occur for a given location and season, is employed as a factor in the coastal model.

For obstruction fading to occur, large positive e gradients must occur over the entire microwave path (a standard path length of 40 km is assumed in this work). Thus, the horizontal homogeneity of the stratified air, which is determined by the flatness of the terrain profile, must be large for the occurrence of obstruction fading. Therefore, a factor, H , is introduced that represents the horizontal homogeneity of the stratified air. Values for this factor are obtained from a map in the United States National Atlas, "Classes of Land Surfaces Forms."¹⁴ In this map, each location is represented by three characteristics: slope, local relief, and profile type. In the present model, each characteristic is defined as a subfactor of H ,

$$H = H_1 H_2 H_3. \quad (2)$$

Numerical values of H_1 , H_2 , and H_3 are acquired, after all other empirically determined parameters are obtained, by comparison of the model with the geographical distribution of microwave transmission performance in the Bell System.⁶ Transmission performance experience is employed since it represents the only existing *path-averaged* obstruction fading information. Table I gives the values obtained for H_1 , H_2 , and H_3 .

It is useful to compare refractivity gradient distributions which are predicted by the present model and represent path averages, with measured distributions compiled at single points. In such cases path

Table I—Air homogeneity subfactors.
Classes are obtained from Ref. 14.

| Slope Class | | Relief Class | | Profile Class | |
|-------------|-------|--------------|-------|---------------|-------|
| h_1 | H_1 | h_2 | H_2 | h_3 | H_3 |
| A | 1.0 | 1 | 1.0 | a | 0.7 |
| B | 1.0 | 2 | 0.9 | b | 0.8 |
| C | 0.8 | 3 | 0.5 | c | 0.8 |
| D | 0.4 | 4 | 0.5 | d | 1.0 |
| | | 5 | 0.5 | None | 1.0 |
| | | 6 | 0.5 | | |

averaging is removed from the predictions. This is accomplished by setting H to its maximum value, 1.0, in those instances.

For the coastal region, production of positive e gradients requires proximity to a large water body. Moist air originating at the coastline at sunset can penetrate a maximum distance inland during the nocturnal stratified air regime. Therefore, we propose the following smoothed-step distance function to represent this effect:

$$D = [1 + (d/d_0)^4]^{-1}, \quad (3)$$

where d is the minimum distance to an ocean, gulf, or one of the Great Lakes and $d_0 = 300$ km, as determined by the maximum distance inland that the air can penetrate during 12 hours when moving at an average speed at a height of approximately 100 m.

For every location there is a wind direction that maximizes the relative advection of moist air over dry air. The appropriate direction varies with distance to and shape of the coastline nearest the site. The procedure for specifying this direction for different sites is discussed in Appendix A. We propose that the production of positive e gradients is proportional to the frequency, f_w , that wind blows along this direction. Values for f_w , in units of fraction of total time, are obtained from frequency distributions of surface wind direction.

The magnitude of the e gradients formed depends upon the water vapor capacity of the air. The water vapor capacity, in turn, depends upon air temperature, T . The saturation vapor pressure function, $e_s(T)$, is a logical choice to represent the dependence of water vapor capacity on air temperature. It is proposed that a water vapor capacity factor be represented by

$$V = \frac{e_s(T)}{e_{s0}}, \quad (4a)$$

where $e_s(T)$ is in millibars and $e_{s0} = 50$ mb, an approximate upper limit for naturally occurring vapor pressures. The function $e_s(T)$ can be represented by⁸

$$e_s(T) = \exp[21.71 - 5433/T], \quad (4b)$$

where T is in kelvin. Temperatures used are those obtained from monthly average temperature maps of the United States.¹⁶

The magnitude of the e gradients produced by advection from large water bodies is reduced if substantial smaller local sources of water exist at the site. This is caused by surface water evaporating into the dry lowest air layer. This effect is modeled with a surface moisture factor

$$M_C = \frac{m_1}{m_0 + m}, \quad (5)$$

where m is a surface moisture index obtained from a map of moisture regions,¹⁷ $m_0 = 70$, and $m_1 = 20$, obtained by comparison with measured N' distributions.¹ The subscript C denotes the coastal region since a different moisture factor is employed for the interior region.

In summary, the six factors used to determine the occurrence of positive e gradients in the coastal region are f_i , H , D , f_w , V , and M_C .

2.1.2 Interior region

The model for interior sites is a modification of that developed for the coastal sites. The two factors representing the production of stratified air layers, H and f_i , are the same as for the coastal model. However, the two factors representing the production of positive e gradients, D and f_w , are combined into a single constant factor, A' . This is required because it is not feasible to take into account the variety of mechanisms thought to contribute to gradient production in interior regions. Determination of A' is discussed below. Finally, of the two factors representing the magnitude of the e gradient produced, V and M , only M need be modified. This modification is required since the surface moisture index, m , provides the necessary information regarding the availability of moisture from local sources. Thus, the proportionality between M and m is inverted with respect to the coastal site model. The interior region model for M is

$$M'_I = \frac{m + m_2}{m_3}. \quad (6a)$$

Since the relative advection factor, A' , is constant, m_3 can be absorbed into it yielding $a = A'/m_3$. We determine empirically that $a = 4.7 \times 10^{-5}$ and $m_2 = 100$ (Appendix B). Equation (6a) can be rewritten as

$$M_I = m + m_2. \quad (6b)$$

The five factors used to determine the occurrence of positive e gradients in the interior region are f_i , H , a , V , and M_I .

The interior model, instead of the coastal model, is applied to the Pacific Coast of the United States. The large-scale moisture-advection

mechanism upon which the coastal model is based is not valid for the Pacific Coast since air masses which cross the coast already contain much moisture. Furthermore, relatively high coastal mountain ranges prevent low-level nocturnal differential advection.

2.2 Refractivity gradient probability distributions

Examination of measured N' cumulative probability distributions¹⁻³ indicates that the shape of each curve is characterized by three features when plotted on normal probability axes (Fig. 2): a central linear segment and positive and negative tails. Ikegami et al.⁴ and Akiyama and Sasaki⁵ have noted these features and have attempted to fit each with different mathematical formulas, but no attempt was made to interpret them meteorologically. We hypothesize that the central segment is produced by an atmospheric regime characterized by at least some vertical air mixing, and the tails are produced by a regime characterized by stratified air. In general, the tails are distributed asymmetrically about the mean. This is expected, since different meteorological mechanisms produce positive and negative gradients in stratified air. Therefore, a mixture of at least three distributions is required to represent the total cumulative probability distribution completely. In view of the complexity of handling a mixture of three distributions relative to two distributions, and since only positive N' gradients are of concern for obstruction fading, we do not attempt to model that half of the cumulative probability distribution where N' is more negative than the mean. Therefore, we employ a mixture of two distributions.

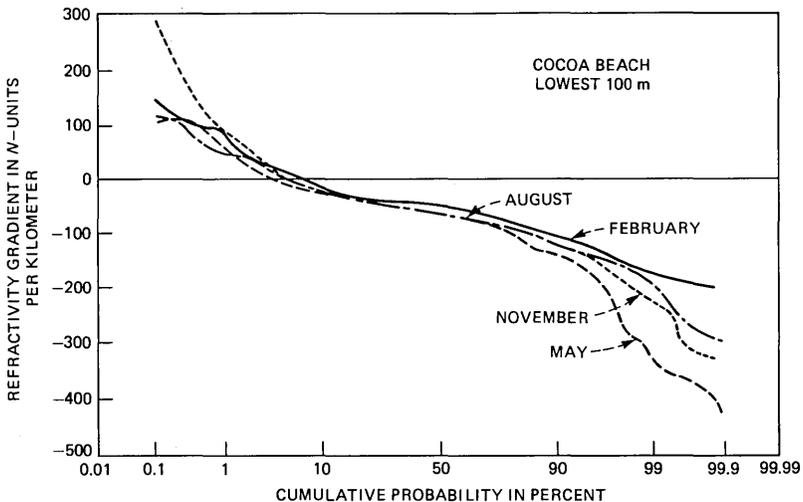


Fig. 2—One of 17 measured refractivity gradient distributions compiled by Samson.¹

In general, the mixed regime distribution comprises approximately 80 percent of the total distribution (see discussion of p^* below). Furthermore, it is well documented that mean refractivity gradients in mixed air are similar to those expected from normal daytime vertical distributions of water vapor, temperature, and air pressure.^{7,18} Therefore, since gradients developed during the stratified regime evolve from those existing during the mixed regime, and since the negative-going half of the total distribution is not modeled (relieving the need to account for asymmetry), it is reasonable to assume that each of the two component distributions can be represented by normal distributions that share the same mean, μ . Figure 3 illustrates this assumption. These component distributions are given by the probability density functions

$$p_m(N') = (2\pi \sigma_m^2)^{-1/2} \exp \left[-\frac{1}{2} \left(\frac{N' - \mu}{\sigma_m} \right)^2 \right] \quad (N' \geq \mu), \quad (7a)$$

$$p_s(N') = (2\pi \sigma_s^2)^{-1/2} \exp \left[-\frac{1}{2} \left(\frac{N' - \mu}{\sigma_s} \right)^2 \right] \quad (N' \geq \mu), \quad (7b)$$

where m and s denote the mixed and stratified regimes, respectively. The total probability density function is given by

$$p(N') = (1 - p^*)p_m + p^*p_s, \quad (8)$$

where p^* is the proportion of mixture for the two component distributions.

We empirically compared a mixture of two distributions with Samson's U.S. data and determined that $p^* = 0.2$ and $\sigma_m = 15$ N -units/km and that both are essentially invariant with respect to location and season. Bean et al. have compiled seasonal maps of the average refractivity gradient in the lowest 1-km depth.² This data is used to specify μ as a function of location.

The standard deviation for the stratified regime, σ_s , is parameterized separately for the coastal and interior regions by

$$\sigma_{sC} = \sigma_0 (f_i H)^{1/2} (Df_w VM_C)^{1/4} \quad (9a)$$

and

$$\sigma_{sI} = \sigma_0 (f_i H)^{1/2} (a VM_I)^{1/4}, \quad (9b)$$

where σ_0 is a constant for a given layer thickness and path length, and the exponents, $1/2$ and $1/4$, are determined empirically. Both σ_{sC} and σ_{sI} are calculated for each site and season, and the larger of the two is employed to calculate p_s for each season. Total probability, $p(N')$, is calculated for each season and the seasonal probabilities are averaged to produce an annual probability density function for each site. Finally,

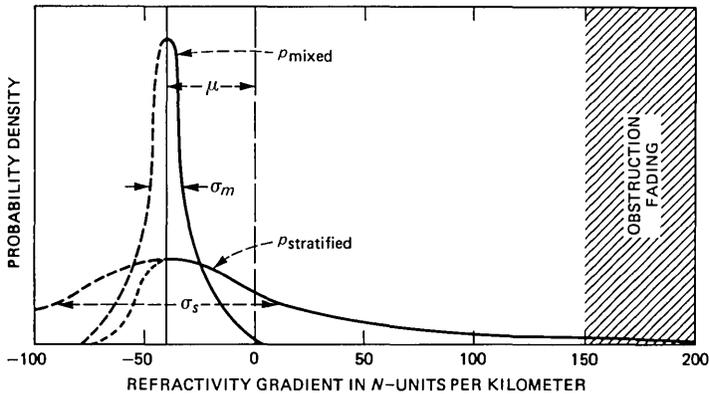


Fig. 3—Illustration of two probability density functions employed to represent a refractivity gradient distribution. Dashed portions of the functions denote halves which are not modeled. The dotted line denotes a possible “true” negative-going half of the stratified distribution which is, in general, asymmetric about the mean.

the probability density functions are integrated from $+\infty$ to μ to produce N' cumulative probability distributions for positive N' .

Samson's 17 measured N' distributions¹ were used to determine empirically that σ_0 for a 100-m-deep layer is 440 N -units/km. N' distributions have been compiled at several sites for surface layer thicknesses (Δz) different from 100 m. Samson compiled annual distributions at Cardington, England for 75-, 150-, and 500-m-thick layers.¹ Dougherty compiled annual distributions at Cocoa Beach, Florida for 50- and 100-m-thick layers.³ Values for the climate parameters appropriate to each site were used in an empirical determination of σ_0 for the five available layer thicknesses. $H = 1$ was used; the reason for this choice has been discussed in Section 2.1.1.

The results are plotted against $(\Delta z)^{-1/4}$ in Fig. 4 and indicate such a relationship. The line in Fig. 4 represents $\sigma_0(\Delta z) = 1400/\Delta z^{1/4}$, where Δz is in meters and σ_0 is in N -units/km. Thus, σ_0 for an arbitrary layer thickness, can be obtained from this plot.

III. RESULTS AND DISCUSSION

Complete specification of a predicted N' probability distribution is made by the parameters p^* , μ , σ_m , and σ_s . The parameters p^* and σ_m are constants; however, μ and σ_s must be specified for each season as a function of location. Bean et al. provide seasonal maps for μ measurements,² and Table II summarizes specification of σ_s .

It is useful to compare a few predicted distributions with measured ones to check the performance of the model. Figures 5 through 7 show examples of N' cumulative probability distributions predicted by the model for 100-m layers (thicknesses for which measured distributions

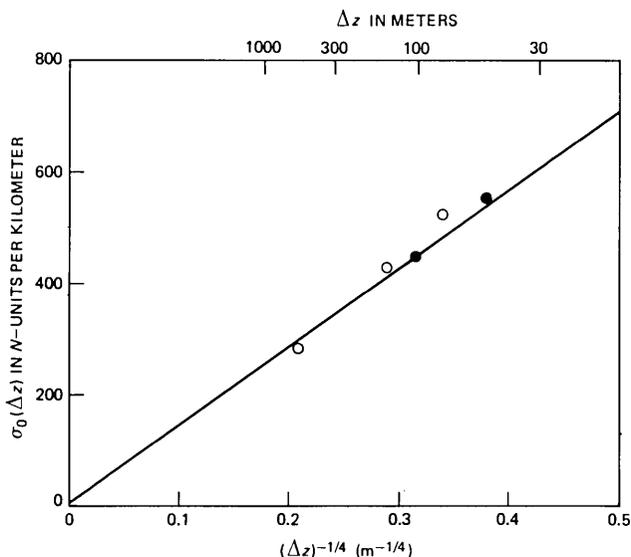


Fig. 4—Dependence of σ_0 upon surface layer thickness, Δz . Data points are derived from Samson's compilations¹ for Cocoa Beach, Florida (●), and Cardington, England (O).

are available); measured curves are also shown. Predicted curves were calculated using $H = 1$ (see Section 2.1.1). The ability of the model to successfully predict large differences in distribution shapes for two sites separated by only 500 km is illustrated by comparing the Cocoa Beach and Key West curves. The model's ability to respond to seasonal variations in both coastal and interior regions is illustrated for Cocoa Beach and Denver.

The agreement between predictions and measurements is generally good; however, it is not perfect. Each measured distribution, with a few exceptions, was compiled from approximately 900 measurements (Table III). Thus, the data point density on the distribution tails is

Table II—Summary of method for evaluating σ_s parameters.

| Parameter for σ_s Calculation | Equation used for Calculation | Equation Constants | Climatological Variables | Sources of Climatological Variables |
|--------------------------------------|-------------------------------|--------------------------|--------------------------|-------------------------------------|
| f_i | — | — | f_i | Ref. 13 |
| H | (2) | Table I | h_1, h_2, h_3 | Ref. 14 |
| D | (3) | $d_o = 300$ km | d | U.S. map |
| f_w | — | — | f_w | Ref. 15 |
| V | (4) | $e_{so} = 50$ mb | T | Ref. 16 |
| M_C | (5) | $m_o = 70, m_1 = 20$ | m | Ref. 17 |
| a | — | $a = 4.7 \times 10^{-5}$ | — | — |
| M_I | (6b) | $m_2 = 100$ | m | Ref. 17 |

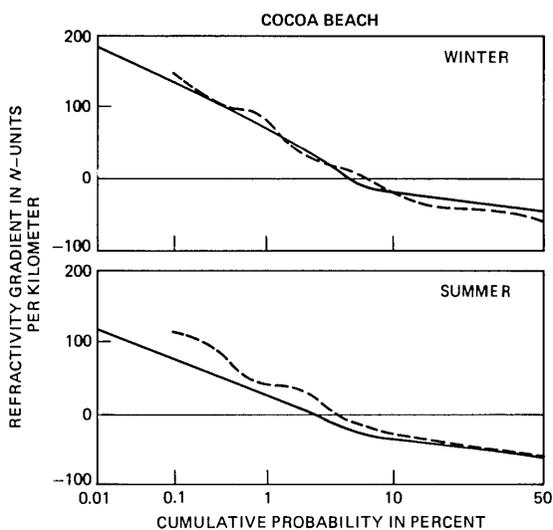


Fig. 5—Comparison of positive refractivity gradient distributions predicted in this work (solid line) with those compiled by Samson¹ from measurements (dashed line) for a 100-m-deep layer for coastal site Cocoa Beach.

low. Therefore, some contribution to the disagreement is expected from statistical noise.

A more complete assessment of the agreement between predictions and measurements has been made by plotting predicted vs measured

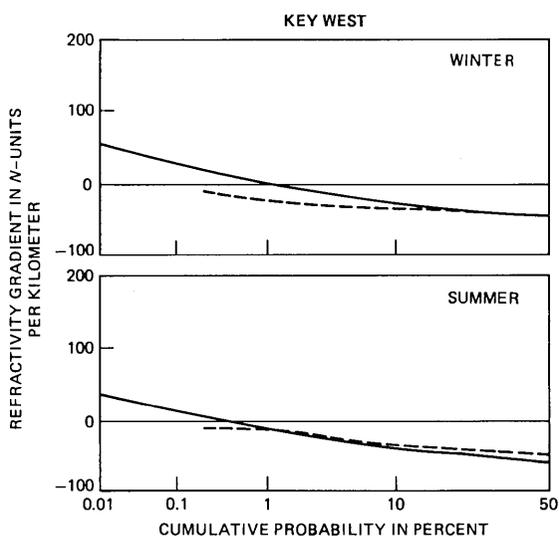


Fig. 6—Comparison of positive refractivity gradient distributions predicted in this work (solid line) with those compiled by Samson¹ from measurements (dashed line) for a 100-m-deep layer for coastal site Key West.

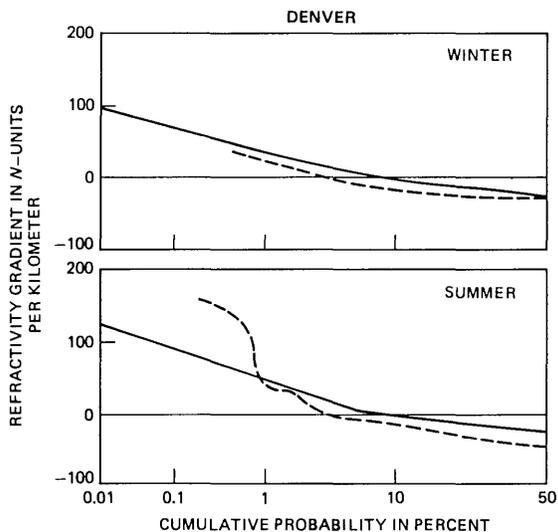


Fig. 7.—Comparison of positive refractivity gradient distributions predicted in this work (solid line) with those compiled by Samson¹ from measurements (dashed line) for a 100-m-deep layer for Denver (an interior site).

annual probability for $N' \geq 100$ N -units/km for Samson's 17 sites (Fig. 8).¹ Such an evaluation requires that path averaging be removed from the predictions for comparison with point measurements. This is accomplished by setting the air horizontal homogeneity factor, H , to its maximum value (1.0) for all sites. Therefore, probabilities calculated in this manner are characteristic of a single point along the 40-km path length.

It is most useful to compare predicted and measured distributions for the largest refractivity gradients for which measured probabilities could be obtained from Samson's data.¹ As the optimum gradient for comparison, 100 N -units/km was selected. However, not all of the measured distributions for each season at a given site intersected 100 N -units/km. Therefore, for those that did not, distribution tails were extrapolated linearly on the original graphs to estimate probabilities at 100 N -units/km for those seasons. Annual values of $P(N' \geq 100$ N -units/km) were obtained by averaging the four seasonal probabilities.

Poorer agreement between predicted and measured distributions is expected for interior sites, since a single specific mechanism has not been modeled as for the coastal sites. However, obstruction fading is not a serious problem in interior regions. For all sites except four, the prediction is near the measurement. Two of the four sites are interior ones, discrepancies for which were anticipated. The four anomalous sites (Denver, El Paso, New York, and Long Beach) all show considerably higher occurrence frequency of positive gradients than pre-

Table III—Sites for which Samson's measured refractivity gradient cumulative probability distributions¹ were employed and approximate number of individual observations associated with each. Symbols in last column identify points in Figure 6.

| Site | Number of Observations | Symbol | Site | Number of Observations | Symbol |
|-----------------|------------------------|--------|--------------------|------------------------|--------|
| Atlanta, GA | 880 | A | Long Beach, CA | 880 | J |
| Brownsville, TX | 880 | B | Miami, FL | 880 | K |
| Charleston, SC | 880 | C | New York, NY | 880 | L |
| Cocoa Beach, FL | 1280 | D | Oakland, CA | 850 | M |
| Columbia, MO | 880 | E | San Diego, CA | 880 | N |
| Denver, CO | 880 | F | Seattle, WA | 880 | O |
| El Paso, TX | 1280 | G | Tatoosh Island, WA | 880 | P |
| Joliet, IL | 360 | H | Washington, DC | 840 | Q |
| Key West, FL | 850 | I | | | |

dicted. This observation suggests that local meteorological effects not incorporated in the model are contributing to producing gradients. Indeed, all four measurement sites are unusual. The Denver site is on relatively flat ground near the foothills of the Rocky Mountains; the El Paso site is on a mesa approximately 60 m above flat, irrigated land; the New York site is located in a complex metropolis-ocean-land-bay region; and the Long Beach site is subjected to Santa Ana winds.¹⁹ Therefore, the anomaly is attributed to local effects on the probability distributions measured at a single point. If it were possible to measure path-averaged distributions for the four anomalous sites, the model predictions for them might, indeed, show better agreement.

IV. SUMMARY

This work presents a climatological model for arbitrary locations in the United States. It provides refractivity gradient occurrence distributions for positive gradients. The results fill in the gaps among the 17 locations for which measured distributions have been compiled. The significance of this additional detail is that climate variations can be accounted for with much better geographic resolution than had previously been possible, in the specification of microwave antenna heights. A method which uses these results to determine antenna heights has been developed by Vigants and is described in a companion paper.²⁰

V. ACKNOWLEDGMENTS

I am very grateful to William T. Barnett and Edward B. Fowlkes for their helpful discussions, and, in particular, to Arvid Vigants for his critique of the manuscript, as well as for helpful discussions.

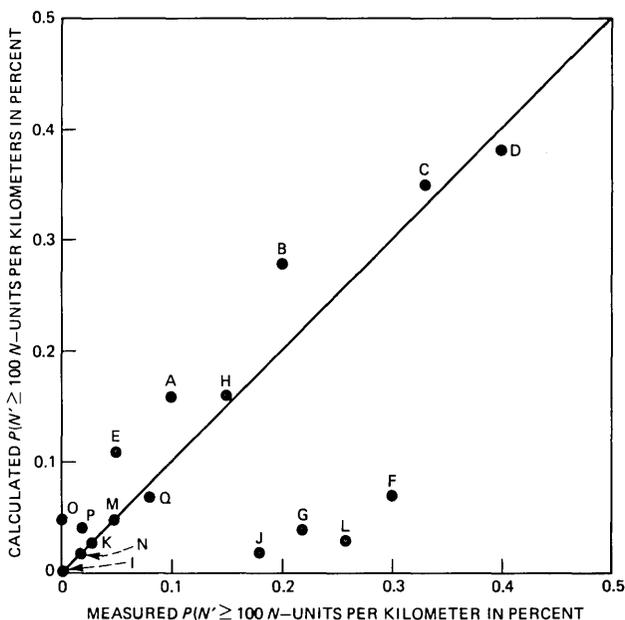


Fig. 8—Comparison of model predictions with Samson's measurements for 17 sites.¹ Values plotted are cumulative probabilities for $N' \geq 100$ N-units/km. The line represents perfect agreement between predictions and measurements. Sites are denoted by symbols which are identified in Table III. The four anomalously low points are attributed to the inability of the model to account for unusual local effects.

APPENDIX A

Critical Wind Direction Choice

Climatological information does not exist on the occurrence frequency of wind shifts from continental to maritime directions as air masses traverse the coastlines. However, there is a certain wind direction relative to a coastline which should be related to the occurrence of large-scale moisture advection.

Before air masses reach the coasts, moderately low-level air flow is roughly parallel to the coast with land to the right of the wind vector. Furthermore, this direction is conducive to producing moisture gradients by the mechanism of wind direction variation with height, as discussed in Section 2.1. Therefore, the surface wind direction that most likely causes flow of marine air over continental air adjacent to the coastline is the direction parallel to the coast with land to the right of the wind vector. This is the critical wind direction at the coastline for an ideal linear coastline of infinite extent. In reality, however, the critical direction choice must account for the curvature of the coastline. This is required to ensure that air assumed to be continental (maritime) in origin indeed is underlain by land (water) for a significant

distance upwind from the site. An appropriate choice of minimum distance is the distance travelled by air in one day at an average speed, approximately 300 km.

Because of unavoidable subjectivity required in selecting critical wind directions for some sites, the selection technique is best illustrated with examples. Figure 9 shows nine sites along the southeast U.S. coast, three each along lines drawn perpendicular to the Atlantic Ocean coast from three different coastal sites: A, B, and C. Critical directions for line AA' are parallel to the coast because land lies upwind along this vector for a distance of at least 300 km. This assures that air arriving at sites along line AA' is indeed continental. Such an assurance is not obtained for sites along line BB' if the critical directions are chosen parallel to the coast at B. The directions are chosen so that there is a 300-km air flow trajectory flow over land upwind from sites along BB' , as shown in Fig. 9.

The situation at site C is complementary to that at site B. Here, the critical wind direction is chosen so that the wind vector is underlain by water for a distance of 300 km upwind from site C. The critical direction at the other two sites along CC' is chosen assuming that maritime air has its source over the Atlantic Ocean rather than the Gulf of Mexico. This is done since a gulf source would require a critical direction at site C' , e.g., as shown by the dotted arrow in Fig. 9. This direction, in turn, would imply continental air originating in the southern half of the Florida peninsula. Obviously, this cannot be considered a source of dry, continental air. Critical direction choices illustrated in the above examples are typical of those made for nearly all locations of the United States.

APPENDIX B

Empirical Determination of a and m_2

The values of a and m_2 for the interior regions are obtained by comparing measured refractivity gradient probability distributions for Columbia, Missouri.¹ The only other interior sites for which measured distributions are available (Denver and El Paso) are characterized by unusual physiographical features (Section III) and are deemed unsuitable for specifying a and m_2 . Values are obtained for μ , f_i , V , and m for Columbia. $H = 1$ is used for comparing the model with point gradient measurements (see Section 2.1.1). The value for σ_0 obtained for coastal regions is used. From eq. (9b), σ_{sI} for Columbia is given by $\sigma_{sI} = \text{constant} \times a(30 + m_2)$, where the constant is evaluated from the above values.

Determining the optimum agreement between calculated N' distributions (using σ_{sI} as specified above) and the measured N' distribution

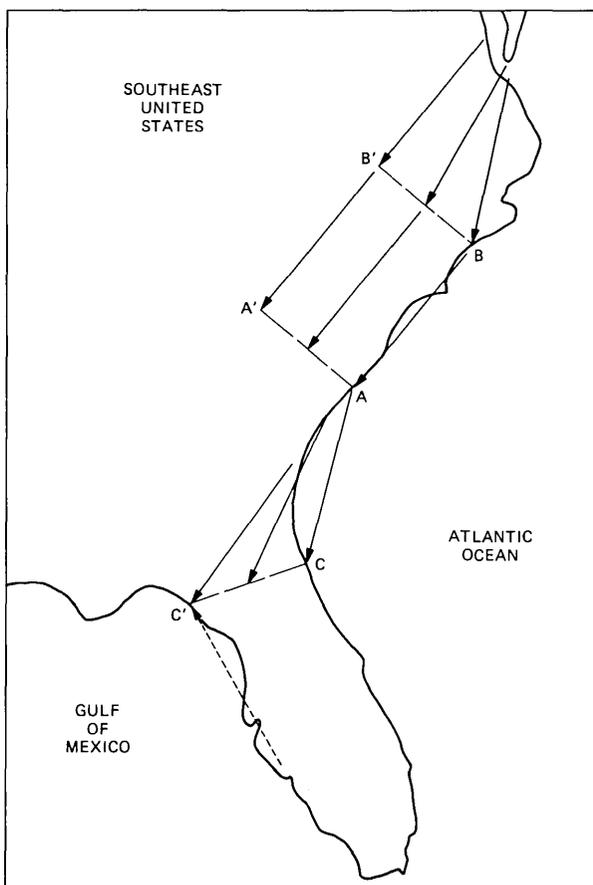


Fig. 9—Examples of critical wind-direction choice. Dashed lines are perpendicular to the Atlantic Coast. Arrows represent critical wind directions chosen for sites at their tips. Arrow length is 300 km (see Appendix A).

for Columbia, yields a relationship between a and m_2 . We determine the remaining variable by subjecting calculated values of $\sigma_{sI}(m)$ to constraints for the possible range of m : $-50 \leq m \leq 120$. Columbia's surface moisture index¹⁷ of 30 is intermediate within this range. A second point to completely specify a and m_2 is obtained by assuming that $\sigma_{sI}(m = 120)$ for Columbia is equal to the smallest value of $\sigma_{sC}(m = 120)$ calculated for all of the 14 noninterior sites. Thus, we determine that $a = 4.7 \times 10^{-5}$ and $m_2 = 100$.

REFERENCES

1. C. A. Samson, "Refractivity Gradients in the Northern Hemisphere," Office of Telecommunications Report 75-59, U.S. Dept. of Commerce, April 1975.

2. B. R. Bean, et al., *A World Atlas of Atmospheric Radio Refractivity*, ESSA Monograph 1, Washington, D.C.: U.S. Govt. Printing Office, 1966, pp. 53-8.
3. H. T. Dougherty, "A Survey of Microwave Fading Mechanisms, Remedies, and Applications," ESSA Tech. Rep., ERL 69-WPL 4, 1968.
4. F. Ikegami, et al., "Variation of Radio Refraction in the Lower Atmosphere," IEEE Trans. Antenn. Prop., *AP-16*, No. 2 (March 1968), pp. 194-200.
5. T. Akiyama and O. Sasaki, "Statistical Distribution and Maximum Critical Value for Refractivity Gradient Variation," Rev. Electrical Commun. Lab., 27 (1979), pp. 841-8.
6. A. Vigants, private communication, 1979.
7. B. R. Bean and E. J. Dutton, *Radio Meteorology*, New York: Dover, 1968.
8. H. Moses, W. C. Ashby, and M. A. Bogner, "Dewpoint Temperature Inversions," J. Appl. Meteorol., 7 (April 1968), pp. 206-16.
9. The Analysis and Forecasting of Atmospheric Radar Refractivity, NAVAIR Report 50-IP-1, Naval Weather Service Command, July 1967.
10. R. M. Holmes and J. L. Wright, "Measuring the Effects of Surface Features on the Atmospheric Boundary Layer with Instrumented Aircraft," J. Appl. Meteorol., 17 (August 1978), pp. 1163-70.
11. J. Neumann, "On the Rotation Rate of the Direction of Sea and Land Breezes," J. Atmos. Sci., 34 (1977), pp. 1913-7.
12. K. M. Zishka, "Climatology of Cyclones and Anticyclones Over the North American Continent and Surrounding Ocean Environs for January and July, 1950-1977," M. S. thesis, Purdue University (May 1979).
13. C. R. Hosler, "Low-Level Inversion Frequency in the Contiguous United States," Mon. Weather Rev., 89 (September 1961), pp. 319-39.
14. *The National Atlas of the United States of America*, Washington: U.S. Depart. Interior Geological Survey, 1970, plate 62.
15. *Weather Atlas of the United States*, Detroit: Gale Research, June 1968, pp. 237, 240, 243, 246.
16. *Ibid.*, pp. 4, 19, 34, 49.
17. C. W. Thornthwaite, "An Approach Toward a Rational Classification of Climate," Geogr. Rev., 38 (1948), pp. 55-94.
18. D. C. Livingston, *The Physics of Microwave Propagation*, Englewood Cliffs, New Jersey: Prentice-Hall, 1970.
19. W. T. Sommers, "Strong Mountain Downslope Wind Field Experiment Using Surface Station, Monostatic Acoustic Sounder, and Rawinsonde Observations," *Fourth Symposium on Meteorological Observations and Instrumentation*, Denver, Colorado, April 10-14, 1978, American Meteorological Society: Boston, 1978, pp. 339-44.
20. A. Vigants, "Microwave Radio Obstruction Fading," B.S.T.J., this issue.

MULGA—An Interactive Symbolic Layout System for the Design of Integrated Circuits

By N. H. E. WESTE

(Manuscript received November 13, 1980)

To aid the design of MOS circuits, a suite of programs residing on the UNIX operating system have been designed and written. These programs allow the interactive editing, layout compaction, circuit connectivity extraction, parasitic audit, and timing simulation of MOS ICs within the symbolic domain. The programs make use of an intermediate circuit description language (ICDL), which captures both geometric placement and circuit connectivity. A convenient interface is provided to enable the procedural definition of symbolic layouts in the C programming language. All design may be carried out at a single low-cost work station which incorporates a high-performance color display. In this paper we summarize the operation and use of these programs. In particular, we describe a new compaction algorithm.*

I. INTRODUCTION

Symbolic layout methodologies are a means of abstracting the detailed and often laborious task of mask design of integrated circuits. They offer the advantages of hand-packed mask design with regard to density of layout, while also having advantages over manual layout with respect to time to design a circuit and reduction in the number of manual errors introduced into a design. In essence, the use of symbolology reduces the complexity of the IC design process, which in addition to the advantages mentioned above, allows experienced designers to undertake more complicated circuits than would otherwise be possible, and, more importantly, allows novice designers to complete designs with a high degree of confidence. This last point is regarded as

* UNIX is a trademark of Bell Laboratories.

especially important as system designers move to use silicon as an implementation medium, rather than more conventional techniques.

A particular IC design may be described in three domains, namely, the structural, physical and behavioral domains. Structurally, a design may be thought of as a graph where components are represented by nodes in the graph and their interconnection modeled by branches. The graph nodes may in turn be represented by subgraphs. The physical domain in an integrated circuit is typified by a collection of geometric areas defined on masks used in the various process steps used to fabricate the chip. Finally, the behavioral domain indicates how the design functions from the level of electrical circuit performance to possibly higher levels of architectural simulation. Manual layout, for instance, captures only the physical attributes of a design. Standard cell layout may achieve consistent descriptions in all three domains at the expense of fixing the physical layout to a set of predefined cells.

This paper describes a symbolic layout system (given the name MULGA) which attempts to achieve consistent descriptions in all three domains in addition to providing a highly interactive environment. The cells designed with this system may be used with standard cell layout systems such as LTX,¹ or combined using structured design techniques as popularized by Mead and Conway² to form super-cells with considerable functionality. In the limit, complete chip descriptions may be completed using the tools that will be described in the body of this paper. All programs are written in the C programming language and run under the UNIXTM operating system. The system has been designed to enhance man-machine interaction, providing a friendly interface to designers with wide ranges of experience in the areas of electrical engineering and computer science.

The paper is divided into four main sections. The first section relates this work to previously published work by examining symbolic design methodologies. We then describe the circuit description language which forms a basis for this system. The design methodology used in this system, the design aids and the hardware that support this methodology are then described. The final section summarizes the work and derives conclusions about the design system and methodology.

II. SYMBOLIC LAYOUT METHODOLOGIES

Symbolic layout methods attempt to abstract the detailed task of designing IC masks to clarify this operation. Normally this is achieved by eliminating or reducing the complexity of the design rules for a given process. These design rules include the minimum spacings and widths of the mask layers used in the technology. They also include

electrical rules for interconnecting layers and the formation of active devices. These simplified rules ideally result in a quicker turn-around of designs and a reduction in errors compared to manual layout. This section illustrates various symbolic design methodologies and their contributions to the IC design process.

2.1 Coarse-grid layout

Coarse-grid layout systems divide the chip surface into a uniformly spaced grid in both the x and y directions. The grid size represents the minimum feature or placement tolerance that is desired in a given process. For each combination of mask layers that exist at a grid location, a symbol is defined. Given a particular design system, these symbols are then placed on the grid to construct the desired circuit much in the same way as one would tile a floor. Symbol sets may be defined as characters or perhaps graphical symbols if a graphics display is used for design.

American Microsystems International (AMI) and Rockwell International have made use of character-based symbolic layout for some time.³⁻⁵ The symbolic interactive design system (SIDS)⁵ uses a color character terminal as a design station which provides a high degree of user feedback. In addition to these character-based systems, Hewlett-Packard has developed an interactive graphics system (IGS),⁶ which is capable of accepting symbolic input on a fixed grid. IGS also uses symbolic representations to reduce the time to display hand-designed layouts.

The design process in these systems consists of laying symbols down on the coarse grid. The use of fixed-size symbols simplifies geometric design rules but does not totally alleviate them. SIDS therefore provides on-line design rule checking for geometric design rule violations, and a "trace" facility to trace circuit nets to visually check electrical connectivity. Similarly, IGS provides "bumpers" which surround symbols to aid designers in the placement of them.

2.2 Gate-matrix layout

Recently, a character-based symbolic layout system has been in use at Bell Laboratories for the design of large CMOS circuits.⁷ This system departs from fixed-grid systems principally because it adopts a design style which involves structuring the way the active and interconnection layers of the process are used. This style has been given the name "gate matrix."

In this layout system, polysilicon is allowed to run in vertical columns of fixed width and pitch. Diffusion is allowed to run horizontally or vertically in "nonpoly" columns. The intersection of a poly column and a diffusion row forms a transistor. The metal layer is

allowed to run in both vertical and horizontal directions to complete connections. The design process involves identifying gate signals and assigning a poly column per signal. Transistors are then placed and interconnected using diffusion columns or Manhattan metal connections. If necessary, extra poly columns may be added to complete a circuit.

This system achieves better density than a comparable fixed-grid layout and moreover has demonstrated "hand-packed" densities on large circuits. This seems to be due to the fact that the designer is provided with a rigid but structured design technique which allows the corresponding "grids" to be placed closer together than in coarse-grid systems.

2.3 Sticks layout

The term "sticks" is a generic term given to symbolic design systems that do not necessarily constrain the designer to a grid and generally require the designer to enter a free-form topological description of a layout via an interactive graphics system. Graphical symbols are placed relative to each other rather than in an absolute manner. Systems representing this form of layout have been reported by Williams,⁸ Dunlop,^{9,10} and Hseuh et al.¹¹ Following the definition phase, the symbolic descriptions are converted to valid mask descriptions using a variety of compaction strategies designed to space symbols in accordance with the process design rules. In this manner, a rough topology is entered prior to an exact geometric description being generated. This is as a result of the designer being totally freed from geometric design rules. However, examples of ICs designed using these systems have not been reported to date.

2.4 Virtual-grid layout

The MULGA system uses a grid-based placement scheme as in coarse-grid layout methods but allows the final geometric spacing between grid lines to be determined by the density and interference of elements on neighboring grid locations. This leads to the notion of a "virtual grid."¹² This concept is best illustrated by a simple example as shown in Fig. 1a. Three vertical wires are shown centered on a grid. The result of using a fixed grid of 10 units per grid unit and a wire width and separation of 10 units leads to the mask description shown in Fig. 1b. By using a grid in which the spacing varies according to topology, the mask description in Fig. 1c is constructed. Grid-based placement allows rapid entry of geometric topology by "snapping" elements to the grid. The use of a grid also aids the capture of the circuit details and the subsequent processing needed prior to preparing a valid set of masks.

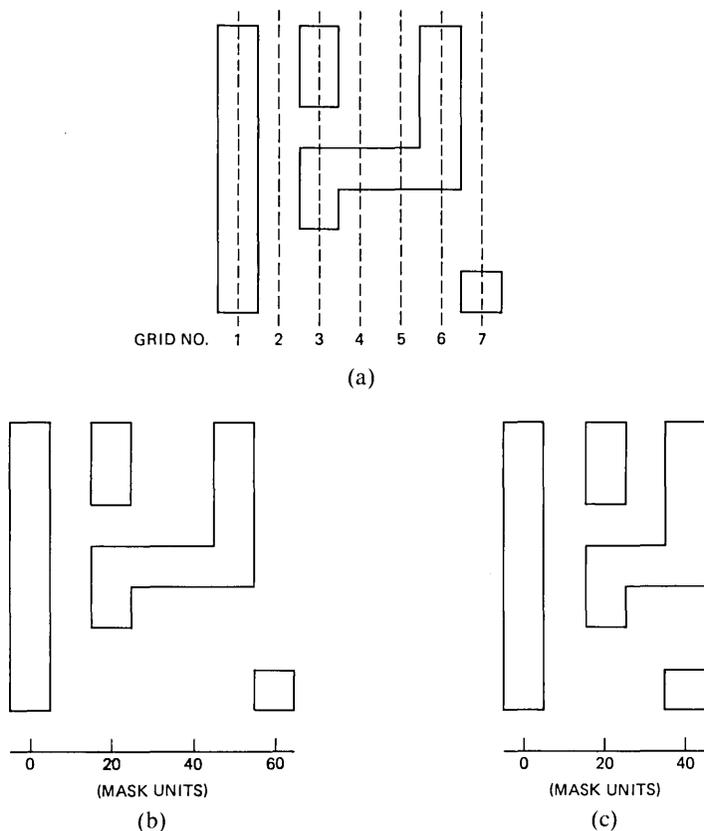


Fig. 1—Comparison of fixed-grid and virtual-grid layouts. (a) Circuit specification based on a grid. (b) Fixed-grid expansion. (c) Virtual-grid expansion.

The notion of a “coordinode” as introduced by Buchanan¹³ is used as an aid to capturing circuit connectivity. A coordinode is defined as, “a named object which has structural, physical, and behavioral significance in the IC design process.”

As its name suggests, a coordinode has the properties of a coordinate, namely some xy position that will eventually map onto the silicon surface. In addition, it may possess the properties of a node in a circuit, perhaps a voltage or a simulation state. Thus both physical and behavioral models may be assigned to the coordinode. The structural significance of a coordinode is that it defines nodes in the graph description of the network that is used to extract the behavioral aspects of the design.

In the MULGA system the nature of a coordinode is altered slightly to suit the interactive symbolic environment in which a designer works. Coordinodes may only exist at coordinates on the conceptual grid, a

result of dealing at the symbolic level rather than the geometric level. In addition, coordinodes need not be specifically named, a result of the desire to have a free-flowing interactive dialog between designer and computer.

2.5 Benefits of symbolic layout

Symbolic layout may aid the design process in two ways. First, the simplification of geometric design rules relieves the designer of detail that can cloud more global and important issues, such as achieving the correct circuit or communication requirements. Transparent design rules also make designs relatively process-independent. If a process design rule changes, the mask descriptions for a circuit may be regenerated with a minimum effort. The second way that symbolic layout may aid the designer is by capturing designer intent. This means that apart from capturing the geometrical details of a design, the design methodology supports capturing the circuit embodied by a design. In other words, the physical and structural aspects of a design may be caught early in the design cycle. An example of this is the case when a designer places a metal path on top of a diffusion region and places a contact at the center of the common area. Physically, a set of rectangular areas on differing mask layers have been specified in size and placed somewhere in the xy plane representing the complete design. Structurally, the net represented by the metal wire on some circuit diagram has now been extended to include the diffusion region, which could, for instance, be the source or drain of a transistor.

The first computer-aided symbolic-layout systems, as represented by coarse-grid layout systems, made only partial use of the full advantages of symbolic layout. In particular, very few if any structural details were captured as the designer was forced to think in terms of tiles of various process layers rather than circuit elements. In addition, circuit verification had to take place on the basis of a program recognizing these tiles in particular configurations that constitute circuits. As noted previously, geometric design rules were only simplified, not totally alleviated.

The more recent sticks systems have, in addition to the perceived benefits of a design rule free environment, the basis for capturing circuit connectivity, although none have treated this benefit in detail. This is due to the fact that specific problems have been addressed, in particular compaction, rather than the complete design cycle and the relevant tools required.

The system described in this paper takes full advantage of the above-cited benefits of symbolic layout in a consistent and logical fashion. Section III describes a convenient language interface.

III. INTERMEDIATE CIRCUIT DESCRIPTION LANGUAGE (ICDL)

As previously discussed, one wishes to capture physical, structural, and behavioral attributes of a design during the specification phase. In addition, in this system a prime requirement was for a highly interactive environment to support "on-line" design. In contrast to the integrated circuit system (ICSYS)¹³ which is implemented in terms of an object-based language, the MULGA system has been implemented in terms of two languages. The first is a primitive but effective intermediate circuit description language (ICDL) that contains physical, structural, and derivable behavioral characteristics. The second level of design supports procedural design in the C language, which enables a designer to write a program that can generate IC DL. All of the software described in this paper uses the IC DL text descriptions of cells as a central data base.

The basic element of the language is a cell, which is composed of elements that may be devices, wires, contacts, pins, or cell instances. This is similar to the set of elements used in Ref. 11. However, rather than being specified in mask coordinates, these elements exist on a virtual grid. As will be seen, this grid serves only as a relative placement framework during the design process and has attributable mask-geometric properties only after further processing.

3.1 Devices

Devices may be a variety of types and are specified by the keyword **device** for a driver type transistor, or **load** for a resistive transistor followed by the device type which may be:

| | |
|----------|------------------------|
| p | <i>p</i> -type device, |
| n | <i>n</i> -type device. |

Devices have attributes of position on the coarse grid, and optional orientation, width, and length parameters. A device is specified as follows:

```
device type xpos ypos <w = width> <l = length>  
< or = orientation >
```

The *type* parameter may define different subclasses of devices, e.g., different implant transistors. Dependent on the position, orientation, width, and length of the device, the position of the source, drain, and gate connections are defined. In this implementation, the position also depends on whether a transistor is a load device or a driver device. The position may be symbolic grid point, line or area, dependent on the connection restrictions that are desired on a device.

For instance, in this implementation a standard driver type device has terminals as shown in Fig. 2. The connection points, lines, or areas, and the basic definitions of the devices are somewhat process depend-

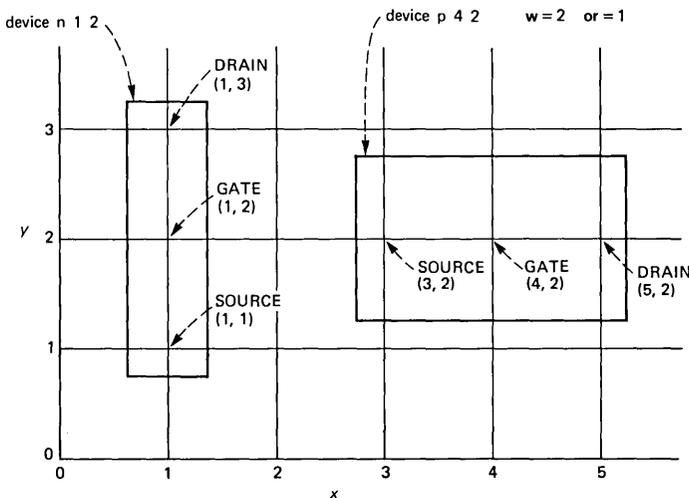


Fig. 2—Driver transistor connections with varying orientations.

ent and must be selected in such a way as to yield a meaningful symbolic representation of the actual layout—both physically and structurally.

The width and length parameters in this implementation are referenced to default minimum width and length devices. Usually the final size of the device gate region is directly scaled using the width and length values. This significance may be changed to suit the process or flexibility required. For instance, increasing the width by a factor of one may increase the actual device size by, say, 2 μm .

It may be seen that transistors dependent on type and parameters have a certain physical realization. The definition of predefined contact points ensures that the structural requirements of a circuit may be met in a logical fashion at the time of capture of the circuit. The behavioral characteristics of a transistor are subsequently found by simulation, based on the physical and structural attributes.

3.2 Wires

Wires can exist on any of the valid interconnection layers of a given process. They are specified by the following statement:

$$\text{wire layer } \langle w = \text{width} \rangle \quad x1 \ y1, \dots, \quad xn \ yn,$$

where $x1 \ y1, \dots, \quad xn \ yn$ are a list of Manhattan vertices. Structurally, wires serve to connect devices and cells together. Crossing two wires on the same interconnection layer provides electrical connection while certain crossings such as poly over diffusion are illegal and result in an error message being generated in a subsequent process. This is as a

result of allowing explicit devices, as defined by the **device** construct. Wires on different layers that intersect at a grid point are only connected if a valid contact is also found at that grid point. Thus wires serve to interconnect coordinodes. Physically, a wire is defined by its width **w**. If no width is given, the default minimum for the interconnection layer is used. Symbolic wires have no absolute length. Their final physical length depends on the compaction process (see Section VI).

3.3 Contacts

A contact joins two or more layers at a grid coordinate. For example, a device may be connected to a wire via the appropriate contact. Specification is as follows:

contact *type xpos ypos*

The range of contacts available depends on the particular process. In CMOS, contacts are of four types: **md**, **mp**, **vdd**, or **vss** which are respectively metal-diffusion, metal-polysilicon, and two substrate contacts. In NMOS depletion load the contacts may be **md**, **mp**, **pd**, or **pmd** for metal-diffusion, metal-polysilicon, poly-diffusion, or poly-metal-diffusion.

3.4 Pins

A pin may be on any interconnection layer and serves to name interconnection points and name specific nets within a circuit. A pin is specified by

pin *type xpos ypos name,*

where *type* may be,

a aluminum,
p polysilicon,
N N-diffusion,
P P-diffusion.

xpos ypos are the symbolic grid coordinates and *name* is the pin name. Although pins have no final geometric significance when it comes to generating mask data, they are of central importance in specifying the circuit corresponding to the physical description. They in essence merge the physical, structural, and behavioral characteristics of a circuit.

3.5 Cell instances

With the above-mentioned four types of element, an MOS circuit may be specified. Hierarchy is added by the cell instance element.

Other cells may be included in a cell by using the following statement:

```
instance    cellname  xpos  ypos  <n = rep> <dx = xx>
            <dy = yy> <"instname">
```

where,

cellname is the name of the cell instanced,

xpos ypos is the symbolic grid location of the origin of the instanced cell,

rep is an optional number of repetitions (default = 1),

xx is the optional *x* grid displacement/repetition,

yy is the optional *y* grid displacement/repetition,

instname is the particular name denoting a particular instance of a cell.

3.6 Graphical representations

As well as the textual descriptions of ICDL, graphical descriptions exist for interactive editing and documentation. Figure 3 gives an example of the symbols used for interactive editing and the associated ICDL text. Note that these symbols mirror the width of wires and the size of different devices, although not necessarily to scale. The object of the representation is to give the designer some idea of *relative* thickness or size to aid in placement of circuit elements. This "thick sticks" or "logs" notation has been found to be more suitable for the symbolic design of ICs than some of the more conventional sticks notations. If one designs a topology using only stick figures, then when the geometrical aspects are taken into account, these neat topologies are often blown apart. The aim in this system is to provide as much support as possible for the designer to structure his design from the start, by taking into account in a symbolic manner the relative geometries of elements.

Symbols used for hardcopy documentation are illustrated in Fig. 4. In addition, a set of characters have been designated so that character terminals may plot layouts in the same manner as coarse-grid symbolic layout systems.

IV. DESIGN METHODOLOGY

4.1 Chip design

The chip design methodology that has been used with this system to date may be described as a top-down plan followed by a bottom-up implementation. The designer starts with the function or functions necessary and completes a chip "floor plan" as illustrated in Fig. 5. This combines a partitioning of the functionality along with some global wiring strategy. In this example of a data path, data runs horizontally in aluminum and control runs vertically in polysilicon.

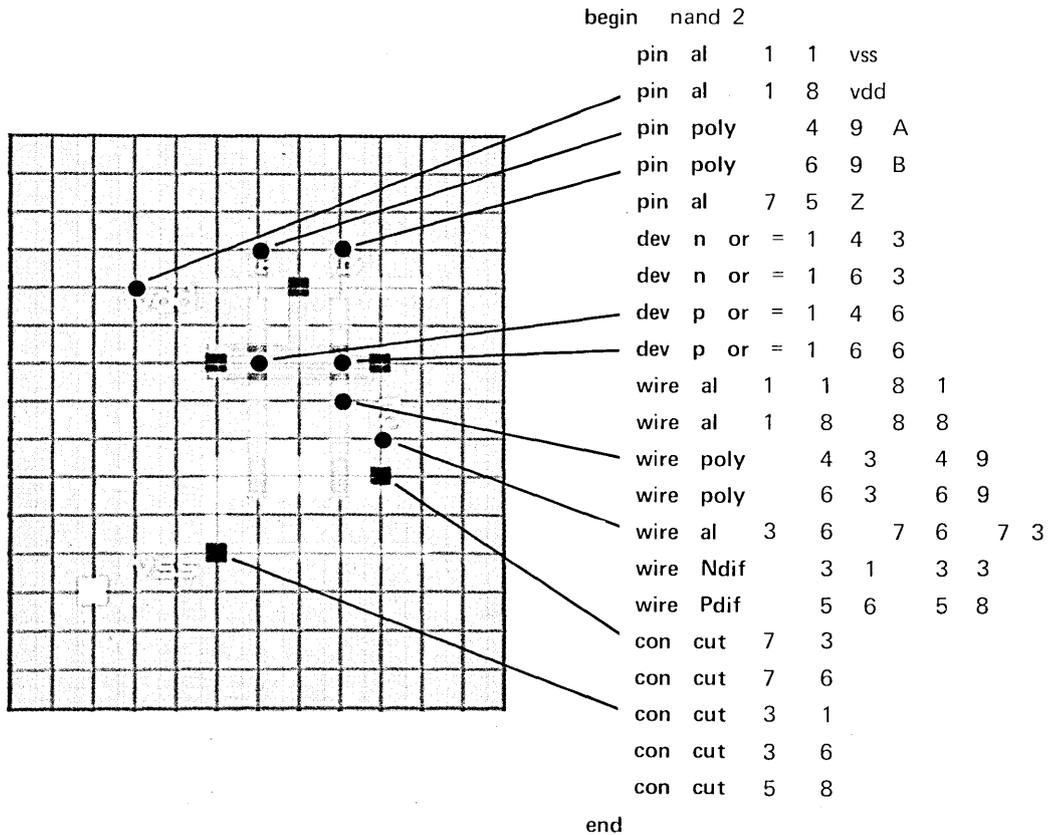


Fig. 3—An ICDL circuit—graphical and textual descriptions.

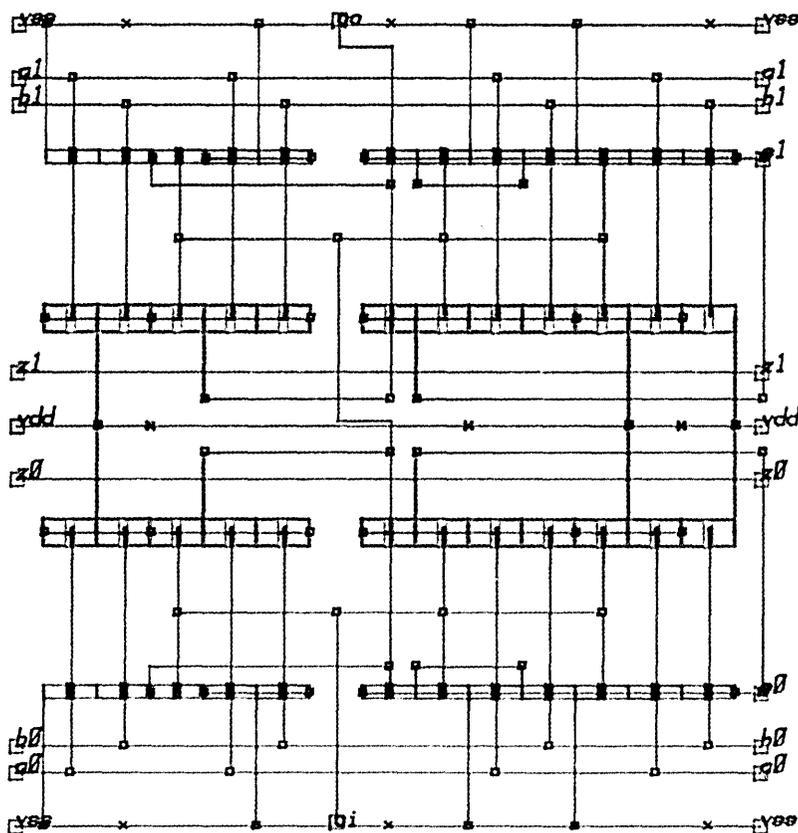


Fig. 4—ICDL output produced on plotter.

Once the overall wiring strategy is defined, the circuit blocks of the "floor plan" are decomposed until manageable cells have been defined. For instance, the Register ALU (RALU) block may be divided into a REGISTER section, CONTROL blocks, and an ALU section. The ALU may be further divided into an XOR, ZERO, ADDER, and SHIFTER section. At this point the ADDER is defined as a 2-bit section taking into account the original data and control directions and interconnection layers. This functional reduction is outlined in Fig. 5 and the resulting ADDER is shown in Fig. 6.

Cells are designed in this manner until the complete circuit block has been defined (RALU). The cell design is then iterated as explained below.

4.2 Cell design

The method of cell design used in this system is as follows:

(i) Cells are first entered into the system. This may be achieved by drawing a rough draft of a cell on squared paper and manually entering an ICDL description via a standard text editor. Alternatively, an interactive graphics editor may be used to edit ICDL files in an on-line

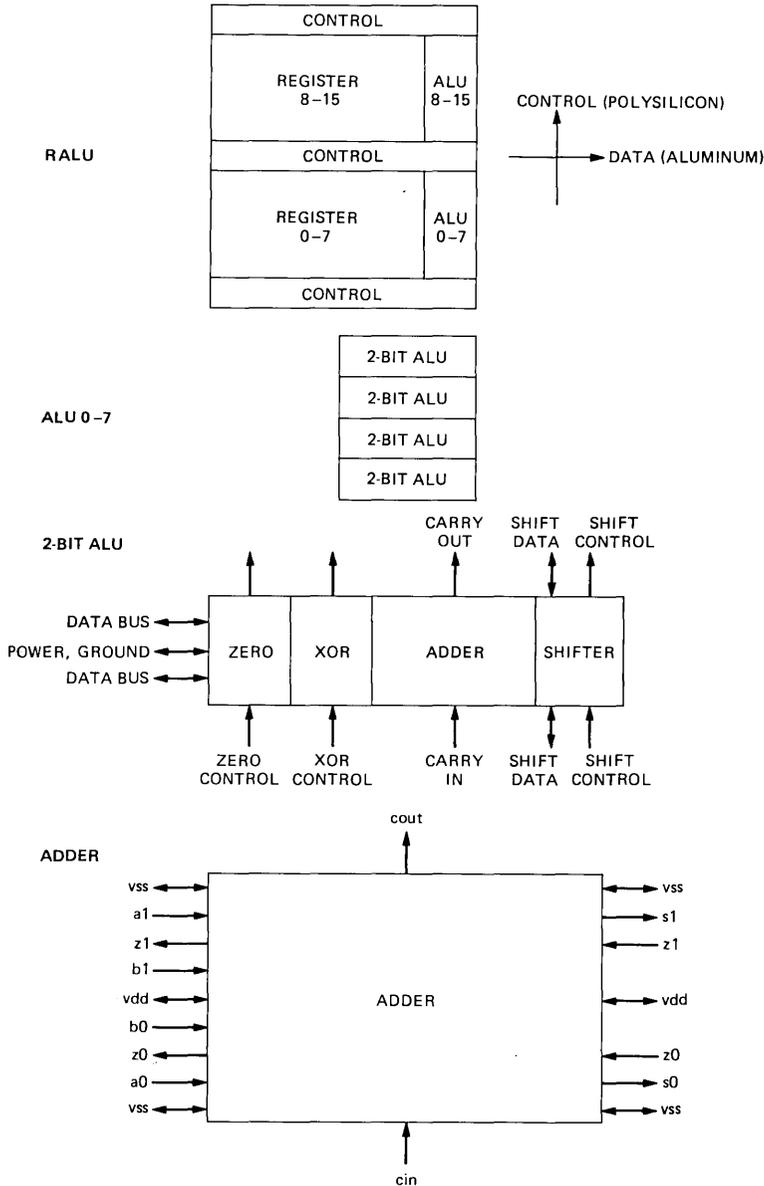


Fig. 5—Chip floor plan showing sample reduction technique to obtain a simple gate.

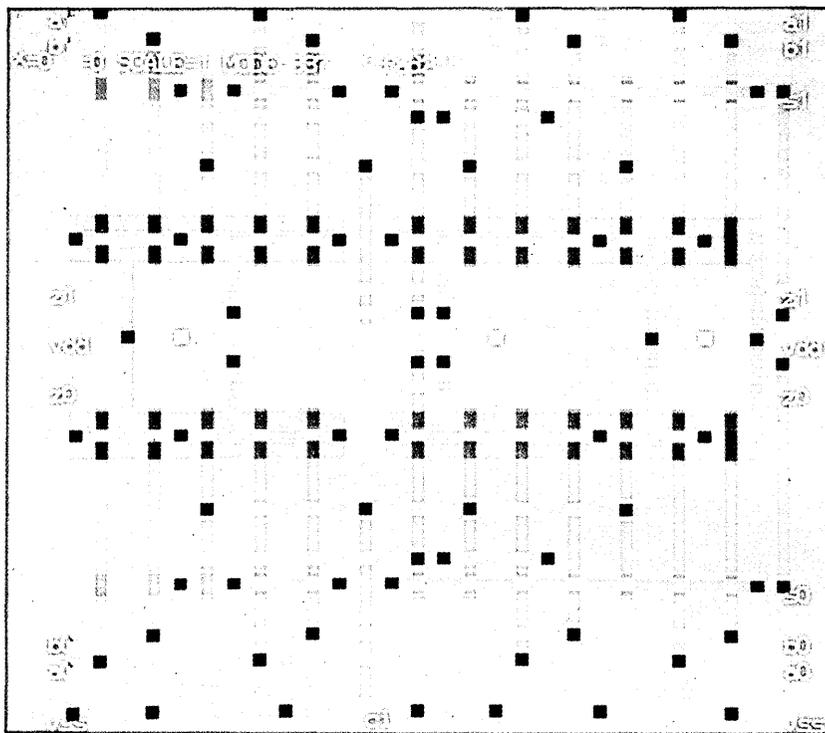


Fig. 6—A 2-bit CMOS adder shown on the color-graphics display.

fashion. Another method of input involves writing programs in the C language to procedurally define ICDL files.

(ii) The structural details of a cell may then be verified via a connectivity extraction program. Cells are re-edited at the symbolic level until the circuit is correct.

(iii) Cells so designed may then be compacted. This relates the grid positions used in the ICDL language to the actual coordinates allowed in the final mask set. Feedback is provided to the designer to enable him to optimize his design at the symbolic level with regard to packing density.

(iv) Behavioral aspects of a cell or group of cells may then be found by using this structural description with an in-built timing simulator. The connectivity extractor now uses the actual physical mask values to calculate parasitic capacitances and transistor sizes which are used by the simulator.

(v) Finally, various methods of expansion to the mask level are used to produce a final physical description of the circuit. If necessary, the

cell may be re-edited at the symbolic level to improve packing or to interface with neighboring cells. Cells are incorporated into larger cells by a combination of repetition, butting, and simple routing.

The following sections describe the design aids needed to support this design style.

V. INTERACTIVE TEXT AND GRAPHICS EDITOR

Because the basis for the system is the ICDL language, we would like to be able to edit this language and see the effects of the edit operation on the layout. This requires a simultaneous text and graphics editor, which is relatively easy to implement with ICDL by a simple interpreter that works on a common data base.¹⁴

A design session begins by reading a cell into the editor. If a cell calls other cells, then they are read in order into the data base. If the design is such that all the cells cannot be read into memory, then cells are thrown out according to a least-recently used algorithm. As cells are read in, their bounding box dimensions are calculated and this information along with the cell name is kept in a global data structure which is maintained for the duration of the edit session.

The cell may then be viewed on a high-performance color graphic display.¹⁵ All details may be plotted or just the bounding boxes of cells shown. The textual output corresponding to the layout displayed on the color display is presented on a standard text terminal. The designer may manipulate the cell description much in the same way that he would use a text editor, except that the actions are mirrored on the graphic display. For instance, if a particular line of text is displayed, then the element it represents is highlighted on the color screen. If the "delete element" command is given, then the element disappears from the screen and is deleted from the ICDL data structure. In addition to being able to perform operations via the terminal keyboard, a designer may use a data tablet for input and receive graphical feedback via the color display. For instance, an element may be identified by placing a tracking cross over the object and invoking the appropriate command. The editor then prints the line of text associated with the object. In the case of multiple objects occurring at the same grid point, each individual element is returned in sequence.

Representative commands are as follows:

view (v)

plots the cell at current scale and x,y origin in current mode;

scale (s)

changes the plotting scale;

append (a)

places the editor into the append mode in which ICDL may be

entered at the keyboard or via a data tablet. Input via the data tablet is directed by a set of menus on the graphics display;

print (p)
prints the current line and highlights the element on the screen;

delete (d)
deletes the current line (which is highlighted on the screen). The element is removed from the screen;

deletearea (D)
deletes all elements in an area defined by keyboard or tablet;

.=
prints the value of the current line;

?
prints the position of the cursor;

move (m)
moves a rectangular section of ICDL by an arbitrary x and y displacement as specified by the tablet or keyboard;

copy (c)
makes a copy of a cell with specified translation, reflection and rotation;

box (b)
flips the editor between the "draw bounding box" and "draw detail" mode;

text (t)
toggles the text flag which enables pin names to be displayed;

identify (i)
identifies a particular element indicated by coordinates typed at the keyboard or by the tracking cursor and data tablet. If so desired, an identified element may be deleted;

Pan (P)
allows the designer to enter a mode which enables him to pan across a layout in real time in any direction. This is done with minimum change to the screen so that the designer does not lose visual perspective of the layout;

grid (g)
turns the grid flag on (or off). This causes a grid to be drawn when *view* is invoked with a grid spacing equal to current *scale* value;

write (w)
writes the cell to disk;

quit (q)
quits the editor.

Figures 7a and 7b show the RALU mentioned in Section 4.1 in symbolic format. Figure 7c shows the ADDER in XYMASK format. Figure 8 shows a section of the RALU at a reduced scale factor.

VI. COMPACTION

6.1 Introduction

Once a symbolic layout has been entered and checked as structurally correct, it is desirable to convert this symbolic description into a set of mask data. This is achieved by what is termed a compaction strategy. In general the symbolic description is examined in, say, the x direction and the minimum geometric spacing between elements or symbols is found. This is then repeated in the y direction. Akers developed a compaction algorithm based on a fixed-grid system.¹⁶ In this algorithm blank grid sites are sought, say, in the y direction. Figure 9 illustrates a simple example. The path of blank cells may be disjoint along so-called shear lines as shown in Fig. 10. When a path is found across the layout, that space is taken out. Dunlop reported a system in which the symbolic input did not have to be placed on a coarse grid but instead used a *node-and-line* form of relative placement.¹⁰ A similar approach is treated in Ref. 11. The reader is referred to these papers for an in-depth discussion of these techniques.

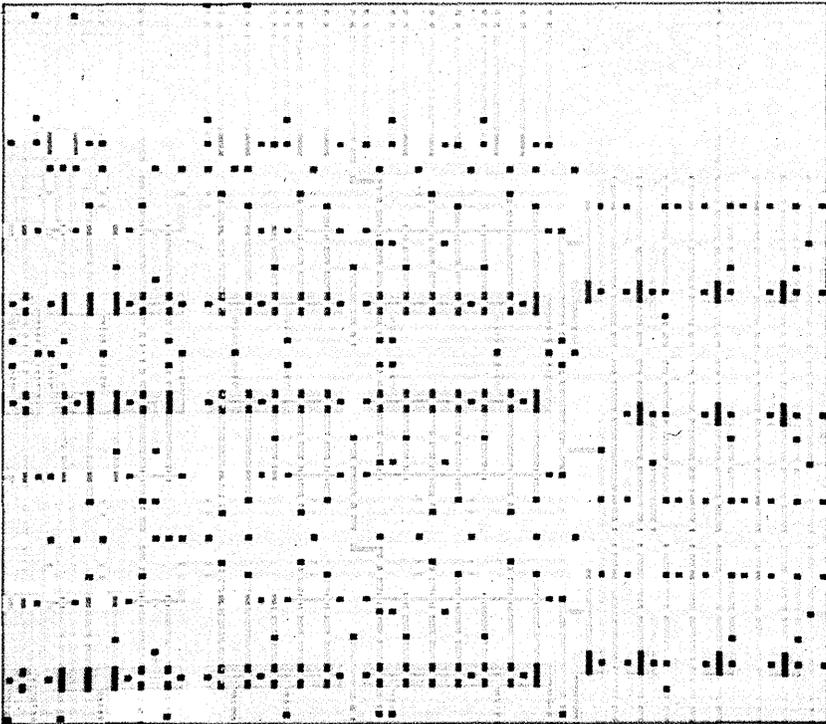


Fig. 7a—Part of the RALU in symbolic format scale = 2.

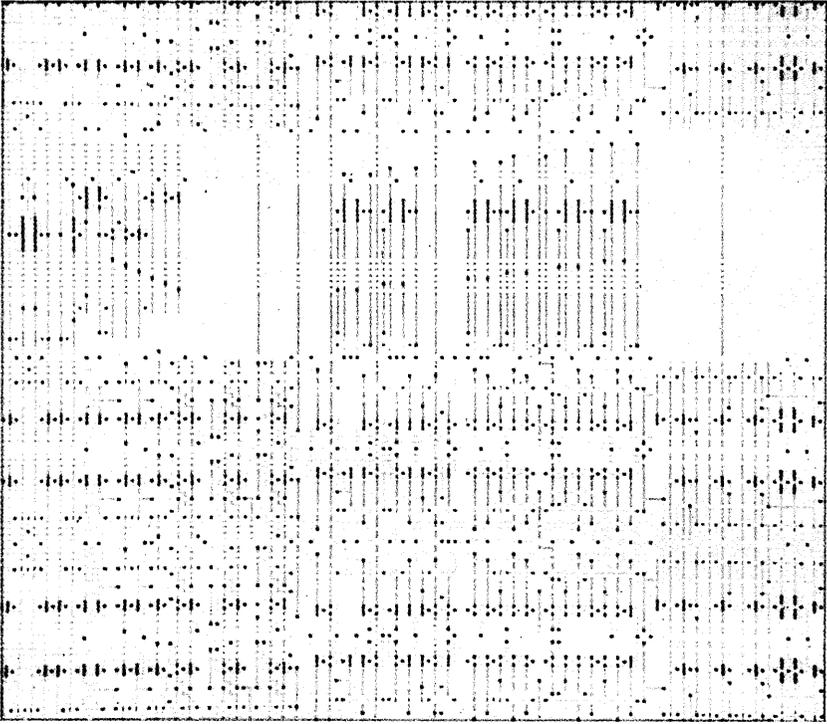


Fig. 7b—RALU in symbolic format scale = 4.

The compaction algorithm used in this system is conceptually very simple and is similar to all those mentioned. Primarily, the object of the algorithm was to be fast while producing adequate compaction. Very little optimization such as jog generation or automatic routing has been included. Rather, copious feedback is provided to the designer to enable optimization of a design interactively at the symbolic level. It is also reasoned that by providing the designer with informative and lucid descriptions (the color display), and a knowledge of compacter performance, a better symbolic design will be entered from the start. Practice has shown this to be a valid design method over the many cells that have been designed on the system.

6.2 *Virtual-grid compaction*

Recall that the data base consists of elements or objects such as transistors and wires that exist on a virtual grid. These two facts are used to form the basis of the compacter. Virtual-grid compaction takes adjacent grid lines and assigns mask coordinates to these grid lines

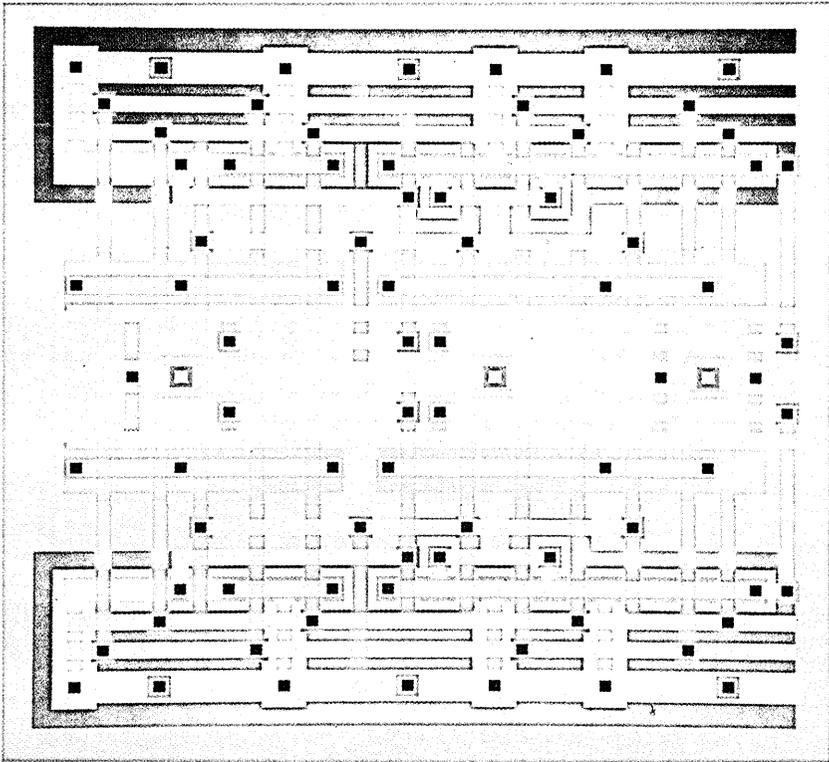


Fig. 7c—2-bit full ADDER in XYMASK format.

according to the relative placement of elements on the grid lines. Thus although elements exist on a grid from the input phase, the spacing of the grid is a function of the particular topology of the circuit being designed, and in fact its environment in the form of neighboring cells.

Elements are initially plotted into a matrix representing the area to be compacted. For instance, wires are plotted as a single line for wire segments parallel to the direction of compaction, and as endpoints otherwise. Transistors are normally plotted as three elements, one each for drain, source, and gate. Pins are not plotted and contacts are plotted as one element. The data structure at each element in the matrix is shown in Fig. 11. Thus in this implementation, at most one device, two wires, and one contact may be recorded at one element. In the case of a conflict the largest of any conflicting element is recorded. From the pointers to the ICDL data structure available at each element, it is relatively easy to calculate not only physical dimensions but also to deduce connectivity. Note that while the data exists on a grid, the

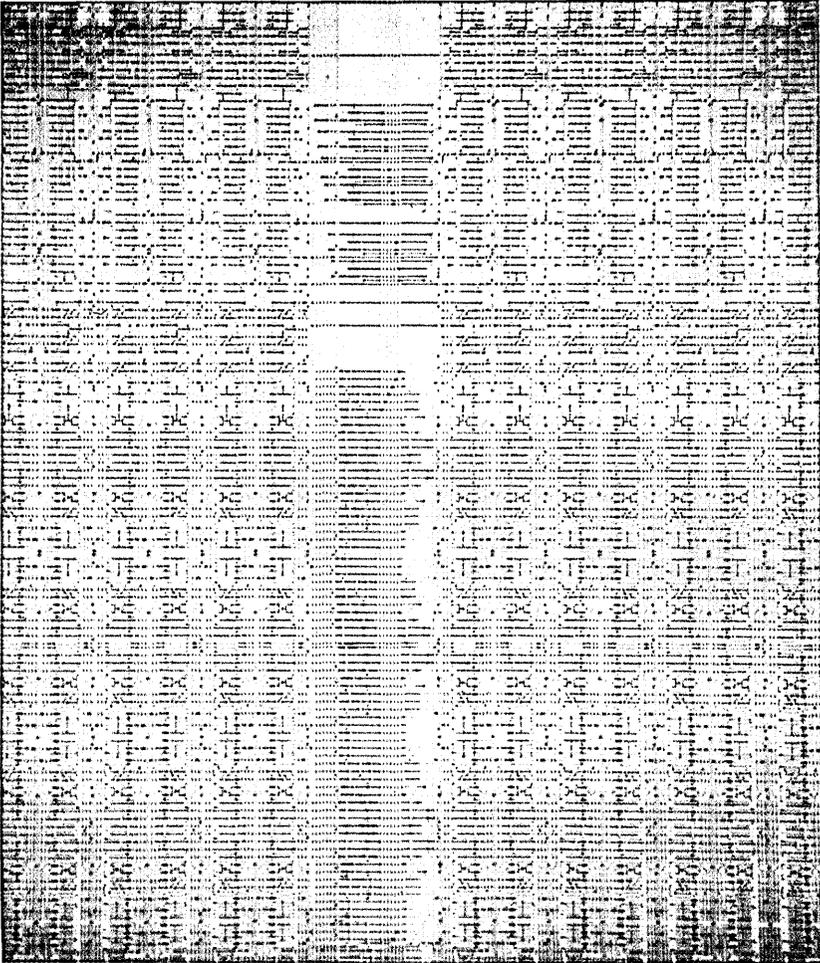


Fig. 8—RALU in symbolic format scale = 8.

grid has no specific geometric relationship to other grid locations until the compaction is concluded.

Having plotted the matrix, it is scanned columnwise, comparing adjacent elements. The worst-case element spacing for a given column is recorded. The mask coordinate corresponding to the current symbolic grid location is this value plus the previous mask grid coordinate. In addition to comparing adjacent columns, the x compaction backtracks to previous columns until the distance between the current column and prior column exceeds some worst-case value. In this way spacings that exert their influence over a number of symbolic grid

locations are discovered. As the array is scanned, the position of the dominating interference element is monitored and stored for future reference.

When the x compaction is completed, the process is repeated for the y direction. However, in addition to backtracking, an arc is swept out from the leading element edge to account for oblique design rule violations. This effect is shown in Fig. 12.

On completion of x and y compaction, the correspondence between the symbolic grid location and the minimum mask coordinate allowed are stored in a "design grid" file. As will be demonstrated this is subsequently used with the original ICDL description to create a valid mask description.

6.3 Performance

The performance of the algorithm is plotted in Fig. 13. Note that the algorithm has essentially a linear execution time with respect to

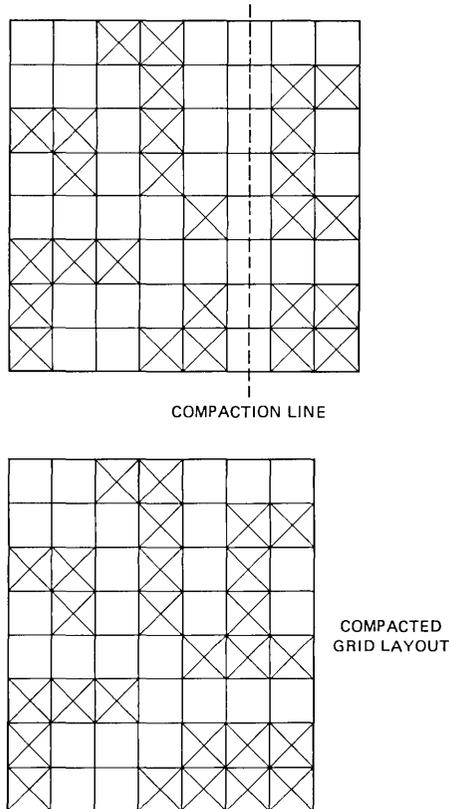


Fig. 9—Simple example of compaction on a coarse grid.

A is then expanded using the y design grid file of pseudocell AB and the x design grid file of cell AC. Cell A has now been pitch matched by virtue of the compaction process to cells B and C which it abuts. Any interconnection point matching in the symbolic description automatically matches in the geometric domain. The other cells are expanded in a similar fashion. This symbolic pitch matching may be attained in this system either procedurally or by visual inspection. The speed of the compacter is important here, as quite often large bit slices have to be compacted.

The second technique was evolved to deal with NMOS circuits, which tend to be less regular than CMOS circuits. Each cell is first compacted individually. Following this, the connection points and direction of connection of adjacent cells are specified in a file. A program adjusts the individual design grid files of common cells to match at the connection points. In the NMOS circuit examples, this saved from 10 percent to 20 percent in area, and will form the basis for future work. Another refinement may include allowing connection points to wander along a connection boundary.

Cells that are only defined at the mask level may also be merged using this technique. Such cells have fixed design grid files and have

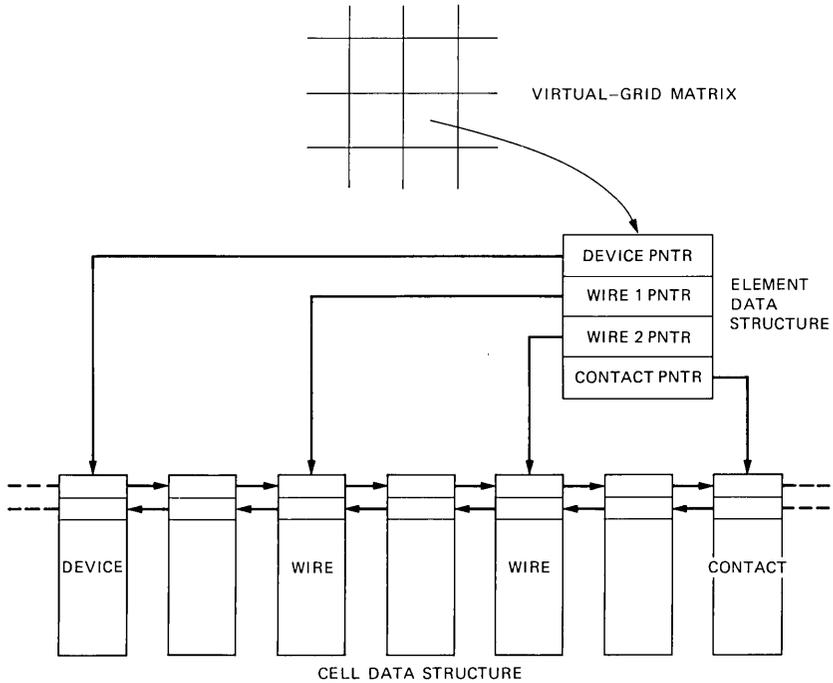


Fig. 11—Virtual-grid data structure.

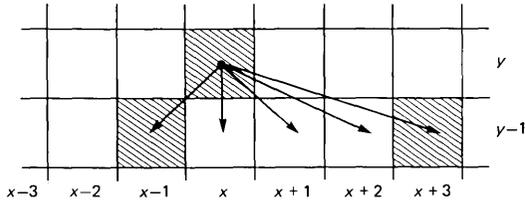


Fig. 12—Oblique violation checking. Arrows show grid points that may be checked during y compaction.

corresponding ICDL files which contain only pin information for interconnection purposes. ICDL cells may be pitch matched to these fixed cells or if necessary a routing channel may be inserted to interconnect them. An example of such a cell is an I/O pad from a standard cell library.

On completion of mask conversion, the output may be viewed in a mask level version of the interactive graphics editor. The same panning abilities are provided. In addition, feedback is provided at the symbolic level showing the critical compaction points in the design. The designer may use this graphical feedback to guide a redesign of the cell at the symbolic level. One advantage of the compacter is its predictable actions which may be used to advantage by the designer to obtain more compact mask layouts.

VIII. PROCEDURAL DESIGN

Situations arise during the course of a design in which it is desirable to procedurally define a design. This may include the parameterization

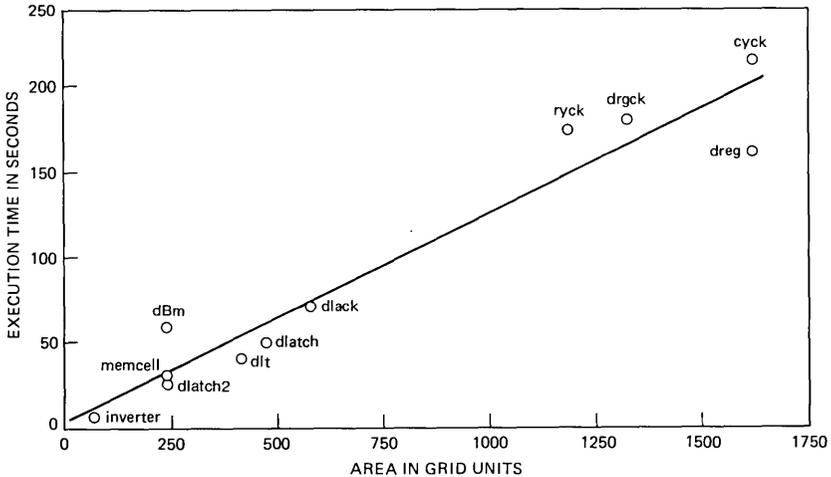


Fig. 13—Compaction algorithm performance.

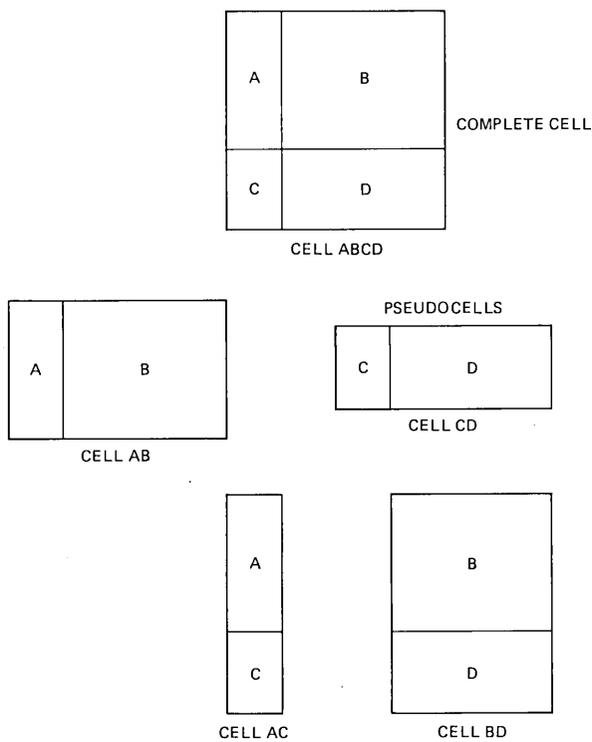


Fig. 14—Mask conversion for butting cells. Subcell A can be expanded using x design grid file of cell AC and y design grid file of cell AB.

of a cell, or the algorithmic specification of cells using looping and conditionals. A higher level language is also useful when it becomes necessary to put large pieces of circuitry together. To obtain this power, a set of C programs were written which generate an ICDL data structure, and may be used to procedurally define ICs.

In brief, subroutines are available to initialize the data structure, to insert devices, wires, contacts, and pins, and to interconnect cells. The subroutines allow the naming of transistor terminals and the subsequent interconnection to these named terminals. As an alternative to this method of procedural design, a user may just use parameterized print statements to generate ICDL files. Of course this method may be programmed in any higher-level language.

IX. CIRCUIT EXTRACTION

One of the benefits of using ICDL is that it carries an implicit circuit-connectivity description of a layout that has been entered either interactively or procedurally. As explained in Section III, devices have designated connection points which are related by simple geometric

rules to the center of a device. Wires serve to connect devices and external connections via interlayer contacts. The pin construct aids the designer in naming specific nodes and connection points. In contrast to circuit extraction programs which work at the mask level, the inherent connectivity in the ICDL description is used to arrive at a transistor node table which does not have to infer the existence of transistors. In addition, values for parasitic capacitance are computed.

The algorithm used is quite simple and is illustrated to indicate the ease with which this may be achieved at the symbolic level by using a representation such as ICDL.

Following input of the ICDL, each pin is given a different node number. Each contact is also given a node number and then each transistor connection not covered by a pin or contact grid position is given a node number. In effect then, each virtual grid position that is occupied by a contact, pin, or transistor has a different node number. These form the coordinodes of the circuit.

Each wire is then taken in sequence and tested for intersection with wires of similar type. At the same time wires on different layers are tested for intersection via interlayer contacts. As each wire is tested a list of nodes that it crosses is maintained. At the end of this step each wire has a record of all wires that connect to it either directly or via valid contacts and a list of coordinodes that it effectively intersects.

A recursive technique is then used to reduce the node numbers common to a particular wire, to one value. All pin, contact, wire, and device node values are adjusted accordingly. For each device the node table is then printed out, with the appropriate pin names used where possible. Unnamed nodes are given an internally generated name.

Wire capacitance values for each named node are then summarized. Mapping from symbolic to mask coordinates may take place by either previously compacting the ICDL description and using values from the design grid file or using a statistically averaged grid spacing value.

The output of the circuit extractor may be piped through various filters to suit a range of circuit analysis and simulation programs. An example of a form suitable for input into a circuit analysis program is shown in Fig. 15.

9.1 Simulation

Once the physical and structural attributes of a cell have been defined, it is necessary to obtain some idea of the behavioral characteristics. To achieve this, a transistor and gate level simulator was written and installed on the MULGA system. It is modeled on a MOS timing simulator (MOTIS) and enables quick turnaround simulation of cells via the circuit extractor.¹⁷ Simulation may take place at the transistor level or transistors may be merged into gate descriptions in

```

•SUBCKT fad ( vss vdd a b _sum ci _co )
MN1 I0 a vss vss N7A
MN2 _co b I0 vss N7A
MN3 I1 ci _co vss N7A
MN4 I1 a vss vss N7A
MN5 _co b vss vss N7A
MN6 I2 _co _sum vss N7A
MN7 I2 ci vss vss N7A
MN8 I2 a vss vss N7A
MN9 I2 b vss vss N7A
MN10 I3 ci vss vss N7A
MN11 I4 a I3 vss N7A
MP1 _co a vdd vdd P14A
MP2 _co b vdd vdd P14A
MP3 _co ci vdd vdd P14A
MP4 I5 a vdd vdd P14A
MP5 _co b I5 vdd P14A
MP6 _sum _co I6 vdd P14A
MP7 I7 ci _sum vdd P14A
MP8 I8 a I7 vdd P14A
MP9 I6 b I8 vdd P14A
MP10 I6 ci vdd vdd P14A
MP11 I6 a vdd vdd P14A
MN12 _sum b I4 vss N7A
MP12 I6 b vdd vdd P14A
CMvss vss 0 CMTOSH 324
CNTvss vss 0 CN + H 48
CMvdd vdd 0 CMTOSH 324
CPTvdd vdd 0 CP + H 48
CPa a 0 CPTOSH 168
CMa a 0 CMTOSH 162
CPb b 0 CPTOSH 168
CMB b 0 CMTOSH 162
CP_sum _sum 0 CPTOSH 30
CM_sum _sum 0 CMTOSH 168
CPT_sum _sum 0 CP + H 12
CPci ci 0 CPTOSH 150
CMci ci 0 CMTOSH 90
CP_co _co 0 CPTOSH 54
CM_co _co 0 CMTOSH 138
CM12 I2 0 CMTOSH 24
CNT12 I2 0 CN + H 12
CM16 I6 0 CMTOSH 72
•FINIS

```

Fig. 15—Circuit extractor output for circuit-analysis program.

the manner that MOTIS operates. A bonafide transmission gate may be modeled and buses are handled inherently by the data structures used. This simulator is called EMU for emulation of MOS circuits on the UNIX system.¹⁸

A filter on the circuit extractor output allows the cell's logic description to be specified in terms of a C subroutine. These subroutines may be then combined hierarchically to suit the high-level functions that the cells constitute. This description may then be compiled to provide an intermediate simulator code which is input to the simulator. A full discussion of the simulator may be found in Ref. 18.

Output from the simulator may be viewed on the color CRT or plotted on a four-color plotter for hard-copy purposes. Typical output is shown in Fig. 16.

The integration of the simulator into the design station is regarded as very important as it allows a designer the same fast feedback

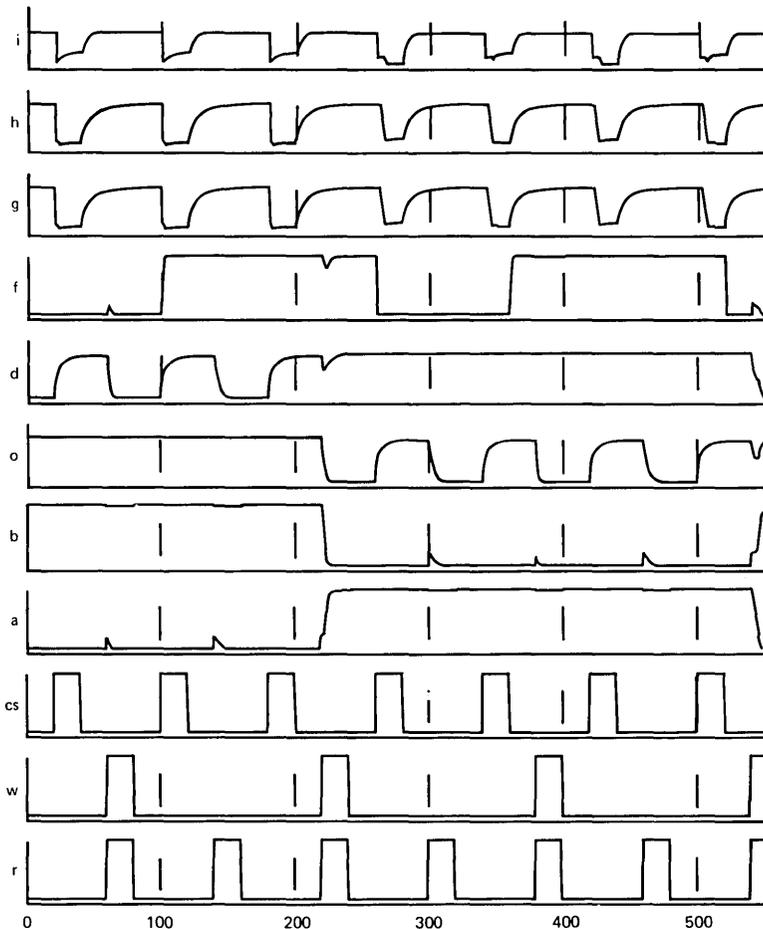


Fig. 16—Typical simulator output.

regarding behavioral aspects of a design, as are available physically and structurally.

X. HARDWARE

10.1 General

Aspects of the hardware have been alluded to during the course of this paper. Figure 17 is a sketch of the hardware used in the design station.

The general purpose host is a DEC LSI 11/23,* with 256 K bytes of

* DEC is a trademark of Digital Equipment Corporation.

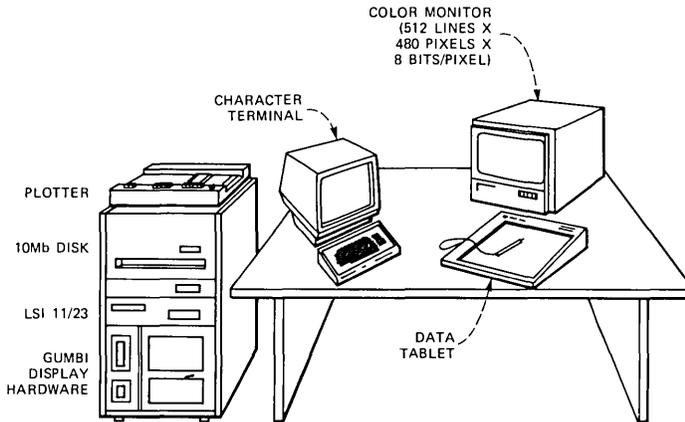


Fig. 17—Hardware used in the design station.

semiconductor RAM. A CDC “Hawk” disc provides 10 M bytes of secondary storage.* Serial ports support a printer, glass TTY, dial-up lines and a four-color plotter. A serial link is also provided to a DEC VAX 11/780 for backup storage and large numeric processing tasks. The *UNIX* multiuser operating system controls all devices.

On the interactive side, an 11-inch data tablet is used in conjunction with a high performance color display. The color display contains a 512 pixel/line by 512 line (480 viewable) frame buffer with 8 bits per pixel. A color map provides 256 colors from a palette of 2^{24} . A 16-bit microprogrammed graphics processor (GUMBI) is responsible for raster conversion into the frame memory.¹⁵ The total cost of the hardware is in the vicinity of 50 thousand dollars, but near term reductions in cost would lead to a total cost in the range 20 to 30 thousand dollars.

10.2 Display attributes

Many of the human-engineered features of the editor rely heavily on the display terminal used, and it is worthwhile to review some of the requirements that have been necessary to obtain rapid response times.

10.2.1 Speed of display

This attribute is important in an interactive situation to reduce operator fatigue and boredom and provide a fluid man-machine dialog. Speed is achieved in two ways:

- (i) A high-speed interpretive graphics processor.

* CDC is a trademark of Control Data Corporation.

(ii) Information bandwidth matching between host and graphics processor.

The first mentioned item is satisfied by a high-speed microprogrammed graphics processor called GUMBI.¹⁵ GUMBI takes commands such as draw box from a host processor and is responsible for plotting the required bit patterns into a frame-store memory. This offloading of menial processing means that the host processor may concentrate on data-manipulation operations rather than time consuming bit-manipulation tasks.

Information bandwidth matching is achieved by:

- (i) Efficient prefiltering of data to be plotted.
- (ii) Direct interpretation of the ICDL data structure.
- (iii) Optimization of the *UNIX* operating system DMA driver.

It is aided by resolving all plotting operations into two steps. First, a symbolic coordinate window of interest is maintained. This is the coordinate representation of the data base and only simple translations need be made to determine actual positions of elements in symbolic coordinates. Second, a screen coordinate window is kept to determine clipped shapes that must be passed to the display processor. As the hierarchical data base is transversed, the bounding box of a cell is first tested against the symbolic window. If it is outside the window, the next cell is considered. Alternatively, the elements within a successful cell are tested against the symbolic window. The successful candidates of this check are transformed into screen coordinates and then clipped against the screen window. Rectangles are then passed to a buffer which is transferred via a DMA interface to the graphics processor. Note that the "display list" for this process is in fact the ICDL data structure, alleviating the need for maintaining intermediate graphic data structures. The normal DMA driver for the *UNIX* operating system was modified to allow direct transfer of data from "user" space to the DMA port. In addition, once a DMA transfer has been initiated the "kernel" process returns rather than waiting for the termination of the DMA transfer as is normal. This change alone improved the speed of data transfer by 10:1.

10.2.2 Selective update

Because a refresh raster scan display is being used as the display medium, alterations to the screen do not necessitate the screen being erased and replotted as is the case for a storage display. Thus in a "delete" operation only the element deleted is removed without disturbing the rest of the screen. In actual fact, the operation first deletes the element in question and then replots the cell for the symbolic window defined by the deleted element. This alleviates any problem

of partial erasure of remaining elements which can occur if the removed element shares screen area with another element of the same color.

The “pan” operation uses this attribute and the ability to scroll the complete screen to give the designer the ability to move across an entire layout as if flying over the final chip in an airplane at low altitude. This operation is summarized in Fig. 18 for moving the display to the left, although any direction movement is possible. First, the display frame memories are scrolled to the left by 16 pixels and a 16-pixel strip is erased at the right vertical border. The symbolic window is adjusted to include elements that will fall within this right border and the data base is searched. Valid elements are then passed to the screen transformation, clipper, and then to the graphics processor. Results have shown that even complete chip layouts may be scrolled in periods commensurate with human response times (100 mS to 1 S).

10.2.3 Color display

The dimension of color is invaluable to the designer. First, all the interconnection layers are shown in different colors and the appropriate cross-overs are colored to indicate transparency or any other subjective effect which enhances designer feedback. Different contacts

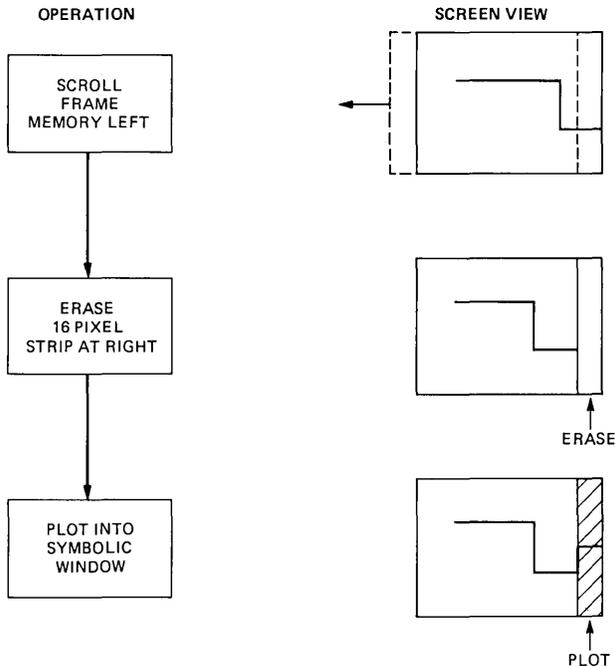


Fig. 18—Flow diagram of the panning operation.

are also coded by their color. Illegal overlaps are designated by eye-catching colors not used in other parts of the layout. In general, a designer may choose the color that he associates with a given layer. This tends to reinforce any recognition tasks that occur. The use of color is also an immense learning aid for novice designers. By giving designers a few rules such as "don't cross green over blue," experience has shown that it takes only a brief familiarization time for candidates to produce useful circuits. Of course, highlighting particular areas of the circuit is trivially accomplished. Even at large-scale factors, when plotting a complete chip the color display yields more information regarding relative wiring densities and placement than storage displays with nominally higher-spatial resolutions. The main reason for this is, despite a lower-spatial resolution (512*480 vs 4096*3000) the higher-color resolution (256 colors) more than compensates. A simple test to show this consists of plotting a monochrome and color layout side by side and reducing the scale until the circuit in each respective case is an unintelligible "blob." Results indicate a factor of four improvement for the color layout.

The use of solid boxes to represent elements alleviates any confusion as to the interior of a box in congested areas. Thus, in conjunction with the "pan" command, a wire may be followed across an entire layout without the designer losing visual perspective, as is the case in more conventional displays.

XI. RESULTS

After using the system, a number of trends in design may be identified. Prior to having the data tablet working, input was primarily via keyboard. At this stage C procedures were used frequently to define circuits. Subsequently, when the data tablet was operational, all cell design was completed at the color terminal with higher-level chip building functions handled by C procedures and *UNIX* operating system shell scripts. In fact, many cells were entered from scratch at the screen, rather than planning a version on paper first. The author has tried interactive input when designing at the mask level and found it to be unsatisfactory without prior planning. However, it would seem that the reduction in complexity afforded by the symbolic notation, and the smooth man-machine interface provided by the editor and color graphics display, yield a truly interactive environment. This leads one to try a number of variations of cells for a particular situation exploring performance and global interface requirements.

The data path of a special purpose 16-bit CMOS processor is shown in Fig. 19. This chip was designed on the MULGA system in 8 man-weeks (without tablet). It contains 5000 transistors in a highly stylized layout modeled on the gate-matrix style.

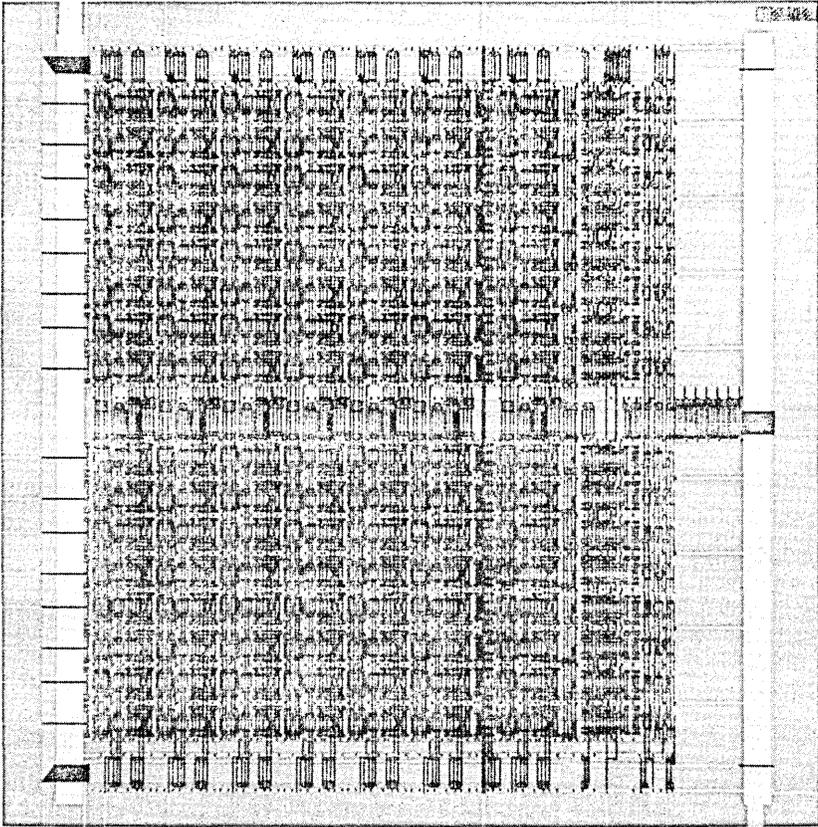


Fig. 19—Five-thousand-transistor CMOS circuit designed with the MULGA system.

In addition to the CMOS chip, two smaller NMOS chips have been designed, each comprising approximately 1000 transistors. These chips were designed in 3 man-weeks.

XII. CONCLUSIONS

We have presented a practical self-contained system to design integrated circuits in the symbolic domain. It is based on a hierarchical circuit description language (ICDL), which describes IC subcircuits in both geometric (physical) and circuit (structural) domains.

The following software supports design at the symbolic level:

- (i) an interactive text and graphics editor based on a high-performance color display,
- (ii) a fast compaction procedure based on the virtual-grid technique,
- (iii) a circuit extraction and parasitic audit program,

(iv) an inbuilt timing stimulator which works at both transistor and gate level,

(v) procedural design in the C programming language.

The efficacy of this system has been demonstrated during the design of chips in both CMOS and NMOS technologies. A predominant reason for the success of the system is the friendly interactive nature that the system provides to designers with a wide range of experience and expertise.

XIII. ACKNOWLEDGMENTS

Bryan Ackland provided many fruitful discussions and ideas in addition to valuable feedback as the first "user" of this system. Steve Law and Alex Lopez provided an introduction to the gate-matrix layout style and an insight into CMOS design techniques. The many discussions with Irene Buchanan and participants of the Caltech Silicon Structures Project about structured design techniques were also valuable to this work. W. H. Ninke provided necessary encouragement and critical input during the development of the system.

REFERENCES

1. G. Persky, D. N. Deutsch, and D. G. Schweikert, "LTX—A Minicomputer-Based System for Automated LSI Layout," *J. Design Automat. Fault-Tolerant Comput.* 1, No. 3 (May 1977), pp. 217-55.
2. C. Mead and L. Conway, *Introduction to VLSI Systems*, Reading, Massachusetts: Addison-Wesley, 1980.
3. D. Gibson and S. Nance, "SLIC—Symbolic Layout of Integrated Circuits," *Proc. 13th Design Automat. Conf.*, June 1976, pp. 434-40.
4. R. P. Larson, "Versatile Mask Generation Techniques For Customer Microelectronic Devices," *Proc. 15th Design Automat. Conf.*, June 1978, pp. 193-8.
5. D. Clary, R. Kirk, and S. Sapiro, "SIDS—A Symbolic Interactive Design System," *Proc. 17th Design Automat. Conf.*, June 1980, pp. 292-5.
6. B. Infante et al., "An Interactive Graphics System for the Design of Integrated Circuits," *Proc. 15th Design Automat. Conf.*, June 1978, pp. 182-7.
7. A. D. Lopez and H. F. Law, "A Dense Gate Matrix Layout Style for MOS LSI," *Proc. ISSSC*, February 1980.
8. J. Williams, "STICKS—A Graphical Compiler for High Level LSI Design," *Proc. 1978 NCC*, May 1978, pp. 289-95.
9. A. Dunlop, "SLIP: Symbolic Layout of Integrated Circuits with Compaction," *Comput. Aided Design*, 10, No. 6 (November 1978), pp. 387-91.
10. A. Dunlop, "SLIM—The Translation of Symbolic Layouts into Mask Data," *Proc. 17th Design Automat. Conf.*, June 1980, pp. 595-602.
11. M. Y. Hsueh and D. O. Pederson, "Computer-Aided Layout of LSI Circuit Building Blocks," *Proc. 1979 Int. Symp. Circuits and Systems*, July 1979, pp. 474-77.
12. N. Weste, "Virtual Grid Symbolic Layout," *Proc. 18th Design Automat. Conf.*, June 1981.
13. I. Buchanan, "Modelling and Verification in Structured Integrated Circuit Design," Ph.D. thesis, University of Edinburgh, Scotland, 1980.
14. S. Trimmerger, "Combining Graphics and Layout in a Single Interactive System," oral presentation, 17th Design Automat. Conf., July 1980.
15. N. Weste and B. Ackland, "GUMBI—A Graphic User Microprogrammable Bit-Slice Interpreter," *Proc. Comcon Fall*, September 1979, pp. 232-7.
16. S. B. Akers, J. M. Geyer, and D. L. Roberts, "IC Mask Layout with a Single

- Conductor Layer," Proc. 7th Design Automat. Workshop, San Francisco, 1970, pp. 7-16.
17. B. R. Chawla, H. K. Gummel, and P. Kozak, "MOTIS—An MOS Timing Simulator," IEEE Trans. Circuits Systems, 22, No. 12 (December 1975), pp. 901-10.
 18. B. Ackland and N. Weste, "Functional Verification in an Interactive Symbolic Design Environment," Proc. 1981 Caltech VLSI Conf., January 1981.

The Effect of Optical Fiber Core and Cladding Diameter on the Loss Added by Packaging and Thermal Cycling

By W. B. GARDNER, S. R. NAGEL, M. I. SCHWARTZ,
F. V. DIMARCELLO, C. R. LOVELACE, D. L. BROWNLOW,
M. R. SANTANA, and E. A. SIGETY

(Manuscript received December 19, 1980)

We conducted an experimental investigation of the effect of fiber cladding and core diameters on the loss added by packaging and thermal cycling. The results confirm the general trends predicted by theory and indicate that fibers with 125- μm claddings and 50- μm cores should have about three times less microbending loss than 110- μm /55- μm fibers. These results contributed to the adoption of the 125- μm /50- μm dimensions for the FT3 metropolitan trunk lightwave system.

I. INTRODUCTION

The Bell System's Atlanta fiber system experiment¹ in 1976 and Chicago lightwave communications project^{2,3} in 1977 to 1979 both used optical fibers with 110- μm cladding diameter and 55- μm core diameter. In 1979, an experimental investigation was conducted into the effect of core and cladding diameter on the loss added by packaging and thermal cycling. The results of that study are reported here. These results contributed to the adoption of 125- μm /50- μm fiber dimensions for the Bell system's new FT3 metropolitan trunk lightwave system.^{4,5}

The conflicting requirements on the core diameter (a) and the cladding diameter (d) are summarized in Table I.

This study is directed at items 1 and 2 of Table I. Items 3 (at least for lasers) and 4 are somewhat more difficult to quantify, and will not be addressed here.

Table I—Requirements on dimensions

| Design Consideration | Dictates that <i>a</i> be | Dictates that <i>d</i> be |
|-------------------------------|---------------------------|---------------------------|
| 1. Splice Loss | large | — |
| 2. Microbending Loss | small | large |
| 3. Source Coupling Efficiency | large | — |
| 4. Cost | small | small |

II. THEORY

2.1 Splice loss

When offset dominates over fiber parameter mismatch, the splice loss γ_s (based on Fig. 9 of Reference 6) obeys to first order

$$\gamma_s \propto (c/a)^{1.5}, \tag{1}$$

where *c* is the transverse offset, *a* is the core diameter, and $c/a \approx 0.4$. Equation (1) gives

$$\gamma_s(a)/\gamma_s(55 \mu\text{m}) = [(55)/a]^{1.5}, \tag{2}$$

where *a* is in μm . Equation (2) is referred to as the predicted splice loss factor in Table II.

2.2 Microbending loss

Using Olshansky's⁷ model for the microbending loss γ_m resulting from fiber packaging,

$$\gamma_m \propto a^4/d^6, \tag{3}$$

where *a* is the core diameter and *d* is the cladding diameter. From this, the predicted microbending loss factor in Table II is calculated,

Table II—Summary of ribbon performance

| Clad Dia. (μm) | Core Dia. (μm) | Clad Core | Ribbon Number | Mean Ribbon Minus Fiber Loss (dB/km) | | Measured Ribbing Loss Factor* (Normalized to 110/55) | | Measured -43°F Loss Factor* (normalized) (from Table III) | Predicted Microbending Loss Factor Eq. (4) | Predicted Splice Loss Factor Eq. (2) |
|-----------------------------|-----------------------------|-----------|---------------|--------------------------------------|---------|--|---------|---|--|--------------------------------------|
| | | | | 1 hr | 100 hrs | 1 hr | 100 hrs | | | |
| 110 | 55 | 2.0 | 84 | 2.63 | 2.24 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 110 | 44 | 2.5 | 81 | 0.79 | 0.70 | 0.30 | 0.31 | 0.30 | 0.41 | 1.40 |
| 110 | 37 | 3.0 | 88 | 1.52 | 0.95 | 0.58 | 0.42 | — | 0.21 | 1.81 |
| 90 | 36 | 2.5 | 89 | 1.77 | 1.60 | 0.67 | 0.71 | — | 0.61 | 1.89 |
| 90 | 30 | 3.0 | 87 | 1.56 | 1.96 | 0.59 | 0.88 | — | 0.30 | 2.48 |
| 125 | 50 | 2.5 | 82 | 0.95 | 1.04 | 0.36 | 0.46 | 0.18 | 0.32 | 1.15 |
| 125 | 42 | 3.0 | 90 | 0.60 | 0.52 | 0.23 | 0.23 | — | 0.16 | 1.50 |

* Normalized to the 110/55 case (Ribbon 84).

$$\gamma_m(a, d)/\gamma_m(55,110) = (a/55)^4/(d/110)^6, \quad (4)$$

where a and d are in μm .

Calculations indicate that based on packaging and splicing losses, the repeater spacing for the Bell system's FT3 laser system would be optimized for clad/core diameter ratios (d/a) in the range of 2.5 to 3.0. These calculations closely parallel those of Murata et al.,⁸ who conclude that the optimum d/a for their laser-based system is 2.5, and for a light emitting diode (LED)-based system, between 1.8 and 2.0.

III. FIBER PREPARATION

To develop confidence in these paper studies, preforms were prepared having d/a ratios of 2.0, 2.5, and 3.0. Standard modified chemical vapor deposition (MCVD) processing techniques⁹ were used to fabricate the germanium borosilicate 0.23 N.A. graded-index preforms. After depositing three barrier layers of fused silica, the SiCl_4 , BCl_3 , and additional O_2 levels were held constant while linearly varying the GeCl_4 concentration during deposition of the core.

For the 2.0 clad/core ratio preform fabrication, 50 passes were used with slight adjustments in the number of passes to compensate for tube cross-sectional area variations. The number of passes had to be reduced to ~ 28 for fabrication of a 2.5 clad/core preform at equivalent traverse speeds and tube sizes. This resulted in an equivalent layer thickness ($\mu\text{m}/\text{pass}$) in the resultant preform. For the 3.0 clad/core preforms, the traverse speed was slightly increased to maintain a reasonable number of grading passes without a marked reduction in deposited layer thickness. The effect of layer thickness on loss is not well understood.

After deposition each tube was collapsed in similar fashion to form a preform. A reverse collapse procedure applying internal pressure was used.

Approximately 1 km of optical fiber was drawn from each preform to 110- μm cladding diameter and coated with silicone. Figure 1 shows the distribution of 820-nm losses in these fibers. The remaining portions of the preforms were drawn to 90-, 110-, and 125- μm cladding diameters and coated with a polyurethane acrylate. The loss trends with this coating were similar to Fig. 1 and suggest that fibers with higher clad/core ratios are less sensitive to process-induced loss. For all values of d , the nominal coating outer diameter was 229 μm .

IV. PACKAGING LOSS

For each of the seven combinations of fiber dimensions shown in Table II, one 12-fiber adhesive sandwich ribbon,¹⁰ approximately

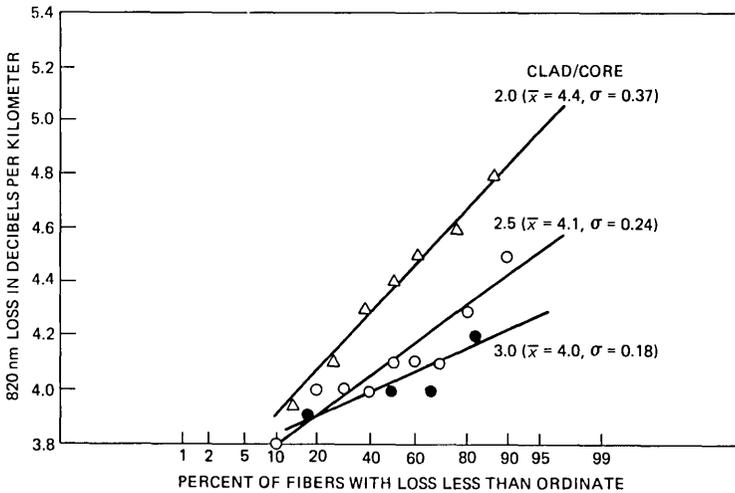


Fig. 1—Impact of fiber dimensions on loss. Distribution of losses in unpackaged silicone coated fibers with 110- μ m cladding diameter.

1-km-long, was made. Since the silicone coating was not compatible with this process, only the acrylate coating was used in the packaging loss experiment. Before making the ribbons, spectral loss was measured¹¹ with the fibers laid in a 58-cm-diameter container to eliminate added loss caused by reel winding. For these measurements, the fiber core (regardless of diameter) was overfilled, and the leaky and higher-order modes were eliminated by winding five turns of the fiber around a 13-mm-diameter mandrel. For each preform used, the core diameter and refractive index profile were checked.

The 633-nm loss of each fiber in each ribbon was measured about one hour after making the ribbon and again about 100 hours later. (Some change occurs over this time because of relaxation in the ribbon structure.) The microbending loss at 633 nm, because of the ribbon making, is shown in Table II for each ribbon as “mean ribbon minus fiber loss.” These values are then normalized to the 110/55 clad/core case under the heading “Measured Ribboning Loss Factor.”

V. THERMAL CYCLING

After determining the added loss in the ribbons, three of the seven ribbons were laid in 58-cm-diameter containers and placed in an environmental chamber. The mean 633-nm loss for the 12 fibers in each ribbon was determined at 75°F and entered as “baseline” in Table III. The remaining values in Table III are changes from this baseline loss after the temperature-time exposure indicated. The changes at the final low-temperature exposure (−43°F) are then nor-

Table III—633 nm loss change with temperature (dB/km)

| Temperature (°F) | Exposure Time (h) | Clad/Core → Clad Dia. → | 2/1 | 2.5/1 | 2.5/1 |
|------------------|-------------------|----------------------------|---------------------|---------------------|---------------------|
| | | | 110 μm Ribbon 84 | 110 μm Ribbon 81 | 125 μm Ribbon 82 |
| 75 | | Baseline | 12.33 ± 0.03 | 12.31 ± 0.10 | 11.78 ± 0.06 |
| -45 | 48 | | +1.09 | -0.07 | -0.09 |
| -15 | 48 | | +0.48 | -0.09 | -0.20 |
| 20 | 48 | | -0.03 | -0.14 | -0.12 |
| 75 | 48 | | -0.10 | 0.00 | -0.05 |
| 190 | 816 | | 0 | +0.59 | +0.57 |
| 75 | 48 | | +0.02 | +0.30 | +0.23 |
| -45 | 48 | | +1.04 | +0.21 | +0.06 |
| -9 | 48 | | +0.45 | +0.12 | 0 |
| 15 | 48 | | +0.11 | +0.11 | +0.19 |
| 75 | 48 | | -0.04 | +0.26 | +0.24 |
| 190 | 1104 | | -0.01 | +0.61 | +0.61 |
| 75 | 48 | | +0.07 | +0.31 | +0.28 |
| -40 | 48 | | +1.33 | +0.35 | +0.19 |
| -15 | 48 | | +0.68 | +0.29 | +0.13 |
| 75 | 72 | | -0.18 | +0.31 | +0.26 |
| 190 | 2376 | | +0.07 | +0.72 | +0.76 |
| 75 | 48 | | +0.33 | +0.59 | +0.42 |
| -43 | 48 | | +2.10 | +0.64 | +0.38 |
| -15 | 48 | | +1.29 | +0.42 | +0.24 |
| +15 | 48 | | +0.52 | +0.36 | +0.27 |
| 75 | 48 | | +0.03 | +0.38 | +0.36 |

malized to the 110/55 ribbon's change at -43°F and recorded in Table II as the "measured -43°F loss factor."

VI. DISCUSSION AND CONCLUSIONS

The microbending loss factor defined by eq. (4) and entered in the second column from the right in Table II is supposed to indicate the relative expected susceptibility of the various fibers to microbending loss. Two occasions when microbending loss appears are (i) in the packaging of the fibers into a ribbon, and (ii) the temperature cycling of the ribbon. The measured susceptibility of the various fibers in those two instances is indicated by the measured ribboning loss factor and the measured -43°F loss factor, respectively, in Table II. In general, these values follow the trends predicted by eq. (4) as a and d vary. Except for the cases of $d/a = 3.0$, where the ribboning loss was consistently larger than predicted, numerical agreement between theory and experiment is good. Although the coatings and ribbons were nominally the same throughout, some randomness in coating roughness, modulus, and ribbon structure undoubtedly weakened the correlation between theory and experiment. Nevertheless, the general trends of loss versus cladding and core diameters are confirmed. Based on these results, the 125/50 fiber is expected to provide about a factor of three, less microbending loss, than the previously used 110/55 fiber. From the last column in Table II, this change in dimensions should

sacrifice only about 15 percent in splice loss. This is the lowest splice loss factor of any of the alternatives to 110/55 investigated here. Thus, the 125/50 combination appears to be a good compromise between microbending loss and splice loss. Since there is more glass per unit length in the 125- μm fiber, its cost will be higher. Where cost is critical, other dimensions will probably be preferred. Thus, it is important that the adoption of the 125/50 dimensions for FT3 not preclude the eventual use of other dimensions as dictated by future needs.

VII. ACKNOWLEDGMENTS

We would like to thank L. L. Blyler, A. C. Hart, and V. A. Foertmeyer for fiber drawing, C. M. Bane for making most of the ribbons, T. W. Givens and L. M. Boggs for measurements, and C. M. Melliar-Smith for helpful discussions.

REFERENCES

1. For articles on the Atlanta system experiment, see B.S.T.J., 57, No. 6 (July–August 1978), Part 1.
2. M. I. Schwartz, W. A. Reenstra, J. H. Mullins, and J. S. Cook, "The Chicago Lightwave Communications Project," B.S.T.J., 57, No. 6 (July–August 1978), pp. 1881–8.
3. T. C. Cannon, D. L. Pope, and D. D. Sell, "Installation and Performance of the Chicago Lightwave Transmission System," IEEE Trans. Commun., COM-26, No. 7 (July 1978), pp. 1056–60.
4. I. Jacobs, "Lightwave Communications Begins Regular Service," Bell Labs Record, 57, No. 11 (December 1979), pp. 298–304.
5. I. Jacobs, "Lightwave Communications—Yesterday, Today, and Tomorrow," Bell Labs Record, 58, No. 1 (January 1980), pp. 2–10.
6. C. M. Miller and S. C. Mettler, "A Loss Model for Parabolic-Profile Fiber Splices," B.S.T.J., 57, No. 9 (November 1978), pp. 3167–80.
7. R. Olshansky, "Distortion Losses in Cabled Optical Fibers," Appl. Opt., 14, No. 1 (January 1975), pp. 20–1.
8. H. Murata, S. Inao, Y. Matsuda, and T. Kuroha, "Optimum Design for Optical Fiber Used in Optical Cable System," Proceedings of Fourth European Conference on Optical Communication, Genoa, Italy (September 12–15, 1978), pp. 242–8.
9. F. V. DiMarcello and J. C. Williams, "Reproducibility of Optical Fibers Prepared by a Modified Chemical Vapor Deposition Process," B.S.T.J., 57, No. 6 (July–August 1978), pp. 1723–34.
10. M. J. Saunders and W. L. Parham, "Adhesive Sandwich Optical Fiber Ribbons," B.S.T.J., 56, No. 6 (July–August 1977), pp. 1013–4.
11. L. Wilson and M. J. Buckler, "Automated Spectral Loss Measurement of Optical Fibers," Optics and Laser Technology, 10, No. 4 (August 1978), pp. 197–9.

Analysis of Maximum-Likelihood Sequence Estimation Performance for Quadrature Amplitude Modulation

By A. S. ACAMPORA

(Manuscript received December 10, 1980)

This paper considers maximum-likelihood sequence estimation (MLSE) for quadrature amplitude modulation (QAM) signaling at rates approaching several baud/Hz. In this regime, intersymbol interference and possibly cross channel coupling are the dominant transmission impairments. We derive the structure of a detector that optimally accommodates both impairments. A bit error rate performance bound is found, and the concept of an error state transition matrix is introduced to facilitate the analysis. We explore a modulation scheme wherein cross-channel coupling is intentionally introduced, and find that it improves detection efficiency. The use of MLSE may be an important consideration for power and spectrally efficient digital radio systems, either terrestrial or satellite, since rates approaching the Shannon limit may be attainable without channel coding, and frequency selective fading is handled in an optimum manner.

I. INTRODUCTION

The search for digital radio modulations which combine power and bandwidth efficiency with ease of implementation has attracted interest for many years.¹⁻⁹ Combined with the performance analysis of receivers which represent a compromise between optimality and practicality, this field remains the focus of much current research activity.¹⁰⁻¹⁵ Quadrature amplitude modulation (QAM) is a particularly simple modulation to implement, and determination of the theoretical bit-error rate (BER) achievable for this modulation at an arbitrary rate of information transfer per unit bandwidth is the subject of this paper. Phase shift keying (PSK) is, of course, a special case of QAM.

The primary impairment to QAM transmission is intersymbol interference caused by bandlimiting at the transmitter and dispersion in

the channel. We investigate maximum-likelihood sequence estimation (MLSE) to optimally detect the impaired signal, which is also corrupted by additive Gaussian noise. In particular, we analyze the Viterbi algorithm for realizing MLSE and then proceed to find an upper bound for the BER. In this regard, the analysis to be presented simplifies, unifies, and expands earlier treatments of the same subject.^{16,17} The primary goal is to obtain a basis against which the performance of any suboptimum receiver may be compared. In so doing, we show QAM to be capable of providing performance close to the Shannon limit, an observation of considerable importance for communication satellite and terrestrial digital radio applications. Additionally, study of the mechanisms responsible for MLSE error generation provides insight into appropriate waveshaping to improve BER performance. One such waveshaping technique, wherein the spectrum is intentionally asymmetric with respect to the transmitting filter passband (e.g., QAM-Single Sideband), is shown to generally improve the BER performance.

In Section II, motivation behind a study of QAM is illustrated, the QAM model is presented, and the MLSE algorithm is derived. Section III is devoted to the derivation of a BER outer bound, and in Section IV, this bound is applied to examples which demonstrate the inherent power of MLSE. We show that a BER of 10^{-4} can be maintained at a transmission rate of 5 bits/s/Hz with an energy per bit penalty no greater than 1 dB compared against ideal nonoverlapping rectangular signaling; a four-pole Butterworth transmit filter is assumed in this calculation.

II. QAM MODEL AND THE MLSE ALGORITHM

Figure 1a gives a general model of a digital communication system. Here, the source produces a stream of binary data $\mathbf{d} = \{d_0, d_1, \dots, d_L\}$ which is converted into an analog waveform by the modulator for transmission to the destination. The modulator may include a band-limiting filter to contain spectral emission. White Gaussian noise is added in the channel, and it is the function of the receiver to reproduce a copy $\hat{\mathbf{d}}$ of the original data sequence with a reasonably low probability of bit error.

The output of the modulator can be represented by the expression

$$S_G(t) = \text{Re} \{ [A(t, \mathbf{d}) + jB(t, \mathbf{d})] e^{-j\omega_0 t} \}, \quad (1)$$

where ω_0 is the carrier radian frequency and $A(t, \mathbf{d})$ and $B(t, \mathbf{d})$ are data-dependent real waveforms.

The form of $S_G(t)$ provides great flexibility to achieve both high power efficiency and high spectral efficiency. However, the modulator to generate the best $S_G(t)$ and the receiver to detect the data in the presence of corrupting noise may be quite difficult to implement, e.g.,

separate waveform synthesizers and receiver matched filters may be required for each of the 2^L possible outcomes of the data stream \mathbf{d} .

To simplify the modulation and detection processes, we consider QAM, a special subset of (1) having the form

$$S(t) = \text{Re} \left\{ \sum_{k=0}^N [a_k(\mathbf{d}) + jb_k(\mathbf{d})] \bar{h}(t - kT) e^{-j\omega_0 t} \right\}, \quad (2)$$

where

$\bar{h}(t) = h_R(t) + jh_I(t)$ is a complex waveshape,

a_k and b_k are real numbers dependent upon the data,

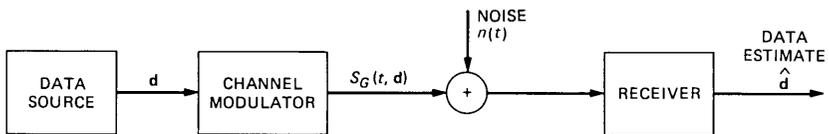
T^{-1} is the channel signaling rate,

and

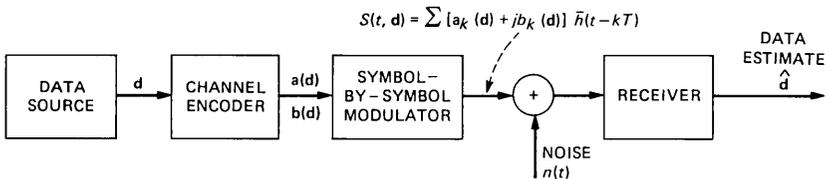
$N + 1$ is the number of complex channel symbols used to transmit the data \mathbf{d} .

Equation (2) can readily be modified to admit the familiar staggered types of modulation for which the in-phase and quadrature pulses are offset from each other by the amount $T/2$.^{6,9} Although MLSE can be applied to this case, we restrict our attention in the following to nonstaggered modulation. We expect that performance similar to what we report here is achievable with staggered modulation as well.

The waveform $S(t)$ may be generated as shown in Fig. 1b. Here, the numbers a_k and b_k are generated by a channel encoder and the



(a)



(b)

Fig. 1—Models for digital communication channels. (a) General digital channel. (b) Symbol-by-symbol QAM digital channel.

numbers $a_k, b_k, a_{k+j}, b_{k+j}$ can be dependent. In the subsequent development we assume that, for $j \neq 0$, the pairs (a_k, b_k) and (a_{k+j}, b_{k+j}) are independent, corresponding to independent symbol-by-symbol signaling. The MLSE algorithm to be derived can be generalized to optimally detect \mathbf{d} when the numbers a_k and b_k are dependent only upon the most recent K data symbols supplied to the encoder. A convolutional encoder is one such case which has been treated for real $h(t)$ and binary signaling.¹⁸

We note that the numbers a_k and b_k may be dependent. Thus, our treatment admits the familiar PSK type of modulation. Further, with no loss in generality, we assume in the following that $\bar{h}(t)$ is the combined impulse response of the transmitter and a possibly dispersive channel.

At the receiver, we observe

$$R(t) = S(t) + \text{Re} \{ [n_1(t) + jn_2(t)] e^{-j\omega_0 t} \}, \quad (3)$$

where $n_1(t)$ and $n_2(t)$ are independent white Gaussian noise processes, each having two-sided spectral density $N_0/2$. Thus, the in-phase and quadrature components of $R(t)$ are as follows:

$$\begin{aligned} \mathcal{I}(t) &= \sum_{k=0}^N [a_k h_R(t - kT) - b_k h_I(t - kT)] + n_1(t) \\ &\triangleq S_R(t) + n_1(t) \end{aligned} \quad (4)$$

and

$$\begin{aligned} \mathcal{Q}(t) &= \sum_{k=0}^N [a_k h_I(t - kT) + b_k h_R(t - kT)] + n_2(t) \\ &\triangleq S_I(t) + n_2(t). \end{aligned} \quad (5)$$

Let

$$\mathbf{a} = \{a_0, a_1, \dots, a_N\}, \quad \mathbf{b} = \{b_0, b_1, \dots, b_N\}. \quad (6)$$

Then, since the noise is white and Gaussian, the optimum receiver selects as detected data that sequence pair (\hat{a}, \hat{b}) for which the likelihood function,

$$L(\mathbf{a}, \mathbf{b}) = \exp\left(-\frac{2}{N_0} \int ([\mathcal{I}(t) - S_R(t)]^2 + [\mathcal{Q}(t) - S_I(t)]^2) dt\right), \quad (7)$$

is maximized. This is equivalent to maximizing, with respect to the hypothesis (\mathbf{a}, \mathbf{b}) , the quantity

$$\begin{aligned} \Lambda(\mathbf{a}, \mathbf{b}) &= 2 \int_{-\infty}^{\infty} [\mathcal{I}(t) S_R(t; \mathbf{a}, \mathbf{b}) + \mathcal{Q}(t) S_I(t; \mathbf{a}, \mathbf{b})] dt \\ &\quad - \int_{-\infty}^{\infty} [S_R^2(t; \mathbf{a}, \mathbf{b}) + S_I^2(t; \mathbf{a}, \mathbf{b})] dt. \end{aligned} \quad (8)$$

From (4) and (5),

$$\begin{aligned}
 \Lambda(\mathbf{a}, \mathbf{b}) &= 2 \sum_{k=0}^N \int_{-\infty}^{\infty} \mathcal{I}(t) [a_k h_R(t - kT) - b_k h_I(t - kT)] dt \\
 &\quad + 2 \sum_{k=0}^N \int_{-\infty}^{\infty} \mathcal{Q}(t) [a_k h_I(t - kT) + b_k h_R(t - kT)] dt \\
 &\quad - \sum_{k=0}^N \sum_{m=0}^N \int_{-\infty}^{\infty} [a_k h_R(t - kT) - b_k h_I(t - kT)] \\
 &\quad \quad \cdot [a_m h_R(t - mT) - b_m h_I(t - mT)] dt \\
 &\quad - \sum_{k=0}^N \sum_{m=0}^N \int_{-\infty}^{\infty} [a_k h_I(t - kT) + b_k h_R(t - kT)] \\
 &\quad \quad \cdot [a_m h_I(t - mT) + b_m h_R(t - mT)] dt. \quad (9)
 \end{aligned}$$

Let

$$\begin{aligned}
 \alpha_k &= \int_{-\infty}^{\infty} \mathcal{I}(t) h_R(t - kT) dt, & \beta_k &= \int_{-\infty}^{\infty} \mathcal{I}(t) h_I(t - kT) dt, \\
 \gamma_k &= \int_{-\infty}^{\infty} \mathcal{Q}(t) h_R(t - kT) dt, & \rho_k &= \int_{-\infty}^{\infty} \mathcal{Q}(t) h_I(t - kT) dt. \quad (10)
 \end{aligned}$$

Then, the quantities in (10) can be generated by processing both $\mathcal{I}(t)$ and $\mathcal{Q}(t)$ through filters matched to $h_R(t)$ and $h_I(t)$, and sampling the output at time $t = kT$.

Let

$$y_k = \alpha_k + \rho_k, \quad z_k = \gamma_k - \beta_k. \quad (11)$$

Finally, let us define

$$\chi_{k-m} = \int_{-\infty}^{\infty} [h_R(t - kT) h_R(t - mT) + h_I(t - kT) h_I(t - mT)] dt, \quad (12)$$

$$\zeta_{k-m} = \int_{-\infty}^{\infty} [h_I(t - kT) h_R(t - mT) - h_R(t - kT) h_I(t - mT)] dt. \quad (13)$$

Then

$$\Lambda(\mathbf{a}, \mathbf{b}) = 2 \sum_{k=0}^N (a_k y_k + b_k z_k) - \sum_{k=0}^N \sum_{m=0}^N (a_k a_m + b_k b_m) \chi_{k-m} - \sum_{k=0}^N \sum_{m=0}^N (a_k b_m - b_k a_m) \zeta_{k-m}. \quad (14)$$

Thus, for each hypothesis under test, the maximum-likelihood receiver forms a linear combination of the received statistics from which is subtracted a stored constant which is independent of the received data. (This constant must be slowly updated if the channel frequency characteristics change as during frequency-selective fading. Our analysis applies to each instantaneous characteristic since the detection epoch is far smaller than the time intervals associated with such changing characteristics.)

We note that the matched filtering operation can be performed in the passband, and that the received statistics y_k and z_k are, respectively, the in-phase and quadrature samples of the passband matched filter response to $\bar{h}(t)$,

$$\bar{v}(t) = \frac{1}{2\pi} \int |H(\omega)|^2 e^{j\omega t} d\omega = \int \bar{h}(\tau) \bar{h}^*(\tau - t) d\tau. \quad (15)$$

Then

$$\bar{v}(nT) = \chi_n + j\zeta_n, \quad (16)$$

and we see that the quantities χ_n and ζ_n are the real and imaginary samples of the matched filter response to $\bar{h}(t)$ at time nT . These are, then, real and imaginary components of the intersymbol interference. From (12) and (13),

$$\chi_{-n} = \chi_n, \quad \zeta_{-n} = -\zeta_n, \quad \zeta_0 = 0. \quad (17)$$

In the following, we assume that the intersymbol interference vanishes for $|n| > M$, a positive integer.

For QAM, we can use the Viterbi algorithm to maximize (14) with respect to (\mathbf{a}, \mathbf{b}) . Rewriting (14),

$$\Lambda(\mathbf{a}, \mathbf{b}) = \sum_{k=0}^N a_k \left[2y_k - a_k \chi_0 - 2 \sum_{m=k-M}^{k-1} (a_m \chi_{k-m} + b_m \zeta_{k-m}) \right] + \sum_{k=0}^N b_k \left[2z_k - b_k \chi_0 - 2 \sum_{m=k-M}^{k-1} (b_m \chi_{k-m} - a_m \zeta_{k-m}) \right], \quad (18)$$

where the convention is adopted that a_k and b_k are identically zero for $k < 0$.

Let the partial sum $\Lambda_n(\mathbf{a}, \mathbf{b})$ be the first n terms in each of the sums appearing in eq. (18). Then

$$\Lambda_n(\mathbf{a}, \mathbf{b}) = \Lambda_{n-1}(\mathbf{a}, \mathbf{b}) + a_n \left[2y_n - a_n\chi_0 - 2 \sum_{m=n-M}^{n-1} (a_m\chi_{n-m} + b_m\zeta_{n-m}) \right] + b_n \left[2z_n - b_n\chi_0 - 2 \sum_{m=n-M}^{n-1} (b_m\chi_{n-m} - a_m\zeta_{n-m}) \right]. \quad (19)$$

Equation (19) is in a form such that the Viterbi algorithm can be directly applied. We see that to calculate $\Lambda_n(\mathbf{a}, \mathbf{b})$, we must add to $\Lambda_{n-1}(\mathbf{a}, \mathbf{b})$ a term which depends on the received data y_n and z_n and the hypothesis subvectors (a_{n-M}, \dots, a_n) and (b_{n-M}, \dots, b_n) . Furthermore, the next calculation to find $\Lambda_{n+1}(\mathbf{a}, \mathbf{b})$ no longer depends on a_{n-M} and b_{n-M} . Thus, we can perform MLSE by means of a trellis diagram containing I^M states, where I is the number of discrete values which can be assumed by each symbol pair (a_n, b_n) . Let us label each state by the $2M$ -tuple $a_{j-M+1}, \dots, a_j, b_{j-M+1}, \dots, b_j$. Then, permissible state transitions take the form shown in Fig. 2, drawn for (a_n, b_n) independent and binary (hence, $I = 4$) and $M = 2$. For each permissible transition, we can compute the branch metric $\Lambda_j(\mathbf{a}, \mathbf{b}) - \Lambda_{j-1}(\mathbf{a}, \mathbf{b})$ [see eq. (19)].

Maximum-likelihood sequence estimation is performed via recursive application of (19) for $n = 0, 1, \dots, j, \dots, N$. Suppose the partial sum of the most likely path leading into each state for $n = j - 1$ is known. We calculate the partial sum of the I competing paths leading into each state for $n = j$ by adding the partial sum of the most likely path into each state at $n = j - 1$ to the branch metric corresponding to that transition. The largest of the partial sums among the paths merging at each state then corresponds to the most likely path leading to that state and is stored for future calculations. The hypothesis (a_{j-M}, b_{j-M}) corresponding to the surviving path at each state is also stored as part of the most likely path leading to that state since a_{j-M}, b_{j-M} does not affect subsequent calculations. Thus, at each node, $I - 1$ out of I possible paths are deleted from further consideration. The path through the trellis with the greatest metric Λ_N then identifies the most likely transmitted sequence $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$.

Unlike the case of the general digital modulation of (1), the complexity of the ML receiver for QAM grows exponentially with M (the one-sided extent of the ISI) rather than exponentially with the message length.

III. BIT-ERROR RATE PERFORMANCE BOUND

To find an upper bound on the BER performance of MLSE, we invoke the concept of an error event. Consider an incorrect path which diverges from the correct path at depth p in the trellis and remerges with it for the first time at depth q . Since the correct and incorrect paths agree over the first p epochs, the difference in path metrics at the remerge point is given by

$$\Lambda_q - \Lambda'_q = \sum_{n=p+1}^q (a_n - a'_n) \left[2y_n - (a_n - a'_n)\chi_0 \right. \\ \left. - 2 \sum_{m=n-M}^{n-1} \{(a_m - a'_m)\chi_{n-m} + (b_m - b'_m)\zeta_{n-m}\} \right] \\ + \sum_{n=p+1}^q (b_n - b'_n) \left[2z_n - (b_n - b'_n)\chi_0 \right. \\ \left. - 2 \sum_{m=n-M}^{n-1} \{(b_m - b'_m)\chi_{n-m} - (a_m - a'_m)\zeta_{n-m}\} \right], \quad (20)$$

where

Λ_q is the metric of the correct path,

Λ'_q is the metric of the remerging incorrect path,

a_n, b_n are the channel symbols along the correct path,

and

a'_n, b'_n , are the channel symbols along the incorrect path.

Let

$$E_n = \frac{1}{2} (a_n - a'_n), \quad (21)$$

$$F_n = \frac{1}{2} (b_n - b'_n). \quad (22)$$

Then, substituting (10) and (11) into (20), and after extensive simplification (see Ref. 18 for a similar derivation), we obtain the result

$$\Lambda_q - \Lambda'_q = D + n_{eq}, \quad (23)$$

where

$$D = \sum_{n=p+1}^q E_n \left[E_n + 2 \sum_{m=n-M}^{n-1} (E_m \chi_{n-m} - F_m \zeta_{n-m}) \right] \\ + \sum_{n=p+1}^q F_n \left[F_n + 2 \sum_{m=n-M}^{n-1} (F_m \chi_{n-m} + E_m \zeta_{n-m}) \right] \quad (24)$$

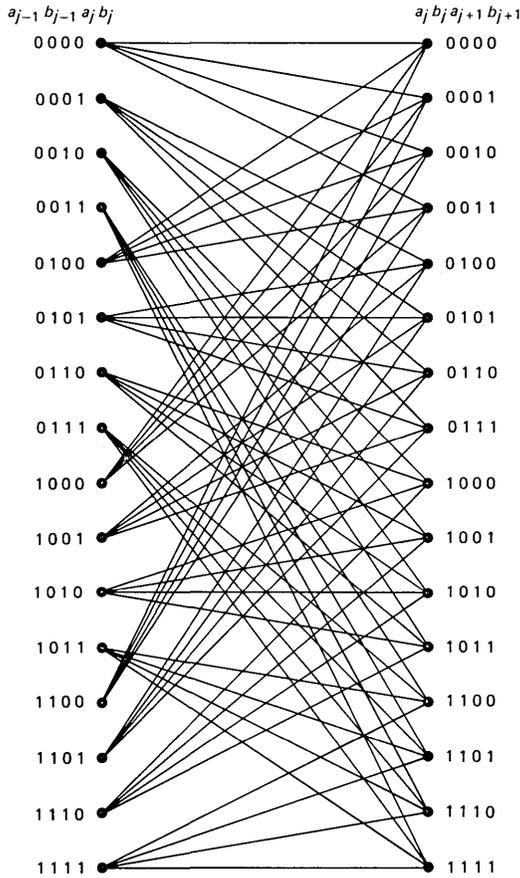


Fig. 2—State transition diagram for four-level QAM with four symbols of ISI.

and n_{eq} is a zero-mean Gaussian noise random variable with variance

$$\sigma^2 = DN_0/2. \quad (25)$$

In (24), we have assumed that the energy contained in $h(t)$ is unity.

At the q th epoch, the MLSE receiver must choose a survivor from among that path agreeing with the correct path over the most recent M epochs and all other incorrect paths remerging with it. An error will be committed if an incorrect path is chosen. Since the correct path may have been deleted by a prior decision, the metric comparisons may be between the remerging incorrect path and a partially correct path with metric greater than that of the overall correct path. The probability of committing an error at the q th epoch by selecting a particular one of the remerging incorrect paths of length $q - p$ is, then, overbounded for a particular correct sequence by

$$P_p \leq P\{\Lambda_q - \Lambda'_q < 0\}. \quad (26)$$

Recognizing that the true energy contained in $h(t)$ is $\chi_0 \triangleq e_b$, we now normalize the intersymbol interference coefficients by dividing each by χ_0 to obtain the result

$$P_p \leq Q(\sqrt{2e_b D/N_0}), \quad (27)$$

where $Q(\cdot)$ is the complementary error function.

Let N_p be the number of incorrect paths which diverge from the correct path at the p th epoch and remerge, for the first time, at the q th epoch, and let $\mathcal{E}_{p,j}$ be an error event associated with one such path. Then the probability of committing an error at the q th epoch is overbounded, for a particular correct path, by

$$P_E \leq P\left\{\bigcup_{p=1}^q \bigcup_{j=1}^{N_p} \mathcal{E}_{p,j}\right\}. \quad (28)$$

Applying the union bound,

$$P_E \leq \sum_{p=1}^q \sum_{j=1}^{N_p} P\{\mathcal{E}_{p,j}\}, \quad (29)$$

where, for a particular event $\mathcal{E}_{p,j}$, $P\{\mathcal{E}_{p,j}\}$ is the probability that $\Lambda_q - \Lambda'_q < 0$.

Let $\mu_{p,j}$ be the number of bit errors associated with the event $\mathcal{E}_{p,j}$. Then, the expected BER for decisions made at the q th epoch, conditioned on a particular correct sequence, is overbounded by:

$$P_B \leq \sum_{p=1}^q \sum_{j=1}^{N_p} \frac{\partial}{\partial \kappa} [\kappa^{\mu_{p,j}} P\{\mathcal{E}_{p,j}\}] \Big|_{\kappa=1}. \quad (30)$$

The particular form chosen for (30) will be useful later.

Finally, we note that the correct path is one of I^{q-p} outcomes over the unmerged span between epochs p and q . The result (30) must, therefore, be averaged over these outcomes to remove the conditional dependence on the correct path.

Thus, evaluation of the upper bound involves a weighted summation of the probabilities of all error events merging with the correct path at the q th node. We will now develop a technique for performing this summation in an orderly fashion.

We note from (25) that for every error event, $D > 0$. Also, for each error event, D is dependent only upon the error sequences (E_n, F_n) , $p < n \leq q$. For one particular error event, the quantity D will assume its minimum value D_{min} . Then, using (27) and the inequality¹ $Q(\sqrt{x+y}) \leq (\sqrt{x})e^{-y/2}$ for $x > 0, y > 0$, we obtain

$$P_B \leq Q\left(\sqrt{\frac{2e_b D_{min}}{N_0}}\right) \exp\{e_b D_{min}/N_0\} \cdot \sum_{p=1}^q \sum_{j=1}^{N_p} \frac{\partial}{\partial \kappa} [\kappa^{\mu_{p,j}} \exp\{-e_b D_{p,j}/N_0\}]. \quad (31)$$

Hence, evaluation of the upper bound involves computation of the set of Euclidean distances $D_{p,j}$ for each correct path. The error-state transition matrix, to be described, provides a systematic way to find these distances, average over correct paths, and perform the sum.

Referring to eq. (24), we see that the contribution to the distance between a correct and an incorrect path incurred between nodes j and $j + 1$, where $p \leq j \leq q$, is given by

$$d = \sqrt{2e_b/N_0} E_j \left[E_j + 2 \sum_{m=j-M}^{j-1} (E_m \chi_{j-m} + F_m \zeta_{j-m}) \right] + \sqrt{2e_b/N_0} F_j \left[F_j + 2 \sum_{m=j-M}^{j-1} (F_m \chi_{j-m} - E_m \zeta_{j-m}) \right]. \quad (32)$$

This contribution can be represented as the distance incurred by a transition from error state $E_{j-M}, E_{j-M+1}, \dots, E_{j-1}, F_{j-M}, F_{j-M+1}, \dots, F_{j-1}$ to error state $E_{j-M+1}, E_{j-M+2}, E_j, F_{j-M+1}, F_{j-M+2}, \dots, F_j$. A particular pair of elements (E_n, F_n) from an error sequence may assume no more than $J \leq I(I - 1) + 1$ discrete values. For example, if a_n and b_n are binary and independent, then the pair (E_n, F_n) may assume the nine values: (0, 0), (0, -1), (0, 1), (1, -1), (1, 0), (1, 1), (-1, -1), (-1, 0), (-1, 1). Excluding the all zeros error state, $J^M - 1$ error states remain. We write down a $(J^M - 1) \times (J^M - 1)$ error-state transition matrix $\mathbf{T} = \{t_{i,l}\}$, where the indices i and l each range over the $J^M - 1$ integers labeling the error states. The element $t_{i,l}$ represents the weighted distance incurred by the $i \rightarrow l$ error-state transition:

$$t_{i,l} = f_{i,l} k^{\mu_{i,l}} e^{d_{i,l}},$$

and

$d_{i,l}$ is the Euclidean distance incurred in permissible transitions $i \rightarrow l$;

$\mu_{i,l}$ is the number of bit errors incurred in the $i \rightarrow l$ transition;

$f_{i,l}$ is a factor representing the fraction of data pairs (a_j, b_j) which can produce the error (E_j, F_j) at any epoch j .

Element $t_{i,l} = 0$ for forbidden transitions.

The weighted distance incurred along any error path between epochs j_1 and j_2 can be uniquely determined from either the error sequence between $j_1 - M$ and j_2 or the error-state transitions between j_1 and j_2 . Using the latter approach, the weighted distance incurred along an error path between epochs j_1 and j_2 is simply the product of the weighted transition distances $t_{i,l}$ along that path.

As an example of the values to be assigned to the factors $f_{i,l}$ and $\mu_{i,l}$, suppose that the l th error state at epoch j is characterized by

$(E_j, F_j) = (1, -1)$, and the channel symbols a_j and b_j are binary and independent. Then, a_j and b_j must equal $+1$ and -1 , respectively, implying that $f_{i,l} = 1/4$, and the number of errors incurred $\mu_{i,l} = 2$. Alternatively, if $(E_j, F_j) = (1, 0)$, then a_j must equal $+1$, b_j can be ± 1 , $f_{i,l} = 1/2$, and $\mu_{i,l} = 1$.

Since each error state can be reached from no more than $J - 1$ error states, each row in the error-state transition matrix contains no more than $J - 1$ nonzero entries, and \mathbf{T} is quite sparse.

The error states are numbered from 1 to $J^M - 1$. Then, at any epoch $j - 1$, an error event may begin by a transition from the correct state ($E_{j-M} = 0, \dots, E_{j-1} = 0, F_{j-M} = 0, \dots, F_{j-1} = 0$) to one of the $J - 1$ states ($E_{j-M+1} = 0, \dots, E_{j-1} = 0, E_j, F_{j-M+1} = 0, \dots, F_{j-1} = 0, F_j$). We introduce the $J^M - 1$ dimensional column vector \mathbf{V}_q . The n th element of \mathbf{V}_q is the accumulated weighted distances of all error paths spanning q epochs which merge with the n th error state at any node. Then, \mathbf{V}_1 contains $J - 1$ nonzero elements, corresponding to the $J - 1$ error states given above. The values assumed by these $J - 1$ elements are

$$t_i = \sqrt{2e_b/N_0 [E_j^2 + F_j^2]} \kappa^{\mu_i} f_i, \quad (33)$$

where

μ_i is the number of bit errors incurred in transitioning from the all-zeros error state to error state i ;

f_i is the fraction of data pairs (a_j, b_j) capable of producing a transition from the all-zeros error state to error state i ;

and

i can assume one of the $J - 1$ values corresponding to the indices of error states which can be reached from the all-zeros error state.

Also, the vector \mathbf{V}_q is given by

$$\mathbf{V}_q = \mathbf{T}^{(q-1)} \mathbf{V}_1. \quad (34)$$

At depth p , a remerger may occur from the $J - 1$ error states ($E_j, E_{j+1} = 0, \dots, E_{j+p-1} = 0, F_j, F_{j+1} = 0, \dots, F_{j+p-1} = 0$). We introduce a second $J^M - 1$ dimensional vector \mathbf{U} containing only $J - 1$ nonzero elements. The n th element of \mathbf{U} is equal to unity only if error state n merges with the all-zeros error state. Then the contribution to the BER bound at depth p from the point of divergence is given by

$$P_{B,p} \leq \mathbf{U}^T \cdot \left[\frac{\partial}{\partial \kappa} (\mathbf{T}^{p-1} \mathbf{V}_1) \right] \Bigg|_{\kappa=1}. \quad (35)$$

Thus, using the error-state transition matrix, the BER bound (30)

can be averaged over the correct sequence and rewritten as:

$$P_B \leq \mathbf{U}^T \cdot \sum_{p=1}^{\infty} \left[\frac{\partial}{\partial \kappa} (\mathbf{T}^{p-1} \mathbf{V}_1) \right] \Bigg|_{\kappa=1} \cdot Q \left(\sqrt{\frac{2e_b D_{\min}}{N_0}} \right) \exp(e_b D_{\min}/N_0). \quad (36)$$

An alternate form is

$$P_B \leq \mathbf{U}^T \cdot \left\{ \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \sum_{p=1}^{\infty} [\mathbf{T}^{p-1}(\kappa = 1 + \epsilon) \mathbf{V}_1 - \mathbf{T}^{p-1}(\kappa = 1) \mathbf{V}_1] \cdot Q \left(\sqrt{\frac{2e_b D_{\min}}{N_0}} \right) \exp(e_b D_{\min}/N_0) \right\}. \quad (37)$$

The bound will converge if the matrix series

$$\mathbf{S} = \sum_{p=1}^{\infty} \mathbf{T}^{p-1}(\kappa = 1) \quad (38)$$

converges. If (38) converges, then

$$\mathbf{S} = (\mathbf{I} - \mathbf{T})^{-1}, \quad (39)$$

and

$$P_B \leq \mathbf{U}^T \cdot \left[\frac{\partial}{\partial \kappa} (\mathbf{I} - \mathbf{T})^{-1} \mathbf{V}_1 \right] \Bigg|_{\kappa=1} \cdot Q(\sqrt{2e_b D_{\min}/N_0}) \exp(e_b D_{\min}/N_0). \quad (40)$$

A necessary and sufficient condition for the convergence of \mathbf{S} is that the largest eigenvalue of \mathbf{T} have magnitude less than unity.

If $J^M - 1$ is small, then the largest eigenvalue of \mathbf{T} can actually be found, and if the convergence criterion is satisfied, then the BER bound for MLSE can be found in closed form by inverting $\mathbf{I} - \mathbf{T}$ and substituting into (40). When $J^M - 1$ is too large to conveniently admit such closed-form solution, then mathematical rigor must be sacrificed by returning to (37), and truncating the series at some point p_{\max} such that the remainder may safely be assumed to be small. In this event, the order of matrix \mathbf{T} is usually quite large, and some concern may arise from the apparent need to repetitively multiply large matrices. This is not the case, however, since the term $\mathbf{T}^{p-1} \mathbf{V}_1$ may be found recursively as

$$\mathbf{T}^{p-1} \mathbf{V}_1 = \mathbf{T}(\mathbf{T}^{p-2} \mathbf{V}_1), \quad (41)$$

i.e., the term $\mathbf{T}^{p-1} \mathbf{V}_1$ may be found by recursively multiplying the matrix \mathbf{T} by a vector.

IV. APPLICATIONS

As a first application, we consider MLSE detection of 4ϕ PSK with overlapping rectangular pulses. Let

$$\bar{h}(t) = p(t)e^{j\Delta t}, \quad (42)$$

where

$$p(t) = \begin{cases} \sqrt{2e_b/\tau}, & 0 \leq t < \tau \\ 0, & \text{elsewhere.} \end{cases} \quad (43)$$

Then,

$$S(t) = \text{Re} \left\{ \sum_{k=0}^L (a_k + jb_k) \bar{h}(t - kT) e^{-j\omega_0 t} \right\}, \quad (44)$$

where $T < \tau$ is the signaling interval, and a_k and b_k are independent and equally likely to be ± 1 . Suppose $T = \tau/4$, i.e., 8 bits are transmitted every symbol duration (τ seconds). If $\Delta = 0$, the four possible signal points for each transmitted symbol are as shown in Fig. 3a, corresponding to the familiar 4ϕ PSK modulation. Similarly, if $\Delta = 2\pi/\tau$, then $\bar{h}(t - kT) = p(t - kT)\exp[j\Delta(t - kT)] = p(t - kT)\exp(-j^{\pi k/2}) \cdot \exp(j^{2\pi t/\tau})$, and the four signal points for each symbol transmitted are again as shown in Fig. 3a, although the points exhibit a progressive 90-degree phase rotation for each successive symbol transmitted. The carrier is also offset to the new value $\omega' = \omega_0 + 2\pi/\tau$. However, since a_k and b_k are independent and equally likely to be ± 1 , the error events for the waveshapes corresponding to $\Delta = 0$ and $\Delta = 2\pi/\tau$ are identical, and both must, therefore, exhibit the same detection performance.

For $\Delta = \pi/\tau$, $\bar{h}(t - kT) = p(t - kT)e^{-j^{\pi k/4}} e^{-j^{\pi t/\tau}}$, the carrier is offset to the new value $\omega' = \omega_0 + \pi/\tau$, and the signal points alternate on

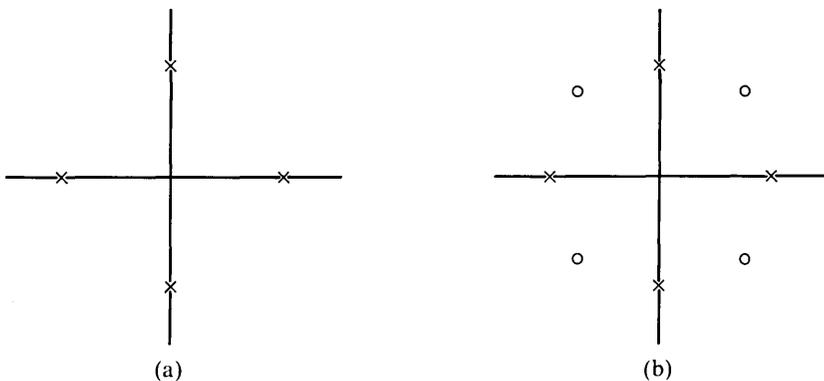


Fig. 3—Signal constellations for four-level QAM. (a) 4ϕ PSK; carrier offset $\Delta = 0$ or π/τ . (b) 4ϕ PSK; carrier offset $\Delta = \pi/\tau$. Signal constellation alternates between crosses and circles.

successive symbol transmissions between the two sets of locations shown in Fig. 3b. As will subsequently be shown, detectability for $\Delta = \pi/\tau$ is superior to that for $\Delta = 0$.

For rectangular signaling with $T = \tau/4$, it is readily established that $M = 3$, and the MLSE receiver for 4ϕ PSK contains $4^3 = 64$ states. The error-state transition matrix needed for analysis, however, is of the order $N = 9^3 - 1 = 728$. We number the error states by means of the six-tuple $(E_1, F_1, E_2, F_2, E_3, F_3)$, where each element can assume 3 values $(0, \pm 1)$. The state where each element is equal to zero corresponds to a merge and is excluded from the error-state transition matrix. For numerical purposes, a convenient way to number the error states is given by

$$B = \begin{cases} 1 + 3^0(F_3 + 1) + 3^1(E_3 + 1) + 3^2(F_2 + 1) \\ \quad + 3^3(E_2 + 1) + 3^4(F_1 + 1) + 3^5(E_1 + 1), & B \leq (9^M - 1)/2, \\ 3^0(F_3 + 1) + 3^1(E_3 + 1) + 3^2(F_2 + 1) \\ \quad + 3^3(E_2 + 1) + 3^4(F_1 + 1) + 3^5(E_1 + 1), & B > (9^M - 1)/2, \end{cases} \quad (45)$$

where B is the state number. If the above numbering scheme for a_k, b_k binary and independent is generalized to arbitrary M , then it can readily be established that the error-state transition matrix possesses the symmetry:

$$t_{i,j} = t_{P+1-i, P+1-j}, \quad P = 9^M - 1. \quad (46)$$

Each row in \mathbf{T} contains no more than nine nonzero elements, corresponding to the permissible transitions.

For the case $\Delta = 0$, D_{\min} was found by search to be 0.5, and for $\Delta = \pi/\tau$, D_{\min} was found to be 0.757. Bit-error rate upper bounds for the two cases as functions of e_b/N_0 are shown in Fig. 4. Also shown, for comparison, is the ideal BER for quadrature signaling with nonoverlapping rectangular pulses, i.e., 2 bits transmitted per symbol duration T . We see that, for $10^{-6} \leq P_B \leq 10^{-4}$, MLSE performance for $t = \tau/4$, $\Delta = \pi/\tau$ is within 1.5 to 2 dB of that achieved with nonoverlapping pulses, although the signaling rate is four times as high. Also, over the same BER range, performance of signaling with $\Delta = 0$ is 1.5 to 2 dB poorer than with $\Delta = \pi/\tau$.

We next consider the transmission of overlapping pulses of finite duration with amplitude tapering to contain spectral emission. The pulse $h(t)$ takes the form:

$$\bar{h}(t) = \begin{cases} [1 + \epsilon \cos(2\pi t/\tau)] \exp(j^{\Delta t}), & -\tau/2 < t \leq \tau/2, \\ 0, & \text{elsewhere,} \end{cases} \quad (47)$$

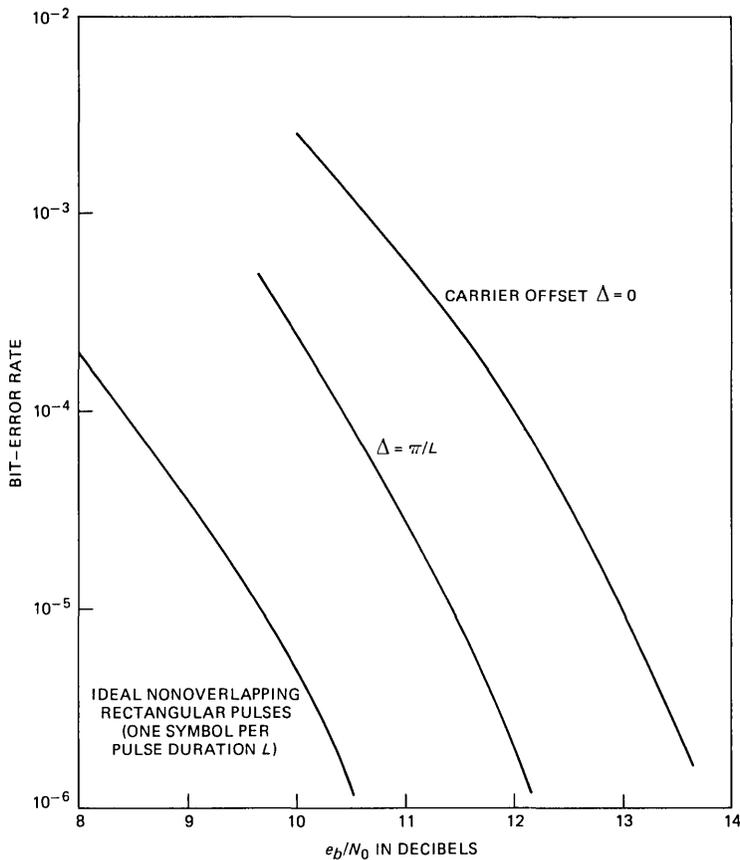


Fig. 4—Bit-error rate vs e_b/N_0 for overlapping rectangular pulses, with no carrier offset and with optimum carrier offset. Four symbols are transmitted per pulse duration L .

where $0 < \epsilon < 1$. The transmitted signal is given by (44) with $\bar{h}(t)$ as above, and the power spectrum for various values of ϵ , with $\Delta = 0$, are as shown in Fig. 5. As ϵ increases in amplitude, the main lobe of the power spectrum tends to increase, but the sidelobes are reduced.

For the particular case $\epsilon = 0.45$, the peak sidelobe is down 24 dB relative to the main lobe height. Let us define the R.F. bandwidth as being twice the baseband frequency at which the power spectrum dips permanently below -24 dB. Then, from Fig. 5, the spectral efficiency of the transmission (bit rate/R.F. bandwidth) is equal to $(2/2.5)(\tau/T)$ bits/s/Hz. where, as before, T^{-1} is symbol signaling rate.

The MLSE error rate bounds for this pulse appear in Fig. 6. For $T = \tau/4$, corresponding to $M = 3$, it was found that D_{\min} occurs at $\Delta = \pi/2\tau$, and for $T = \tau/5$, corresponding to $M = 4$, it was found that D_{\min} occurs

at $\Delta = 5\pi/8\tau$. Both cases are compared against $\Delta = 0$. Results for $T = \tau/4$, corresponding to a transmission rate of 3.2 bits/s/Hz, show that for $P_B = 10^{-4}$, optimum carrier offset is about 0.5 dB superior to no offset, and that e_b/N_0 must increase by about 4 dB, compared against nonoverlapping pulses, to maintain the error rate. However, for the same pulse shape, transmission of nonoverlapping pulses provides a rate of only 0.8 bits/s/Hz. Thus, the transmission rate is higher by a factor of four. For $T = \tau/5$, the transmission rate is 4 bits/s/Hz, (a factor of five higher than for nonoverlapping pulses), and the required increase in e_b/N_0 is about 5.1 dB for optimum carrier offset. Performance without carrier offset is, again, about 0.5 dB inferior.

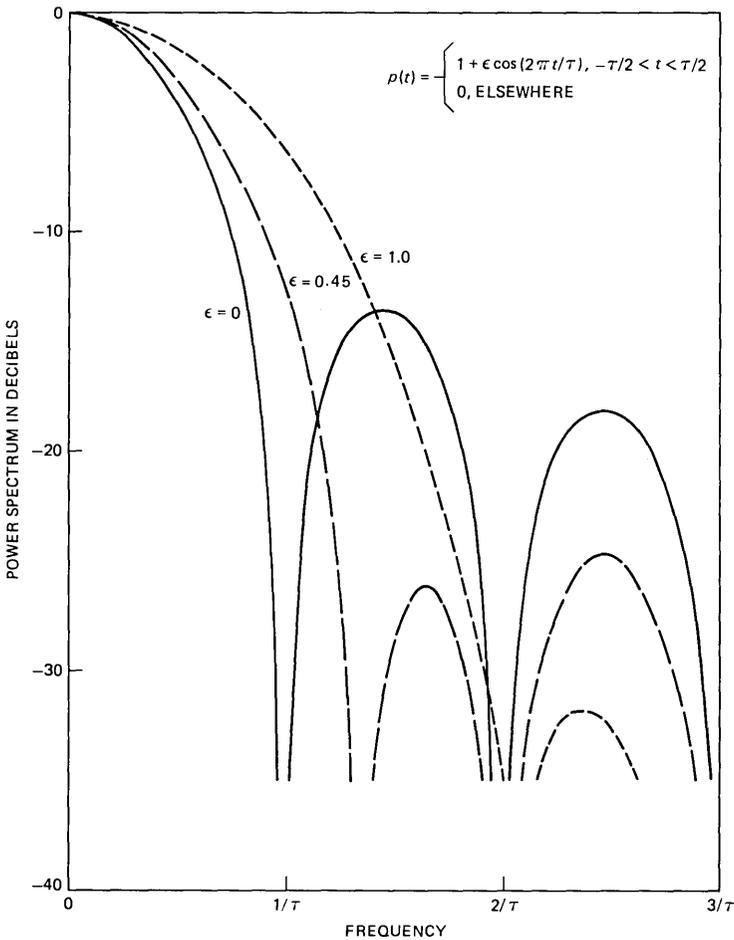


Fig. 5—Spectrum of raised-cosine pulses. $p(t) = 1 + \epsilon \cos(2\pi t/\tau)$ for $-\tau/2 < t < \tau/2$, and $p(t) = 0$ elsewhere.

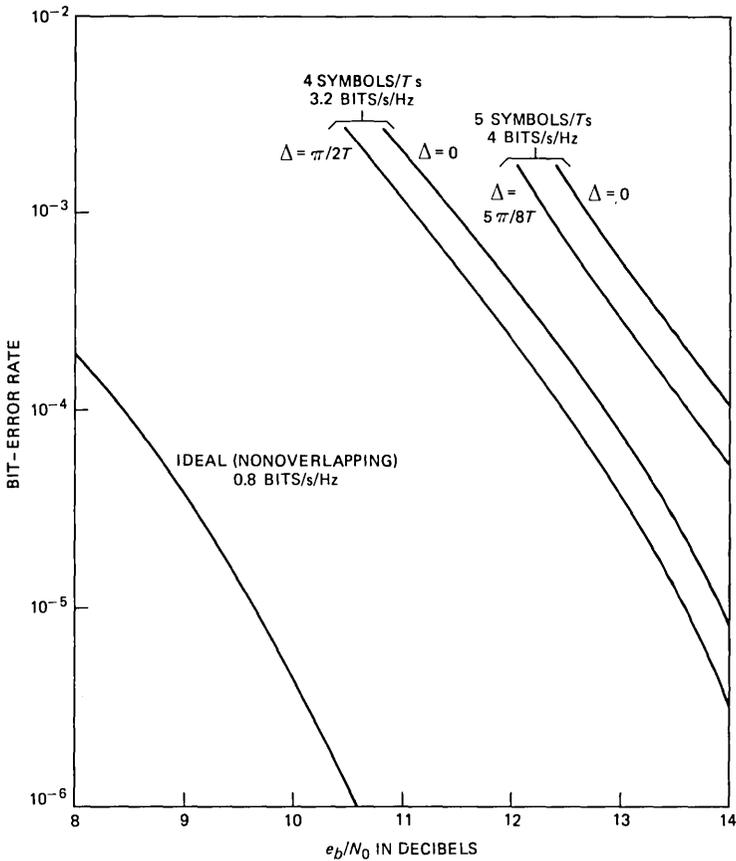


Fig. 6—Bit-error rate for overlapping raised-cosine pulses with $\epsilon = 0.45$, for no carrier offset and optimum carrier offset.

Let us compare these results against M-ary PSK signaling with nonoverlapping pulses. We define the carrier-to-noise (CNR) ρ required to achieve a given BER by

$$\rho = Re_b/N_0, \quad (48)$$

where R is the transmission rate in bits/s/Hz. For nonoverlapping binary pulses (4ϕ -PSK), the transmission rate is 0.8 bits/s/Hz, and to maintain a BER of 10^{-4} , $e_b/N_0 = 8.4$ dB and $\rho = 7.4$ dB. If we double the transmission rate to 1.6 bits/s/Hz by employing 16ϕ PSK with nonoverlapping pulses, the required CNR ratio increases by 12 dB to $\rho = 19.4$ dB. By contrast, MLSE with optimum carrier offset provides a transmission rate four times higher (3.2 bits/s/Hz) with $\rho = 17$ dB, and a transmission rate five times higher (4 bits/s/Hz) with $\rho = 19.7$ dB. Thus, for $\rho = 19.7$ dB, quadrature modulation with overlapping

pulses, optimum carrier offset, and MLSE can more than double the capacity compared against 16 ϕ PSK with nonoverlapping pulses.

As a final example, we give perhaps our most dramatic demonstration of the inherent power of MLSE. We consider that $\bar{h}(t)$ is the impulse response of a four-pole Butterworth filter whose center frequency is offset from the carrier frequency. At baseband, the matched-filtered response to $\bar{h}(t)$ has a spectrum

$$H(\omega) = \left[1 + \left(\frac{\omega - \omega_1}{\omega_{3\text{dB}}} \right)^8 \right]^{-1}, \quad (49)$$

where $\omega_1 = \alpha\omega_{3\text{dB}}$ is the carrier offset, and $\omega_{3\text{dB}}$ is the cutoff frequency of the Butterworth filter. We arbitrarily define the R.F. bandwidth as twice the 3 dB cutoff frequency for no carrier offset. Transmission of $S(t)$, (44), at a rate $T^{-1} = \omega_{3\text{dB}}/2\pi$ then corresponds to $R = 1$ bit/s/Hz. We consider transmission of overlapping pulses at a rate five times as high. The various χ_n 's and ζ_n 's are found from (15) and (16) via contour integration. The impulse response is truncated at $M = 4$, and $|\chi_n|$ and $|\zeta_n|$ are both less than 0.02 for $n > 4$.

Results are as shown in Fig. 7. The optimum carrier offset was found to be $\alpha = 0.625$, yielding $D_{\min} = 0.9$; for $\alpha = 0$, $D_{\min} = 0.8$. We see that, for both cases, results are close to ideal rectangular signaling with nonoverlapping pulses, and that the optimum carrier offset is superior to no offset by about 0.5 dB. To maintain a BER of 10^{-4} , $e_b/N_0 = 9.5$ dB. Since the transmission rate is 5 bits/s/Hz, the required CNR is then $\rho = 16.5$ dB. At this CNR, Shannon's channel capacity $C/W = \log_2(1 + \rho)$ is 5.5 bits/s/Hz. Thus, for $P_e = 10^{-4}$, MLSE performance with the simple Butterworth signaling chosen is quite close to the Shannon limit. Even without carrier offset, the required CNR is about 17 dB, for which the Shannon limit is 5.7 bits/s/Hz.

V. CONCLUSION

Maximum-likelihood-sequence estimation has been shown to be a powerful technique to permit high rates of data transmission with straightforward channel signaling. In this treatment, channel coding techniques were not used, and Quadrature Amplitude Modulation was employed. Implementation of MLSE by means of the Viterbi Algorithm requires a detector with I^M states, where I is the number of levels which may be assumed by each channel symbol and M is the extent, in symbol intervals, of intersymbol interference. Complex signaling waveforms, which are generated by means of carrier offset modulation, can be accommodated and may improve detection efficiency; MLSE equalizes both the in-phase and quadrature components of the intersymbol interference.

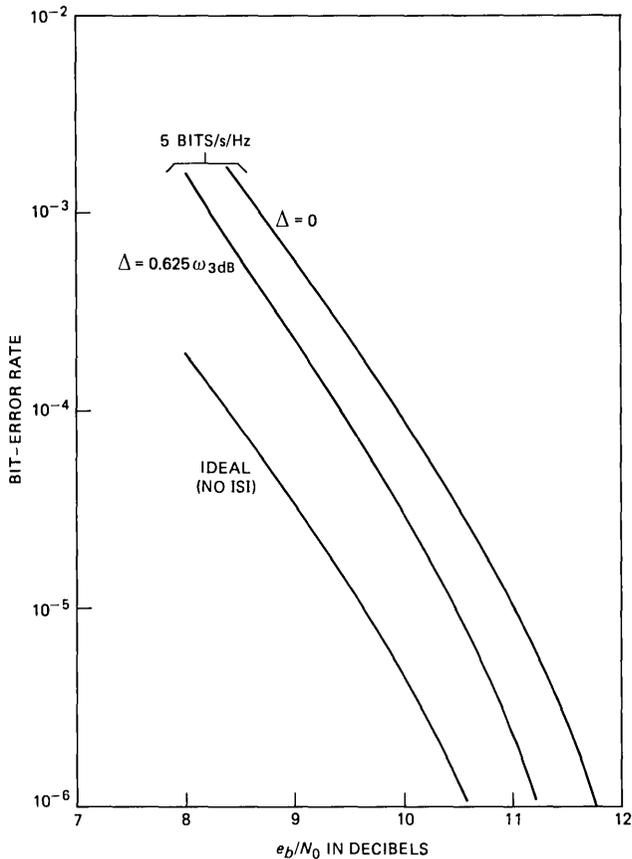


Fig. 7—Bit-error rate vs e_b/N_0 for four-pole Butterworth transmit filtering for no carrier offset and optimum carrier offset.

An error-state transition analysis was used to systematically compute a BER bound for MLSE. Applying this bound, it was shown by example that, with MLSE, transmission rates approaching the Shannon limit may be possible without channel coding.

The power of MLSE detection makes this technique a candidate for radio systems constrained in bandwidth. Since the MLSE approach admits complex waveshapes, channels admitting frequency-selective fading can readily be accommodated, provided that the slowly varying channel frequency characteristics are monitored to update the detection constants continuously. The real limitation on the application of MLSE involves the speed at which the detector can operate, and this places a constraint on the symbol rate of the radio channel. For rates under 30 Megabaud, MLSE may be quite feasible; implementation at real-time rates of 300 Megabaud, such as may apply to satellite

channels, is beyond the current state-of-the-art. However, if the satellite system employs time-division multiple access (TDMA), then the low-duty cycle for any ground station may permit MLSE implementation in nonreal time between TDMA burst arrivals. For duty cycles lower than 10 percent, this would appear to be feasible today. At higher duty cycles, nonreal time detection with parallel processors is another possibility.

REFERENCES

1. A. J. Viterbi, "Convolutional Codes and their Performance in Communication Systems," *IEEE Trans. Commun. Tech.*, *COM-19* (October 1971), pp. 751-771.
2. J. A. Heller and I. M. Jacobs, "Viterbi Decoding for Satellite and Space Communication," *IEEE Trans. Commun. Tech.*, *COM-19* (October 1971), pp. 835-848.
3. L. C. Tillotson, C. L. Ruthroff, and V. K. Prabhu, "Efficient Use of the Radio Spectrum and Bandwidth Expansion," *Proc. IEEE*, *61* (April 1973), pp. 445-452.
4. I. M. Jacobs, "Practical Applications of Coding," *IEEE Trans. Inform. Theory*, *IT-20* (May 1974), pp. 305-310.
5. V. K. Prabhu, "Spectral Occupancy of Digital Angle-Modulation Signals," *B.S.T.J.*, *5* (April 1976), pp. 429-454.
6. S. A. Gronemeyer and A. L. McBride, "MSK and Offset QPSK Modulation," *IEEE Trans. Commun.*, *COM-24* (August 1976), pp. 809-820.
7. V. K. Prabhu, "PSK-Type Modulation with Overlapping Baseband Pulses," *IEEE Trans. Commun.*, *COM-25* (September 1977), pp. 980-990.
8. F. deJager and C. B. Dekker, "Tamed Frequency Modulation, A Novel Method to Achieve Spectrum Economy in Digital Transmission," *IEEE Trans. Commun.*, *COM-26* (May 1978), pp. 534-542.
9. W. J. Weber III, P. H. Stanton, and J. T. Sumida, "A Bandwidth Compressive Modulation System Using Multi-Amplitude Minimum Shift Keying (MAMSK)," *IEEE Trans. Commun.*, *COM-26* (May 1978), pp. 543-551.
10. S. Benedetto, G. DeVincentiis, and A. Luvison, "Error Probability in the Presence of Intersymbol Interference and Additive Noise for Multilevel Digital Signals," *IEEE Trans. Commun.*, *COM-21* (March 1973), pp. 181-190.
11. S. Benedetto and E. Biglieri, "On Linear Receivers for Digital Transmission Systems," *IEEE Trans. Commun.*, *COM-22* (September 1974), pp. 1205-1215.
12. R. Fang and O. Shimbo, "Unified Analysis of a Class of Digital Systems in Additive Noise and Interference," *IEEE Trans. Commun.*, *COM-22* (October 1973), pp. 1075-1091.
13. D. D. Taylor, H. C. Chan, and S. S. Haykin, "A Simulation Study of Digital Modulation Methods for Wideband Satellite Communications," *IEEE Trans. Commun.*, *COM-24* (December 1976), pp. 1351-4.
14. K. Feher et al., "Comparison of PSK and Multilevel AM Data Modulation," *Int. J. Electron.*, *41*, No. 2 (1976), pp. 153-8.
15. L. J. Greenstein and D. Vitello, "Digital Radio Receiver Responses for Combating Frequency-Selective Fading," *IEEE Trans. Commun.*, *COM-27* (April 1979), pp. 671-681.
16. G. D. Forney, Jr., "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Trans. Inform. Theory*, *IT-18* (May 1972), pp. 363-378.
17. G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier Modulated Data Transmission Systems," *IEEE Trans. Commun.*, *COM-22* (May 1974), pp. 624-636.
18. A. S. Acampora, "Maximum-Likelihood Decoding of Binary Convolutional Codes on Band-Limited Satellite Channels," *IEEE Trans. Commun.*, *COM-26* (June 1978), pp. 766-776.



Characterization for Series-Parallel Channel Graphs

By X. M. CHANG, D. Z. DU, and F. K. HWANG

(Manuscript received January 21, 1981)

Channel graphs are often used to describe the connecting paths between a pair of terminals in a link system. Series-parallel (SP) channel graphs and Takagi graphs are special classes of channel graphs heavily studied in the literature. In this paper, we give a noninductive definition of SP channel graphs which we prove to be equivalent to the usual inductive definition. We then use this new definition to characterize SP Takagi graphs and regular SP channel graphs.

I. INTRODUCTION

A graph G is called an s -stage graph if the set of vertices can be partitioned into s disjoint subsets V_1, \dots, V_s , and the set of edges partitioned into $s - 1$ disjoint subsets E_1, \dots, E_{s-1} , such that edges in E_i connect vertices in V_i to vertices in V_{i+1} . G is called an s -stage channel graph if each of V_1 and V_s contain a single vertex, called the source and the sink, respectively, and if each vertex in V_i , $1 < i < s$, is incident to at least one edge each in E_{i-1} and E_i . For convenience, we assume that vertices in the same subset form a column and the columns are ordered from left to right in the natural way.

A channel graph is series-parallel (SP) if it can be obtained either by a series combination or a parallel combination of two smaller SP channel graphs, with the smallest such graph being an edge. A series combination of an s -stage channel graph G and a t -stage channel graph H is a horizontal union of G and H , with the last stage of G being identified with the first stage of H (thus, the new channel graph has $s + t - 1$ stages). A parallel combination of two s -stage channel graphs G and H is a vertical union of G and H , with the first stage of G being identified with the first stage of H , and the last stage of G being identified with the last stage of H (the new channel graph remains s -stage).

A channel graph is *regular* if for every fixed i , $i = 1, \dots, s$, every vertex in V_i has x_i edges in E_{i-1} and y_i edges in E_i , where x_i and y_i depend only on i .

As first illustrated by Lee,¹ the blocking probability of an SP channel graph can be computed inductively. On the other hand, it is well known that the blocking probability of a nonSP channel graph can be hard to compute in general.² This difference in computability explains the special importance of SP channel graphs in switching network studies.

Historically, SP channel graphs have always been defined inductively. In this paper, we give a simple noninductive definition of SP channel graphs and show that it is mathematically equivalent to the inductive definition. The noninductive definition sometimes facilitates the recognition of an SP channel graph and other times it facilitates the use of its properties. Then, we apply this new definition to obtain characterizations of SP Takagi graphs and regular SP channel graphs, which are critically used in Ref. 3 for proving a result on regular SP channel graphs.

II. A NONINDUCTIVE DEFINITION OF SP CHANNEL GRAPHS

Material in this section has close analogy with the study of "series-parallel connection" and "bridge connection" in the theory of circuit design.^{4,5}

Define a *bridge* (see Fig. 1) as a channel graph which consists of two paths p and q from source A to sink B that do not intersect anywhere else and a third path r from a point D on p to a point C on q .

Theorem 1: A channel graph G is SP if and only if no bridge is a subgraph of G .

Proof: We first prove that if G is SP, then G cannot contain a bridge. Our proof is by induction on the number of vertices. An edge, the smallest SP channel graph with two vertices, clearly cannot contain a bridge. Suppose G has n vertices. Since G is SP, there exist two SP channel graphs, G_1 and G_2 , such that G is obtained by a series or a

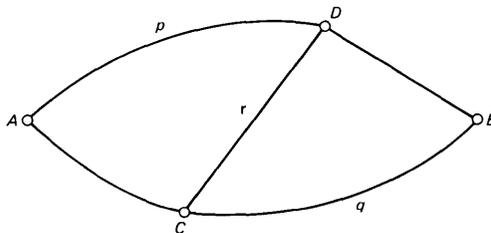


Fig. 1—A bridge channel graph.

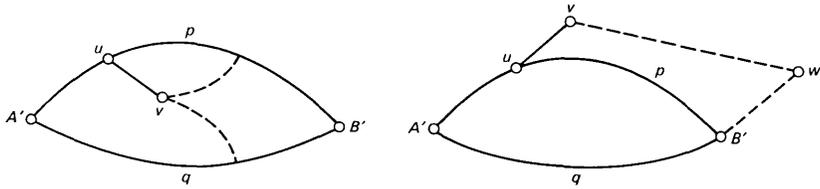


Fig. 2—Consequence of edge $[u, v]$.

parallel combination of G_1 and G_2 . By the inductive assumption, neither G_1 nor G_2 contains a bridge. Furthermore, it is clear that neither a series combination nor a parallel combination of two bridgeless graphs can produce a bridge, hence, G contains no bridge.

Next we prove that if G does not contain a bridge, then G is SP. The proof is by induction on the number of edges in G . Clearly the result holds when G has one edge. For the general case, define a *circuit channel graph* as a channel graph consisting of two nonintersecting paths. Let G' be the *smallest circuit channel subgraph* of G , i.e., one containing the least number of vertices, and hence, the least number of stages and edges. Let $G \setminus G'$ denote the complementary graph of G' in G . We prove that no edge of $G \setminus G'$ includes a vertex of G' , other than possibly the source or sink of G' .

Let p and q be the two paths of G' from source A' to sink B' . Suppose to the contrary that there exists an edge $[u, v]$ in $G \setminus G'$ with vertex u on p (see Fig. 2). If v is on p or q or if an extension of $[u, v]$ meets p or q , then a circuit channel graph smaller than G' exists, which is a contradiction to the definition of G' . If v is not on p or q and no extension of v intersects p or q , then since both v and B' must be connected to the sink of G , there exists a vertex w in a later stage than that of v and B' such that the two vertices v and B' are joined by paths to w , but to no such vertex in an earlier stage. However, we now see that the portion of path p from u to B' forms a bridge between two disjoint paths joining A' to w , a contradiction to the assumption made on G .

Therefore, no edge of $G \setminus G'$ includes a vertex of G' other than possibly its source or sink. Thus, the deletion of one path of G' from G certainly does not affect the SP property of G . Since G without that path contains no bridge and has fewer edges than G , the inductive assumption implies that it must be SP, hence, G must also be SP.

III. A CHARACTERIZATION OF REGULAR SP CHANNEL GRAPHS

An (i, j, r) *multiplex*, $1 \leq i < j \leq s$, of an s -stage channel graph G is an s -stage channel graph consisting of r copies of G where the subgraphs from stage 1 to stage i , and from stage j to stage s are

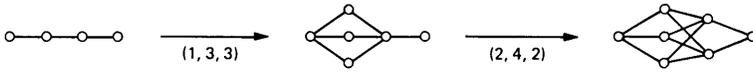


Fig. 3—Multiplexing operations.

merged into one copy (see Fig. 3). A channel graph is called a *Takagi graph* if it can be obtained by multiplexing a smaller Takagi graph where the smallest Takagi graph (with given number of stages) is a multistage line.⁶⁻⁸ Corresponding to each (i, j, r) multiplex operation is a *multiplex index* $m_{ij} = r$. It has been proven (see Ref. 6) that a set of multiplex indices uniquely determines a Takagi graph, i.e., the order of the operations is not important. Let m_{ij} and m_{pq} denote two multiplex indices. Then m_{ij} is said to *cross* m_{pq} if $i < p < j < q$, and to *contain* m_{pq} if the interval $[i, j]$ contains the interval $[p, q]$. We now use Theorem 1 to give a necessary and sufficient condition that a given Takagi graph be SP.

Theorem 2: A Takagi graph is SP if and only if it does not contain two crossing multiplex indices.

Proof: Let G denote the given Takagi graph. Suppose G contains two crossing multiplex numbers m_{ij} and m_{pq} . Since the order of the multiplex operations is not important, we may assume the m_{ij} is the first operation and m_{pq} is the second. First, we assume $m_{ij} = m_{pq} = 2$. Then, the channel graph after each operation is as given in Fig. 4. Clearly, a bridge emerges after the second operation.

Since the graph corresponding to $m_{ij} \geq 2$ and $m_{pq} \geq 2$ contains the graph in Fig. 4b as a subgraph, our argument remains unchanged for this case. Finally, since adding other multiplex operations to the graph in Fig. 4b still results in graphs containing Fig. 4b as a subgraph, we conclude that two crossing multiplex indices always cause a bridge, hence, the graph cannot be SP.

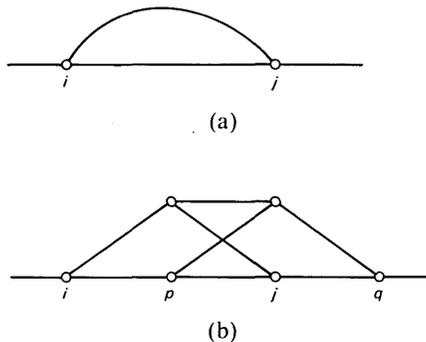


Fig. 4—The $m_{ij} = m_{pq} = 2$ Takagi graph.

Next, we prove that if G does not contain two crossing multiplex indices, then G is SP. Let m_{ij} be a multiplex index of G such that if m_{pq} is any other multiplex index of G , then either $i < p$ (hence $j \geq q$) or $i = p$ but $j > q$.

Case (i) $i = 1$ and $j = s$. Then, G is a parallel combination of m_{ij} smaller Takagi graphs. Using an induction argument, it follows that these smaller graphs are SP since they do not contain crossing multiplex indices, hence, G is SP.

Case (ii) $i \neq 1$. Then, G is a series combination of two smaller Takagi graphs, one from stage 1 to stage i and the other from stage i to stage s . Again by induction, these two smaller graphs are SP since they do not contain crossing multiplex indices, hence, G is SP.

Case (iii) $i = 1$ but $j < s$. Similar to Case (ii).

The proof is complete.

Corollary: A channel graph is regular SP if and only if it is an SP Takagi graph.

Proof: It is easily seen that every Takagi graph is regular and that every regular SP channel graph is a Takagi graph. Furthermore, a Takagi graph is a regular SP channel graph if and only if it does not contain two crossing multiplex indices. The corollary now follows immediately from Theorem 2.

REFERENCES

1. C. Y. Lee, "Analysis of Switching Networks," B.S.T.J., 34 (1955), pp. 1287-1315.
2. F. K. Hwang, "Computing the Blocking Probabilities for Spiderweb Channel Graphs," unpublished work.
3. D. Z. Du, and F. K. Hwang, "Comparisons on Blocking Probabilities for Regular Series-Parallel Channel Graphs," unpublished work.
4. R. J. Duffin, "Topology of Series-Parallel Networks," J. Math. Anal. Appl. 10 (1965), pp. 303-18.
5. J. Valdes, R. E. Tarjan, and E. L. Lawler, "The Recognition of Series-Parallel Digraphs," *Conference Records of the Eleventh Annual ACM Symposium on Theory of Computing* (Atlanta, Ga., 1979), New York: ACM, 1979, pp. 1-12.
6. F. K. Hwang, "Superior Channel Graphs," Proceedings of the 9th ITC, Spain, 1979, p. 543.
7. K. Takagi, "Design of Multistage Link Systems With Optimal Channel Graphs," Rev. Elect. Commun. Lab., 17 (1969), pp. 1205-26.
8. K. Takagi, "Optimal Channel Graph of Link System and Switching Network Designs," Rev. Elect. Commun. Lab., 20 (1972), pp. 962-85.

On the Effects of Varying Analysis Parameters on an LPC-Based Isolated Word Recognizer

By L. R. RABINER, J. G. WILPON, and J. G. ACKENHUSEN

(Manuscript received December 10, 1980)

For practical hardware implementations of isolated-word recognition systems, it is important to understand how the feature set chosen for recognition affects the overall performance of the recognizer. In particular, we would like to determine whether hardware implementations could be simplified by reducing computation and memory requirements without significantly degrading overall system performance. The effects of system bandwidth (both in training and testing the recognizer) on the performance must also be considered since the conditions under which the system is used may be different than those under which it was trained. Finally, we must take account of the effects of finite word-length implementations, on both the computation of features and of distances, for the system to properly operate. In this paper we present the results of a study to determine the effects on recognition error rate of varying the basic analysis parameters of a linear predictive coding (LPC) model of speech. The results showed that system performance was best with an analysis parameter set equivalent to what is currently being used in the computer simulations, and that variations in parameter values that reduced computation also degraded performance, whereas variations in parameter values that increased computation did not lead to improved performance.

I. INTRODUCTION

When faced with the problem of building hardware for a speech processing system, the practical problems of deciding how to implement the system are often solved based on insufficient information of the effects of system parameters on performance. Generally the hardware designer is given a "working system" and asked to devise hardware that performs the same signal processing operations. The designer

often sees potential reductions in hardware complexity (price, etc.), but without a good understanding of the tradeoffs between complexity and performance, he cannot utilize his design knowledge in an efficient manner.

The situation described above is applicable to a number of areas of speech processing. This is especially the case for speech recognition, in which performance scores for a number of different systems have been reported, but for which there is no good experimental data showing how performance degrades (or improves) as system variables are changed in value.¹⁻⁴ Perhaps the closest that investigators have come to obtaining such performance data are the studies by White and Neely comparing two feature sets (linear predictive parameters and bandpass filter parameters) and two time warping methods (linear warping and dynamic time warping),² and the one by Silverman and Dixon, comparing spectral analysis and classification techniques.⁵

In this paper, we present results of a systematic study of the effects on performance of the parameters of a linear predictive coding (LPC)-based, isolated word recognition system. The major emphasis is on studying the effects of the LPC-analysis parameters, namely the number of poles, the frame length, the shift between frames, the use of a preemphasizer, and the anti-aliasing bandpass filter. (These parameters primarily affect computation.) However, results are also presented on the effects of using different system bandwidths for training and testing the recognizer, and on direct quantization of the word reference templates. (Most of the storage in the recognizer is for word reference templates.)

The organization of this paper is as follows. In Section II, we review the LPC recognition model. In Section III, we discuss the experimental method used to evaluate the performance of the recognizer as the LPC parameters were varied. In Section IV, we present results of the performance evaluation. Finally, in Section V, we discuss the implications of the results on a proposed hardware structure.

II. THE LPC-BASED ISOLATED WORD RECOGNIZER

Figure 1 shows a block diagram of the overall isolated word recognition system, and Fig. 2 shows an expanded block diagram of the signal processing for feature analysis. As shown in Fig. 1, the system operates in three modes, namely training, template creation, and testing. For training, the speech signal (recorded off a dialed-up telephone line) is analyzed into features, the endpoints of the spoken isolated word are located, and the features (within the word boundaries) are stored. For template creation, either a clustering procedure,^{6,7} or a robust averaging method⁸ is used to create a set of word reference templates. Finally, for testing, the feature sets for each reference

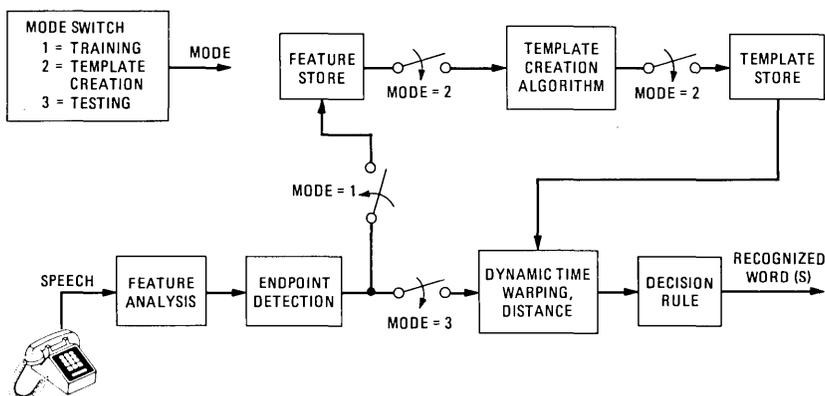


Fig. 1—Block diagram of isolated word recognition system.

template and the feature set for the unknown utterance are compared, using a dynamic time-warping time-alignment algorithm,³ and a decision rule is used to give a set of word candidates, ordered by recognition distance scores.

In this paper, we are concerned primarily with the effects of feature analysis variables on the performance of the word recognizer. Hence, we will concentrate primarily on the analysis operations, shown in Fig. 2, as required for an LPC-based system. As shown in this figure, the basic analysis operations consist of:

(i) Bandpass filtering to bandlimit the speech signal prior to analog-to-digital conversion to minimize aliasing in the signal. The bandpass filter typically bandlimits the signal to the range $100 \text{ Hz} \leq f \leq 3200 \text{ Hz}$ with an analog Butterworth filter with slopes of 24 dB to 48 dB per octave. Thus, two potential analysis parameters are the high-frequency cutoff of the filter (F_H), and the slope of the attenuation characteristic with frequency.

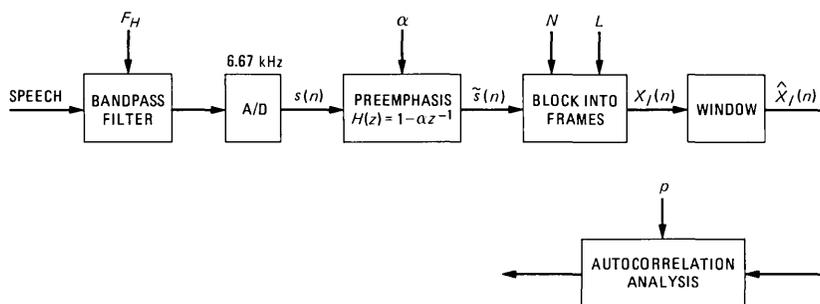


Fig. 2—Expanded block diagram of the signal processing in the analysis part of the system.

(ii) Analog-to-digital conversion at a sampling rate of 6.67 kHz, giving the digital signal $s(n)$.

(iii) Preemphasis of $s(n)$ by a first-order digital network with a transfer function

$$H(z) = 1 - \alpha z^{-1}, \quad (1)$$

giving the signal

$$\tilde{s}(n) = s(n) - \alpha s(n - 1). \quad (2)$$

For the preemphasizer it is important to understand the effects of the parameter α on the performance of the system.

(iv) Blocking the signal into analysis frames, giving the signal $x_l(n)$, defined as

$$x_l(n) = \tilde{s}(lL + n), \quad 0 \leq n \leq N - 1, \quad 0 \leq l \leq Q - 1, \quad (3)$$

where L (the shift between frames) and N (the size of the frame) are parameters which must be investigated, and Q is the number of frames in the recording interval.

(v) Windowing the signal to give

$$\hat{x}_l(n) = x_l(n) w(n), \quad (4)$$

where $w(n)$ is an N -point window used in the analysis.

(vi) Performing an autocorrelation analysis to give

$$R_l(m) = \sum_{n=0}^{N-1-m} \hat{x}_l(n) \hat{x}_l(n + m), \quad 0 \leq m \leq p, \quad (5)$$

where p , the number of poles used in the analysis, is again a parameter of the system. The set of vectors $R_l(m)$, $0 \leq l \leq Q - 1$, define the feature set used for both training the system, and for performing the comparisons between reference and test patterns.

Based on the above discussion, we chose to investigate the following analysis parameters:

- (i) F_H = the high-frequency cutoff frequency of the anti-aliasing bandpass filter,
- (ii) S = the shape of the attenuation characteristic of the high-frequency slope of the bandpass filter,
- (iii) α = the value of the preemphasis constant,
- (iv) N = the size of the analysis frame,
- (v) L = the shift between consecutive analysis frames, and
- (vi) p = the number of poles in the analysis system.

To carefully investigate and understand the effects of the bandpass filter on system performance, we modified the analysis system of Fig. 2 so that all processing was done digitally. Figure 3 shows the modified

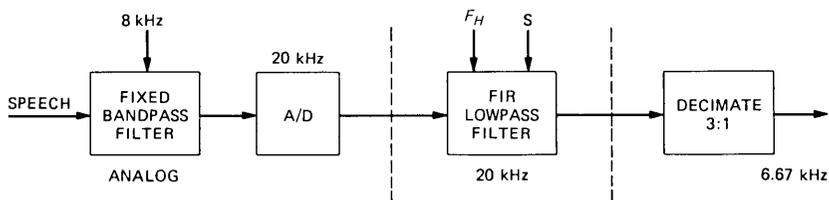


Fig. 3—Block diagram of modified analysis system.

front end of the analysis system. The analog speech signal was filtered with a fixed bandpass filter with a high-frequency cutoff of 8 kHz (24 dB/octave slope) and digitized at a 20-kHz rate. We then used a finite impulse response (FIR) lowpass digital filter (with cutoff frequency F_H , and slope factor S —corresponding to narrow or wide transition band) to filter the signal to the appropriate bandwidth, and a decimator to reduce the sampling rate to 6.67 kHz (i.e., a 3-to-1 decimation of the signal). Because of the use of digital, rather than analog, filters, greater control could be exercised on the filter parameters, and the effects of aliasing could be more easily studied.

2.1 Range of variation of the analysis parameters

For the digital low-pass filter, we designed a set of six equiripple, linear phase, FIR low-pass filters with the following specifications:

$$\begin{aligned}
 F_0 &= \text{sampling frequency} = 20000 \text{ Hz}, \\
 F_p &= F_H = \text{variable} (2400, 3200, 4200 \text{ Hz}), \\
 F_s &= F_p + S = \text{variable} (S = 133, 600 \text{ Hz}), \\
 \delta_p &= 0.0316, \\
 \delta_s &= 0.00316 (-50 \text{ dB}).
 \end{aligned}$$

The above set of specifications were met by either 69-point filters ($S = 600$ -Hz transition band), or by 301-point filters ($S = 133$ -Hz transition band) for all three values of F_H . Thus, six distinct low-pass filters were designed. Figure 4 shows plots of the frequency response (on a dB scale) for the two low-pass filters with $F_H = 3200$ Hz. Part (a) of this figure shows results for the case $S = 600$ Hz and part (b) shows results for $S = 133$ Hz.

For the first order preemphasis network, two values of α were considered, namely $\alpha = 0.95$ (standard preemphasis) and $\alpha = 0$ (no preemphasis).

For the analysis frame length and shift parameters (N and L), six sets were considered, namely:

1. $N = 300, \quad L = 100,$
2. $N = 300, \quad L = 150,$
3. $N = 300, \quad L = 300,$

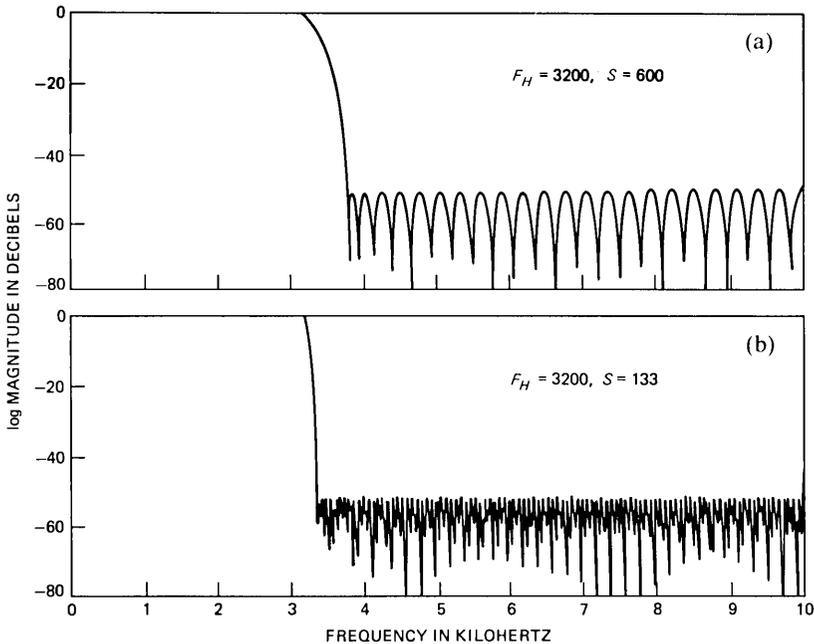


Fig. 4—Frequency responses of the two 3200-Hz cutoff filters.

- 4. $N = 200$, $L = 100$,
- 5. $N = 200$, $L = 200$,
- 6. $N = 100$, $L = 100$.

These six sets span the range from 67 ($L = 100$) to 22 ($L = 300$) frames per second, with frame lengths from 15 ms ($N = 100$) to 45 ms ($N = 300$). We considered longer analysis frames to be impractical, and we always chose the shift length as a rational fraction of N .

Finally, values of p from 2 to 10 were considered in increments of 2, i.e., $p = 2, 4, 6, 8, 10$. Increments of two were used since speech poles tend to occur in complex-conjugate pairs. The range of p went down to two since speech models for recognition based on a 2-pole LPC analysis have been proposed.⁹

The current “operating point” of the isolated word recognizer is the set of values

$$\begin{aligned}
 F_H = 3200 \text{ Hz}, & \quad S = 600 \text{ Hz}, & \quad \alpha = 0.95, \\
 N = 300, & \quad L = 100, & \quad p = 8.
 \end{aligned}$$

A major point of interest is the performance of the recognizer around this operating point. If the performance does not change too much around this operating point, it is an indication that the system is fairly robust.

2.2 Quantization of template variables

As discussed previously, the word reference templates are created from the sets of autocorrelation coefficients. The templates are stored as sets of autocorrelations of LPC coefficients, i.e., the signal autocorrelation coefficients $R_i(m)$ are converted to LPC coefficients, $a_i(m)$, and the autocorrelation of the a 's [$R_a(m)$] is stored as the reference template. [This form is used since the distance computation required in the dynamic time warping is reduced to a simple $(p + 1)$ -point dot product.]

Although a great deal is known about optimal quantization methods for LPC parameters,¹⁰⁻¹³ we are primarily interested in the effects of linear PCM-coding of the reference templates. We are not able to use the sophistication of coding theory ideas since additional processing hardware would be required to convert the templates from coded form to the appropriate recognition format for the distance computation. Hence, we are mainly interested in determining the smallest number of bits that can be used to represent the templates in the given format, while maintaining the overall system performance at the same level as for no template quantization.

Since the distance required for the LPC-based recognizer is implemented as a dot product, we can derive a simple expression for the effects of quantization of the reference templates on this dot product. We denote the (unquantized) reference template coefficients (for a single frame) as $R_i = R_a(i)$, $i = 0, 1, \dots, p$, and the test coefficients (again for a single frame) as T_i , $i = 0, 1, \dots, p$. Then the dot-product distance between frames R and T can be expressed* as

$$d(R, T) = \sum_{i=0}^p R_i T_i. \quad (6)$$

If we uniformly quantize R_i to a B -bit number, then we can express the quantized value, \hat{R}_i , as

$$\hat{R}_i = R_i + \epsilon_i, \quad (7)$$

where ϵ_i is a uniformly distributed random variable over the range $-\Delta/2 \leq \epsilon_i \leq \Delta/2$, where

$$\Delta = \frac{(R_i)_{\max} - (R_i)_{\min}}{2^B}, \quad (8)$$

$(R_i)_{\max}$ is the anticipated maximum value of coefficient R_i , and $(R_i)_{\min}$ is the anticipated minimum value of coefficient R_i .

* For the overall LPC distance, a log of the dot-product distance is taken. We ignore the log for this error analysis.

Based on the above model, the distance between the quantized reference and the test becomes

$$\hat{d} = d(\hat{R}, T) = \sum_{i=0}^p \hat{R}_i \cdot T_i. \quad (9)$$

The distance error, e , caused by quantization, can be expressed as

$$e = d - \hat{d} = \sum_{i=0}^p (R_i - \hat{R}_i) \cdot T_i = \sum_{i=0}^p \epsilon_i T_i. \quad (10)$$

If we make the simplifying assumption that $T_i = k$ (i.e., the test pattern is a constant), then the error becomes

$$e_k = k \sum_{i=0}^p \epsilon_i. \quad (11)$$

The variable e_k is therefore, by the central limit theorem, a Gaussian-distributed variable of mean 0, and variance

$$\sigma_{e_k}^2 = k^2 \sum_{i=0}^p \overline{\epsilon_i^2} = k^2 \sum_{i=0}^p \sigma_{\epsilon_i}^2. \quad (12)$$

Thus, the effect of quantizing the reference templates is to add a small, random component to the distance score. Although we cannot state precisely how the random component of the distance will affect recognition scores, it seems clear that the only cases affected will be those words whose recognition distance is close to the distance of another word in the vocabulary. For those cases, the random distance errors will generally increase the error rate, especially as B becomes smaller (i.e., larger quantization errors). When B is sufficiently large we would expect almost no change in recognition error rates (in fact, the accuracy could go up since some cases with errors and close distance scores could be improved).

To implement the model of eqs. (7) and (8), we need to know the ranges $[(R_i)_{\max}]$ and $[(R_i)_{\min}]$ of each reference template coefficient. Table I gives these ranges, as measured from an actual set of 468 speaker-independent templates (over 17000 frames). Figure 5 shows measured histograms of the nine coefficients. From Table I and Fig. 5 we see that the range of each coefficient is different, whereas the shapes of each distribution are similar.* We show later that one practical implication of this result is that the quantized reference coefficients need a different scale factor for each coefficient. For hardware realizations an alternative, essentially equivalent, quantiza-

* A check on the distribution shape and range was made by measuring it again on a different set of templates. Essentially equivalent results were obtained.

Table I—Range values for the reference template coefficients

| i | $(R_i)_{\min}$ | $(R_i)_{\max}$ |
|-----|----------------|----------------|
| 0 | 1 | 18 |
| 1 | -28 | 28 |
| 2 | -6 | 24 |
| 3 | -21 | 12 |
| 4 | -4 | 13 |
| 5 | -9 | 6 |
| 6 | -3 | 6 |
| 7 | -4 | 3 |
| 8 | -1 | 1 |

tion scheme (to be described later) can be used to approximately provide the different quantization ranges.

III. PERFORMANCE EVALUATION OF THE RECOGNIZER

To evaluate the performance of the recognizer as the analysis parameters were varied, a set of four talkers (three male, one female) were asked to make recordings. The recognition vocabulary consisted of the 26 letters (A to Z), the 10 digits (0 to 9), and the 3 command words STOP, ERROR, and REPEAT.

For training purposes, each talker recorded the 39-word vocabulary seven times over a dialed-up telephone connection. For each talker the digitized acoustic waveform (at a 20-kHz rate) for each word was stored on a training file. During the evaluation, the robust training procedure of Rabiner and Wilpon⁸ was used to provide one (speaker dependent) reference template per vocabulary word, once the set of recognition parameters was chosen. The number of tokens per word required for training varied from two to seven; hence, for many words all seven training tokens were unnecessary. Tokens not used in the training were effectively discarded.

For testing purposes, each talker spoke the 39-word vocabulary 10 times over a new dialed-up telephone connection. Again, the acoustic waveforms for each word were stored on a testing file.

Three evaluation experiments were run. For the first experiment, all combinations of the six parameters F_H , S , α , N , L , and p were tried; i.e., a total of $3 \times 2 \times 2 \times 3 \times 2 \times 5 = 360$ recognition runs were made for each talker. For each run an overall error rate score (on the top candidate) was measured.

For the second experiment, the values of α , N , L , and p were set at 0.95, 300, 100, and 9, respectively (the operating point), and all combinations of F_H and S were tried for *both* testing and training, i.e., the filter for testing could be different than the filter for training. A total of 36 runs was made for each talker.

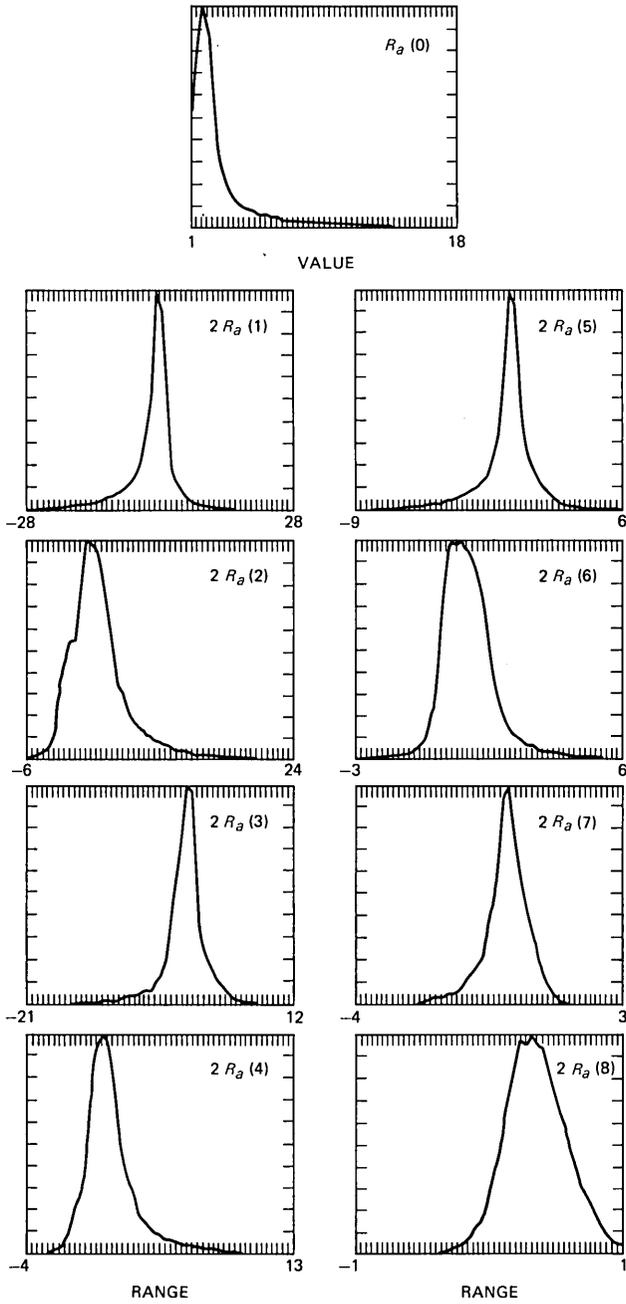


Fig. 5—Histogram of reference template coefficient distributions.

Finally, a recognition experiment was run (using speaker-independent templates) to determine the effects (on word error rate) of linear quantization on the word reference templates. The number of bits in the quantizer (B) ranged from 8 to 12. The recognition scores without quantization ($B = \infty$) were also measured.

IV. RESULTS OF THE PERFORMANCE EVALUATION

The results of the performance evaluation are given as a series of plots of word error rate for each talker as a function of some system parameter, averaged over all other system parameters. Minimal interaction between system parameters is necessary for this data to be meaningful. For some parameters this is indeed the case; for others, this assumption is not valid. We shall endeavor to point out such cases as they appear.

4.1 Variations of analysis parameters

The results of the first experiment are given in Figs. 6 to 9. Figure 6 shows plots of average word error rate versus p , the number of LPC poles, for each talker. Two general trends emerge from these curves. First, we see that the word error rate curves differ greatly among talkers—i.e., talker 1 achieved about an 8.6 percent word error rate ($p = 10$), whereas talker 4 had a 22.6 percent word-error rate. This variation in error scores is typical for this complex a vocabulary,¹⁴ and the scores of the four talkers fall within the normal range.

The second trend noted in Fig. 6 is that the error rate falls rapidly as the number of poles in the analysis is increased from 2 to 8; however, the error rate scores for 8 and 10 poles are comparable. This result reflects the well-known fact that *at least* 6 poles are required to adequately represent the three formants of speech, and that 8 or 10 poles provide an extra margin of safety for cases when four or more formants are present, or when several real poles are required for the optimum all-pole fit to the signal.

Figure 7 shows plots of average word-error rate versus L , the number of samples between frames, for the three values of N , and for each talker. Again we see that the general trends in the data are the same for all talkers, even though the absolute error rate scores are different. As L increases from 100 to 300 (i.e., as we measure fewer frames per second), the error rate scores increase by about 5 percent. We also see that the size of the analysis frame (N) affects the error score considerably less than the shift rate. For talkers 2 and 4, there were essentially no differences in error rate as N varied (for fixed L); whereas for talkers 1 and 3, we obtained differences in error scores of about 2 percent. However, these differences in error score were not consistent between talkers 1 and 3; i.e., talker 1 had better scores for $N = 200$

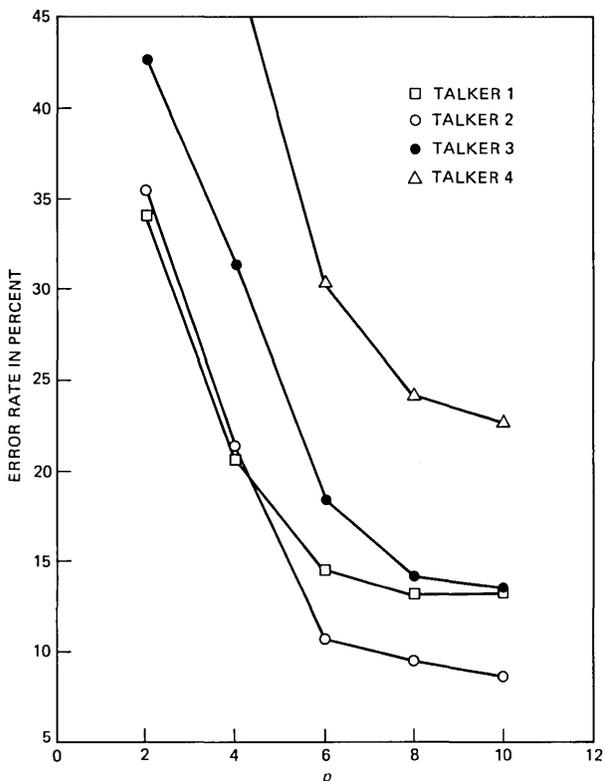


Fig. 6—Plots of word-error rate versus number of LPC poles (p) for each talker.

than for $N = 300$, whereas talker 3 exhibited the opposite effect (for $L = 100$ samples). Hence, we conclude that a value of $L = 100$ (67 frames per second) is important for highest recognition scores; however, the exact value chosen for N is not as important a factor.

Figure 8 shows a plot of the average error rate as a function of the preemphasis constant, α , for each talker. Interestingly, these curves (at least for two talkers) show lower-error rates for $\alpha = 0$ (no preemphasis) than for $\alpha = 0.95$. This result is somewhat misleading since the α parameter strongly interacts with the parameter p (the number of LPC poles). This is because for small values of p (namely 2 and 4) the effect of preemphasis is to boost the high-frequency components of the signal spectrum and thereby cause the all-pole model to match higher formants for some sounds, rather than the lower formants. For systems where p was large enough ($p = 6, 8, 10$) this did not occur. A detailed analysis of the error rate scores as a function of p and α combined showed that the above observation was indeed the case. For values of $p \geq 6$, the effects of preemphasis were insignificant in terms

of average error rate. Of course, preemphasis is important in an LPC analysis system in which finite-precision arithmetic is used in the analyzer.¹⁰

Finally, Fig. 9 shows plots of the average word error rate as a function of the filter cutoff frequency (F_H) for each of the four talkers. We see that there is essentially no variation in error rate versus filter cutoff frequency. Similarly, we found that the average error rate scores were independent of the filter transition bandwidth (S). We discuss these results more thoroughly in Section 4.2.

The results given in this section indicate that the operating point chosen in the LPC analysis system is robust, i.e., the error rate scores are essentially as low as they can be made, and small changes in parameter values which increase computation change the recognition scores only slightly, whereas changes in parameter values which decrease computation also tend to degrade performance.

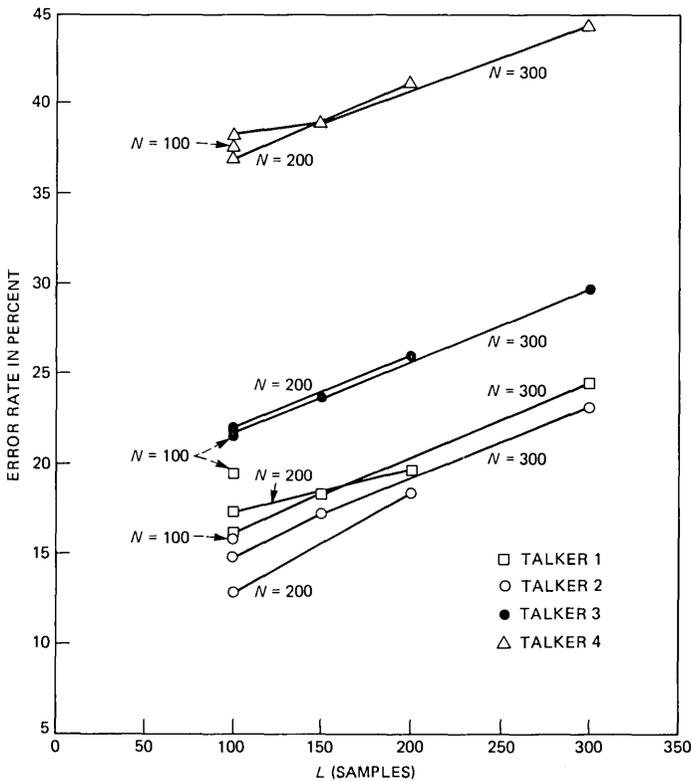


Fig. 7—Plots of word-error rate versus the shift (L) and the frame length (N) for each talker.

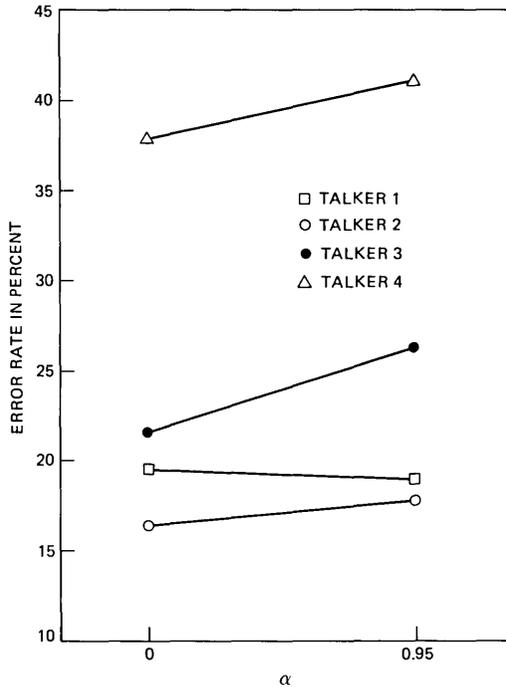


Fig. 8—Plots of word-error rate versus the preemphasis constant (α) for each talker.

4.2 Variations of system bandwidth for training and testing

Figures 10 and 11 give the results of the second experiment. In this experiment, one of the six low-pass filters was used for the training system and another of the six low-pass filters was used in the testing system. Figure 10 shows the word error rate (for each talker) as a function of the training filter, where the testing filter is the same as the training filter (the solid curve), and where the testing filter gives the lowest error rate (the dashed curve).^{*} Figure 11 shows the curve of error rate (for each talker) as a function of testing filter for the training filter giving the lowest error rate. The data in these curves were obtained with the analysis parameters set to the operating point values as discussed previously.

The curves in Figures 10 and 11 show the following:

- (i) A training filter cutoff frequency of at least 3200 Hz is required to provide low error rates.
- (ii) As the training filter cutoff frequency is raised above 3200 Hz,

^{*} In Figures 10 and 11, the notation F_H+ is used to denote the low-pass filter with a cutoff frequency of F_H hertz, and with a transition width of $S = 133$ Hz. If no + is used, the transition width is 600 Hz.

the error rate remains substantially the same as long as the testing filter cutoff frequency remains at 3200 Hz or higher.

The key result here is that a system bandwidth of at least 3200 Hz is required for both training and testing, and that broader bandwidths for training or testing (leading to some signal aliasing) seem to have little effect on the error rates.

4.3 Uniform quantization of reference templates

Table II gives the results on word error rate using uniformly quantized (over the variable ranges of Table I) reference templates. It gives word error rates for the top 1, 2, and 5 candidates using a value of $K_{NN} = 2$ for the K -nearest neighbor rule. We see that the error rates for 12-bit coding are comparable to the error rates for ∞ -bit (unquantized) templates. For 10-bit coding, a 2 percent loss in accuracy for the top candidate results, and for 8-bit coding a 5 percent loss in accuracy results. Thus, we conclude that about 12 bits per word are required for high accuracy using a uniform quantizer.

We made some informal tests using a logarithmic quantizer, since the distributions of the reference template coefficients were highly nonuniform, as shown previously in Fig. 5. However, the results indicate that no significant reduction in the number of bits can be obtained.

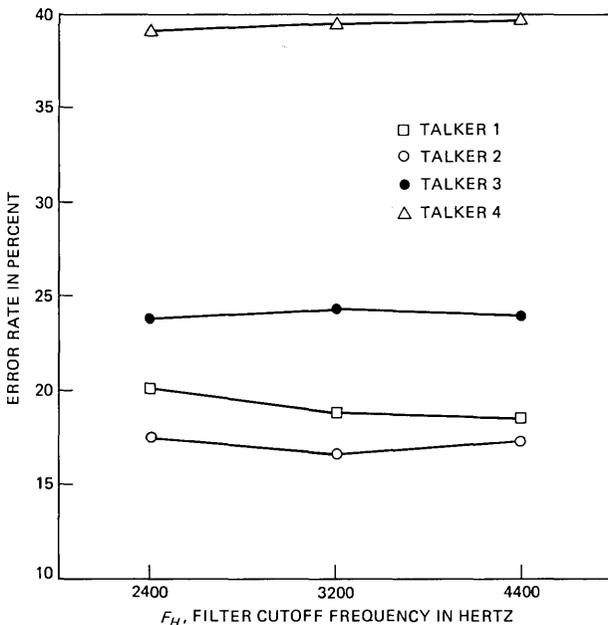


Fig. 9—Plots of word-error rate versus filter cutoff frequency (F_H) for each talker.

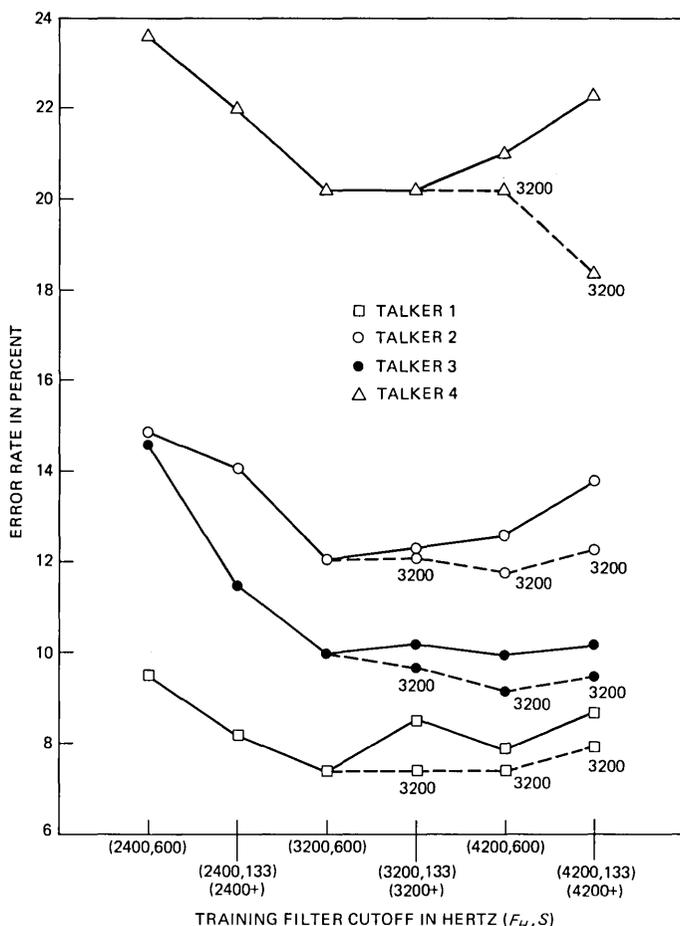


Fig. 10—Plots of word-error rate as a function of training filter cutoff frequency for each talker.

Therefore the extra cost of converting from log to linear quantization was not justified.

V. DISCUSSION AND SUMMARY

One of the purposes of this investigation was to determine the effects of variations of the LPC-analysis parameters on the overall performance of an LPC-based isolated word recognition system to understand the potential tradeoffs for hardware implementation. The two key-analysis parameters that are directly related to cost (computation) are the shift rate (L) and the number of poles (p). The results showed that the tradeoff between performance (error rate) and L is undesirable, namely doubling L (halving the computation) increases the error rate by about

4 percent. This large an increase in error rate is unacceptable for many applications. Similarly, the tradeoff between performance and reductions in p shows that even small reductions in p (from 8 to 6) lead to increases in error rate of 1.5 to 5.5 percent for different conditions. Thus, from the point of view of simplifying a hardware implementation of the system, there appears to be no such simplifications.

However, the results show that the recognition system is robust around the analysis operating point, and that the system bandwidth for training and testing interact somewhat. We have also found that uniform quantization of the reference templates to a 12-bit word size is sufficient for all practical purposes.

Since a uniform quantization of reference templates to 12-bit accuracy was sufficient, we devised a strategy for approximately realizing the required variable range quantization of the templates in hardware.

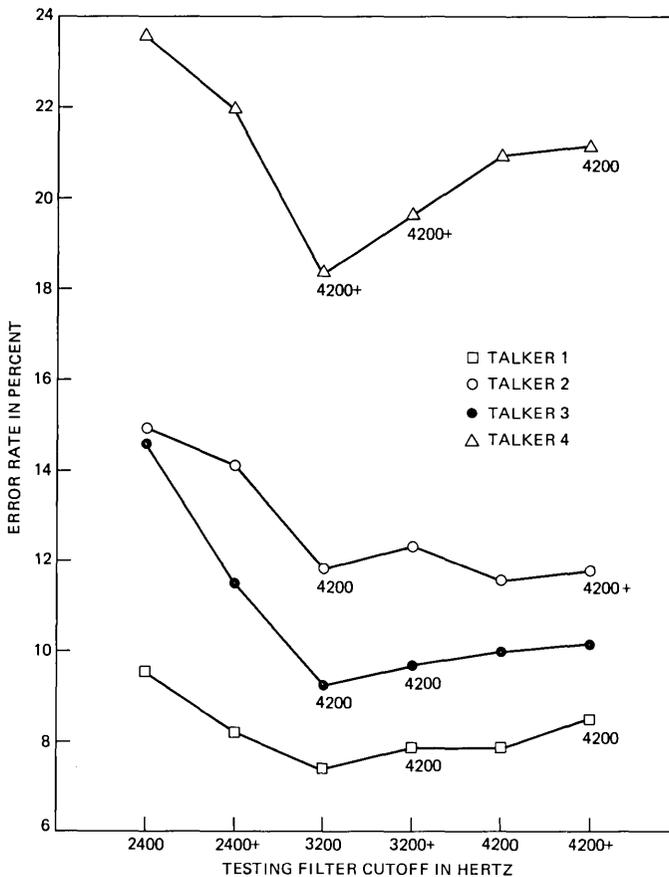


Fig. 11—Plots of word-error rate as a function of testing filter cutoff frequency for each talker.

Table II—Word error rates as a function of word candidate position and the number of bits used in the linear quantizer

| Number Bits | Candidate Position | | |
|----------------|--------------------|------|-----|
| | 1 | 2 | 5 |
| 8 | 25.9 | 11.3 | 2.6 |
| 10 | 22.4 | 10.5 | 2.6 |
| 12 | 20.1 | 9.0 | 2.1 |
| ∞ | 21.6 | 9.3 | 1.8 |

The key problem is that the range of each coefficient is different; therefore, before computing distance, the normalized (quantized) range of the template coefficients has to be converted back to the unnormalized (correct) range. A simple way to implement the variable range quantization is to represent the i th reference parameter as R_i , and to let

$$R_i = S_i \bar{R}_i, \quad (13)$$

where

$$|\bar{R}_i| \leq 1, \quad (14)$$

and S_i is a normalization constant* to guarantee that eq. (14) is valid. Then *each* of the \bar{R}_i variables can be uniformly quantized to B -bit accuracy, and the S_i correction can be applied to the test coefficients, since this only needs to be done once per test. Thus, the distance is computed as

$$\begin{aligned} D(R, T) &= \sum_{i=0}^p R_i \cdot T_i = \sum_{i=0}^p \bar{R}_i (S_i T_i) \\ &= \sum_{i=0}^p \bar{R}_i \tilde{T}_i, \end{aligned} \quad (15)$$

where \tilde{T}_i is the weighted test coefficient. In this manner, the quantized (range normalized) reference coefficients are used directly in the distance computation.

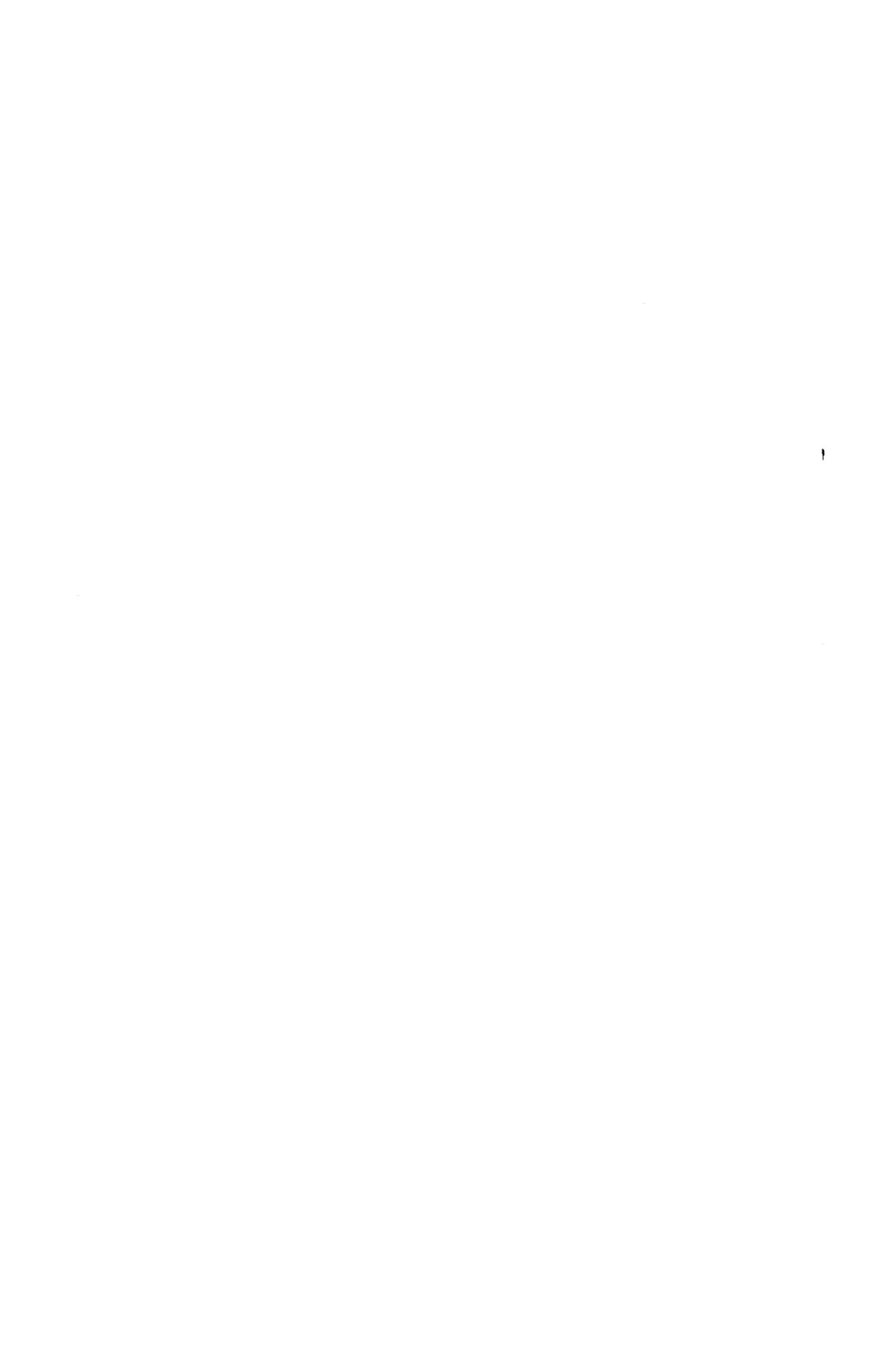
In summary, we have found that the recognition system can be reliably trained using a wide variety of analysis frame sizes, LPC-system orders, and system bandwidths, and that good recognition scores can be obtained for a reasonable range of the analysis parameters. The results showed that system performance was best with an analysis

* For convenience, S_i could be shown to be a power of 2 such that eq. (14) holds. In this case the required scaling of the test coefficients is implemented as a shift.

parameter set equivalent to what is currently being used in the computer simulations, and that variations in parameter values which reduced computation also degraded performance, whereas variations in parameter values which increased computation did not lead to improved performance.

REFERENCES

1. T. B. Martin, "Practical Applications of Voice Input to Machines," *Proc. IEEE*, *64* (April 1976), pp. 487-501.
2. G. M. White and R. B. Neely, "Speech Recognition Experiments with Linear Prediction, Bandpass Filtering, and Dynamic Programming," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, *ASSP-24*, No. 2 (April 1976), pp. 183-8.
3. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, *ASSP-23*, No. 1 (February 1975), pp. 67-72.
4. L. R. Rabiner et al., "Speaker-Independent Recognition of Isolated Words Using Clustering Techniques," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, *ASSP-27*, No. 4 (August 1979), pp. 336-49.
5. H. F. Silverman and N. R. Dixon, "A Comparison of Several Speech-Spectra Classification Methods," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, *ASSP-24*, No. 4 (August 1976), pp. 289-95.
6. L. R. Rabiner and J. G. Wilpon, "Considerations in Applying Clustering Techniques to Speaker-Independent Word Recognition," *J. Acoust. Soc. Am.*, *66*, No. 3 (September 1979), pp. 663-73.
7. S. E. Levinson et al., "Interactive Clustering Techniques for Selecting Speaker-Independent Reference Templates for Isolated Word Recognition," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, *ASSP-27*, No. 2 (April 1979), pp. 134-41.
8. L. R. Rabiner and J. G. Wilpon, "A Simplified Robust Training Procedure for Speaker Trained, Isolated Word Recognition Systems," *J. Acoust. Soc. Am.* *68*, No. 5 (November 1980), pp. 1271-6.
9. J. Makhoul and J. Wolf, "The Use of a Two-Pole Linear Prediction Model in Speech Recognition," Report 2537, Cambridge, Massachusetts: Bolt, Beranek, and Newman Inc., September 1973.
10. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, New York: Springer, 1976.
11. R. Viswanathen and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, *ASSP-23*, No. 3 (June 1975), pp. 309-21.
12. A. H. Gray, Jr., R. M. Gray, and J. D. Markel, "Comparison of Optimal Quantization of Speech Reflection Coefficients," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, *ASSP-25*, No. 1 (February 1977), pp. 9-23.
13. R. M. Gray et al., "Distortion Measures for Speech Processing," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, *ASSP-28*, No. 4 (August 1980), pp. 367-76.
14. B. Aldefeld et al., "Automated Directory Listing Retrieval System Based on Isolated Word Recognition," *Proc. IEEE*, *68*, No. 11 (November 1980), pp. 1364-79.



Three-Dimensional Radiation Characteristics of a Pyramidal Horn-Reflector Antenna

By P. E. BUTZIEN

(Manuscript received January 29, 1981)

A widely used Bell System pyramidal horn-reflector antenna, used in terrestrial radio relay systems, has been characterized over all space. We give the dB contours for principal (as opposed to cross-polarized) measured radiation patterns and compare them to contours of calculated patterns for this antenna at 4 GHz. In addition, we present the contours for 6-GHz measured principal patterns, describe the measurement method, and discuss the causes of departure of the measured from the calculated patterns.

I. INTRODUCTION

Since its conception in the late forties¹ and its implementation in the fifties,^{2,3} thousands of pyramidal horn-reflector antennas have been installed in terrestrial radio systems. The antenna is a broadband device currently used in the 2-, 4-, 6-, and 11-GHz common-carrier bands. Antenna dimensions are given in Fig. 1, along with coordinates as used below. Discussion of its performance will be limited in this paper to the 4- and 6-GHz bands which constitute the major usage.

Although the performance of this antenna is well known in the xz and in the yz planes (see Fig. 1), its exact performance in directions other than these principal planes was unknown until the measurements reported here were made. The three-dimensional responses are of interest in any case where an interfering signal impinges upon the antenna outside of these planes, e.g., direct interference from other terrestrial stations or satellites, or reflections from buildings or the ground.⁴ Calculated three-dimensional responses based on aperture fields have been reported for small angles from the main beam.⁵ Here, we also include calculated three-dimensional responses for the 4-GHz band out to angles of 30° from the main beam.

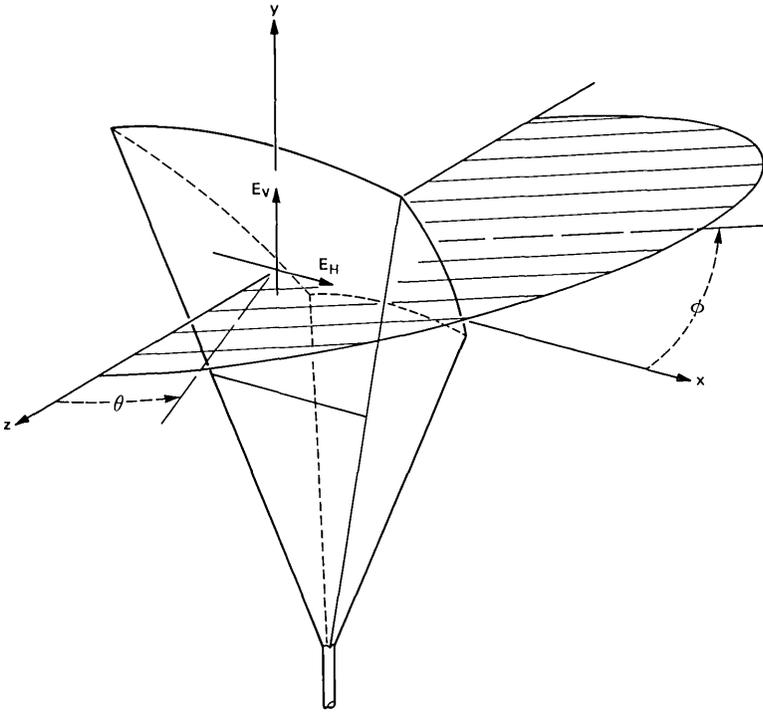


Fig. 1—Pyramidal horn-reflector antenna with coordinate system. Aperture dimensions are: height in center of aperture 2.7 m, width in center of aperture 2.4 m. Apex half-angle is $14^{\circ}31'$.

II. MEASUREMENT SYSTEM

A computer-controlled measurement system was designed by the author to deal with the extensive data involved in these measurements. The signal level was measured accurately to 60 dB below the main beam, using a pulse technique (using pulses of less than 50-ns effective length) to exclude reflections from objects adjacent to the antenna range. Linear polarizations depicted by E_V and E_H in Fig. 1 were recorded as principal patterns simultaneously with their respective cross-polarized components (cross-polarized patterns), using two receivers covering a range of 360° in θ in each of 91 cuts defined by 1° intervals in ϕ . Symmetry of the antenna about the yz plane obviates the necessity of changing ϕ by more than 90° . In all cuts, the antenna range source was copolarized with either E_V or E_H , as the case may be, when θ was zero. This measurement method has the distinct advantage of each cut going through the main beam so that the effects of drifting signal levels or other anomalies can be eliminated.

III. SPATIAL CONTOURS

We give the measured and calculated three-dimensional responses of the pyramidal horn-reflector antenna in the form of contour maps, using 10-dB contour intervals with the exception of 3 dB near bore-

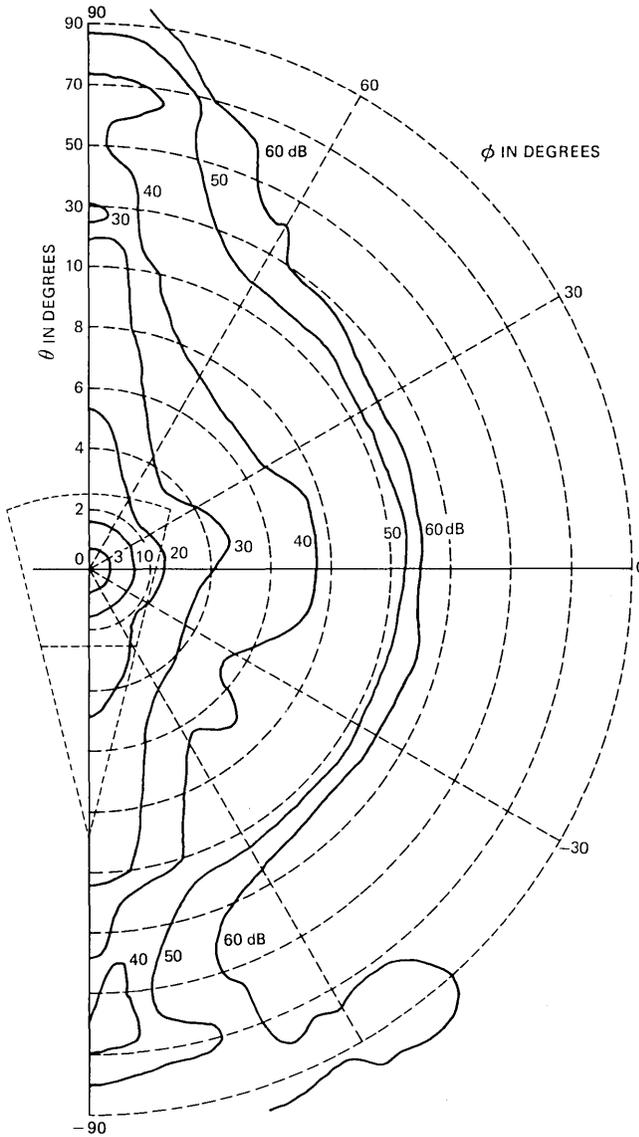


Fig. 2—Measured radiation envelope contours for pyramidal horn-reflector antenna. Principal pattern E_V at 3.9 GHz.

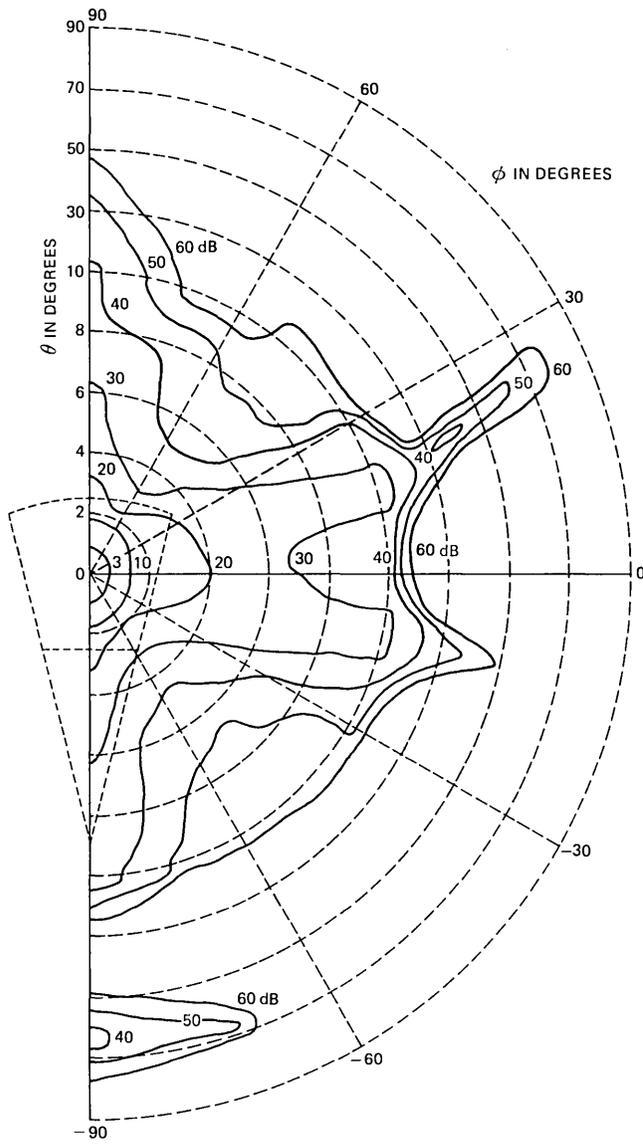


Fig. 3—Measured radiation envelope contours for pyramidal horn-reflector antenna. Principal pattern E_H at 3.9 GHz.

sight. Isometric plots of the measured data have been published for the 4-GHz band.⁴ The 6-GHz three-dimensional patterns are quite similar but with narrower lobes. In Figs. 2 through 7, we show the contour maps for the principal patterns at 4 and 6 GHz. Notice that a scale change is implemented at 10° in θ to give more detail near

boresight. Data for the rear hemisphere of the antenna is not shown since only the 4-GHz vertical polarization has any contours below 60 dB in that space. Both the top ($\phi = 90^\circ$) and the bottom ($\phi = -90^\circ$) exhibit levels between 50 and 60 dB for a small angular region behind the antenna at 4 GHz (vertical polarization) only. Figures 2 and 3 depict the radiation contours of measured performance in the 4-GHz

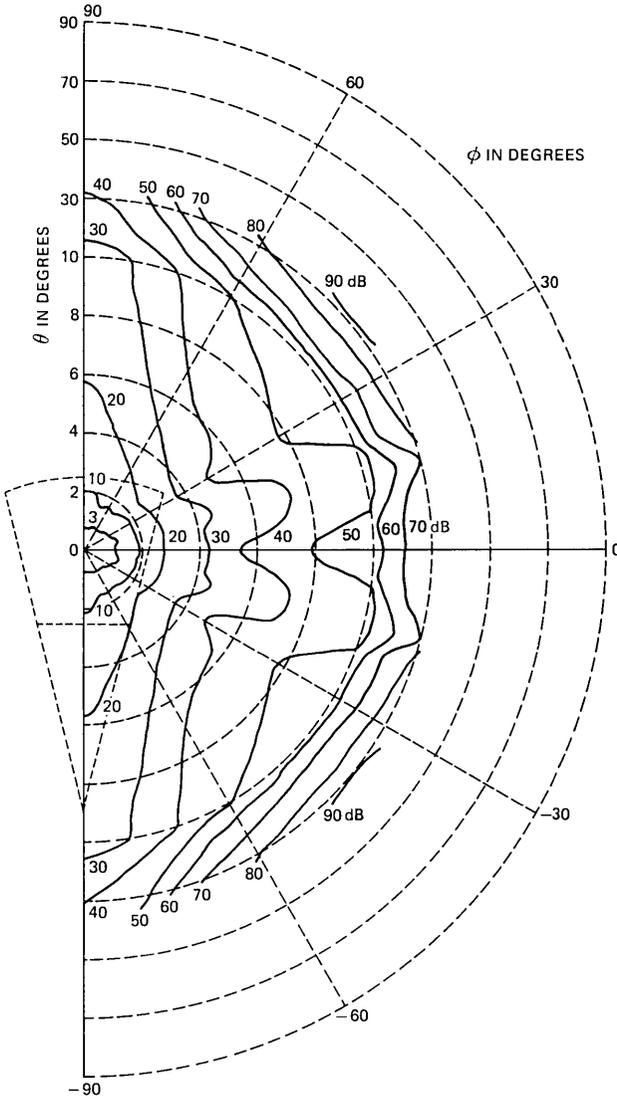


Fig. 4—Calculated radiation envelope contours for pyramidal horn-reflector antenna. Principal pattern E_V at 3.9 GHz.

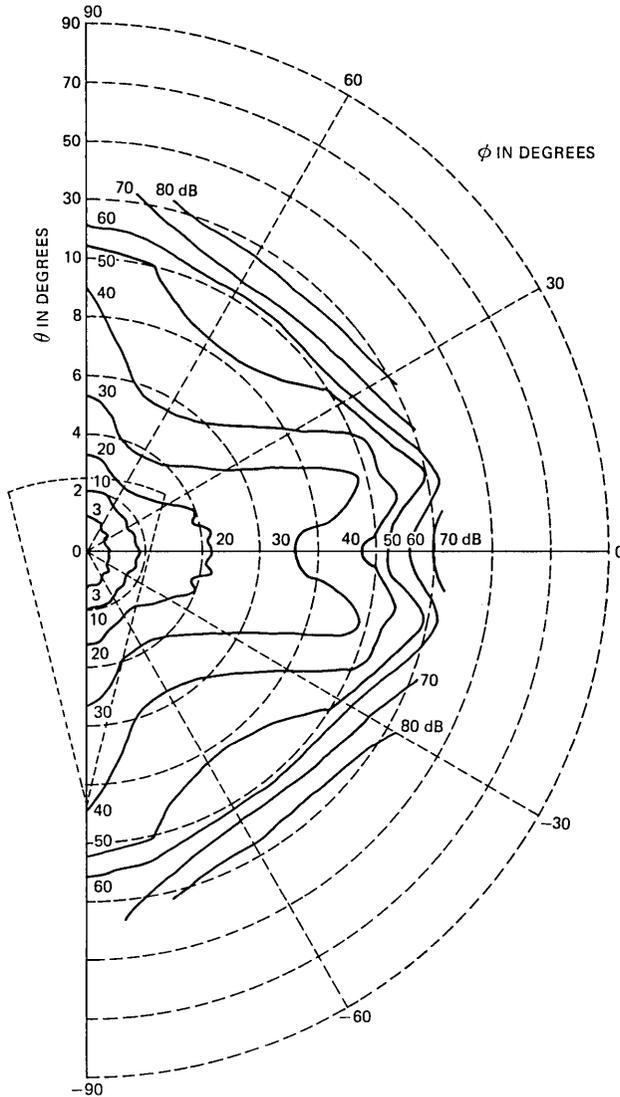


Fig. 5—Calculated radiation envelope contours for pyramidal horn-reflector antenna. Principal pattern E_H at 3.9 GHz.

band. These contours represent the envelope of the rapidly oscillating sidelobe structure where the envelope is always greater than or equal to the actual sidelobe levels. Envelopes are formed by going through the sidelobe peaks with smoothed interpolation. Thus, some very low-level lobes may be masked by larger lobes on each side. The contours may be thought of as giving the worst case response of the antenna at

any three-dimensional point. Since the envelope changes vary slowly with frequency, only one measurement frequency was chosen in the middle of each band.

Figure 2 shows the measured principal pattern E_V . The cross-polarized pattern is not given, but may be found qualitatively in Ref. 4.

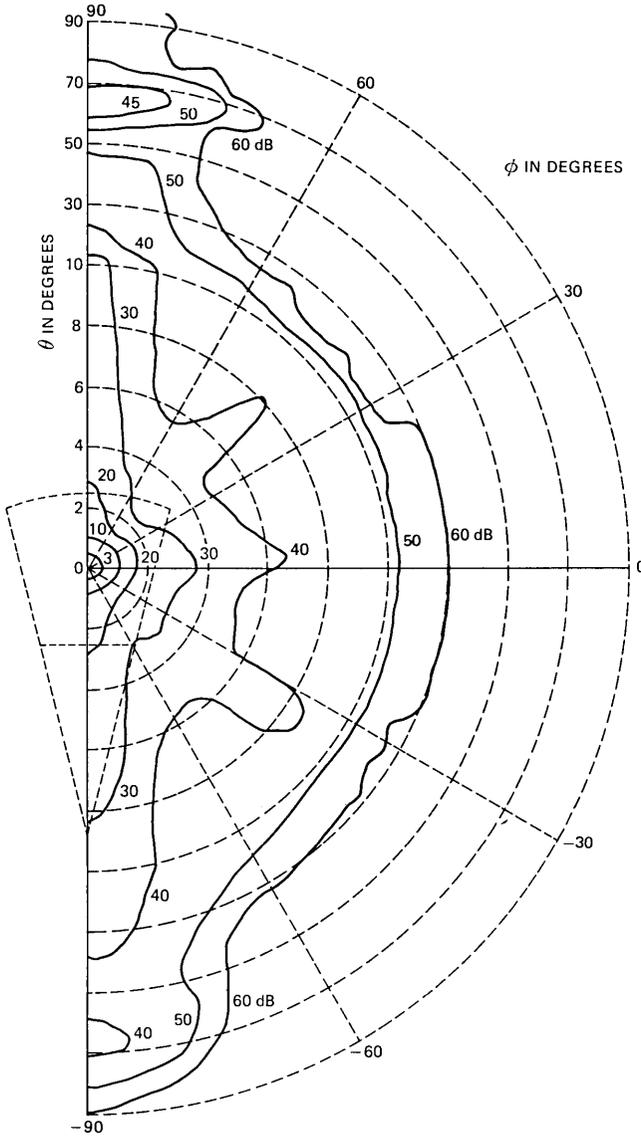


Fig. 6—Measured radiation envelope contours for pyramidal horn-reflector antenna. Principal pattern E_V at 6.175 GHz.

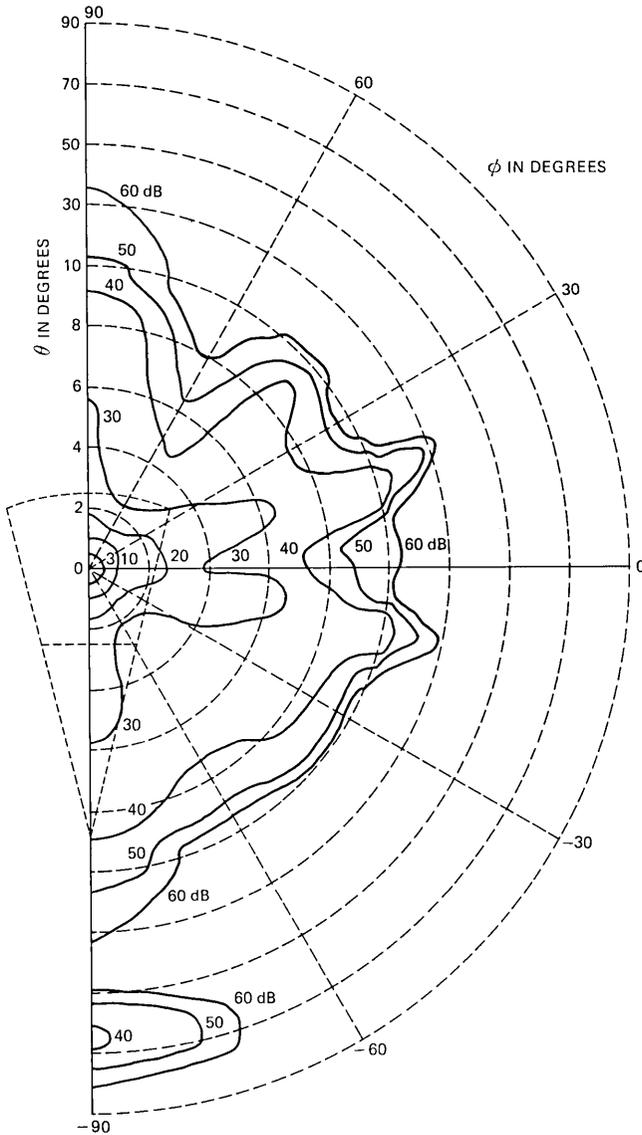


Fig. 7—Measured radiation envelope contours for pyramidal horn-reflector antenna. Principal pattern E_H at 6.175 GHz.

Figure 4 is the calculated response of the same antenna, with the same polarization used in Fig. 2, again, depicting the envelope of the response. The calculation is based on the electric-field aperture distribution resulting from the dominant mode launched by the antenna feed. Responses measured and calculated near the main lobe (θ near

zero) and near the horizontal plane (ϕ near zero) are in quite good agreement. Notice, however, that in the vertical plane (ϕ near $\pm 90^\circ$) broad lobes of approximately 40 dB level at their peak exist in both the upper and the lower regions of the measured, but not the calculated, data. The upper lobe, around $\theta = 70^\circ$, is caused by spillover from the pyramidal feed and the lower lobe is caused by reflections from the flat weather cover. The "spillover lobe" is especially strong for vertical polarization, but the "weather cover lobe" is almost independent of polarization. A bottom edge blinder consisting of a flat (1.784×1.626 m) aluminum plate attached to the bottom rim of the aperture is now available to eliminate the weather cover lobe. The principal response for the horizontal polarization is shown as measured in Fig. 3 and as calculated in Fig. 5. The agreement is, again, good except for the spillover region and the weather cover lobe. The pronounced ridges near ϕ angles of $\pm 15^\circ$ are caused by aperture shape effects (Ref. 5).

Measured responses in the 6-GHz band are given in Figs. 6 and 7. The contours in general tend to be closer to the main beam and the main beam is smaller, as would be expected for the larger gain of the antenna at this higher frequency. Again, the spillover and weather cover effects are evident.

IV. SUMMARY

We have given contours of the measured three-dimensional principal responses of the Bell System pyramidal horn-reflector antenna in the two major usage bands of 4 and 6 GHz, along with contours of the calculated response at 4 GHz. These contours are useful in determining the magnitude of interfering signals at any direction.

V. ACKNOWLEDGMENTS

It is with pleasure that I acknowledge contributions of C. P. Bates, E. R. Fladager, A. J. Giger, M. J. Pagonos, J. R. Richard, C. A. Siller, Jr., and H. C. Wang in this effort.

REFERENCES

1. H. T. Friis, "Microwave Repeater Research," B.S.T.J., 27, No. 2 (April 1948), pp. 183-246.
2. A. T. Corbin and A. S. May, "Broadband Horn Reflector Antenna," Bell Labs Record, 33, No. 11 (November 1955), pp. 401-4.
3. R. W. Friis and A. S. May, "A New Broad-Band Microwave Antenna System," AIEE Trans., Part I, 77 (March 1958), pp. 97-100.
4. P. E. Butzien, "Radio System Interference from Geostationary Satellites," IEEE Trans. Commun., COM-29, No. 1 (January 1981), pp. 33-40.
5. C. A. Siller, Jr., "The Influence of Aperture Shape on the Radiation Patterns of Horn-Reflector Antennas," IEEE Trans. Ant. Prop., AP-27, No. 3 (May 1979), pp. 435-8.

Noise Caused by GaAs MESFETs in Optical Receivers

By K. OGAWA

(Manuscript received January 8, 1981)

In the application of GaAs metal-semiconductor field effect transistors (MESFETs) in ultra low-noise lightwave receivers, the channel noise is often the dominant effect in determining sensitivity. This paper analyzes for the first time the excess channel-noise factor Γ for GaAs by considering the effect of circuit capacitance, as well as gate-to-source capacitance on the correlation of gate and channel fluctuations, and derives a useful and analytic expression for Γ . For example, we find that Γ for practical GaAs MESFET amplifiers can be much larger than 1.1 as is usually assumed. The multiplication factor, Γ is approximately 1.75 for the practical GaAs MESFET with 1- μm gate length, which explains the discrepancy between the optical sensitivity from the noise calculation and experiments.

I. INTRODUCTION

The GaAs metal-semiconductor field effect transistor (MESFET) originally designed for microwave applications has become an important component of lightwave receivers used in communication applications. Unlike most microwave applications, the lightwave-receiver application requires a consideration of induced gate noise and correlation with the channel noise. Van der Ziel's original evaluation of the noise contribution from this component¹⁻³ was later extended by Baechthold to include effects present in MESFETs with short gate length, as well as the intervalley scattering in GaAs.^{4,5} The following computation extends this earlier work to determine the noise factor Γ .

This factor relates the input noise current i_{nt} resulting from all noise sources of the FET to the FET transconductance g_m such that

$$\langle i_{nt}^2 \rangle = 4kT\Gamma(\omega C_T)^2 \Delta f / g_m, \quad (1)$$

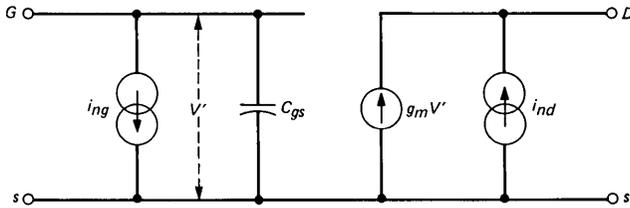


Fig. 1—Equivalent circuit of the intrinsic field-effect transistor with channel noise current i_{nd} and induced gate noise current, i_{ng} .

where C_T is the total input capacitance, k is Boltzman's constant, and T is absolute temperature.

Figure 1 shows an equivalent circuit of the intrinsic field-effect transistor with two noise sources, the channel thermal noise and the induced gate noise sources. We neglect other well-known noise sources, such as the gate leakage current noise and the flicker noise, because our main purpose is to evaluate the noise factor Γ .

The channel noise is described in the equivalent circuit by the noise current i_{nd} having the mean square

$$\langle i_{nd}^2 \rangle = 4kT \cdot P \cdot g_m \cdot \Delta f, \quad (2)$$

where P is a factor depending on various FET parameters and the gate bias.⁴

As explained in Ref. 2, every disturbance in the channel potential introduces a gate voltage disturbance and, in turn, a channel current fluctuation. The mean square of the induced gate noise current is

$$\langle i_{ng}^2 \rangle = 4kT \cdot R \cdot \frac{(\omega C_{gs})^2}{g_m} \cdot \Delta f, \quad (3)$$

where C_{gs} is the gate source capacitance and R is a factor depending on various FET parameters and the gate bias.⁴ Since these two noise sources have the same origin, a correlation exists. It can be expressed in the form

$$\langle i_{ng}^* i_{nd} \rangle = j4kT \cdot Q \cdot (\omega C_{gs}) \cdot \Delta f, \quad (4)$$

where Q is a factor depending on various FET parameters and the gate bias.⁴

Figure 2 shows P , Q , and R for GaAs MESFETS of various gate lengths parameters and gate bias parameters.⁵

We evaluate the total input noise by transferring i_{nd} to the input as shown in the circuit of Fig. 3. In this circuit, the input admittance Y is defined by

$$Y_{in} = G + j\omega C_T, \quad (5)$$

capacitance C_T consists of the gate capacitance C_{gs} and a capacitance C_s comprising the photodiode capacitance of the circuit stray capacitance. The mean square of the total input noise current becomes

$$\begin{aligned} \langle i_{nt}^2 \rangle &= \langle (i_{ng} + i_{nd} Y_{in}/g_m)(i_{ng}^* + i_{nd}^* Y_{in}^*/g_m) \rangle \\ &= 4kT \left[P - 2Q \left(\frac{C_{gs}}{C_T} \right) + R \left(\frac{C_{gs}}{C_T} \right)^2 \right] \frac{(\omega C_T)^2}{g_m} \Delta f \\ &\quad + 4kTP \frac{G^2}{g_m} \Delta f. \end{aligned} \tag{6}$$

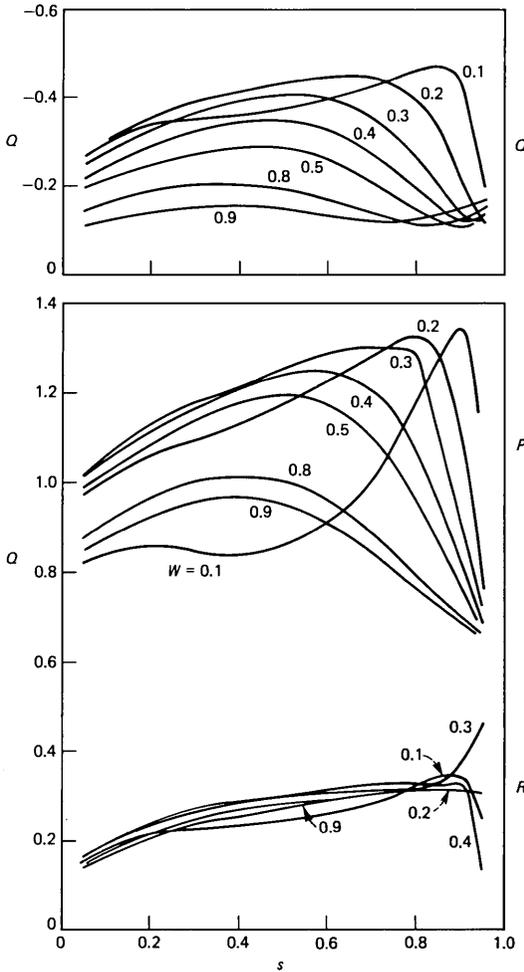
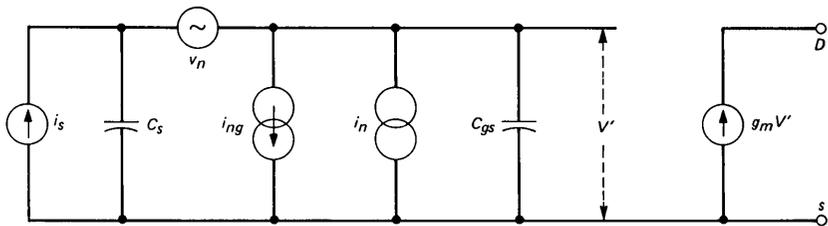


Fig. 2—The dependence of P , R , and Q on normalized gate voltage $s = [(0.8 - V_g)/W_0]^{1/2}$. The normalized gate length $W = E_s L/W_0$ is the parameter used. W_0 is the pinch-off voltage, E_s is the saturation field (4kV/cm), and L is the gate length.⁵



$$\langle v_n^2 \rangle = \langle i_{nd}^2 \rangle / g_m^2$$

$$\langle i_n^2 \rangle = \langle i_{nd}^2 \rangle (\omega C_{gs})^2 / g_m^2$$

Fig. 3—Equivalent circuit as in Fig. 1 but with channel noise i_{nd} transformed into input current and voltage noise sources.

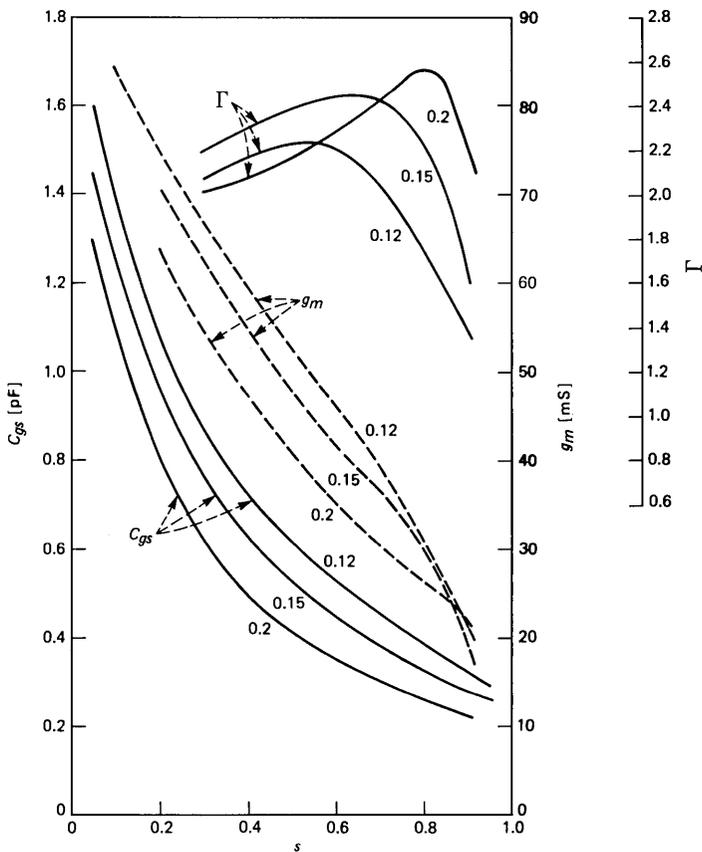


Fig. 4—Transconductance g_m , capacitance C_{gs} and Γ as functions of the normalized gate bias voltage s for $C_{gs} = C_T$. The channel depth in μm is the parameter.

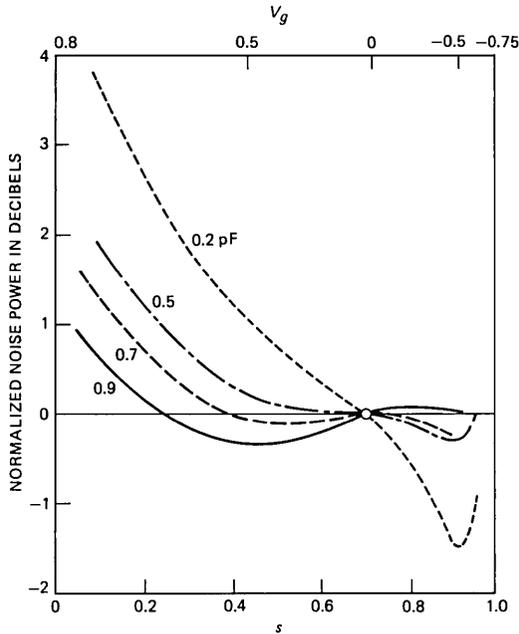


Fig. 5—FET noise power normalized with respect to noise power at zero bias voltage as a function of the normalized bias voltage s . The additional input capacitance C_s in pF is the parameter.

Since commonly $G \ll g_m$ in optical receiver design, the comparison of (1) and (6) yields

$$\Gamma = P - 2Q \left(\frac{C_{gs}}{C_T} \right) + R \left(\frac{C_{gs}}{C_T} \right)^2. \quad (7)$$

As shown in Figs. 2 and 4, P , R , Q , g_m , and C_{gs} are functions of the bias voltage and the FET gate length. However, as Fig. 5 shows, the noise

Table I—Noise factor Γ for 1- μm gate length and various C_s ($C_{gs} = 0.5$ pF)

| C_s (pF) | Γ |
|------------|----------|
| 0.1 | 2.164 |
| 0.2 | 2.009 |
| 0.3 | 1.897 |
| 0.4 | 1.814 |
| 0.5 | 1.749 |
| 0.6 | 1.697 |
| 0.7 | 1.655 |
| 0.8 | 1.619 |
| 0.9 | 1.589 |

$\langle i_{nt}^2 \rangle$ changes little in the bias voltage range between -0.5 V and $+0.5$ V as long as C_s is between 0.5 pF and 0.9 pF in spite of the voltage dependency of P , Q , and R .

As a good approximation, one can use the noise parameters determined at zero gate bias voltage for the entire operating range of the transistor. Table I shows Γ at zero bias voltage for various capacitance values C_s , in the case of an FET gate length of $1 \mu\text{m}$, a gate width of $400 \mu\text{m}$ which have the following parameters: $P = 1.24348$, $Q = -0.42384$, and $R = 0.30329$.

REFERENCES

1. A. Van der Ziel, *Noise: Sources, Characterization, Measurement*, Englewood Cliffs, New Jersey: Prentice-Hall, 1970.
2. A. Van der Ziel, "Gate Noise in Field Effect Transistors at Moderately High Frequencies," *Proc. of IEEE*, March, 1963, *51* (March 1963), pp. 461-7.
3. A. Van der Ziel, "Thermal Noise in Field Effect Transistors," *Proc. IRE* *50* (August 1962), pp. 1808-12.
4. W. Baechtold, "Noise Behavior of Schottky Barrier Gate Field-Effect Transistors at Microwave Frequencies," *IEEE Trans. Electron. Device*, *ED-18*, No. 2 (February 1971), pp. 97-104.
5. W. Baechtold, "Noise Behavior of GaAs Field Effect Transistors with Short Gate Lengths," *IEEE Trans. Electron Device*, *ED-19*, No. 5 (May 1972), pp. 674-80.

T-CARRIER Characterization Program—Overview

By J. R. DAVIS and A. K. REILLY

(Manuscript received January 22, 1980)

This is the first of a series of four papers describing a characterization program aimed at building a better understanding of the makeup and performance of the T-carrier network. In this paper, we provide a brief historical background of T-carrier and show the need for an increased understanding of the network and its performance. We also describe the characterization program conducted in conjunction with several operating telephone companies.

I. T-CARRIER EVOLUTION

T-carrier was first put into service in 1962 to provide interoffice trunks in the operating telephone company (OTC) plant. The carrier consisted of the T1 digital repeatered line and its associated terminal, the D1 channel bank. The channel bank encodes 24 voice-frequency (4kHz) channels by pulse-code modulation and multiplexes them by time-division into a 1.544-Mb/s pulse stream for transmission over the T1 digital line, which operates in the paired cable trunk plant. T1 has grown rapidly in the metropolitan trunk plant and today there exist more than 160,000 T1 systems operating on more than 5,000 interoffice spans. Nearly half of all metropolitan interoffice trunks are now provided by T-carrier.

Since the introduction of T1, two additional paired cable systems have been deployed in the trunk plant. The first of these is T2—introduced in 1971—which provides 96 voice-frequency channels and is configured for intercity applications. The second is T1C, which was introduced in 1975 and provides 48 voice-frequency channels primarily in the metropolitan area trunk network. T1C accounts for a large fraction of the growth in the metropolitan trunk plant. New versions of digital terminals have been developed as well. Direct digital interfaces between digital lines and switching machines have been deployed with certain switches. Still others are being developed. Such arrange-

ments can, for example, avoid the need to convert digital carrier signals to voice-frequency before interfacing with digital switching machines.

II. INCREASING IMPORTANCE OF T-CARRIER PERFORMANCE

The original service for which T1 was engineered and used was POTS, or "plain old telephone service." Now, special services account for approximately 40 percent of the T-carrier capacity. In addition, however, the T-carrier plant is considered more and more as a part of an evolving overall switched digital network. The present metropolitan digital trunk and switching plant is interfacing with the emerging digital technologies in the loop plant and intercity plants (long-haul and short-haul) as well. Various new services, some of which have more stringent performance requirements than needed for voice are being planned for this switched digital network. While performance objectives can be specified for new transmission facilities consistent with overall objectives, all of these services and networks must use existing T-carrier for the metropolitan trunking portion of the network. At the same time, OTCS want to use the in-place trunk plant more effectively, expanding the carrier capacity of existing cables to defer new cable installations as long as practical. The emergence of new and sometimes more demanding services, as well as the desire to exploit fully the existing trunk plant, place greatly increased importance on an understanding of the performance level of the T-carrier plant, on what elements control that performance, and on what might be done to improve poorly performing systems.

The performance of a T1 digital system—from terminal to terminal—is the sum of the performance of the component repeater sections that are placed in tandem to comprise that system. Thus, given a performance model for a repeater section, the overall network performance can be modeled with a knowledge of the parameters and statistics of repeater sections, interoffice spans, and systems. In addition, these statistics are useful in examining in a more general way the nature of the existing metropolitan trunk plant, the magnitude of the remaining exploitable capacity, and the need and market for various new trunk transmission technologies in relation to the existing plant and its projected growth.

III. CHARACTERIZATION PROGRAM

Given the increasing importance of understanding T-carrier performance, as well as the need for knowledge of the physical makeup of the T-carrier plant, a T-carrier characterization program was chartered in the summer of 1976. Since the basic element of overall system performance is the performance of the individual repeater sections

that comprise the system, we decided to start the performance characterization with the individual repeater section. The objective of this program was to determine the statistical distribution of margin and other parameters of the repeatered line. (Margin is the amount of signal-to-noise degradation that can be tolerated before the error-rate objective is violated.) Among the other parameters to be determined were those properties of the cable, repeaters, apparatus cases, and section layout which affect section margin.

To allow a variety of measurements to be made rapidly and accurately on a repeater section, we developed an automated measurement system. This system consists of various transmission equipment (i.e., digital signal sources, meters, etc.) placed in two vehicles, one located at each end of the repeater section under test. A minicomputer controls the switching of various equipment for different measurements at its location and at the other location as well, via telemetry provided over a voice-frequency pair. In addition, the minicomputer controls the entire sequence of tests and records the results on both paper printout and cassette tape. A special computer language was developed to permit a flexible program structure that is easy for people not experienced with computer programming to understand, and that allows very rapid change of the measurement algorithms and sequences if the need arises. The automated measurement system provides rapid, accurate measurements to be made by personnel with a minimum of experience and training.

To provide a measure of repeater end-section performance, a companion set of equipment was developed. (The end section is the section adjacent to a central office.) Since switching noise in central offices was known to be a potential problem when T1 was first engineered, the end sections are limited in loss or length to approximately two-thirds of the value allowed for intermediate sections. Thus, sources of crosstalk, which potentially limit T1 performance in intermediate sections, are much less influential. The central office measurement equipment was designed specifically to examine the characteristics of the noise on the incoming T1 pairs. A complete description of both measurement systems and the measurements made are provided in the paper on the automated measurement system.

The automated measurement system has been used to examine the performance of more than 2000 T1 repeatered lines in 22 different intermediate repeatered sections in three OTCs, as well as end sections terminating at central offices in two companies. These sections were chosen on the basis of having characteristics that should limit the performance of digital systems containing those sections, based upon the a priori understanding of T1 system performance; they are generally maximum-length sections operating on full-sized (900- and 1100-

pair) 22-gauge pulp cables, by far the most common cable type in the present metropolitan T-carrier plant. The third paper describes field measurements results, details a model that accurately characterizes the T1 repeater section performance, and establishes parameters for the model.

The description of the physical makeup of the T-carrier plant is given in the paper on the metropolitan digital trunk plant. This paper includes basic statistics of the interoffice T-carrier spans that comprise the digital systems. Various OTC records, both mechanized and manual, were obtained to provide this information. The data represent a sample of seven companies that have more than 25 percent of the metropolitan digital trunk plant.

The Metropolitan Digital Trunk Plant

By J. R. DAVIS, L. E. FORMAN, and L. T. NGUYEN

(Manuscript received January 22, 1980)

This paper statistically describes the physical makeup of the T-carrier metropolitan trunking plant. The basic building block of a digital carrier system is the interoffice span. This span is the collection of all the repeatered lines, or span lines, between two adjacent central offices. Span lines are connected in tandem to form systems. Statistics for interoffice span characteristics of length, cross section, percentage fill and growth are given and discussed for seven metropolitan areas.

I. INTRODUCTION

The metropolitan trunk plant consists of a combination of central office equipment and outside plant cables and electronics. These provide a collection of voice-frequency (VF) circuits to interconnect central offices. At first these circuits were provided solely by loaded VF wire pairs in underground cables located in ducts. As the network grew and traffic increased, ducts became congested and ways were developed to provide additional circuits on the same wire pairs. In 1962, T1 carrier was introduced to provide these circuits and now accounts for a large fraction of the metropolitan trunk plant capacity. In 1975, T1C, operating at twice the bit rate of T1, was introduced and provides a growing fraction of the plant capacity.

The T-carrier plant now contains approximately 160,000 T-carrier systems consisting of some 6,000 interoffice spans providing interoffice connections throughout metropolitan areas. Each T-carrier system provides 24 voice circuits by means of a digital channel bank at each end interconnected by a T-carrier span line (or several span lines in tandem traversing intermediate offices). A D-channel bank converts the 24-voice circuit inputs to a DS1-level, 1.544-Mb/s pulse stream, which is transmitted over the T-carrier span line and is reconverted to 24 voice circuits by the channel bank at the other end. T1C provides 48 channels at the DS1C level, 3.152 Mb/s. The recently introduced

D4 channel bank provides an option for 48 voice circuits for direct interconnection to T1C, as well as 24 circuits for connection to T1.

For the past 3 years, studies have been conducted to characterize both the physical configuration and electrical performance of the metropolitan digital plant to:

(i) Allow performance analysis of existing systems and examine the potential for higher-capacity systems in the existing plant.

(ii) Provide a base of information to be able to examine the applicability of new transmission facilities.

This paper statistically describes various aspects of the physical makeup of the plant and its short-term growth as projected by current planners. Companion articles in this issue provide field measurement results and a description of automated measurement instruments developed and deployed to obtain them. To study the physical plant, span data were collected from six operating telephone companies (OTCs) covering seven metropolitan areas.

These data represent approximately 1,400 spans or about 25 percent of the total digital spans in the Bell System. However, not all data were available for each span, since each company has its own collection of records, both manual and mechanized. Portions of the span data were obtained from two sources: the physical layouts and cable information were obtained from the outside plant staff, while the growth information was provided by the current planning organization. Since requests for these data were made over a two-year period, the effective date of the information varies from company to company. This paper gives the results of an examination of the physical characteristics of the spans in the metropolitan network. Limited system data from two areas show that the T1 systems average 21 miles long and pass through 4 to 5 spans.

Section II gives statistics of the fundamental parameters of spans for the entire sample, and compares these statistics for the individual metropolitan areas. Section III gives length statistics for individual repeater sections and Section IV relates various aspects of short-term growth projected by the OTC current planners and examines the nature of the growth. Section V summarizes the data. Since individual OTCs are represented only on a sampled basis, the data may not be representative of the entire plant. Thus, each metropolitan area is indicated by a number rather than a name to avoid specific identification. The individual area statistics give an indication of the variation expected over individual metropolitan areas, while the composite statistics give a measure of the overall Bell System metropolitan plant.

II. SPAN STATISTICS

Nearly all metropolitan central offices are interconnected by T-

carrier. Span lines, interconnected as needed between terminal offices, form digital systems. Thus, from the planning and provisioning standpoint, the span is the basic building block of T-carrier. This section examines the following span characteristics: length, cross section, cable and apparatus case fill, and the composition of the cable plant in terms of cable size, gauge and insulation type. Results are given for the total sample and the seven metropolitan areas are compared.

The spans and cables included in this sample are qualified in several ways. Only those spans with at least one intermediate repeater are counted in this sample. Short cables connecting two close adjacent buildings or two floors within the same building, called tie cables, are omitted. Only cables that have T-carrier presently operating in them are included. Cables containing only vF trunks often run in parallel with T-carrier cable, but these vF cables are not included. While not all parameters were available for all spans, the maximum number of available data points was used to generate each of the statistics.

2.1 Span length

Span length, or central office-to-central office spacing, is a fundamental span characteristic. The number of manhole repeaters, powering voltage required, etc., are directly dependent upon span length. The extent of the application for a transmission facility with span length restrictions can be determined from a knowledge of the distribution of span lengths.

Figure 1 shows the cumulative probability distribution of span lengths for the total sample. This distribution and others are shown on a normal probability scale so that if the sample were Gaussian or normally distributed, the points would lie on a straight line. Since the curve shows a rather constant bow over the entire distribution, the distribution is not normal; the mean of the distribution is 5.8 miles and the median is 5.4 miles. The standard deviation is 3.2 miles and 68 percent of the spans lie between 2.6 and 8.6 miles, the mean plus and minus the standard deviation. Eighty percent of the spans have lengths between 2 and 10 miles, and 50 percent have lengths between 3.7 and 7.5 miles, the lower and upper quartiles.

The span length distributions for each of the seven metropolitan areas are not normal as well, so quartiles and medians (25, 50, and 75 percent points) are used to compare them. Figure 2 shows the span length distribution for these seven areas, identified by number. The upper and lower bars at the end of the vertical lines indicate the maximum and minimum length in each area. The bar at the center of the box indicates the median, and the end of the boxes are the upper and lower quartiles of the distribution. The width of each box is proportional to the square root of the number of spans in that distri-

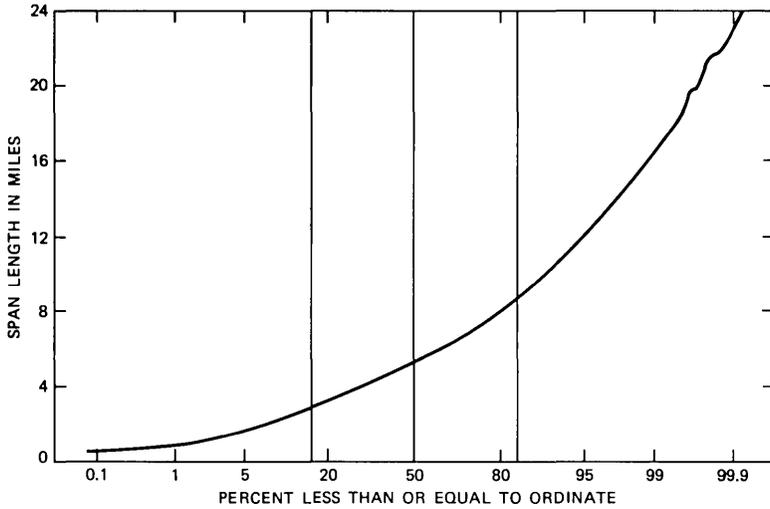


Fig. 1—Distribution of span lengths in seven metropolitan areas. Number of spans = 1398; average 5.8 miles; standard deviation = 3.2 miles.

bution to give an indication of the relative number of spans in each area.

There appear to be two groups of areas: 2, 5, 6, and 7 with medians of about 5.9 miles, and areas 1, 3, and 4 with medians of about 4.6 miles. The 4.6-mile median group has more densely populated metropolitan areas, and thus, has shorter spans. The overall quartiles and median are indicated by the arrows.

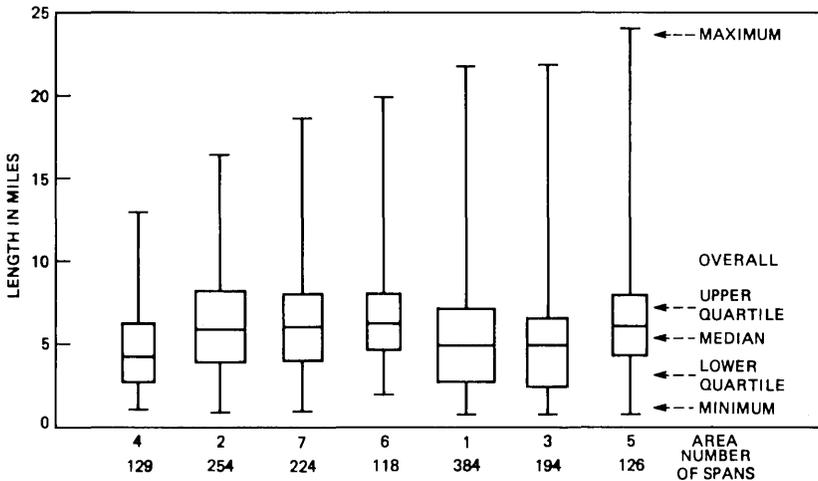


Fig. 2—Distribution of span lengths in seven metropolitan areas.

Six out of seven areas have spans less than one mile long. These short spans generally use 24- or 26-gauge cable and have one intermediate repeater.

2.2 Span working cross section

Span working cross section is defined as the total number of DS1-level signals currently carrying traffic in a span. Maintenance and spare lines are not included. This cross section is on a span basis, and may include systems in several parallel cables. DS1 signals are used as a common denominator; thus, one T1C line with a DS1C-level signal is counted as two DS1 signals.

Figure 3 shows the distribution of span cross sections for the collected sample. For purposes of comparison, each area's cross section was adjusted, using current planning information, to a common date of January 1980, the middle of the period for which growth data have been obtained. As can be seen, this distribution is not normal; the mean is 137, but the median is 90. This mean corresponds to a cross section of approximately 3300 vF trunks, 24 per DS1. Eighty percent of the spans have cross sections between 10 and 330 DS1s, while 50 percent lie between 20 and 180 DS1s. Note that 78 percent of the spans have cross sections less than 200, which can be supplied by a single 900-pair cable.

As with span length, the seven metropolitan areas are compared as shown in Fig. 4. The areas fall into two groups: area 7 with relatively

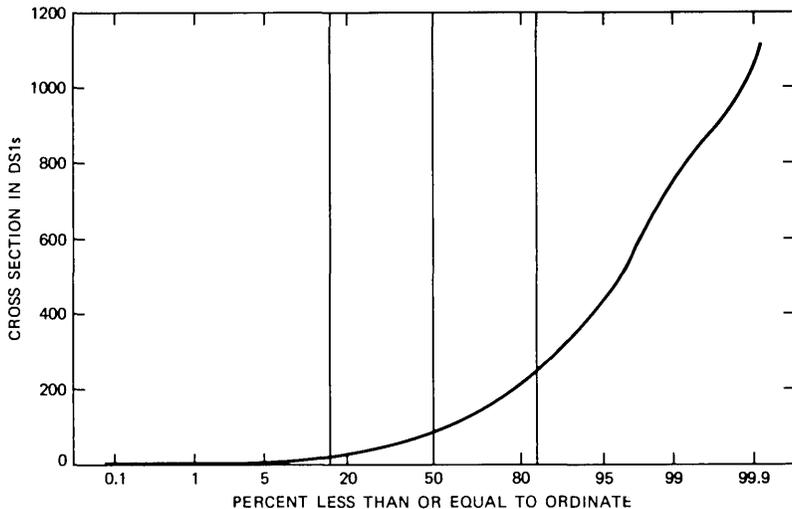


Fig. 3—Distribution of cross section in DS1s in seven metropolitan areas in January 1980. Number of spans = 1339; average = 136.8 DS1s; standard deviation = 151.3 DS1s.

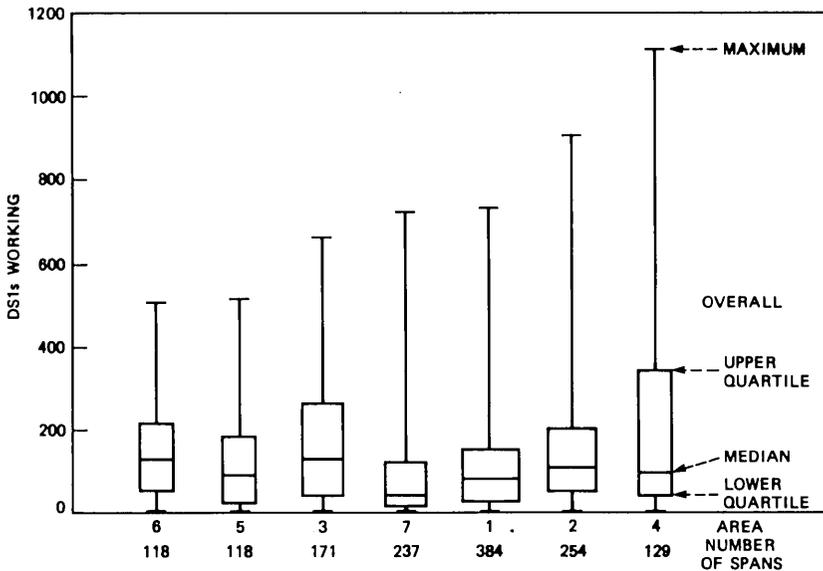


Fig. 4—Distributions of cross sections in seven metropolitan areas.

thinly populated suburbs as well as metropolitan areas resulting in a median of 41, and the other 6 areas with medians ranging from 80 to 130. In addition, areas 3 and 4 stand out by having many spans with cross sections greater than 200, more than twice the overall median. Area 4 has the maximum cross section span in the total sample of 1100 DS1s and a natural corridor of population causing the many high cross-section spans.

2.3 Span length versus working cross section

Combining span length and cross section gives a more complete picture of T-carrier spans than either quantity alone. Figure 5 is a scatter plot of span length versus working cross section, with spans having cross sections greater than 500 DS1s or lengths greater than 16 miles identified by area number.

The plot shows a wide variation of cross section for any given span length. The large cross-section spans are dominated by area 4 as mentioned in Section 2.2. The four area 2 spans near the center of the plot are part of a long backbone route being constructed in this area primarily with T1C. All areas have one or two spans longer than 12 miles which are disjoint from the rest of that area's spans.

The average cross section per 1-mile interval is shown, with the spans longer than 13 miles averaged together. The average cross section shows a slight decreasing trend with increasing span length.

2.4 Theoretical span carrier-pair fill

Various aspects of existing and potential cable occupancy are important when considering the effective use of existing cable to serve future OTC needs. Among the important characteristics are the number of pairs

currently in use for T-carrier,
spliced to T-carrier apparatus cases,
that can ultimately support T-carrier,
that are spare (no VF, no T-carrier),
used for VF trunks,
loaded for VF.

Data collected provide very little VF information, so the cable-fill evaluation is limited to the examination of T-carrier fill.

The first aspect of fill examined is the current occupancy of potential T-carrier pairs. This theoretical fill is defined as the ratio of working DS1s to the maximum number of DS1s a cable can support, according to the engineering rules in effect when the cable was placed. The use of middle-ring binder groups for T-carrier is not used in this analysis. Thus, in accordance with T-carrier engineering practice, a maximum of 200 systems in a 900-pair cable, and 250 systems in an 1100-pair cable is assumed, with corresponding numbers for other cable sizes.

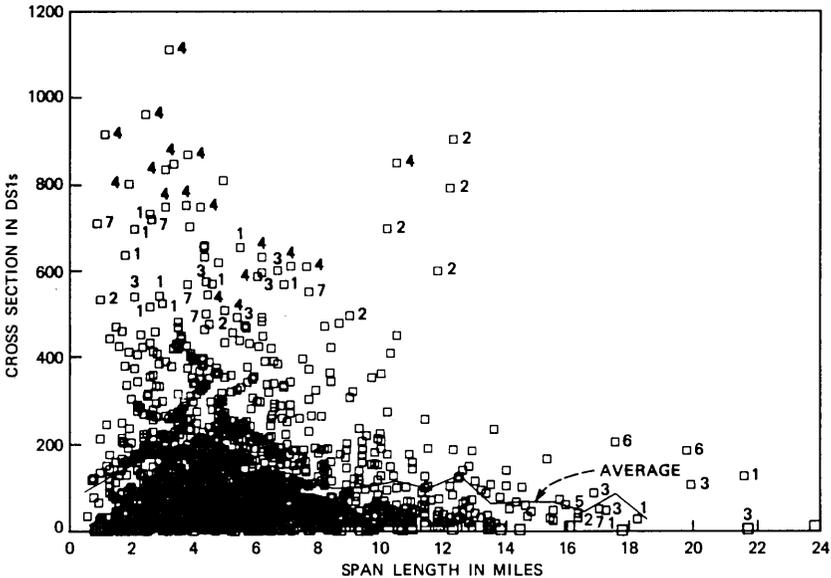


Fig. 5—Span line cross section vs. span length in seven metropolitan areas in January 1980.

Cables in a span are combined so that carrier fill is examined on a span basis. The ultimate capacity of each of the cables was calculated using their current configuration. Resplicing to increase capacity, such as changing from single-cable operation to dual-cable operation, or converting T1 cables to T1C or T1D, was not considered. However, cables engineered for T1C are assumed to fill with T1C, even though they may currently contain T1 electronics.

Four areas provided the type of data needed to calculate carrier fill. Figure 6 shows the distribution of carrier fill on a span basis for these areas. The distribution is nearly normal, with a mean and median of approximately 41 percent, and a standard deviation of 24 percent. Thus, half of the spans are using less than 41 percent of their theoretical T-carrier capacity. In fact, low percentage fill would be expected in small cross-section spans. Recall from Fig. 3 that half of the spans in this seven-area sample have cross sections smaller than 90 DS1s. Most spans use full-size, 900- or 1100-pair cables to make efficient use of duct space. Assuming that all the spans having cross sections of less than 200 DS1s use 900-pair cable, the resulting carrier fill for the seven-area sample is less than 90/200 or 45 percent for half of the spans, reasonably close to the observed value of 41 percent for this four-area sample.

Larger cross-section spans are expected to use more of the available capacity and, thus, exhibit higher-percentage carrier fill. Figure 7 examines the average span carrier fill per 50-DS1 interval versus the

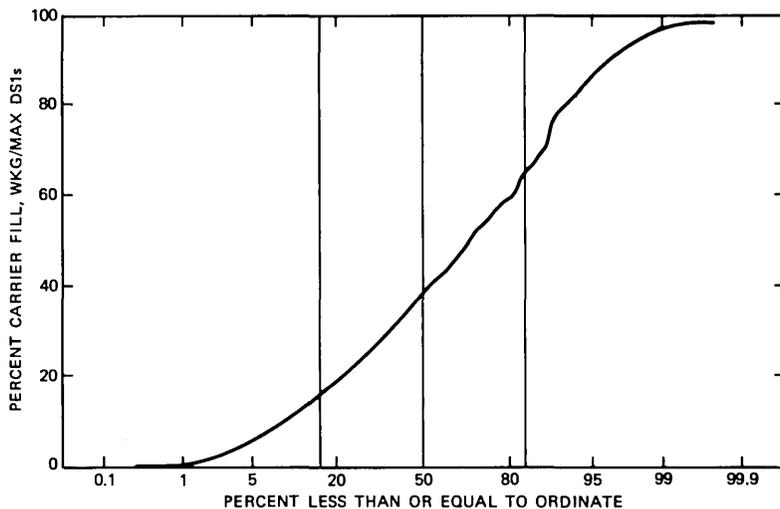


Fig. 6—Distribution of theoretical percent carrier fill on a span basis in four metropolitan areas. Number of spans = 266; average = 40.6; standard deviation = 23.7.

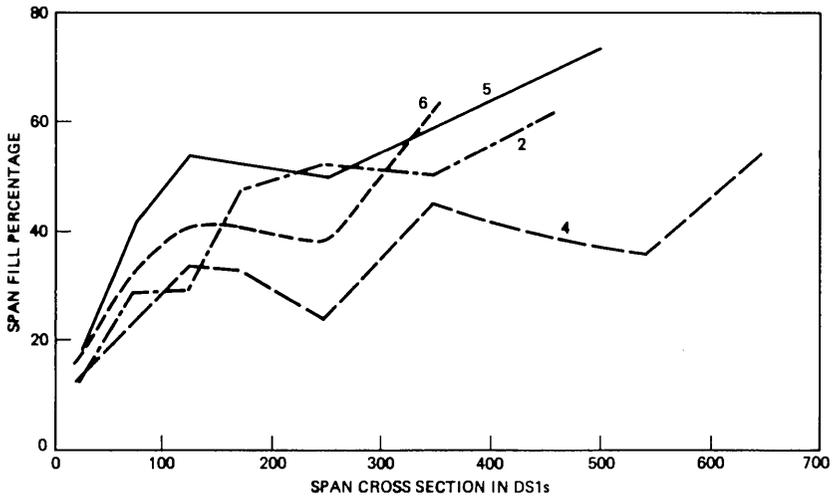


Fig. 7—Theoretical span fill (percent) vs. cross section.

cross section for the same four areas. All four areas show the expected trend of increased-percentage carrier fill with increasing cross section. However, since only a small percentage of these spans have large cross sections where higher-percentage fills lie, the average carrier fill is only 41 percent. The fill of area 4 is lower than the others because dual-cable operation predominates in this area, whereas single-cable operation is prevalent in the other three areas. Although the cables can be completely filled with T-carrier in dual-cable operation, and the fill here is calculated on that basis, the cables were placed when large νF trunk requirements existed, and the cables were used for both purposes. Most dual-cable operations are not and will not be filled with T-carrier, except for newer cables specifically dedicated to T-carrier.

Combining the spans in these four areas results in the distribution of average theoretical carrier fill versus cross section of Fig. 8. Omitting the spans of area 4 (see the dashed line), with their predominant dual-cable operation, gives an increased average carrier fill for all cross sections as shown in the figure. To examine the reasonableness of this carrier fill, assume that all T-carrier cables are 900-pair, single-cable operation. Also, assume that each cable is fully used for carrier (200 systems) before a new 900-pair cable is placed. Then, the span carrier fill, versus cross section, would be as shown by the straight line rays in Fig. 8. The first cable would be completely filled at 200 systems, then a second cable would reduce the percentage fill to 50 percent. The span would then be completely full at 400 systems, requiring a third 900-pair cable, at which time the fill would drop to 67 percent, and so forth.

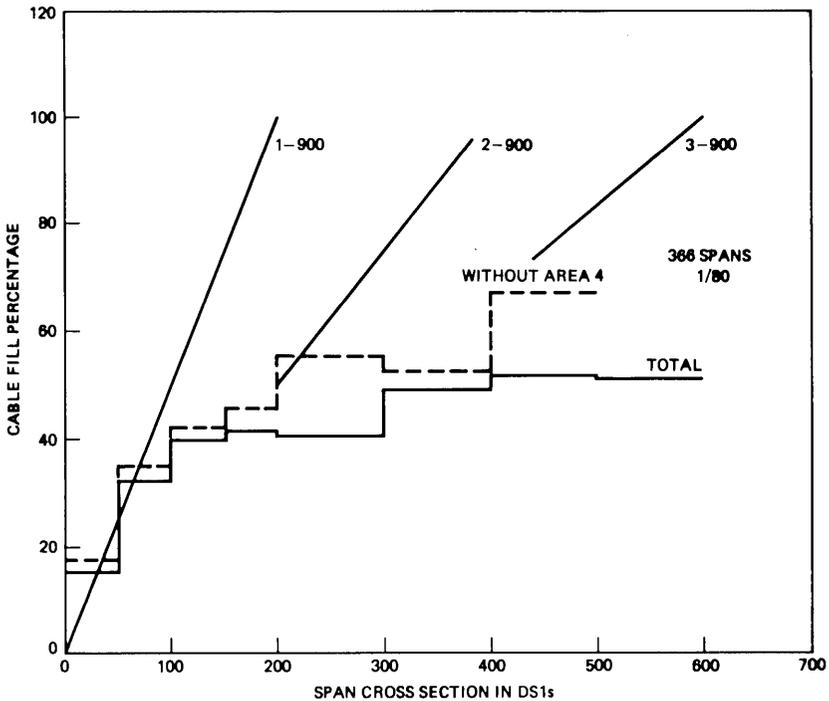


Fig. 8—Theoretical span carrier fill vs. cross section.

Spans with cross sections less than 100 (recall from Fig. 3 that this amounts to 58 percent of the total) follow the one 900-pair cable curve fairly well. However, for cross sections greater than 100, almost no spans have percentage fills as high as expected from this model. This deviation suggests that new cable placement for the metropolitan trunk plant has not been triggered by exhausting T-carrier pair capacity, but rather by VF trunk needs. Pairs remaining for use when additional cable is placed—generally T-carrier pairs—will be enough to reduce T1 fill measurably, but not enough to supply VF trunk needs and delay significantly the placement of additional cable. When cables are added, T-carrier is generally placed on the additional cable, as well as the original for reliability purposes, yielding lower carrier fill for the span than would be the case if the carrier had triggered the exhaust. Limited data show that large working VF complements frequently occupy potential T1 binder groups, in addition to VF binder groups, supporting this hypothesis.

While the carrier fill is relatively low, the percentage of pairs used in the cables is much higher. On the average, more than half the active pairs in a cable and, thus, in a span, are not capable of supporting T-carrier, and are generally loaded for VF trunks. As indicated above,

limited data indicate that the fill of the vF pairs is high, frequently such that binder groups that could support T-carrier are used for vF in addition to the others. Areas 2 and 7 have indicated that they are willing to roll, and have rolled, active vF trunks to free pairs for T-carrier. Thus, the carrier fill as derived here is consistent with the oTC view for these areas, since pairs currently used for vF trunks could be made available for T-carrier. However, area 4 does not usually roll active vF. In addition, although the other areas will roll vF, certain cables may have been designated for no additional T-carrier for other reasons, such as poor cable condition. This factor would increase the carrier fill from the oTC standpoint, since a particular cable with relatively few systems may be considered fully used for T-carrier by the company.

2.5 Apparatus case fill

The previous section examined the use of the ultimate span capacity, i.e., the fraction of the maximum number of T1 lines a span could support that are currently in use. This section examines the apparatus-case fill on a span basis, defined as the ratio of the number of working DS1s to the number that could be supported by the existing apparatus cases. This percentage measures how effectively apparatus cases are being used, and also indicates the growth that can be accommodated without additional splicing.

The distribution on a span basis of apparatus-case fill is shown in Fig. 9 for the four areas that supplied these data. The average and median are both 56 percent, with a standard deviation of 26 percent. This median of 56 percent implies that half of these spans can accommodate a 79-percent increase in working DS1s before a new apparatus case is needed. Of course, some spans are nearly full; 12 percent have apparatus-case fills greater than 90 percent. However, most spans have significant reserve apparatus-case capacity.

Area 6 splices the maximum number of cases the cable can support when initially equipping for T-carrier. Removing these spans from the population results in the distribution shown by the dashed line. The average is now 64.9 percent, and the median 67 percent, both an increase in use. However, the median of 67 percent implies that half the spans in the remaining three areas can accommodate a 49-percent increase in working DS1s before a new case is needed, which still indicates a significant reserve.

The low-percentage apparatus-case fill is likely to occur on low cross-section spans, where growth may be slow and a case may not fill for some time. Figure 10 shows this to be the situation. The higher cross-section spans tend to have higher-percent apparatus-case fill, as shown by the line indicating the average fill per 200-DS1 step. How-

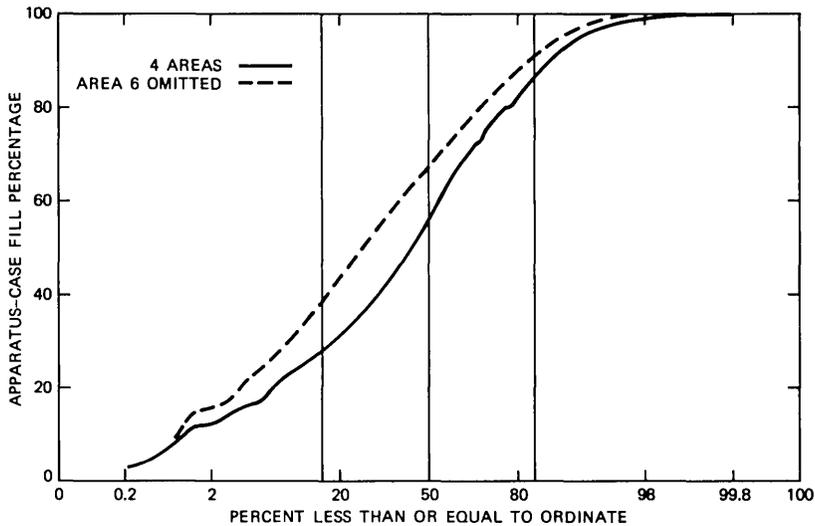


Fig. 9—Distribution of percent of apparatus-case fill on a span basis in four metropolitan areas. Number of spans = 445; average = 55.9; standard deviation = 25.5.

ever, there is a wide variation in percentage apparatus-case fill for a given cross section. There are some high cross-section spans that have many spare apparatus cases as shown by the family of curves that indicate the number of spare apparatus cases for a given cross section and percentage fill. For example, consider the 8 spare-case curve. A working cross section of 200 DS1s requires a minimum of eight apparatus cases; if there were eight additional cases on such a span, the percentage fill would be 50 percent. This number of additional cases assumes all unused slots are combined into the minimum number of apparatus cases; actually, there may be no completely empty apparatus cases in a particular span, rather, many partially filled ones.

Fifteen spans with working cross sections greater than 400 have eight or more spare apparatus cases. In fact, 39 spans (11 percent) have 16 or more spare cases. Most of these low percentage-fill spans have smaller cross sections, as previously indicated. The points representing spans from area 6, which have more than 8 spare apparatus cases, are identified by the number 6. Almost all spans with more than 16 spare cases are from this area. However, there are spans (unmarked) from the three other areas which have from 8 to 16 spare cases. Thus, all areas contain spans with significant reserve apparatus-case capacity.

Three factors contribute to this 56-percent average fill. First, system rearrangements can cause a decrease in span carrier requirements which results in one or more T1 span lines being turned down. These

lines will probably be reused in the future, but this churn causes some formerly active span lines to be depowered and left inactive. In this examination of apparatus-case fill, any such lines are counted as empty case slots. Second, most ORCs will splice several apparatus cases at a time when equipping a cable for T-carrier. As mentioned, area 6 splices the maximum number of cases the cable can support when equipping for T-carrier. This procedure increases reliability by eliminating future disturbances of the cable sheath and splice cases, but also results in empty apparatus cases for some time. Third, cables originally placed in T-spans to satisfy loaded VF requirements are used to diversify T-carrier. This diversity leads to more partially equipped apparatus cases. For example, instead of one case 100-percent full on one cable, there could be two cases 50-percent full, one on each of two parallel cables.

2.6 Cable plant

T-carrier was originally engineered for use in most common VF trunk cables. Thus, T1 is found in a variety of cable gauges—19-, 22-, 24-, and 26-gauge cable, and on sizes of from 50 to 1800 pairs. Spans which change cable gauge and size between offices are used, as well as spans with cables containing more than one gauge in the same sheath, called composite cable. Both directions of transmission can be in the same

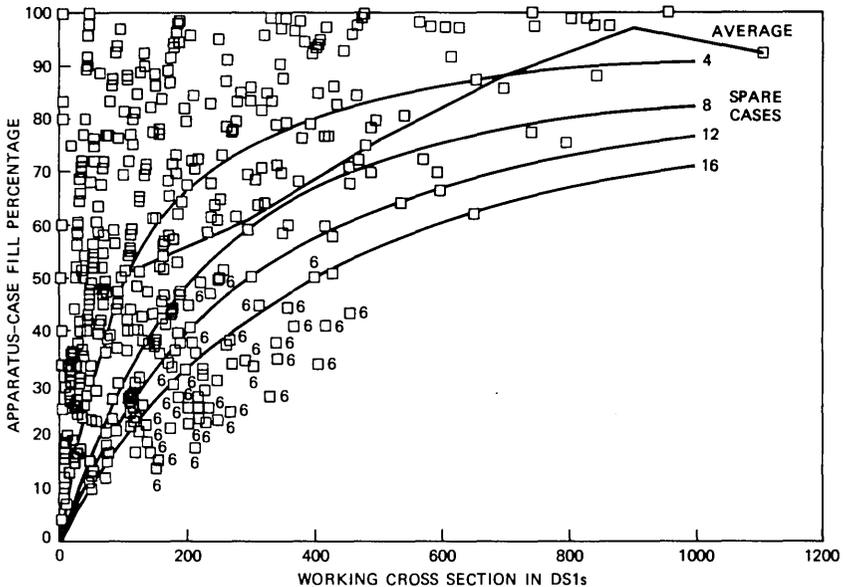


Fig. 10—Apparatus-case fill vs. working cross section in four metropolitan areas.

sheath, called single-cable operation, or two separate cables can be used, one for each direction, called dual-cable operation. This section examines the nature and use of the outside plant cables in the metropolitan T-carrier plant. The cable size, gauge, type, insulation, and type of operation are described in the following paragraphs.

Cable size by plant mileage is displayed in Fig. 11 for the six metropolitan areas for which data are available. Spans that use cables of different sizes in tandem between central offices, or odd cable sizes such as 707 or 812-pair, are classified as "other." Area 2 has the most uniform plant with 63 percent 900-pair, 26 percent 1100-pair, and very little other. Area 7 is at the other extreme with 65 percent other. This area, unlike the others in this figure, combines subscriber pairs with VF and T-carrier-supplied trunks in a common sheath to such an extent that the cables generally decrease in size with distance from the central office as subscriber pairs leave the cable. Other areas shown here maintain separate cables for trunk circuits. Area 1 (not shown) also combines subscriber and trunk in common sheaths and cable records are dispersed throughout individual districts, making collection and analysis impractical. The large 1500- or 1800-pair cables are generally

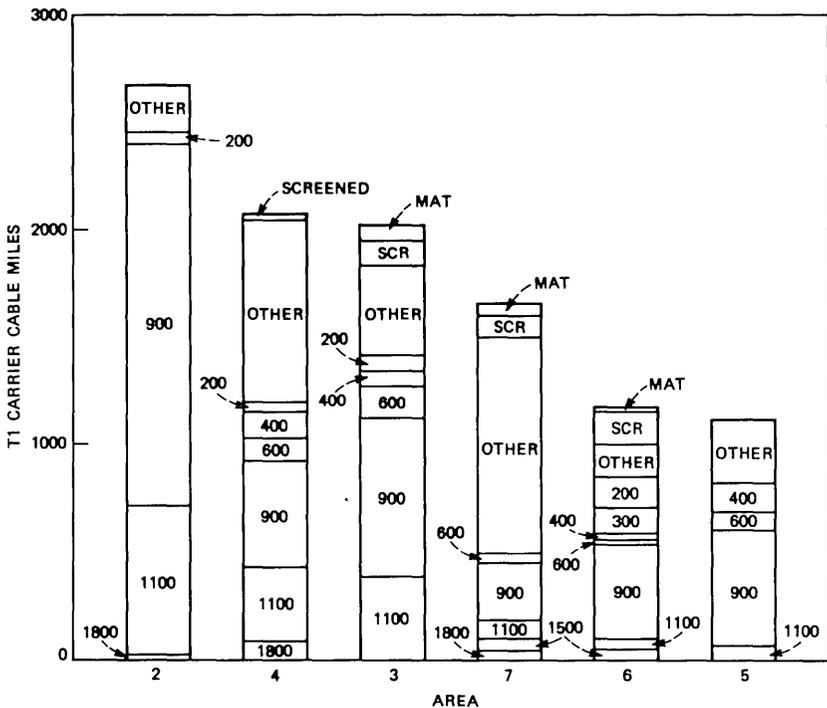


Fig. 11—Cable sizes and types in six metropolitan areas.

used in short spans connecting large downtown offices. The 1100-pair cables are used in spans where slightly more VF trunk requirements are anticipated than could be supplied by 900-pair. The most common single cable is 900-pair, accounting for 39 percent of the total cable mileage. The smaller-size 200- through 400-pair cables are often used in aerial dual-cable operation in the suburbs.

Metropolitan Area Trunk (MAT) and other screened cables have been counted separately since their built-in shield permits full-fill T-carrier operation in a single sheath. These screened cables account for 4 percent of the total cable mileage. MAT cable, comprising approximately 1 percent of the plant, is 1200- or 1400-pair. Of the remaining 3 percent accounted for by other screened cables, 63 percent is 900-pair, 22 percent is 600-pair, and the rest smaller sizes. Recent cable installations are primarily MAT or other screened cables, with sizes of 1200-pair or more being MAT, and 900-pair or less being other screened. Since MAT is now available in smaller sizes, new smaller size cable installations will probably use MAT. This sample contains 10,644 cable miles. Since this cable sample is about 20 percent of the total T-carrier plant, the total T-carrier cable mileage is estimated to be 50,000 miles.

Cable gauge by plant mileage is displayed in Fig. 12 for the six metropolitan areas for which data are available. Spans that use more than one gauge in tandem between central offices, as well as 25 (MAT) and 26-gauge cables, are also classified as other. On this basis, area 2 again has the most uniform plant with 93 percent being 22-gauge cable. area 7 is still the most heterogeneous with 40 percent other. Again, this irregularity is a result of subscriber and trunk pairs occupying common sheaths. The 19-gauge cable is typically used when extra-long intermediate sections are needed or to permit full-fill T1 operation at 6000 foot spacings. (See Section III for intermediate section details.) The 22-gauge cable accounts for 80 percent of the total mileage for the aggregate sample, and at least 75 percent for each of the areas except area 7. Much of the other cable mileage contains 22-gauge sections. Often one of these other cables will have a 24-gauge section adjacent to the central office, then 22-gauge cable until the end section adjacent to the next office, where 24- or 26-gauge is again used. Spans consisting of entirely 24-gauge cable are generally short, downtown spans which usually contain 1500- or 1800-pair cables.

Cable insulation type is nearly all pulp. Of the total mileage, 92 percent is pulp-insulated and the remaining 8 percent is PIC. Metropolitan Area Trunk cables, 1 percent of the total mileage, are included with the other PIC cables. The amount of PIC by area ranges from 11 percent in areas 3 and 4 to less than 1 percent in area 2.

Cable operation by plant mileage is shown in Fig. 13 for all seven metropolitan areas. Single-cable operation, having both directions of

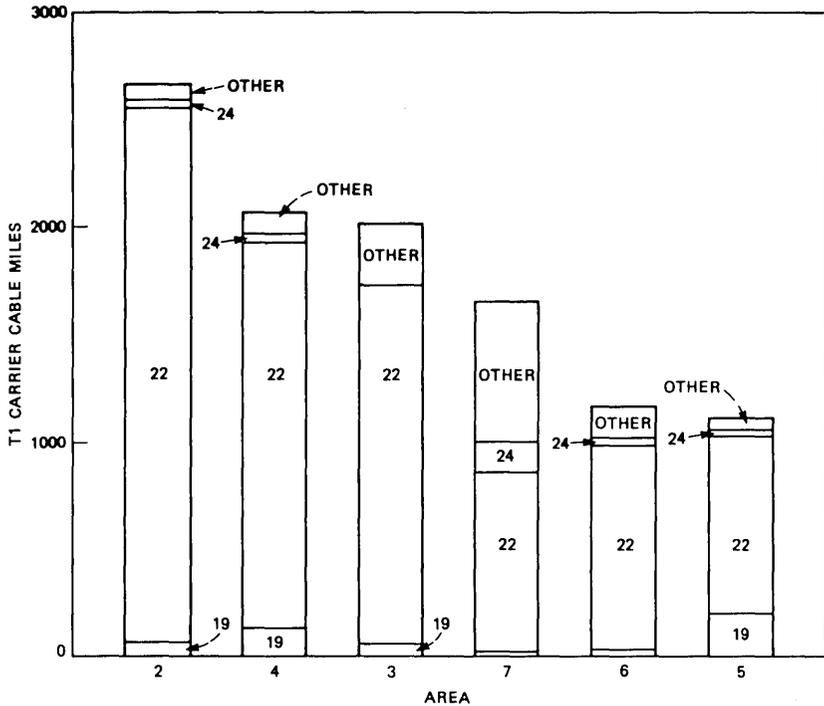


Fig. 12—T1 carrier cable gauges in six metropolitan areas.

transmission in one cable sheath, limits carrier fill to somewhat less than half of the pairs for most cables. Dual-cable operation, having each direction of transmission in a separate cable, allows full carrier fill. MAT and other screened cables are again counted separately since they provide both directions of transmission in a single sheath and allow full carrier fill. Single-cable operation accounts for 66 percent of the mileage, dual-cable 30 percent, with MAT and other screened cable accounting for only 4 percent. However, since new installations are predominantly MAT or other screened cable, this percentage is increasing with time. Higher cross-section areas 3 and 4 frequently use dual-cable to support their dense routes.

III. REPEATER SECTION INFORMATION

T-carrier system performance is determined by the performance of individual repeater sections, which in turn depends on section parameters, chiefly section loss. In this section, the repeater-section statistics of the T-carrier network are investigated. We discuss distributions of section lengths which include intermediate sections of all gauges, 22-

gauge pulp sections only, the longest section per span, and the number of repeaters per span.

3.1 Intermediate section distribution

Intermediate sections are defined as those between two outside plant repeater locations; sections adjacent to a central office (end sections), which are limited in length to less than the maximum allowed for the intermediate sections, are excluded. Figure 14 shows the distribution of intermediate section lengths combined for all metropolitan areas. As seen from the figure, the average section length is 5.3 kft and the standard deviation is 1.03 kft. Ninety-six percent of the sections are less than 6.3 kft, the engineering limit for T1 systems operated on 22-gauge pulp. The remaining 4 percent longer than 6.3 kft are comprised of either 19- or 22-gauge air core PIC cables for which the maximum allowable section length is 9.5 and 6.9 kft, respectively.

The separate distribution of intermediate section length for each of the seven metropolitan areas are coplotted in Fig. 15, showing the contribution of each area to the overall distribution. The mean section lengths vary from 5.0 to a maximum of 5.75 kft and the standard deviation from 0.90 to 1.12 kft. All sections above 7 kft are from only

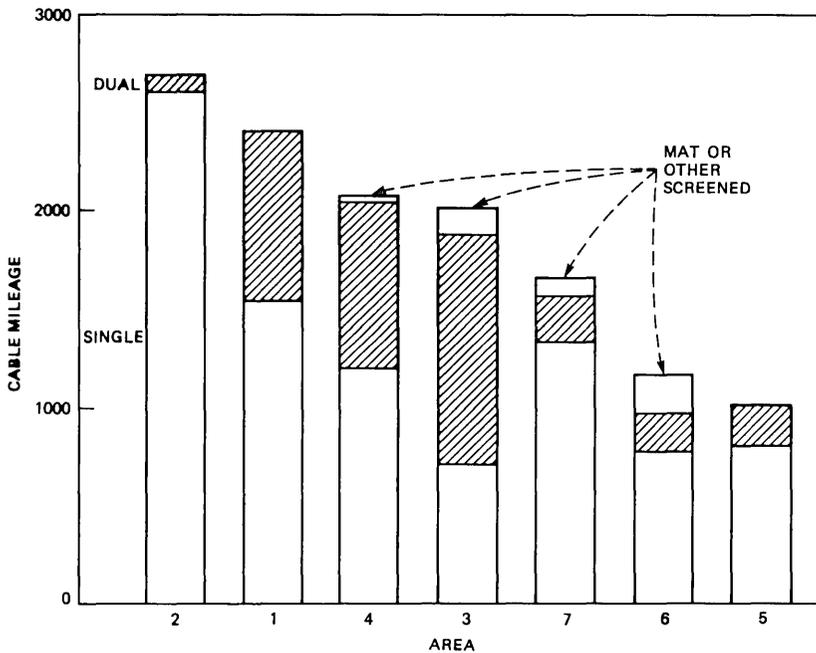


Fig. 13—Single- and dual-cable operation in seven metropolitan areas.

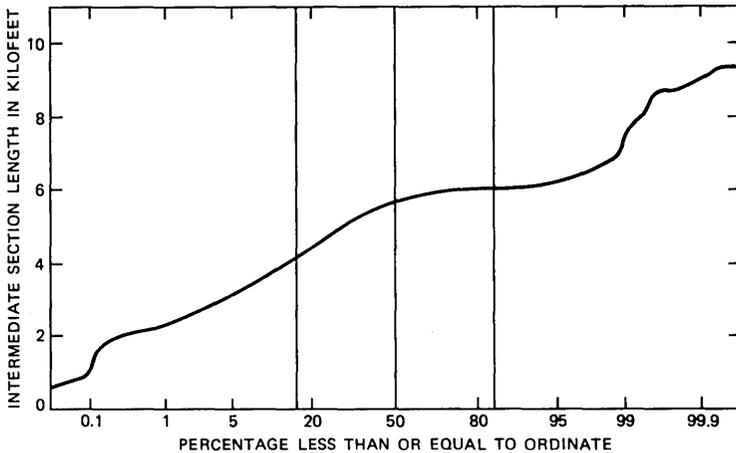


Fig. 14—Distribution of intermediate section length in seven metropolitan areas. Number of points = 4812; average 5.30; standard deviation = 1.03.

three areas, each of which utilize a larger than average percentage of PIC cables. Note that all areas have sections under 4 kft, ranging from 4 percent to 20 percent of the individual areas' distributions.

As indicated in Section 2.6, over 70 percent of the metropolitan digital plant in this survey is made up of 22-gauge pulp cable sections; the distribution of these section lengths is shown in Fig. 16. Assuming a cable loss of 5.2 dB per kft at T1's Nyquist frequency, the average length of 5.3 kft translates to 27.6 dB of insertion loss. As expected, no sections exceed 6.3 kft, the maximum repeater spacing allowed for 22-gauge pulp. Except for two areas with highest means (5.61 and 5.73 kft), the rest of the samples have similar characteristics with mean values ranging from 5.2 kft to 5.38 kft and standard deviations from 0.65 kft to 0.93 kft, as shown in Fig. 17. Only 10 to 30 percent of the sections are less than 5.0 kft.

3.2 Longest intermediate-section distribution

The performance of a set of digital span lines is largely controlled by the highest loss (usually longest) section in the span. Figure 18 shows the distribution of the longest section per span for all seven areas combined. The shape of the longest-length distribution is very similar to the overall length distribution presented previously. The mean is slightly higher (5.63 kft vs. 5.30 kft) and the standard deviation lower (0.84 kft vs. 1.04 kft). More than 80 percent of all spans have at least one section longer than 5.0 kft. Figure 19 displays the distribution of the highest-loss section per span, where the loss data are either provided by the participating operating companies or derived by

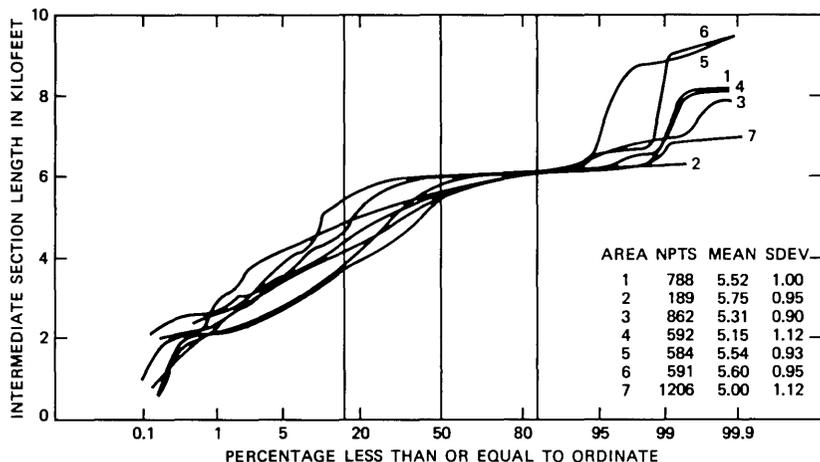


Fig. 15—Distribution of intermediate-section length in seven metropolitan areas. Number of points = 4812.

applying the appropriate cable-engineering loss to each cable section. Note that all sections have less than 33 dB of loss and 50 percent of the spans have no section loss greater than 30 dB. There are, however, remarkable differences among the areas at the lower-loss portion of the distribution, as shown in Fig. 20. For example, the percentage of spans with longest section loss less than 28 dB varies from as little as 2 percent in one area to about 30 percent in others.

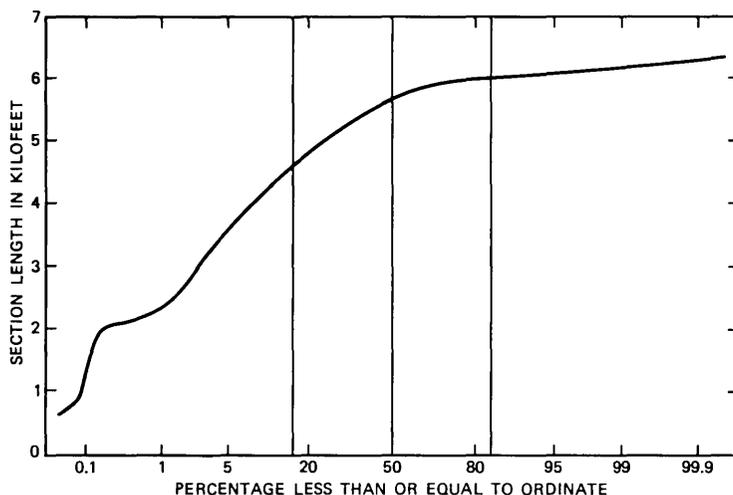


Fig. 16—Distribution of intermediate section length (22-gauge cable) in six metropolitan areas. Number of points = 2538; average = 5.37; standard deviation = 0.83.

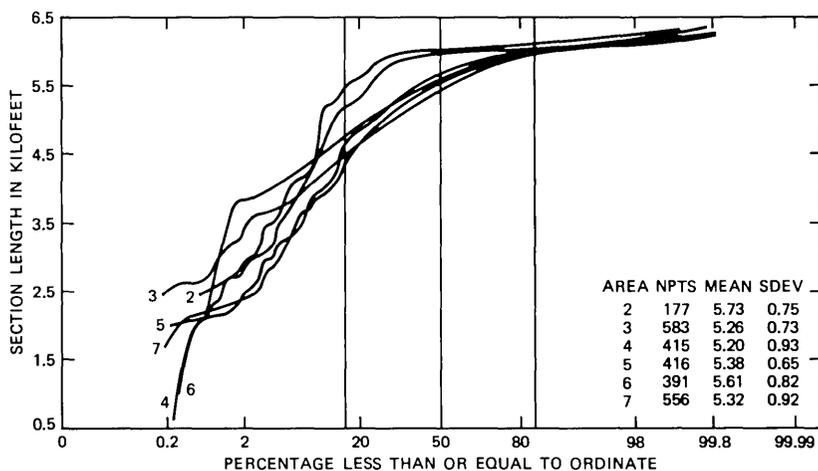


Fig. 17—Distribution of intermediate-section length (22 gauge cable) in six metropolitan areas. Number of points = 2538.

3.3 End-section distribution

Figure 21 shows the distribution of the end sections—those adjacent to a central office. The mean is 2.86 kft and the standard deviation is 0.65 kft. As with the intermediate section lengths, the mean value is well within the engineering limitation for a T1 end section with 22-gauge pulp cable (4.5 kft). Approximately 5 percent of sections use 19- or 22-gauge PIC cable and exceed the maximum 22-gauge pulp spacing.

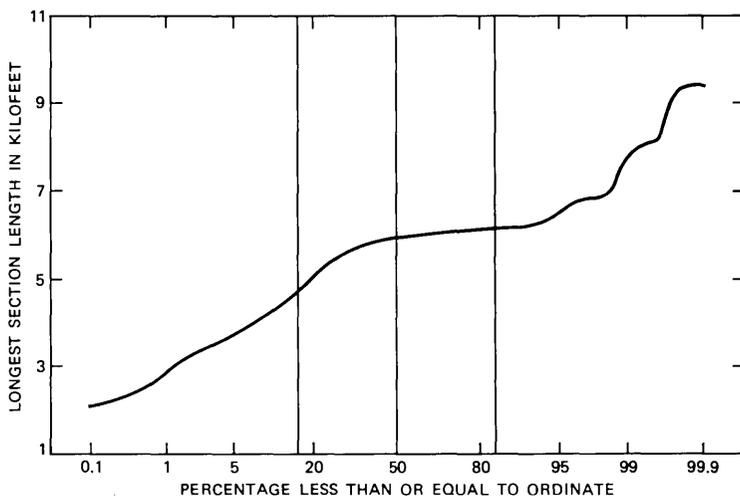


Fig. 18—Distribution of longest section per span in seven metropolitan areas. Number of points = 1016; average = 5.63; standard deviation = 0.86.

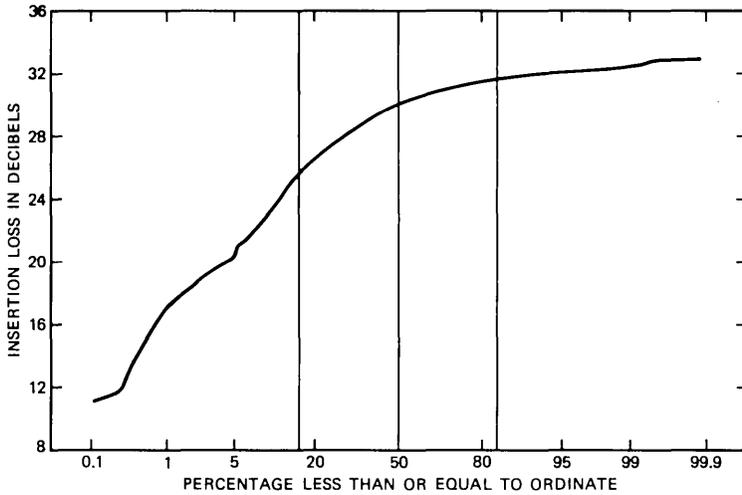


Fig. 19—Distribution of maximum intermediate-section insertion loss per span in six metropolitan areas. Number of points = 817; average = 28.77; standard deviation = 3.59.

All areas have similar end-section distributions, as shown in Fig. 22. The mean and standard deviations are within a few hundred feet of each other.

3.4 Distribution of the number of sections per span

Figure 23 shows the distribution of the number of repeater sections per span (i.e., number of tandem outside plant repeaters plus one).

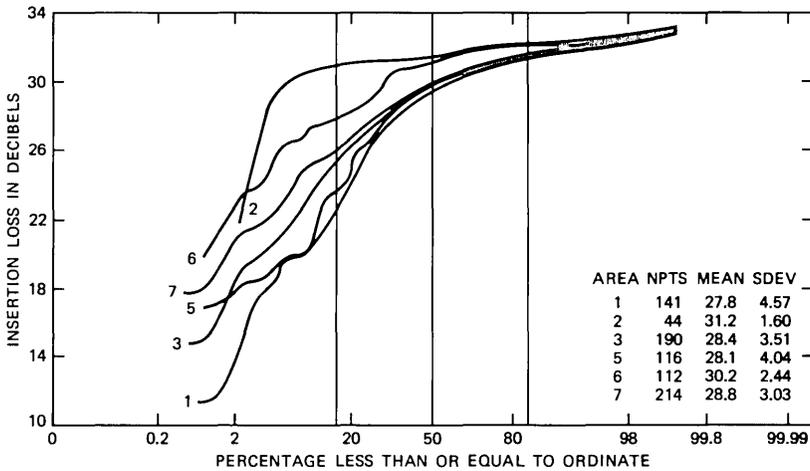


Fig. 20—Distribution of maximum intermediate-section insertion loss per span in six metropolitan areas. Number of points = 817.

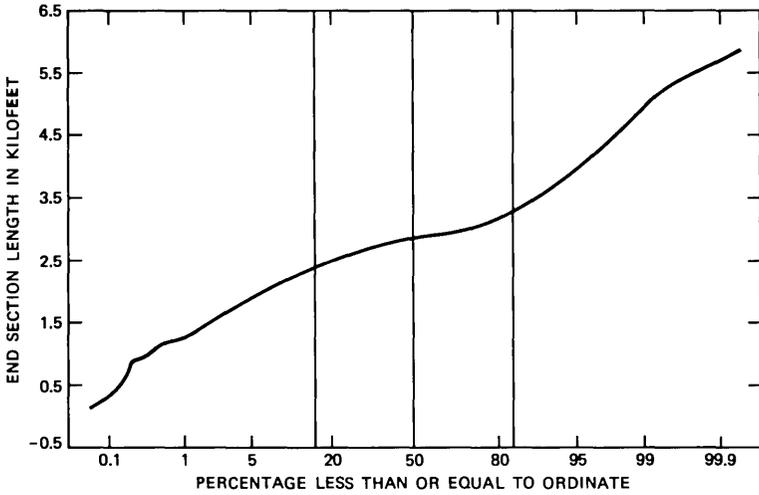


Fig. 21—Distribution of end-section length in seven metropolitan areas. Number of points = 2029; average = 2.86; standard deviation = 0.62.

The average number of sections per span is 6.62, with the longest span having 21 sections. Nearly 85 percent of the spans have between 4 and 11 sections. Only two areas differ slightly from the rest of the sample—as shown in Fig. 24—having average values of 7.1 and 7.38 sections per span. One area, area 6, has longer than average span length, as would be expected. The other, area 7, not only has slightly longer than average distances between central offices, but also uses conservative

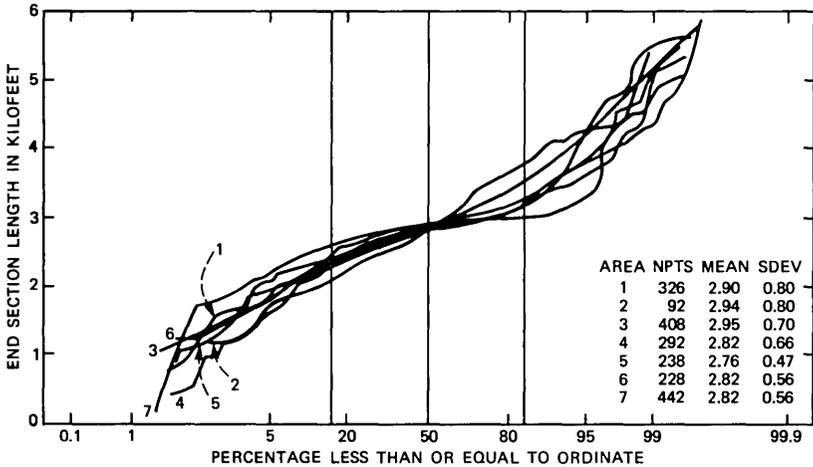


Fig. 22—Distribution of end-section length in seven metropolitan areas. Number of points = 2026.

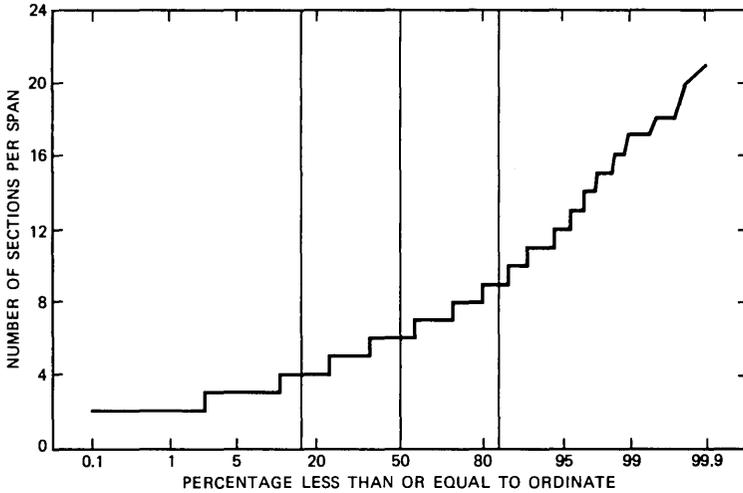


Fig. 23—Distribution of number of sections per span in seven metropolitan areas. Number of points = 1041; average = 6.62; standard deviation = 2.90.

engineering rules, forcing shorter section lengths. Both of these factors increase the number of sections per span.

IV. SPAN GROWTH

The efficient use of the present plant, and the alternatives and applications for various possible new transmission facilities require an understanding of not only the present state of the plant but also of the nature of its growth. Not only is the amount of growth important, but an understanding of the pattern of this growth is needed as well. In this section, span growth means the increase in span lines of the network, usually provided by adding span lines in existing spans. New spans are being added at a rate of less than three percent per year, and have a small effect on the total number of span lines added.

It is important to note that span-line growth is not equivalent to system growth. Since many systems with diverse originating offices use the same span, growth in that span may be due to increased requirements for many different systems, none of which may originate in that span. Conversely, since a single system passes through many spans, the placement of one new system requires an additional span line in several spans. Since the span is the fundamental building block of the T-carrier network, span-line growth is described in the following sections.

To examine the span-line growth, data were requested from six OTC current planning groups. These requests occurred over two years so that the growth projections begin at a different time for each area. In

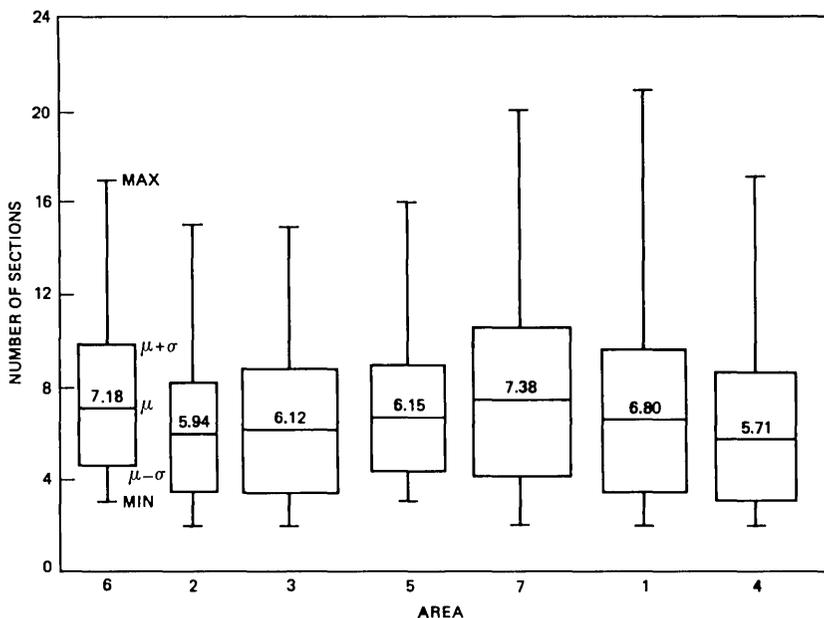


Fig. 24—Number of sections per span.

addition, each planning group projects their growth for different time intervals. Starting times ranged from January 1977 to January 1978, while projection intervals varied from 2 to 5 years. For five areas, this span-line-growth data represent the projected number of working DS1s each year for each span. For area 5, span-line growth was derived from the data given for the number of voice circuits to be provided by T-carrier. For this growth, 100-percent circuit fill of T-carrier line additions was assumed.

The number of span lines versus time for each area, the span-line growth average on a per-year basis, and network growth patterns are discussed in the following section.

4.1 Growth versus time by area

The span-line growth for all spans in each area was combined to produce a plot of growth for each area. Since maintenance lines and backbone spare lines are not included in this growth, the total span-line additions are higher than shown by 4 percent. Figure 25 shows the number of span lines versus time for six areas. Note the different starting dates and growth projection intervals. The number of span lines has been normalized to unity at the starting date for each area.

Reference lines of 5 and 10 percent growth per year are given. Areas 1, 2, 6, and 7 have an average growth rate over their respective years

of from 7 to 8 percent. Area 4 planned a year of substantial growth in 1979 to provide service for two No. 4 ESS machines. The low-growth rate of about 4 percent for area 5 was derived from the number of voice circuits to be provided on T-carrier, and may be conservative. This method assumes 100-percent channel-bank fill—or that only one additional T1 span line is required for each group of 24 new voice circuits, which is often not the case. Limited data indicates channel-bank fills of about 80 percent, which would bring the growth of area 5 to about 5 percent.

4.2 Average yearly growth

Combining span growth data from different areas is complicated by the different starting dates and projected growth intervals for each area. Also, individual spans vary widely in year-to-year growth. One reasonable way to compare individual spans is to use an average yearly growth, the average number of DS1s added per year for each span. This average yearly growth is a measure of the growth of each span and smooths year-to-year variations, enabling comparison of spans from all areas. An average yearly growth of 24 DS1s per year is equivalent to adding one T1 apparatus case or one-half T1C apparatus case per year.

The examination of average yearly growth gives insight into the pattern of span-line growth in the operating companies. Knowledge of

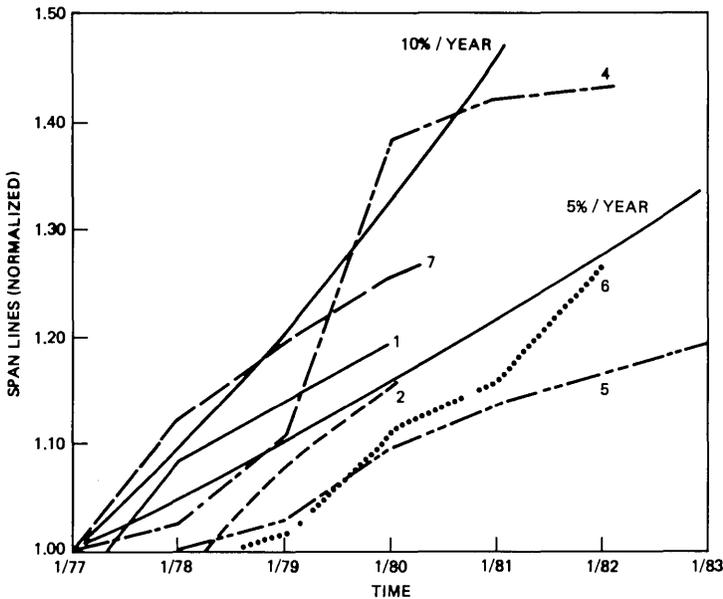


Fig. 25—Digital network growth on a span basis for six metropolitan areas.

the growth and its distribution is a valuable aid to study the applicability of new transmission systems. Usually in market studies, an estimate of the length and cross section of the span population, along with a predicted average growth rate applied to all spans, is used to predict the total network growth and potential applicability of individual facility types. Since both the actual distribution of average yearly growth and the distribution of cross section are known in this study, the validity of the method of growing all spans at the same rate can be evaluated.

The average growth rate of 6.0 percent was obtained by dividing the sum of the average yearly growth in DSIs of all spans by the total cross section of those spans as of January 1980, the middle of the growth projection period. Figure 26 shows the distribution of average yearly growth and the distribution predicted by applying the average growth rate of 6 percent to each cross section. The spans average a growth of 7.6 DSIs per year. About 8 percent of the spans are adding more than 24 span lines per year, and 32 percent are adding 1 span line or less. In fact, 21 percent of the spans are not growing or are turning down lines, in spite of averaging over the several year period. The predicted distribution differs from the actual distribution in that it underestimates the growth of the faster-growing spans (adding more than 15 DSIs) and overestimates the growth of the slower-growing spans. The predicted distribution does not show any spans adding

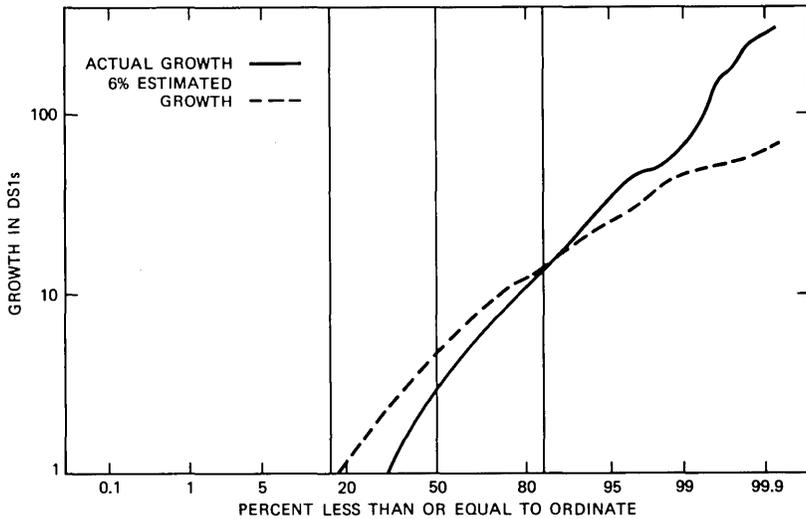


Fig. 26—Distribution of average yearly span growth in six metropolitan areas. Actual growth: average = 7.6; standard deviation = 18.5. Estimated growth of 6 percent: average = 7.6; standard deviation = 8.9.

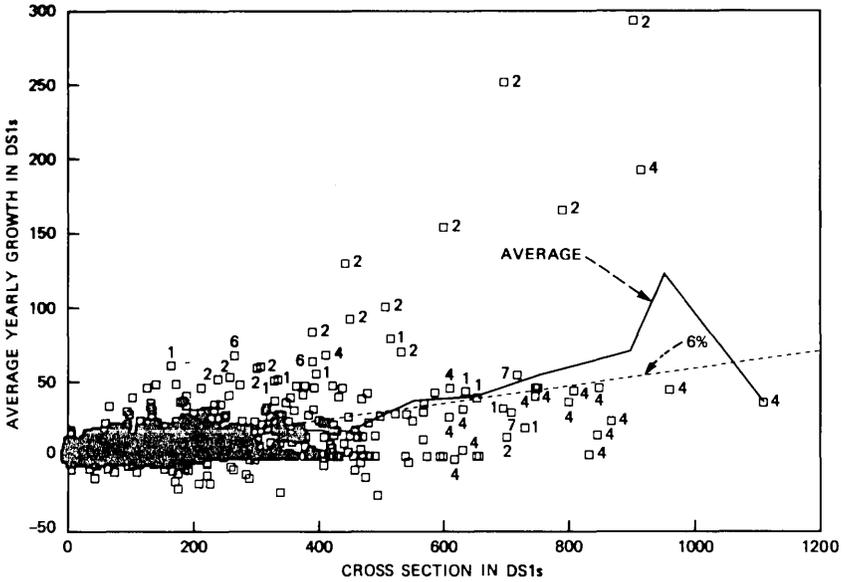


Fig. 27—Average yearly growth vs. working cross section in six metropolitan areas.

more than 65 DS1s per year, but growth projections indicate 12 such spans. One contribution to the difference between the actual distribution of projected growth and that predicted by applying the average network growth method to all spans is that the latter predicts positive growth for each span, whereas 21 percent of the spans have zero or negative growth. The concentration of growth in a small fraction of the spans is examined in Section 4.5.

4.3 Average yearly growth versus working cross section

Spans with large cross sections might be expected to add many DS1s each year. Hence, there may be a trend of increasing growth with increasing cross section. The relationship of average yearly growth to the January 1980 cross section is shown in Fig. 27. The overall average of 6 percent (indicated by the dashed line) is the ratio of the total average yearly growth to the total January 1980 cross section.

Three features are noteworthy. First, there is a trend of increasing average yearly growth with increasing cross section as indicated by the solid line, which is the average over 100 DS1 steps. Second, there is a wide variation of average yearly growth for a given cross section, including some negative values. Some spans have a net decrease in span lines over a period of 2 to 5 years. And third, area 2 has 14 spans of the 24 with average yearly growth greater than 50. These spans

represent a major backbone route into the downtown metropolitan area.

4.4 Average yearly growth versus span length

Since downtown spans tend to be short, and suburban spans tend to be long, there may be a relationship of average yearly growth to span length. Figure 28 shows that this is not the case. There is no significant trend of average yearly growth with span length, as indicated by the solid line, which is the average over one-mile intervals up to 13 miles, with spans longer than 13 miles combined.

4.5 Span-line growth patterns

To understand the patterns of growth, the geographic location of high-growth spans was examined. All high-growth spans—those having average yearly growth greater than 50 DSIs—were found to be concentrated in the vicinity of No. 4 ESS digital switching machines. To quantitatively examine the extent to which T-carrier growth is influenced by No. 4 ESS switching machines, a No. 4 ESS span—or 4E span—is defined to be a span which has one of the following characteristics:

It is directly connected to a No. 4 ESS machine.

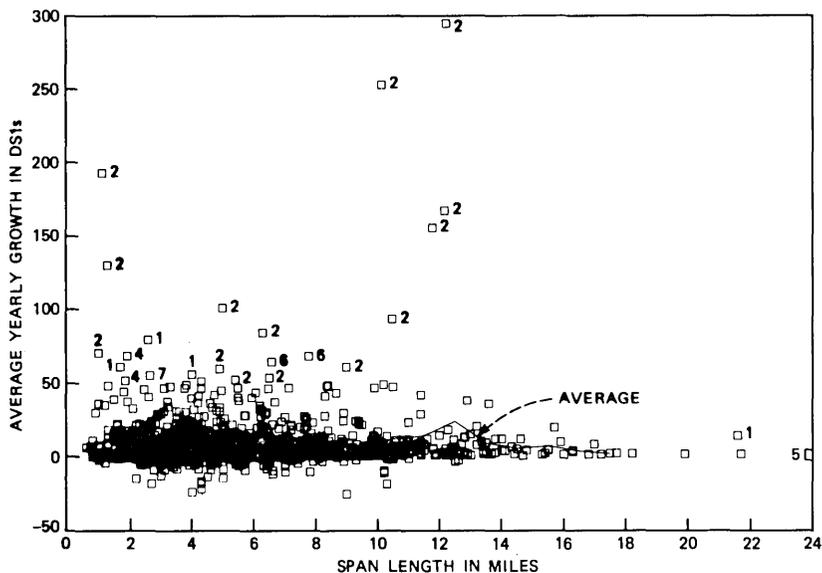


Fig. 28—Average yearly growth vs. span length in six metropolitan areas.

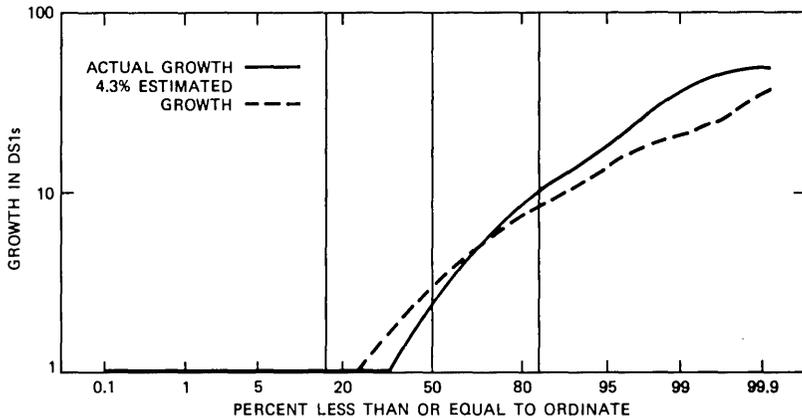


Fig. 29—Distribution of average yearly span growth in six metropolitan areas for 1151 non-4E spans. Actual growth: average = 4.4; standard deviation = 8.0. Estimated growth of 4.3 percent: average = 4.4; standard deviation = 4.6.

It is in a tandem path between two No. 4 ESSs.
 It is part of a major backbone route to a No. 4 ESS.

Spans meeting one of these criteria usually have large cross sections. For example, Fig. 5 shows 66 spans with cross sections greater than 400 DS1s. Fifty-five of these high cross-section spans are 4E spans according to the above definition. Although these 4E spans were first identified by high growth, other spans which meet one of the three criteria have been included. A total of 88 spans in this study have been identified as 4E spans.

Since the 4E spans were first identified on the basis of large growth, examining growth for two separate populations, 4E and non-4E spans, may give insight into the growth pattern. Figure 29 shows the growth distribution for non-4E spans, together with the estimate based on growing all span cross sections at the average growth rate for non-4E spans of 4.3 percent. This estimate is a slightly better match than that of the total population in Fig. 26. However, the same discrepancies occur: underestimating the growth of the faster-growing spans and overestimating the growth of the slower-growing spans. The average non-4E span is adding 4.4 DS1s per year, about half of the 7.6 average for the total population, and no non-4E span is adding more than 48 DS1s per year.

Next, the population of the 88 4E spans is considered separately. The distribution of average yearly growth for the 4E spans, together with the estimate based on growing all 4E span cross sections at the 4E span average rate of 10.8 percent, are shown in Fig. 30. In this case, the estimate is a much better match to the actual data. In fact, this

estimate is the closest obtained, mainly because none of the 4E spans have zero or negative average yearly growth. The average 4E span is adding 48 DS1s per year, about six times the total population average and 11 times the non-4E average.

Thus, the spans defined as 4E spans dominate the current T-carrier span line growth; Table I summarizes the characteristics of the 4E spans. Growth data are available for more than 1,200 spans. For the sum of all spans, the average growth rate is 6.0 percent. Although the 88 4E spans comprise only seven percent of the spans, they account for 25 percent of the total currently working cross section, and for 46 percent of the total average yearly growth.

This growth pattern is based on short-term current planning data. In the longer term, as No. 4 ESS machines become established, the growth on 4E spans may not be as dramatic, although some new 4E spans may be created that will grow rapidly. For the present network, however, this separation of spans into 4E and non-4E populations yields insight into the short-term growth behavior.

Another aspect of the growth is shown in the last two rows of this table. Based on the average yearly growth, there are 983 spans adding, and 256 spans losing, DS1s. The 79 percent of the spans gaining represent 86.8 percent of the total January 1980 cross section. Twenty-one percent of the spans have zero or negative average yearly growth in spite of averaging over several years. None of the 4E spans are losing lines, as indicated in Fig. 30.

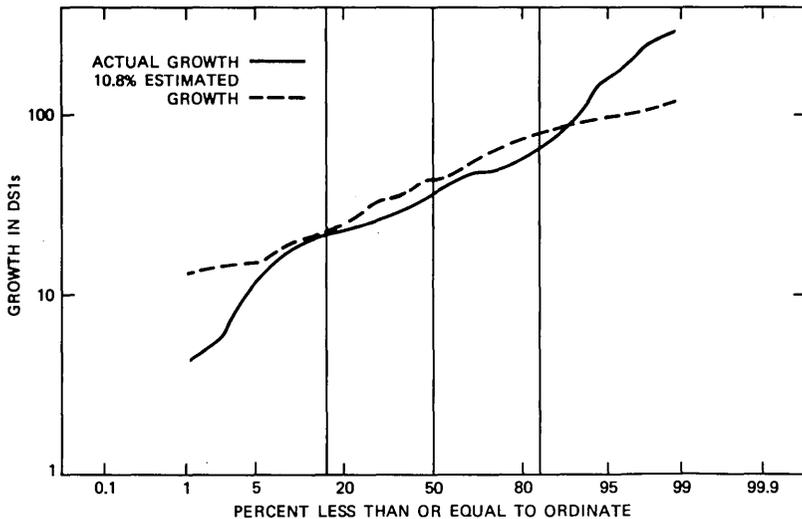


Fig. 30—Distribution of average yearly span growth in six metropolitan areas for 88 4E spans only. Actual growth: average = 48.4; standard deviation = 47.0. Estimated growth of 10.8 percent: average 48.4; standard deviation = 24.9.

Table I—Characteristics of the No. 4 ESS spans in the digital metropolitan network

| Category | No. of Spans | Total Cross Section | Average Yearly Growth | Percent |
|----------|--------------|---------------------|-----------------------|-------------|
| Total | 1,239 | 157,233 | 9,354 | 6.0 Growth |
| Not 4E | 1,151 | 117,758 | 5,094 | 4.3 Growth |
| 4E only | 88 | 39,475 | 4,260 | 10.8 Growth |
| Gaining | 983 | 136,499 | 10,039 | 86.8 Base |
| Losing | 256 | 20,734 | (-685) | 13.2 Base |

V. SUMMARY

The metropolitan digital trunk plant is estimated to contain from 5,000 to 6,000 spans providing approximately 160,000 T-carrier systems. Fundamental aspects of seven metropolitan areas, representing a 25-percent sample, have been presented and discussed. All areas are similar in span length distribution and have a 5.9-mile average and a 5.4-mile median. The average cross section is 137, with a median of 90 equivalent DS1 signals. One area has significantly higher cross sections generated by a natural geographic corridor of population.

On a span basis, an average of 41 percent of the ultimate engineering capacity of T-carrier in the present cable plant is used. Higher cross-section spans tend to have higher-percentage carrier fills.

Again, on a span basis, the use of apparatus-case slots averages 56 percent. Larger cross-section spans tend to have higher percentage apparatus-case fills.

The cables used for T1 cover a wide range of size and gauge, but the dominant type is single-cable operation on 900-pair, 22-gauge pulp. This 900-pair cable accounts for 39 percent of the cable mileage in the total sample, and as high as 63 percent in one area. However, recent and future installations are dominated by MAT or other screened cable.

Since the cable plant is mostly 22-gauge pulp, the average intermediate section length is 5.32 kft, which corresponds to a T1 insertion loss of 27.7 dB. Four percent of these sections use 19- or 22-gauge PIC and are longer than 6.3 kft. All sections have less than the 34-dB insertion loss allowed by T1 engineering rules.

The average end section is 2.86 kft, which corresponds to a T1 loss of 15.4 dB. Five percent of the end sections use 19- or 22-gauge PIC and exceed 4 kft. Except for the extremes of the distributions, all seven areas do not differ significantly in either the intermediate or end-section statistics.

The number of working span lines in the sample obtained is projected to grow at an average rate of six percent in the near future. The actual amount of growth varies somewhat from year to year in each area. Growth on individual spans varies considerably and is more

concentrated in a small fraction of spans than is predicted on the basis of a constant percentage growth for all spans. Examining the geographic locations of the high-growth spans showed them to be associated with No. 4 ESS digital switching machines.

T1 Carrier Characterization—Field-Measurement Results

By T. C. KAUP, D. G. LEEPER, A. K. REILLY, and
P. E. SCHEFFLER

(Manuscript received February 8, 1980)

We undertook a field-measurement program for T1 carrier to gain an understanding of crosstalk performance in the outside plant and office environments. This would help us to use the existing plant more efficiently and to predict the performance of future systems. Repeater-section lines were measured in the trunk plant of three Bell System operating companies during 1977 and 1978. Results given here include the distributions of repeater section crosstalk margin, noise, and various system and cable properties. We describe a T1 crosstalk engineering model which explains the significant contributions to section margin. The measurements and analysis show that repeater apparatus-case crosstalk dominates intermediate repeater-section performance, while the performance of sections adjacent to central offices (end sections) is limited by cable crosstalk with evidence of impulse noise present on some lines. Maximum use of cable and equipment for T1 and future digital transmission systems has been made much easier with the knowledge gained from these field measurements. For instance, the results of this program have already been used to redefine engineering rules for bidirectional cable operation, allowing more wire pairs in the cable to be used for T1 operation.

I. INTRODUCTION

The T1 repeatered line transmits a 1.544×10^6 b/s bipolar signal on paired cable.¹ Since its introduction in 1962, its largest application has been as the transmission medium between D-channel banks that digitally encode and time-division-multiplex voice-frequency signals. A D/T1 system consists of two D-bank terminals and the transmission path of tandem T1 span lines interconnected at central offices along

the route between these terminals. Each span line is a series connection of T1 repeater sections between two central offices. Those sections adjacent to the offices are termed end sections; all others are referred to as intermediate sections.

The T1 error-rate objective specifies that at least 95 percent of 50-section systems shall have average transmission error rates less than 10^{-6} errors per bit.² Since the digital signal is regenerated at each repeater location, the error performance of a T1 system is approximately equal to the accumulated performances of the individual repeater sections. Therefore, an understanding of T1 carrier repeater section performance and the elements that control that performance is needed to establish a model for system behavior.

Measurements were conducted in the trunk plants of three Bell System OTCS during 1977 and 1978 using the automated equipment described in Ref. 3. In all, approximately 2000 repeater section lines were examined in 30 different repeater sections.

Most of the measurements were performed on nearly full-length repeater sections (measured section lengths were about 6,100 feet for intermediate sections and about 4,000 feet for end sections), on 22-gauge pulp-paired cable, terminated with apparatus cases, each of which holds 25 discrete-component T1 repeaters. Other situations examined included sections terminated with repeater cases that hold 25 smaller-size integrated-circuit versions of the T1 repeater. These two measurement situations are identified in this paper as 466/201 and 475/208, respectively.

The primary measure of performance was *repeater section margin*, which is defined as the additional noise at the regenerator input that can be tolerated before a 10^{-6} error rate is observed.

A margin model is shown here and the margin components identified from observations of various noise and section characteristics (e.g., noise spectrum and amplitude statistics, and cable-pair insertion loss). The individual measured quantities and the measurement techniques used are outlined in Ref. 3.

Section I of this paper briefly describes the measurement program. Section II gives the results of measurements of the primary performance parameter-repeater section margin. Section III outlines observed characteristics of crosstalk and office noise and the contribution of different types of crosstalk noise to repeater-section margin. Section IV describes a model for section margin and gives the measured characteristics of each element identified in this model. Section V characterizes plant parameters, such as cable insertion loss and examines the performance of repeater sections other than those having pulp cable with 466 apparatus cases and 201 repeaters; Section VI summarizes the key results of the measurement program.

II. T1 REPEATER-SECTION MARGINS

Repeater-section margin, which is the incremental noise a section can tolerate before producing a 10^{-6} error rate, was measured directly for intermediate repeater sections by attenuating the signal at the beginning of the section and amplifying the signal and noise by the same amount at the end of the section (just before the repeater input), as shown in Fig. 1. The amplification required to produce a 10^{-6} error rate was recorded as the margin.

A modified procedure was used to measure end section margins, mainly due to the inconvenience of transporting the rather elaborate van-based measurement system into and out of central offices. In the absence of a T1 signal, the noise at the input of a T1 office repeater was amplified until errors (pulses) at a 10^{-6} error rate were observed at the repeater output. The repeater, whose input section was set to equalize for the proper end-section cable loss, was clocked externally at a 1.544×10^6 rate to simulate the clock that would normally be derived from a T1 signal. Thus, the margin for an "all zeroes" pulse sequence was obtained.

2.1 Intermediate section margins

Figure 2 displays the measured margin distribution obtained for intermediate sections for the standard 466/201 22-gauge pulp case. The margins were measured using the original service repeaters plugged into the apparatus cases. (Note: a second set of margins was measured using a particular Bell Laboratories test repeater. The difference in margins measured for test and service repeaters is discussed in Section 4.8.)

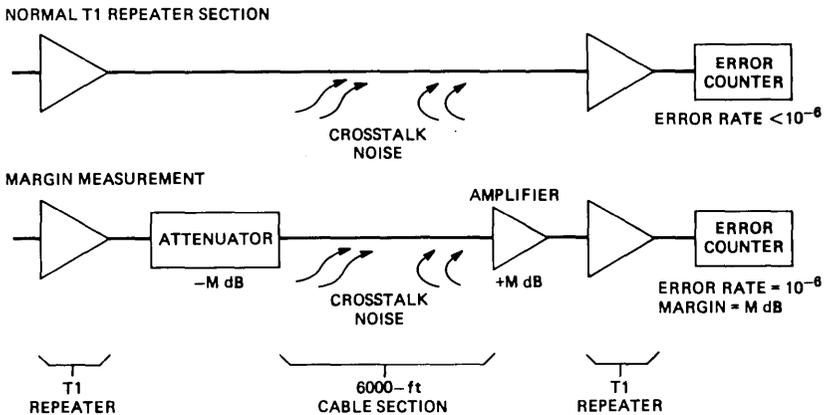


Fig. 1—Margin measurement by noise enhancement.

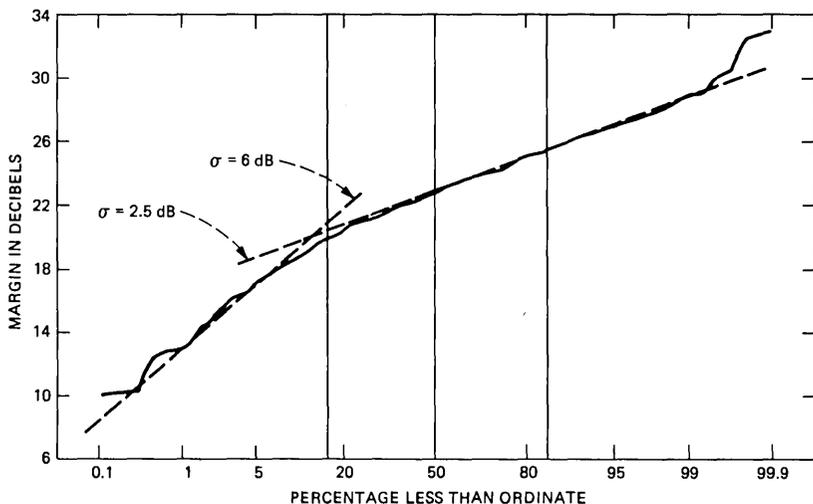


Fig. 2—Service margins (1977-78) for the 466/201 situation. Number of points = 915; average = 22.6; standard deviation = 3.03.

Each point on the curve shows the percent of repeater sections that have the indicated margin or less. The horizontal axis is a probability scale so that Gaussian distributions should appear as straight lines on the plot. The average section margin was 22.6 dB, for sections whose average cable insertion loss was 30.7 dB at 772 kHz (772 kHz is the half-baud for T1 transmission, which is near the major peak in the transmitted energy spectrum).

The minimum margin observed was 10 dB, a value for which essentially no errors due to intersystem crosstalk should be observed under normal conditions, even on lines comprising 50 such sections in tandem. However, the current measurements do not rule out widely separated impulses or noise bursts as a cause of errors, since the margin measurements were made for a period of only about 30 seconds on each repeater-section line.

The 0.1 percent point of the distribution may be estimated by drawing a straight line through the lower end of the curve as shown. This point on the distribution must have at least 3 dB of margin according to the repeater-section objective (Section IV). The estimated value is about 8.5 dB, which is 5.5 dB better than the objective.

To account for the *longest possible* repeater sections, the margins can be corrected to values expected for a section loss of 34 dB, the maximum loss allowed by engineering rules. Since the average loss for the margin measurements was 30.7 dB, the 0.1 percent margin is expected to be about 4.9 dB for systems composed of maximum length

sections.* This is still 1.9 dB above the objective and represents a conservative estimate of actual system performance in the field, since most sections are less than full length.

The measured margin distribution may be divided into two regions, as is shown by the two straight lines drawn through the measured points. These regions correspond to different types of intersystem crosstalk as discussed in Section III. It should be noted that the low (worst margin) end of the distribution has poorer margins than would be expected, assuming a Gaussian distribution with mean and standard deviations equal to those values determined for the overall set of measurements.

In summary, intermediate repeater-section margins measured in the field easily meet section objectives. Therefore, intersystem crosstalk on intermediate sections is not expected to be a major source of errors on T1 lines.

2.2 End-section margins

The amount of data collected for end sections (about 300 lines measured) in the field measurement program are much less than that collected for intermediate sections. Also, considering the wide variety of office environments possible, the end-section data are not likely to represent all T1 layouts that exist in the field today. Therefore, these data should be treated as a possibly biased sample of end-section performance.

End-section margins were measured in two OTCS on sections using 22-gauge pulp cable with 772-kHz cable losses ranging from 11.4 dB to 21.4 dB (section lengths from 2.2 to 4.1 kft). As shown in Fig. 3, the mean margin determined from the office-noise-amplification technique is 25.0 dB for all lines measured, with a standard deviation of 1.9 dB. The 0.1-percent point trend of the distribution is about 18 dB, if the one poor margin point at 14 dB is ignored.

All of the margins, however, should be adjusted downward to take into account intersymbol interference in normal repeater operation. (Remember that the end-section measurements were made in the absence of a T1 signal on the line.) This adjustment, if the office repeater degradation is the same as that found for the intermediate-section test repeater (see Section 4.7) would be about 4.3 dB. If the margins are adjusted by this amount, the average margin would be 20.7 dB, the minimum margin would be 9.7 dB, and the estimated 0.1 percent point (trend of the distribution) would be about 13.7 dB.

* The correction of 3.6 dB in margin for a cable loss change of 3.3 dB assumes a proportionality factor of 1.09 between margin and cable loss. See discussion of noise power, I , in Section 4.5.1.

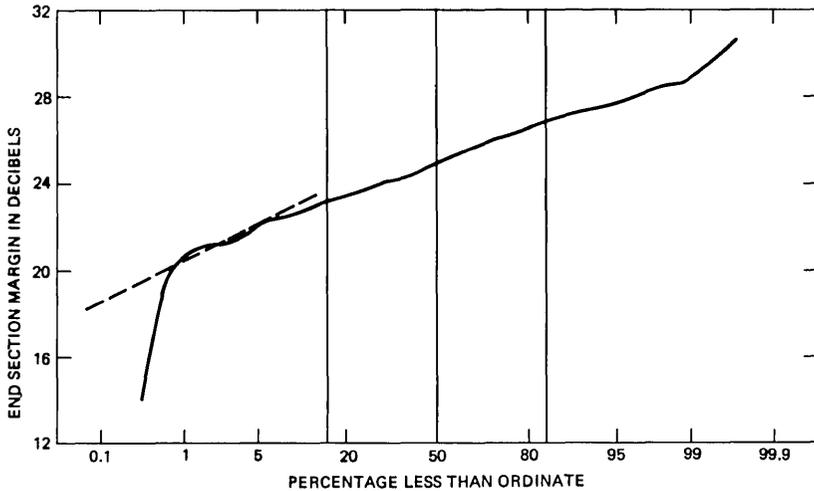


Fig. 3—End-section margins for the 466/201 situation. Number of points = 303; average = 25.0; standard deviation = 1.88.

The adjusted mean and minimum end-section margins are similar to the values found for intermediate-section measurements. However, the estimated 0.1-percent point of the end-section margin distribution is much higher than that for intermediate sections. Also, the standard deviation is smaller than for intermediate section measurements, despite the wider range of cable lengths. A possible explanation of these differences in terms of different types of crosstalk noise present is given in the next section.

Whether these data are truly representative of all end sections remains an open question. One conclusion from the above discussion is that the average performance of (these) end sections and intermediate sections are roughly the same; both layouts appear to have some reserve crosstalk margin that can possibly be exploited to pack more transmission channels into the same physical cable medium or, alternatively, to raise performance requirements.

III. SOURCES OF NOISE FOR T1 CARRIER

In the original design analysis for T1 carrier,² allowances were made for noise due to near-end crosstalk (NEXT) and far-end crosstalk (FEXT) on intermediate repeater sections and for impulse noise due to office switching transients on end sections. Results from the field-measurement program reported here indicate that the dominant noise source for intermediate-repeater sections is crosstalk noise generated in the repeater apparatus case and stub cable (designated ACXT), rather than

NEXT or FEXT generated in the main cable. Also, the character of the noise for most of the end sections measured does not appear to be strongly impulsive, at least for measurement periods of one minute or less.

3.1 Intermediate repeater sections

Figure 4 illustrates the types of noise-coupling paths found in intermediate-repeater sections of a T-carrier span. Intersystem crosstalk (noise due to other T1 systems in the same cable) is dominant in intermediate sections; other sources of noise, such as thermal, may be neglected. Intersystem crosstalk of three types are shown:

(i) Cable near-end crosstalk (NEXT) caused by coupling between wire pairs in the main cable where the output (high level) pulse stream of a T1 repeater interferes with the input (low level) of the same or a nearby repeater.

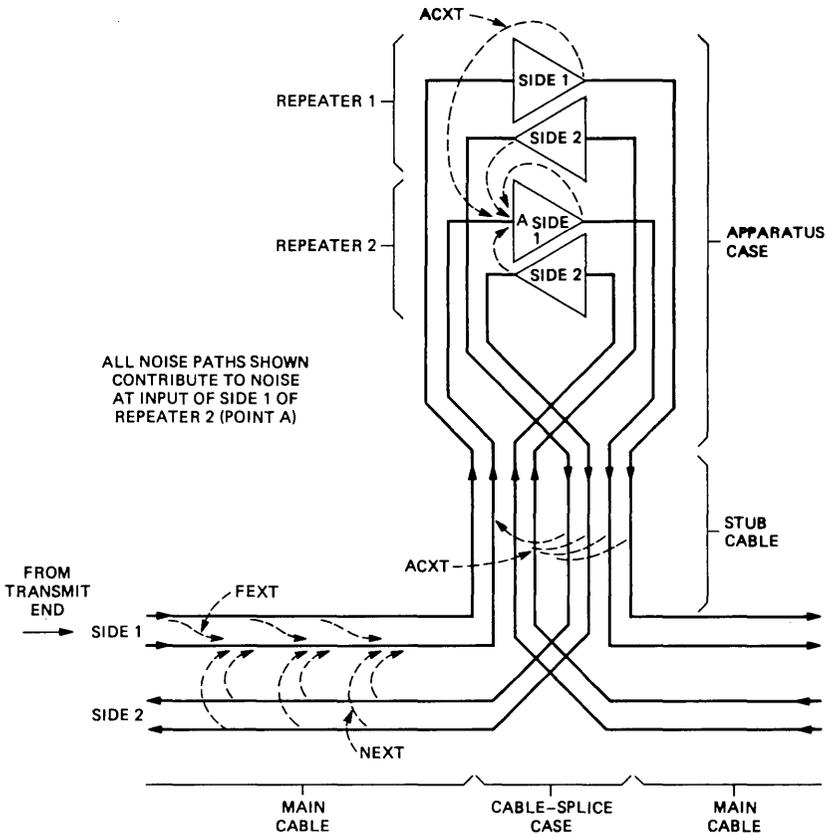


Fig. 4—Noise-coupling paths for intermediate-repeater sections. All noise paths shown contribute to noise at input of side 1 of repeater 2 (point A).

(ii) Cable far-end crosstalk (FEXT) caused by coupling between wire pairs in the main cable whose signals are propagating in the same direction and at the same level. Occasionally, T1 signals of different levels appear at the same point in the cable because two or more different routes merge at a point between repeaters. These are called incidental junctions. The increased FEXT noise resulting from this situation is not treated here.

(iii) Apparatus-case crosstalk (ACXT) caused by coupling in the wiring inside the apparatus case or in the case stub cable.

In standard T1 engineering layouts, main cable NEXT is negligible compared to ACXT and FEXT because opposite directions of transmission are segregated into binder groups on opposite sides of the cable as illustrated in Fig. 5. However, NEXT can be significant in some less commonly used configurations such as those described in Section 5.3.2. Assuming main-cable NEXT to be negligible for the standard configurations measured, the noise appearing at the repeater input was due to T1 sources with individual ones densities coupling through ACXT and

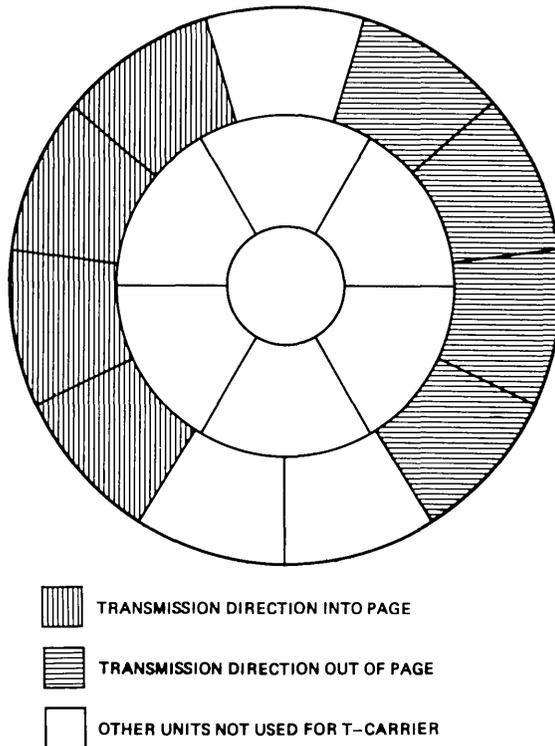


Fig. 5—Cable cross section showing use of cable units in a 900-pair, 22-gauge pulp cable with 200 active T1 systems.

FEXT paths of various losses. In the field measurement program, a simple means was found to separate the noise into ACXT and FEXT components and to determine an “effective” ones density for the disturbers.

3.1.1 Ones density (p) and fraction ACXT (α)

Based on measurements of the crosstalk noise-power density at the repeater input (before equalization), expected values for two parameters, p and α , have been calculated from a fit of the data to eq. (15) of Appendix B where:

(i) p is the effective ones density of all of the disturbers defined as that single ones density that best fits the shape of observed noise spectrum.

(ii) α is the fraction of the crosstalk noise density at 772 kHz at the repeater input that is attributable to ACXT.

The effective ones density of the disturbers proved to be a useful concept, even though it was known that all disturbers did not have the same ones density. Besides providing an indication of the average ones density (weighted by crosstalk coupling) in the T1 environment of the line being measured, it enabled a cleaner separation of the components of the noise spectrum, caused by ACXT and FEXT (i.e., the determination of α), than would otherwise be possible.

Distribution plots of p and α are given in Figs. 6 and 7. Ones density, p , ranges from 0.5 to 0.95 with an average of 0.69. This is consistent with a combination of ones densities from the older D1 banks (<0.5)

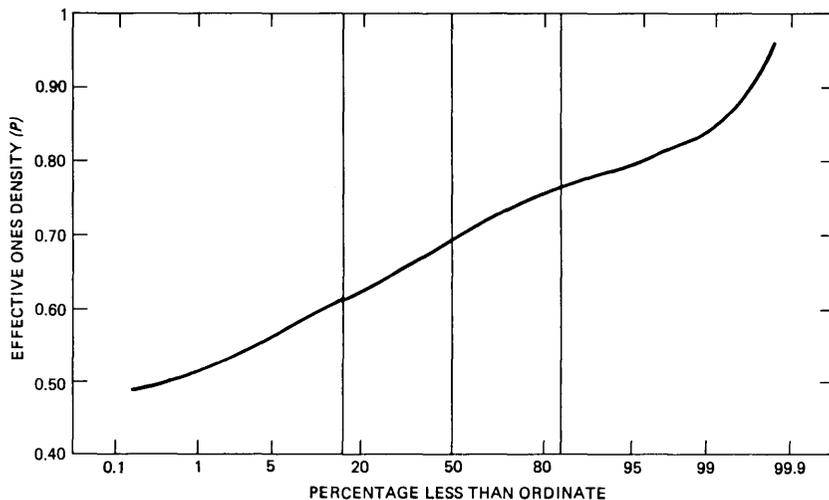


Fig. 6—Effective ones density, p , for the 466/201 situation. Number of points = 596; average = 0.69; standard deviation = 0.07.

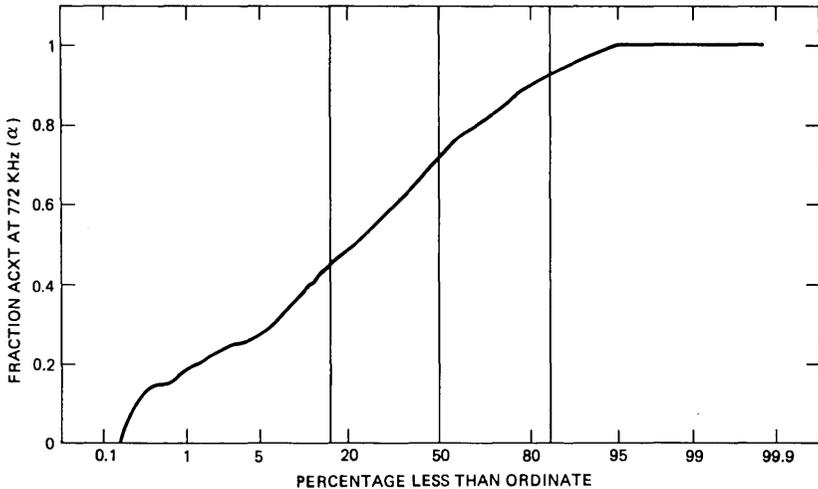


Fig. 7—Fraction ACXT at 772 kHz, α , for the 466/201 situation. Number of points = 596; average = 0.69; standard deviation = 0.22.

and from the newer D1D or D3 banks (>0.8). (In a few cases, it was known which bank predominated, and in these situations, the fitted value of p was seen to track the bank type.)

From Fig. 7, α is seen to range from 0 to 1 with a mean value of 0.69. The fraction represented by α is larger than 0.5 on 80 percent of the lines measured, indicating that ACXT tends to dominate over FEXT, at least at 772 kHz. Furthermore, in at least 5 percent of the measured lines, there is no observable FEXT at all ($\alpha = 1$).

3.1.2 Predicted intermediate section margins

In Fig. 8, predicted margin distributions due to ACXT, FEXT, and NEXT, and their composite, are compared to measured repeater margins for the standard layout of 22-gauge pulp multipair unit cable and 466-type apparatus cases.

The ACXT and FEXT curves are based on calculated distributions of ACXT and FEXT noise powers for those T1 systems measured in the field. (See Section 4.4.) The NEXT distribution is based on published NEXT pair-to-pair statistics for remote- and alternate-unit coupling in a 900-pair, 22-gauge pulp cable, with 200 active T1 systems. The numerical details of the crosstalk and margin calculations appear in Appendix A.

The predicted margin distributions were calculated using the margin equation, (see Section 4.2), with the assumption that crosstalk noise was the only variable. Other parameters, such as the cable section loss, ones density of signals on the crosstalk disturbers, repeater degrada-

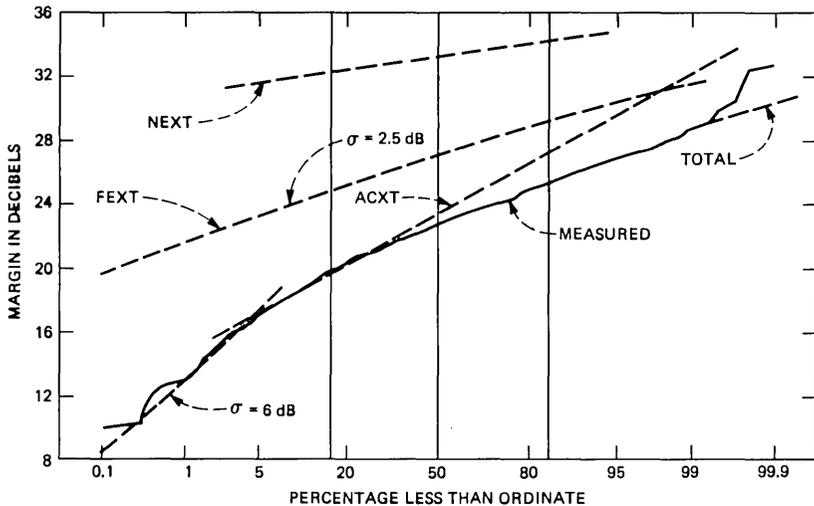


Fig. 8—Predicted service margins due to NEXT, FEXT, and ACXT. Number of points = 915; average = 22.6; standard deviation = 3.03.

tion, and peak factor of the noise, were held to the mean values observed over the course of the field measurements. (See discussion of these parameters in Section IV.)

From Fig. 8, it can be seen that the total derived margin distribution is dominated by apparatus-case crosstalk at the low margin end, which is the significant region for T1 engineering purposes. (The margin at the 0.1-percent point is the engineering parameter used to account for the longest metropolitan T1 systems, which consist of 50 repeater sections in tandem.) It is only at higher margins that FEXT makes a significant contribution to the total margin. Near-end crosstalk has a negligible effect on the total margin for all lines.

Since, for engineering purposes, it is only necessary to know the behavior of the distribution at the low-margin end, a model that only considers the contribution of ACXT is sufficient for predicting T1 performance. New system layouts, however, will use apparatus cases with reduced internal crosstalk coupling so that cable FEXT may become the limiting factor on performance for these systems. Also, any layout that allows opposite directions of transmission to be routed through adjacent binder groups or closer must take into account cable NEXT. (See Section 5.3.2.)

The close agreement of the predicted margin curves with measured data provides confidence that the major noise sources and coupling mechanisms that dominate margin performance of intermediate repeater sections are understood.

3.2 End-section noise environment

The noise sources and coupling paths for end sections vary more than those for intermediate sections, because of the multiplicity of cabling and equipment layouts, as well as switching type and activity occurring at offices at which through or terminating T-carrier systems appear. Figure 9 is a sketch of possible coupling paths for noise that eventually appears at the decision point of a T1 office repeater, but such noise paths are not completely known or characterized.

The end-section margin distribution (Fig. 3) has a standard deviation (1.9 dB) that is consistent with the predicted FEXT margin distribution (Fig. 8), but the average margin for end sections is worse than that predicted for intersystem FEXT alone. Other noise sources, perhaps office switching or traffic noise coupling through NEXT paths in the end-section cable, might be responsible for the difference.

Since we know that the office noise environment tends to be impulsive in nature because of switching activity, an effort was made to further characterize end-section noise observed in T1 office repeaters as either impulsive, Gaussian, or truncated by observing some characteristics of the noise-amplitude distribution. Figure 10 will help to illustrate the relationship between two parameters measured on each line.

The first parameter, the noise peak factor, is defined as the ratio (expressed in dB) of the repeater-error threshold voltage (instantaneous voltage that must be exceeded by the noise for the repeater to

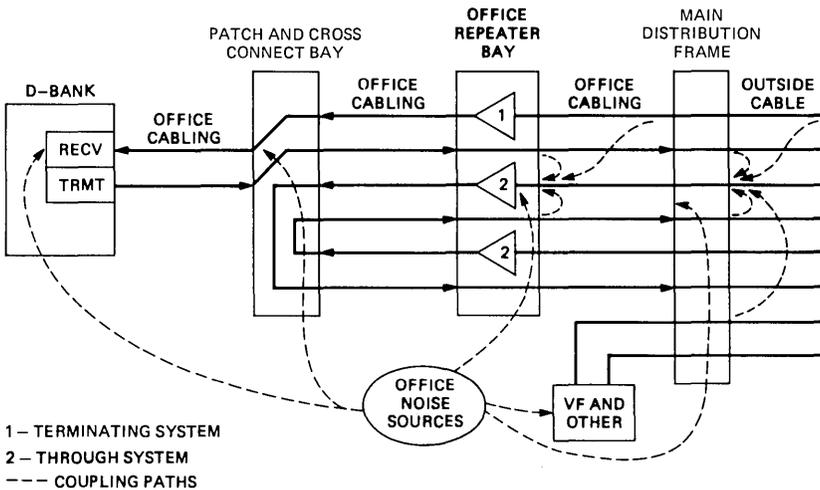


Fig. 9—Noise paths for end sections. Noise sources: cable crosstalk (from T1 and other services); frames, grounds, and power supply lines; radiative transmission; and unknowns.

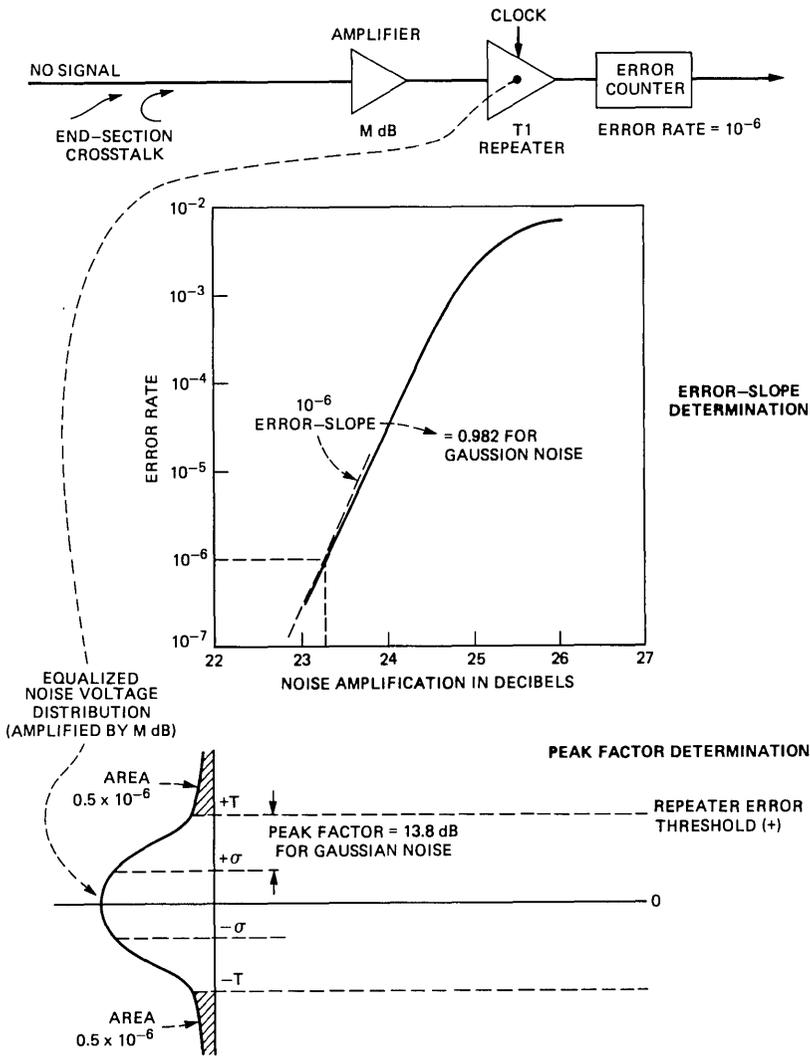


Fig. 10—End-section margin measurement for the 466/201 situation.

make errors) to the rms noise voltage which results in a 10^{-6} error rate. The second parameter, error-rate slope, represents the change in the log of the error-rate-per-dB change of noise amplification at the repeater input, near the 10^{-6} point (log of error rate = -6).

Both the peak factor and error-rate slope deviate predictably from the known values for a Gaussian-noise-amplitude distribution when impulses are present, or when a few disturbers (T1 or periodic in nature) are dominant. The presence of impulses tends to lengthen the

tail of the noise-voltage distribution, thus, raising the noise peak factor and reducing the error-rate slope. The dominance of a single disturber sets a limit on the peak of the noise (truncates the noise-amplitude distribution), with consequent reduction of the peak factor and increase of the error-rate slope at the 10^{-6} point. For example, the expected peak factor for Gaussian noise is about 13.8 dB, while the peak factor of the noise from a single T1 disturber (coupled through a FEXT path) is about 7 dB. The expected error-rate slope for Gaussian noise is 0.982, while Cravis and Crater (Ref. 2) measured slopes close to 0.1 (decades per dB) for office switching noise. See Fig. 10.

Figure 11 shows the distribution of peak factors observed in the end-section measurements. The mean value, 13.6 dB, indicates a slight tendency towards truncation, rather than impulses. However, about 15 percent of the lines show peak factors greater than 13.8, indicating impulsive effects. These points, however, were entirely from one measured cable, so that impulsive behavior was not the norm for the end sections measured.

Figure 12 shows the distribution of error-rate slope for all end sections measured. The mean slope, 1.0, indicates a tendency towards truncation of the noise distribution, in agreement with the peak factor results. The lower 15 percent of the distribution, indicating impulsive noise behavior, is again entirely due to the same particular end-section cable mentioned above.

In summary, noise on these central office pairs appears, overall, to be very nearly Gaussian in nature, with a slight tendency towards

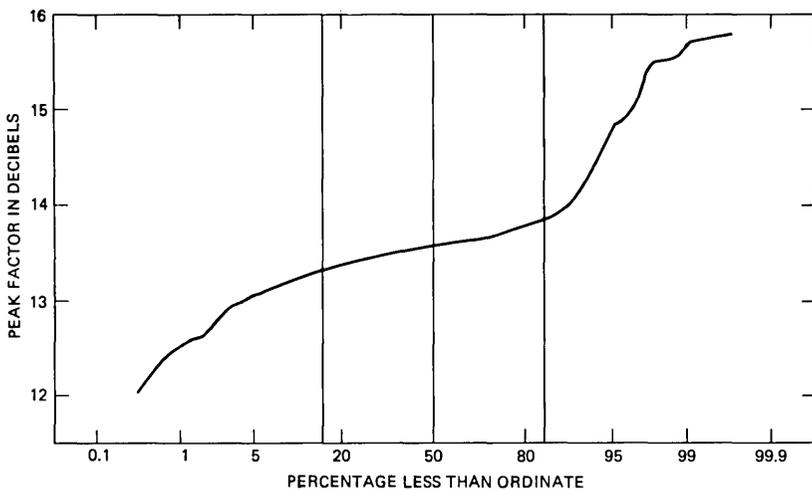


Fig. 11—End-section peak factors for the 466/201 situation. Number of points = 303; average = 13.6; standard deviation = 0.50.

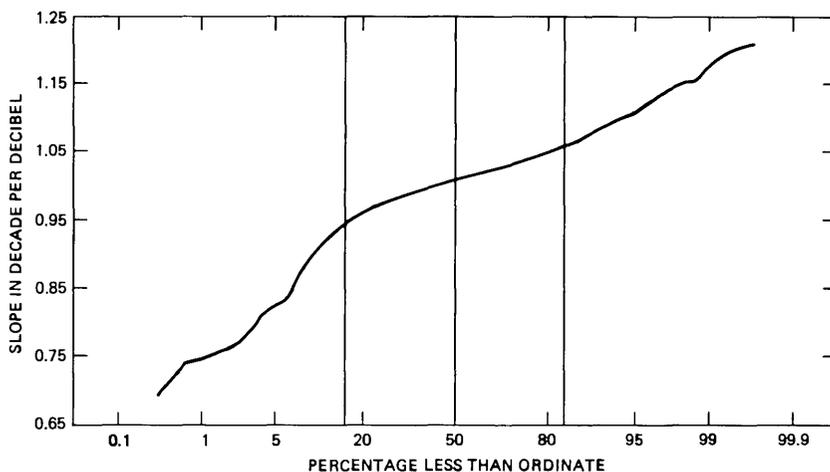


Fig. 12—End-section slopes for the 466/201 situation. Number of points = 300; average = 1.0; standard deviation = 0.08.

single-disturber effects (truncation). Only one cable was observed to have a noticeable impulsive noise background.

IV. ENGINEERING MODEL FOR T1 CARRIER

The performance objectives for individual T-carrier repeater sections depend on the maximum number of sections that are to be connected in tandem, as well as on the end-to-end performance objective for the complete T-carrier system. In metropolitan areas, where the maximum number of sections in tandem is expected to be 50, and where 95 percent of end-to-end systems must have error rates less than 10^{-6} , the repeater-section objective can be stated: 99.9 percent of repeater sections must have error rates less than 10^{-6} .

This objective is based on the assumption that the overall error rate of an end-to-end system is dominated by a single repeater section, such that the probability for the whole system of exceeding the specified error rate is the sum of the probabilities for each repeater section of exceeding the same error rate.

By definition, the margin of a repeater section that has an error rate of 10^{-6} is zero, and the margin of a section with a smaller error rate is the amount of amplification of the noise at the repeater input that can be tolerated before an error rate of 10^{-6} is reached. Since digital transmission systems are designed to have at least 3-dB reserve margin, the repeater-section objective can be stated: 99.9 percent of repeater sections must have margins better than 3 dB.

In other words, the margin at the lower 0.1-percent point of the repeater-section margin distribution must be greater than 3 db. The

margin distributions obtained from field measurements meet this objective with 4 to 6 dB to spare, as was shown in Fig. 2. If there is reserve margin in most of the T1 plant, it is conceivable that engineering rules could be modified, allowing more wire pairs in a cable to be reserved for use by the T-carrier. The following model may be used to predict the margin at the 0.1-percent point for various situations of cable-section layout and crosstalk noise.

4.1 Factors affecting repeater-section margin

In a complete engineering analysis, one must be concerned with all of the following:

(i) Characteristics of the signal and noise sources (ones density of interfering T1 signals, mean time between impulsive events, etc.).

(ii) Amount and character of noise coupling (ACXT, FEXT, impulse noise, etc.).

(iii) Effects of system layout (cable loss, junctions) on both signal and noise.

(iv) Equalization properties of repeater (intersymbol interference and ALBO setting).

Most of the above elements are accounted for in the margin equation used for section engineering calculations.

4.2 Margin equation

From the definition given above, a margin equation may be written

$$M = N(10^{-6}) - N(\text{actual}), \quad (1)$$

where M is the repeater-section margin in dB, $N(10^{-6})$ is the average noise power in dBm at the repeater decision point (eye) which would produce an error rate of 10^{-6} , and $N(\text{actual})$ is the actual average noise power in dBm at the repeater eye for the given situation. It is assumed that the character of the noise (peak factor, spectral shape) in the first term above is identical to the actual noise encountered on the given repeater section, except for a flat gain.

The margin eq. (1), when applied to margin measurements involving noise amplification, also assumes that the noise appearing at the (internal) repeater decision point increases or decreases linearly with external amplification or attenuation at the repeater input. This assumption does not necessarily hold for repeaters with automatic gain adjustment (provided by an ALBO), since when the noise becomes a significant fraction of the signal voltage, the ALBO may be misled into assuming that a larger signal exists and reduce its gain accordingly. However, the adjustment of the margin equation (1) (by the addition of a new term to account for nonlinearity) is offset by a reduction in the peak signal power S (see below) by roughly the same amount so

that eqs. (1) and (2) may be used as given, if it is assumed that the repeater gain is always properly adjusted for the correct cable loss. (The above statement is true only for those repeaters for which the effective error threshold (noise voltage at which errors are made) is reduced by the same amount in dB that the ALBO gain has changed.)

For engineering analysis the margin equation may be broken up into five terms as follows:

$$M = (S - B - A) - (I + Q), \quad (2)$$

where

$$N(10^{-6}) = S - B - A, \text{ and } N(\text{actual}) = I + Q.$$

S = the nominal peak signal power at the repeater decision point, which for the ideal T1 repeater equals 13.6 dBm. (See Section 4.6.)

B = the ratio of peak signal voltage to rms noise voltage that would produce an error rate of 10^{-6} with an ideal bipolar repeater, that is one with no intersymbol interference. For Gaussian noise, and assuming 50-percent ones density for the signal pulse stream, $B = 19.7$ dB. (Note: this peak factor is defined slightly differently from the end-section peak factor of Section 3.2. It relates the rms noise level to the *peak signal* level, rather than to the error-threshold level. If the error threshold is set to one-half the peak signal height, then the difference in definitions amounts to 6 dB. The 0.1-dB residual difference between the end-section Gaussian peak factor 13.8 and the Gaussian value for B , $13.7 + 6.0$, is because of the assumption of an "all zeroes" pulse stream for the end-section measurements, but a 50-percent ones density in the calculation of B .)

A = the repeater degradation, an allowance for the nonideal performance of the repeater caused by intersymbol interference, threshold offsets, and jitter. A has traditionally been assigned a value of 6 dB in engineering calculations.² Empirical values for A , smaller than 6 dB, have been found in the present measurements.

Then, $S - B - A$ is equal to the average noise power in dBm at the repeater decision point that will produce a 10^{-6} error rate in a real repeater, where B depends on the characteristics of the noise present and A depends on the equalization properties and threshold setting of the particular repeater.

The actual (measured) noise power at the repeater decision point comprises two terms, I and Q , where:

I = average noise power in dBm at the repeater decision point

which would be caused by a *single disturber* coupled through a *net coupling path loss of 0 dB at 772 kHz*. The "path" may include components from any or all of the ACXT, FEXT, or NEXT type.

Q = scaling factor in dB which accounts for the fact that there are *many disturbers* of different types (ACXT, FEXT, NEXT) with a *distribution* of coupling path losses. As an "example," for ACXT alone, Q is the power sum of the apparatus-case slot-to-slot crosstalk coupling coefficients from all outputs (disturbers) into the slot being measured for margin.

$I + Q$, called the *noise power sum*, is equal to the actual average noise power in dBm at the repeater decision point.

In Fig. 8, where the margin components (because of ACXT, FEXT, and NEXT) were separated, the coupling power sum, Q , was assumed to be the only variable. We will now investigate all terms in the margin eq. (2) to determine their actual variances as measured or derived in the field-measurement program.

4.3 The noise power sum

In the field-measurement program, the average noise power ($I + Q$) for all lines was measured directly in the absence of the T1 signal at the decision point of a test repeater, whose ALBO was held fixed to equalize for a particular length of cable (31.7-dB section loss at 772 kHz). Since the margin M is inversely related to the noise power ($I + Q$) in the margin eq. (2), one expects a high degree of correlation between margin and noise power ($I + Q$) from line to line, if the peak signal power S , noise peak factor B , and repeater degradation A do not vary appreciably from one line to the next.

A scatter plot of loss-adjusted* test margin versus noise power is presented in Fig. 13. The solid line represents the best first-order fit to the data, where A_0 and A_1 are the intercept and slope of the fit. The dashed line represents the best fit with the slope constrained to -1.0 , where B_0 is the intercept obtained from the fit. The plot shows that the correlation between margin and noise power sum is almost dB for dB, with only a slight tendency towards increased margins at the low-margin (high-noise) end of the plot. The rms error in assuming a perfect one-to-one relationship between margin and noise power sum is 0.75 dB as indicated at the bottom of the plot; this may be considered to be the maximum error in assuming the other terms of the margin equation, S , B , and A to be constants. (The actual error is less because

* The margins are adjusted to the value they are expected to have for a section cable loss of 31.7 dB (at 772 kHz) which was the loss for which the fixed ALBO was set to equalize in the noise power measurements.

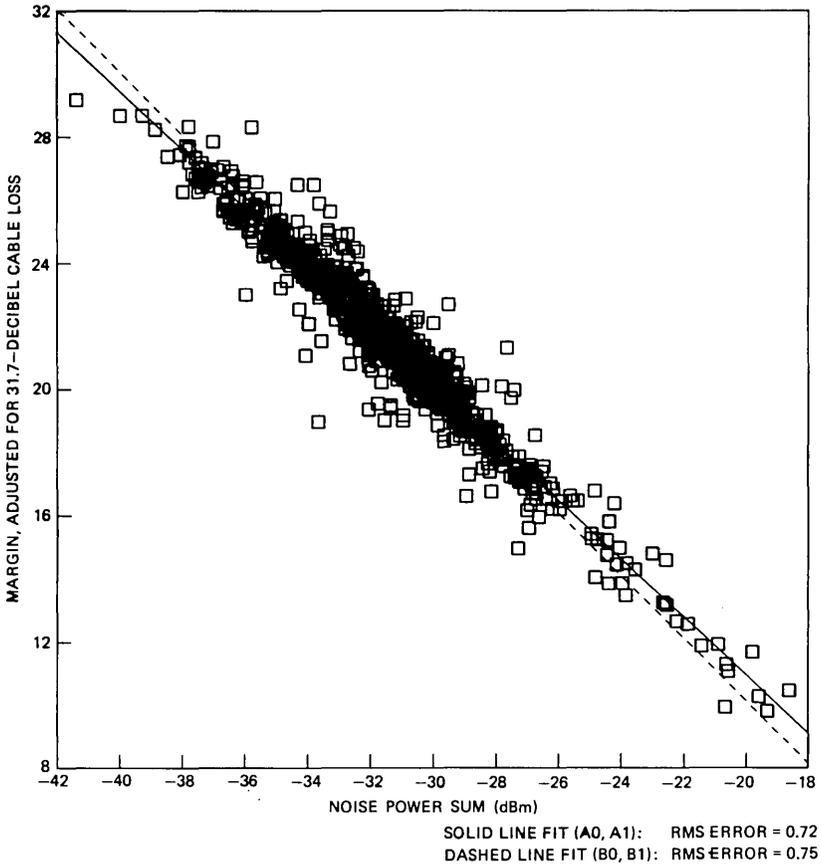


Fig. 13—Loss-adjusted test margins vs noise power sum for the 466/201 situation. Number of points = 850; $A_0 = -7.60$; $A_1 = 0.93$; $R = 0.97$; $B_0 = -9.86$; $B_1 = 1.00$.

the rms deviation on the scatter plot includes the effects of measurement error on margin, noise power, and cable loss.)

4.4 The crosstalk coupling power sum

The coupling power sum is derived in Appendix B and is given by

$$q(l) = a_A + a_F(l)c(l, f_0) \quad (3)$$

and

$$Q = 10 \log_{10} q, \quad (4)$$

where the quantities (except l) in (3) are power transfer ratios and

a_A = the power sum of the ACXT crosstalk coupling coefficients from all disturbers.

$a_F(l)$ = the power sum of the equal-level FEXT crosstalk coupling coefficients for a section of length l from all disturbers.

$c(l, f_0)$ = the cable power transfer function at 772 kHz.

The following sections describe the distributions of a_A and $a_F(l)$ found for the measured repeater sections.

4.4.1 Apparatus case crosstalk

The ACXT power sum a_A was calculated for each measured line using eq. (23) of Appendix B. The distribution of a_A for the field measurements is shown in Fig. 14. Since apparatus-case crosstalk is dominant, the properties of this distribution are reflected into the properties of the distribution of repeater-section performance (margin). In particular, the bend observed in the ACXT distribution at the 5-percent point is directly reflected into a bend in the margin distribution at the same point, causing the worst margins to be lower than would be expected if a normal distribution were assumed with the measured values for the mean and standard deviation. The 6-dB slope of the low end of the ACXT distribution matches the 6-dB slope observed at the low end of the margin distribution (Fig. 2).

Of special interest is the 0.1-percent point of the distribution, where the repeater-section margin objective is defined. In the simplified margin model, where all parameters except Q are held constant, the value of Q_{ACXT} at the 0.1-percent point determines how much the T1 plant exceeds the section objective. The value at this point, determined by extrapolation of the trend of the high coupling end of the distribution, is about -55 dB, which translates to about 8.5 dB of margin for the measured sections in Fig. 8.

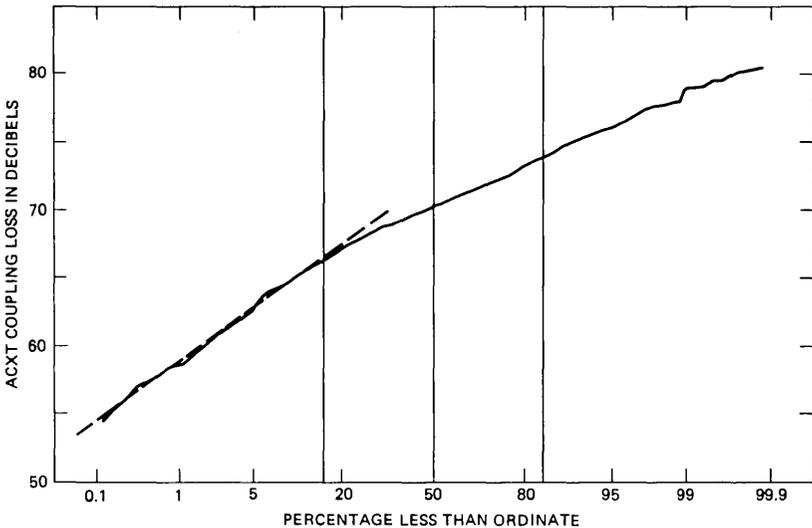


Fig. 14—ACXT power sum—all points measured for p and α —for the 466/201 situation. Number of points = 8.0; average = -70.0 ; standard deviation = 3.94.

4.4.2 Far-end crosstalk

The secondary source of crosstalk noise in the standard T1 system layout is FEXT occurring between cable pairs that have T1 signals travelling in the same direction and at equal levels. Coupling is most severe between pairs in the same binder group.

Far-end crosstalk power sums $a_F(l)$ were calculated using eq. (24) in Appendix B. The distribution of $a_F(l)$ for all lines is given in Fig. 15, where l is approximately 6000 feet.

The difficulties of separating ACXT and FEXT components of the noise for α 's close to unity is illustrated by the turn-up of the distribution at high coupling losses (right side of Fig. 15). The error in the fit for α is very high for these points and, therefore, they should be ignored in the characterization of the distribution properties.

Ignoring these points, the distribution may be roughly characterized as having a mean power sum of about 44.5 dB, with a standard deviation of about 2.5 dB. This is slightly less than the power sum mean for one cable measured directly for FEXT, for which the data (scaled to 772 kHz) are also plotted in Fig. 15. The standard deviation of the distribution of a_F is also slightly larger than that for the measured cable (1.5 dB). These differences imply less FEXT disturbers on average for the field measurements than for the laboratory-measured cable. However, the difficulties in calculating FEXT for different cable situations (different lengths, different splicing arrangements) make a meaningful comparison difficult.

The main distinguishing feature of the FEXT power sum distribution from the ACXT distribution is its smaller standard deviation, which

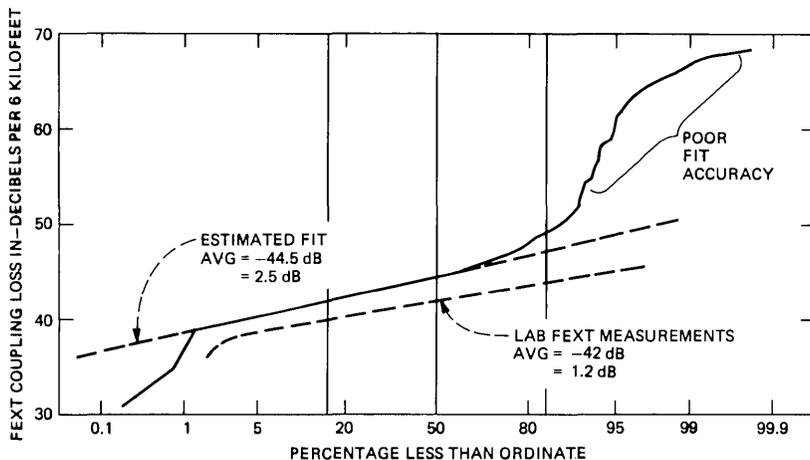


Fig. 15—FEXT coupling power sum for the 466/201 situation. Number of points = 514; average = 45.9; standard deviation = 5.68.

results from the greater number of noise sources and coupling points for FEXT. This lower standard deviation is reflected into the standard deviation of the upper end of the margin distribution (Fig. 2), where FEXT is strongest.

The predicted FEXT-alone margin distribution of Fig. 8, obtained from the α_F distribution and the margin equation, shows that FEXT is only a secondary source of noise for a majority of the T1 plant.

4.5 Single interferer noise power

By definition, I is the noise power at the repeater decision point attributable to a single interferer with 0-dB crosstalk coupling loss at 772 kHz. I depends on the spectral shape of the interferer's signal, the crosstalk coupling paths, the insertion loss of the line, and the repeater cable equalization. For the present analysis, the interferer's spectral shape is parameterized by p , the ones density, and the coupling path is parameterized by α , the ratio of ACXT to total noise (ACXT + FEXT) at 772 kHz. From Appendix B, an expression for I is written:

$$I = 10 \log \left[\int_0^\infty w(p, f) (f/f_0)^2 \left[\alpha + (1 - \alpha) \frac{c(l, f)}{c(l, f_0)} \right] r(l, f) df \right], \quad (5)$$

where

$w(p, f)$ = the one-sided interferer's signal power spectrum, which depends on the ones density p .

$(f/f_0)^2 \left[\alpha + (1 - \alpha) \frac{c(l, f)}{c(l, f_0)} \right]$ = the power transfer function of the crosstalk coupling path, which is defined to be 1 at $f = f_0$, and where $c(l, f)$ is the power transfer function of the cable section of length l at frequency f .

$r(l, f)$ = the power transfer function of the repeater cable equalization where l , the length of the cable section, determines the overall amount of gain through the action of the ALBO.

4.5.1 Dependence of I on cable insertion loss

The dependence of section margin on cable insertion loss is entirely through the I term of the margin eq. (2). The insertion loss power ratio $1/c(l, f)$ appears in both the repeater gain shape and in the FEXT part of the crosstalk coupling expression. For ACXT-dominated lines, I decreases by about 1.09 dB for every dB increase in insertion loss at

772 kHz. This is entirely through the action of the ALBO acting to change the overall gain of the repeater to bring the equalized signal up to a given peak level. The field measurements allowed the calculation of the quantity $I(L) - I(L_0)$ for each line measured, where L was the measured insertion loss in dB for the given line and L_0 is equal to 31.7 dB, which was the loss equalized for by the test repeater for all noise power measurements. This quantity eliminates the large variation in I for a given loss due to variation in Q over the lines measured. It is plotted versus measured insertion loss in Fig. 16. The one-for-one dependence of I on insertion loss is clearly seen.

4.5.2 Dependence of I on p and α

The single interferer noise power I also depends on the ones density of the interferer and on the fraction of interference due to ACXT and FEXT. A plot of I versus p for $\alpha = 0, 0.5,$ and 1 calculated for the test

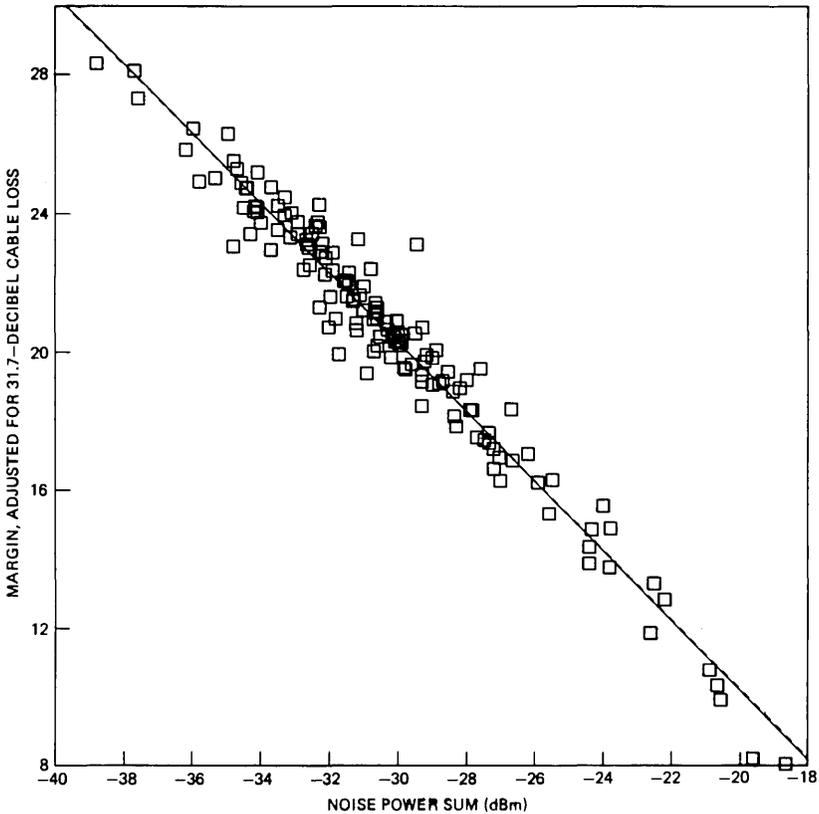


Fig. 16—Relative noise power integral vs. insertion loss for the 466/201 situation. Measurements relative to $L = 31.7$ dB. Number of points = 864; $A_0 = -33.56$; $A_1 = 1.05$; $R = 0.90$.

repeater for a 772-kHz section loss of 30.7 dB is given in Fig. 17. (The value of I for the theoretical 100-percent cosine roll-off repeater defined by eqs. (19) and (22) ranges about 1.5 to 3.0 dB greater than the equivalent value of I for the test repeater.)

The ones density dependence reflects the fact that as more pulses are sent on the disturbers, more crosstalk power is generated. Roughly speaking, twice the pulse density on the disturbers will produce twice the noise power (3-dB increase), except where adjacent pulse interference or coherent effects over several disturbers cause a large degree of cancellation or reinforcement.

The change of I with change in α is small and may be positive or negative depending on the repeater preamplifier gain shape. For example, I increases by about 0.6 dB for the theoretical 100-percent cosine roll-off repeater when α changes from 0 to 1. However, I decreases by 0.7 dB for the test repeater for the same change in α .

4.6 Peak signal power at the repeater decision point

For an ideal repeater, which perfectly equalizes the cable-loss shape, the T1 signal spectrum is operated on by a 100-percent cosine roll-off low-pass filter with compensation for 50-percent duty cycle pulses as shown in eq. (19). The overall gain at the midpoint of the roll-off

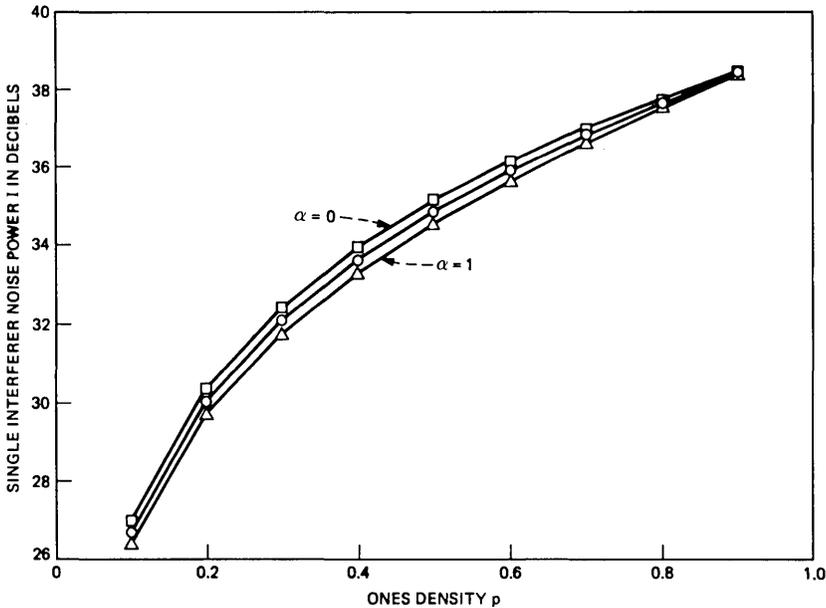


Fig. 17—Single interferer noise power vs. p and alpha measured for the test repeater—772-kHz loss = 30.7 dB.

function is $-6.0 + 0.91$ dB, where the second term is an adjustment for 50-percent duty-cycle pulses. Since the power of a 3-volt, 50-percent duty cycle, all ones T1 signal at 772 kHz is 18.7 dBm, the peak power at the repeater decision point may be estimated as $18.7 - 6.0 + 0.9 = 13.6$ dBm.

4.7 Peak signal-to-rms-noise ratio

The term B is defined as the ratio of the peak signal voltage to the rms noise voltage that would produce a 10^{-6} error rate in an ideal repeater. B will change depending on the ones density of the line being measured, since the probability of error given a zero (no pulse) is not necessarily equal to the probability of error given a one (positive or negative pulse). For Gaussian noise, B is 19.6 dB, assuming all ones, or 19.7 dB, assuming 50-percent ones, where the error threshold is assumed to be exactly one-half the peak equalized pulse height. Since the average ones density observed in the field measurements was about 0.7, we will, in general, assume the Gaussian peak signal-to-rms-noise ratio to be about 19.7 dB.

B was calculated for a subset of lines for which noise voltage distribution measurements were made in the field. This subset was not random, and tended to emphasize those lines that had low margins or obvious single disturber effects, as seen by a visual inspection of the crosstalk noise at the repeater decision point. The resultant distribution of B is plotted in Fig. 18.

The average value of B calculated for the selected lines was 19.3 dB,

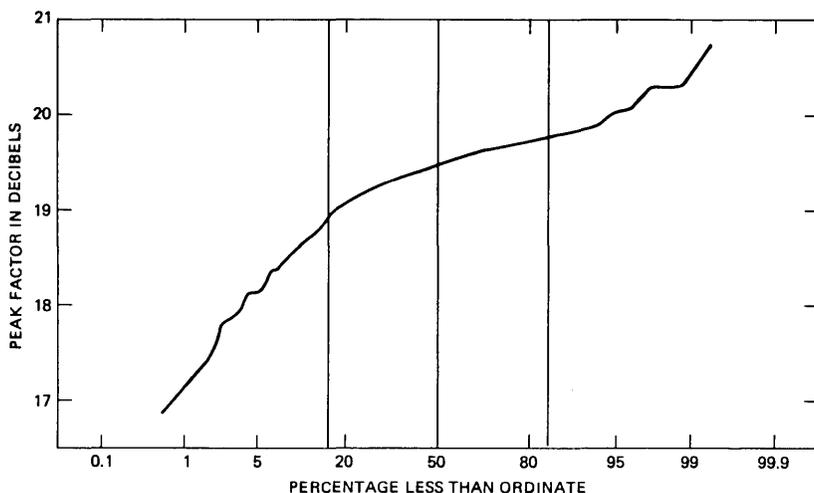


Fig. 18—Peak factor for the 466/201 situation. Number of points = 165; average = 19.3; standard deviation = 0.56.

quite close to the Gaussian value. The lowest value determined for B was about 17 dB, which is closer to the Gaussian value than the value expected for a single T1 disturber (13 dB for F_{EXT}).

A slight correlation of peak signal-to-rms-noise ratio with margin has been observed such that the lowest margins are increased by a small amount (by about 0.7 dB on average) over what would be expected if the noise were Gaussian with the same rms voltage, as assumed in the simplified model of Section III. This effect is shown in Fig. 13 by the fact that the best fit to the margin versus noise power data (solid line) has a slope less than one (dashed line), and that the fitted curve is about 0.7 dB above the unit-slope curve at the low-margin end. That the correction for peak factor indeed reduces this discrepancy is illustrated in Fig. 19, which shows the same plot as Fig. 13 (but only for those lines for which B was measured), except that the

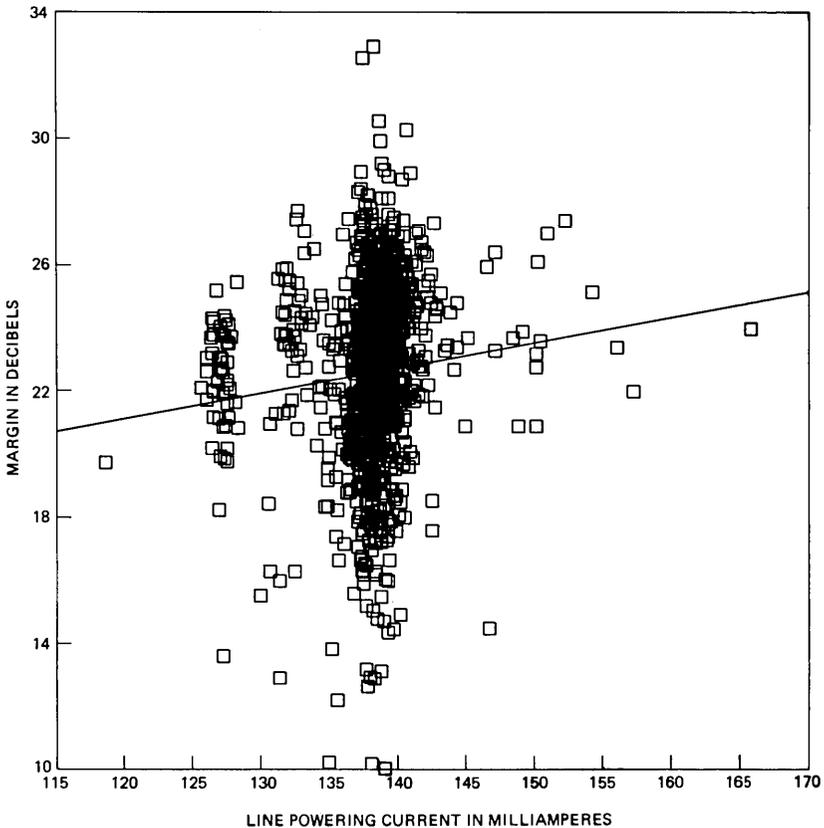


Fig. 19—Loss-adjusted test margin vs. power sum for the 466/201 situation. Number of points = 144; $A_0 = -9.89$; $A_1 = -1.00$; $R = 0.98$; $B_0 = -9.77$; $B_1 = -1.00$.

margins have been corrected to a constant value for B . The coincidence of the fitted and unit-slope lines is almost exact.

The main reason for the reduction of B for low-margin lines appears to be that apparatus-case crosstalk is controlled by within-slot coupling. With only one or two disturbers, the noise-amplitude distribution is truncated, and the peak factor is reduced. The reduction is quite small, however, and even in situations where one noise source was observed to be visually dominant at the repeater decision point, only a slight reduction of peak factor from Gaussian was seen, indicating that residual noise was present.

4.8 Repeater impairment

The value of A , the repeater impairment, is defined by the margin eq. to equal:

$$A = S - B - (I + Q) - M, \quad (6)$$

where $S = 13.6$ dBm and M and $(I + Q)$ are measured. The value of B , the peak signal-to-rms-noise ratio at a 10^{-6} error rate, is less well determined, since voltage-distribution measurements were made on only a small subset of the total lines measured.

If we assume the mean measured value, 19.3 dB, for B , we obtain a distribution for A for the test repeater as shown in Fig. 20. The distribution of calculated A values for the test repeater runs from 2 to 9 dB, a fairly large range. Other than variations caused by differences in cable loss versus frequency shape from pair to pair (resulting in differences in equalization from line to line, even with the same

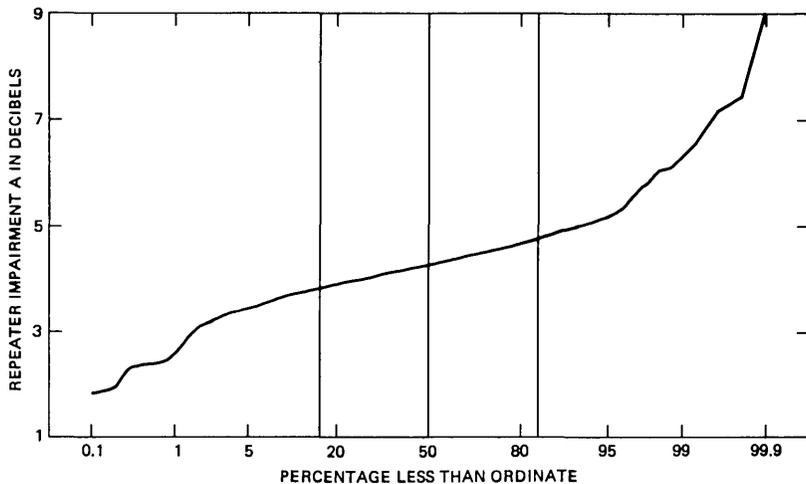


Fig. 20—Repeater impairment for the 466/201 situation. Number of points = 962; average = 4.31; standard deviation = 0.61.

repeater), one expects the degradation A for a given repeater to be constant. Indeed, although the extreme ends of the measured distribution of A are widely separated, 95 percent of the points lie between 3 and 6 dB, resulting in an overall standard deviation of only 0.61 dB about the mean of 4.3. Much of the variation is likely because of measurement error for the margin M and the noise power sum ($I + Q$), as well as the fact that the A values are uncorrected for variations in B .

The expected A for the ensemble of service repeaters measured in the field may be related to the differences in margin observed between test and service repeaters. The distribution of test-minus-service margin is plotted in Fig. 21. The mean difference is -0.2 dB, indicating the average service-repeater performance was slightly better than the test-repeater performance. The standard deviation of test-minus-service difference is 0.91 dB, indicating corresponding variation in repeater properties (equalization gain shape, threshold offsets, etc.).

Since the distribution of test-minus-service margin has a mean near zero for the field measurements, the expected A value for service repeaters is also 4.3 dB, if one assumes the average I value (which depends on the repeater gain shape) to be the same for test and service repeaters. If I is not the same, then at best we know that the quantity ($I + A$), which is a measure of the overall repeater performance, is the same on the average for the test and service repeaters.

Since the 4.3 dB value for A is 1.7 dB less than has been previously assumed for section engineering,² and the value of I for the test repeater is 1.5 to 3.0 dB less than I for the 100-percent cosine roll-off repeater, which has been used in the past for system engineering, the

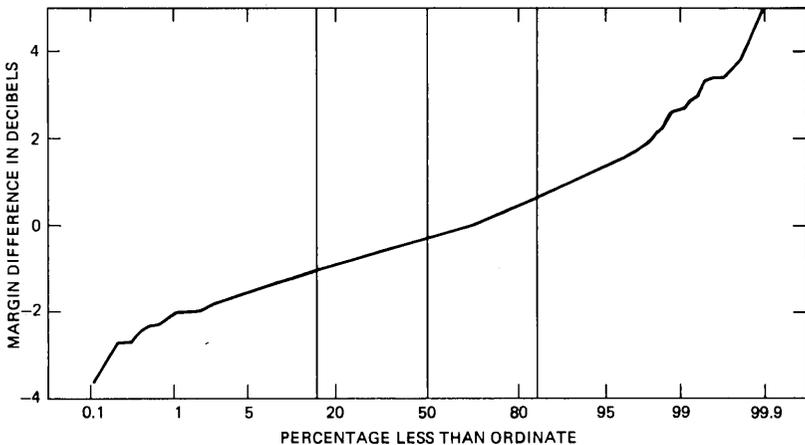


Fig. 21—Test-minus-service repeater margin for the 466/201 situation. Number of points: 913; average $0 = -0.20$; standard deviation = 0.91.

overall repeater performance observed in the field averages about 4.0 dB better than has been previously assumed.

4.9 Summary of margin model as applied to field measurements

The margin eq. (2) has been used to account for the margin performance of intermediate-repeater sections of known cable loss L , in a noise environment characterized by peak signal-to-rms-noise ratio B , effective disturber ones density p , fraction apparatus-case crosstalk α , and total equalized noise power $(I + Q)$. The independent measurement or calculation of these quantities, along with the measurement of actual margin itself allowed the determination of a value for the repeater degradation A for the test repeater, and by inference, limits on A for the population of service repeaters in the field.

Since the range of variation of the parameters L , A , and B was very small for the particular set of lines measured (e.g. almost all lines were 6 kft long), it was possible to obtain a very good fit to the measured margin distribution by setting L , A , and B to their mean values and assuming that the crosstalk coupling power sum Q was the only random variable from line to line. The distribution of Q then mapped directly onto the distribution of M .

However, it is desirable to go beyond the limited set of repeater-section layouts studied in the field measurements. If the total population of T1 repeater sections is to be considered, values for model parameters, especially section-loss L , must be taken from distributions of all possible values in order to predict current performance, or must be taken from expected extreme limits in the engineering of sections to perform properly under all conditions.

V. ADDITIONAL CHARACTERIZATION RESULTS

Measurements of basic cable and system parameters (such as insertion loss and dc powering current) and of margins for less common or nonstandard situations were made in the field-measurement program. Some of the results are described in this section.

5.1 Cable insertion loss

The insertion loss of each cable section measured has a direct impact on the margin expected for the section, mainly because the amount of signal attenuation is the main determining factor of the expected signal-to-noise ratio at the receiving repeater input.

Figure 22 is a plot of the distributions of cable-section insertion loss at 772 kHz for all of the 22-gauge pulp sections measured for margin. The different average section losses reflect the different section lengths. The standard deviation of loss about the mean is approximately the same for all sections measured and is about 0.5 dB. This

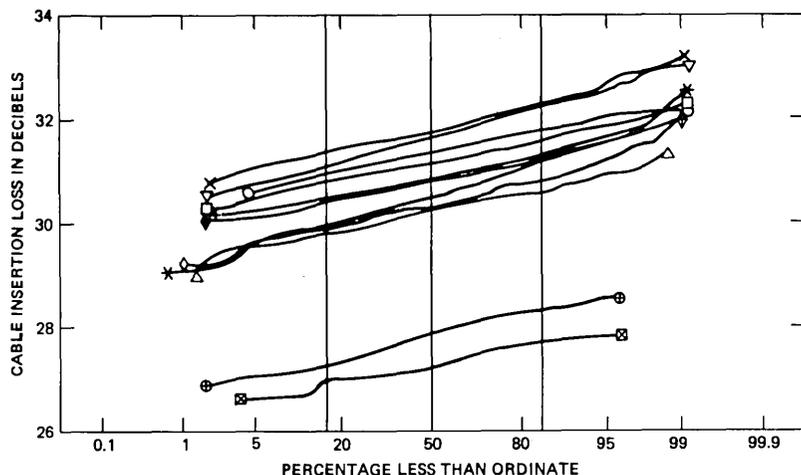


Fig. 22—Distributions of 772-kHz insertion loss showing separate results for each cable for the 466/201 situation. Number of points = 923.

variation is reflected into an expected variation of service margins for each section, which, however, is usually not directly apparent since variations due to crosstalk noise differences from line to line predominate over variations due to differences in insertion loss.

If insertion losses are normalized to 1 kft, the distribution of losses in dB per kft of Fig. 23 is obtained. The mean result, slightly less than 5.2 dB per kft, is in agreement with the 772-kHz losses of previous measurements of 22-gauge pulp cable.

5.2 Dc simplex powering current

One of the measures of system integrity is the dc current used to power the manhole repeaters. This current is sent in simplex mode (longitudinally over the same wire pair used for the signal) from one of the span offices through all the side-1 regenerators, is looped back at some intermediate point or at the other span office, and returns to the original office through all the side-2 regenerators of the given line.

The nominal value for this current is 140 mA dc for the older repeater types. Figure 24 shows the distribution of observed currents for the repeater sections that were measured for margin. While the mean is close to the nominal value, some outliers are strongly apparent. (The set of points at about 125 mA all come from the same span.) However, the repeater powering voltage is not strongly dependent on the line powering current, and therefore, most repeaters will operate without observable margin degradation over a wide range of current. A scatter plot of margin versus powering current, Fig. 25, shows no obvious correlation; however, since the margin variations are domi-

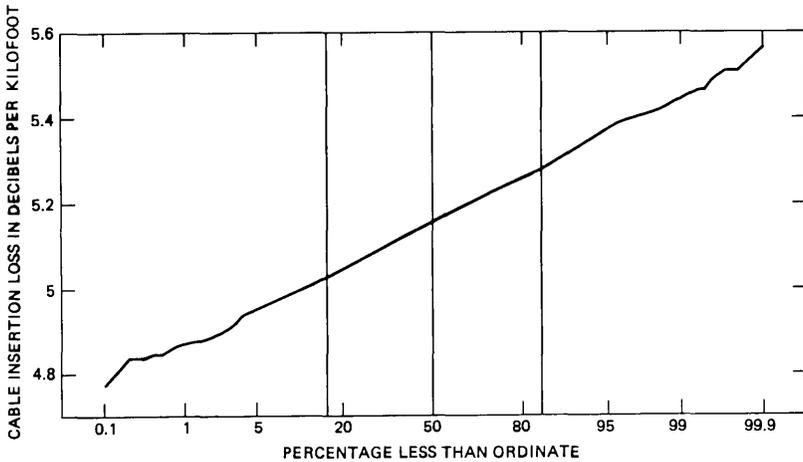


Fig. 23—Distributions of 772-kHz insertion loss in dB/kft for the 466/201 situation. Number of points = 923; average = 5.16; standard deviation = 0.12.

nated by crosstalk noise variations, small correlations between margin and powering current are not easily observed.

5.3 Other margin measurements

While we have concentrated on a “standard” configuration of 22-gauge pulp cable with 466-type apparatus cases and 201-type repeaters, it is important to note that a significant fraction of the physical plant does not have this configuration. Perhaps the most common difference are the use of lightning-protected 205-type repeaters in 468-type ap-

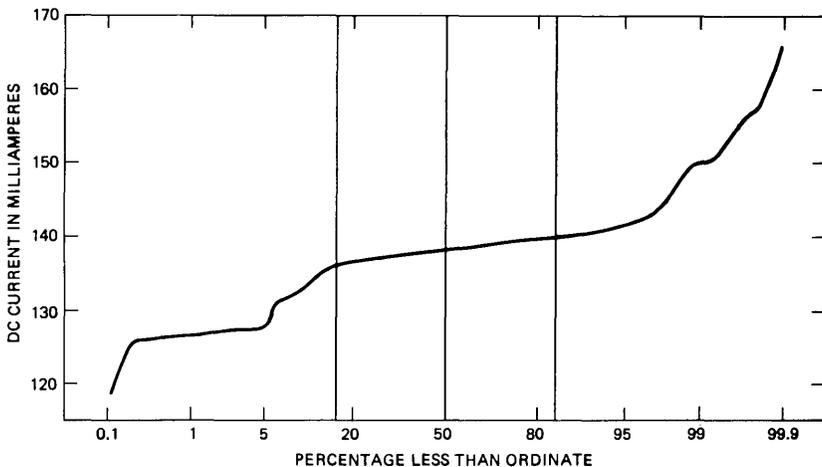


Fig. 24—Line-powering currents for the 466/201 situation.

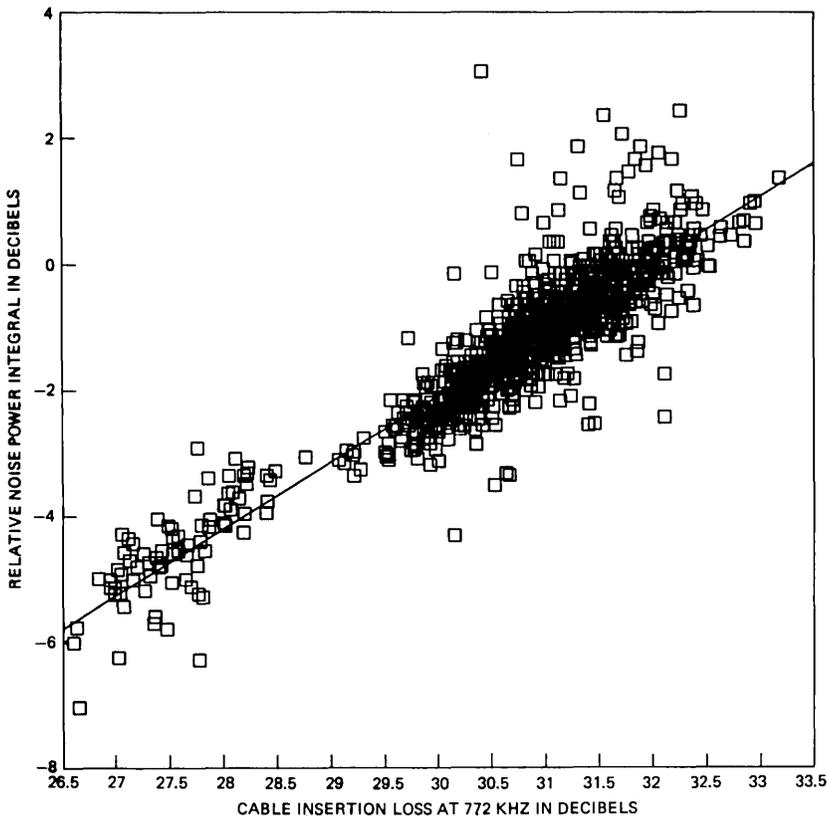


Fig. 25—Margin vs. line-powering current for the 466/201 situation. Number of points = 922; average = 137.9; standard deviation = 3.85.

paratus cases, and the deployment of the newer and smaller 208- and 209-type integrated circuit repeaters in the 475-type apparatus case. Also, the use of 19- and 24-gauge instead of 22-gauge and the use of polyethylene-insulated conductor (PIC) instead of pulp insulated cable is not uncommon. A relatively small subset of field measurements treated some of these “nonstandard” situations.

5.3.1 Apparatus case type 475

About one sixth of the lines measured for margin had 475-type apparatus cases rather than the more usual (for these measurements) 466-type. These cases house the 208- or 209-type integrated circuit repeater. A comparison of case performance may be made by observing margins measured using the test repeater for both case types.

Figure 26 shows the measured margin distributions for the test repeater for the two types of cases. The mean margin for the 475 distribution is about 1 dB worse than the mean 466 margin. However,

differences in FEXT contributions for the two measurement groups are responsible for at least part of this difference.

The lower portions of the two distributions, most likely dominated by ACXT effects, have estimated asymptotes indicated by the straight lines shown. Although the data available for the 475 cases are small, it appears that they may be treated equivalently to 466 cases in margin calculations, since the estimated effects at the 0.1-percent point are the same.

5.3.2 Nineteen-gauge cable, NEXT exposure

One of the special conditions encountered in the field measurements was a single 455-pair 19-gauge pulp cable using 22-gauge section lengths, which was completely filled with T-carrier so that opposite directions of transmission sometimes appeared in adjacent binder groups. This configuration made cable NEXT the dominant source of noise at the repeater input, at least for those binder groups that were adjacent to T1 groups with transmission in the opposite direction.

Figure 27 shows the measured distributions of test repeater margins for each of the six binder groups observed for this cable. The lowest margins occur for binder groups in the outer part of the cable, where adjacent-unit NEXT was dominant. The next lowest margins occurred for inner binder groups that were subject to adjacent- or alternate-unit NEXT from other inner or outer binder groups. The best margin distribution was observed for a binder group that had no adjacent-unit NEXT exposure.

The situation illustrates the engineering tradeoff between number

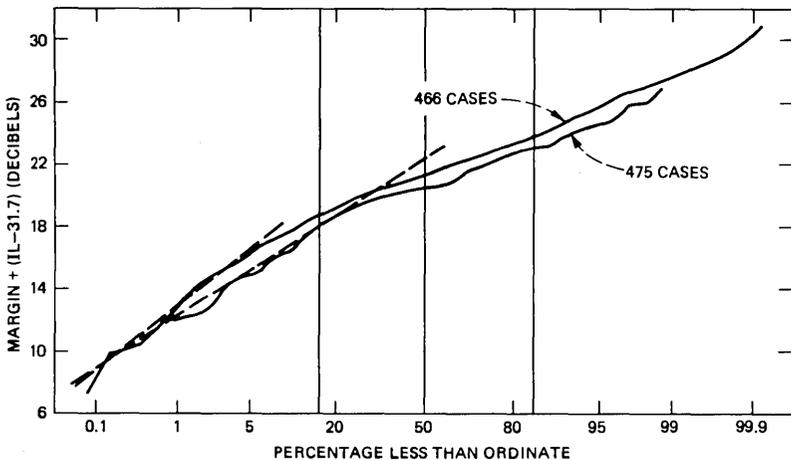


Fig. 26—Test margins (loss corrected) for the 466 and 475 apparatus cases. Number of points = 1401; $A_0 = 11.47$; $A_1 = 0.08$; $R = 0.10$.

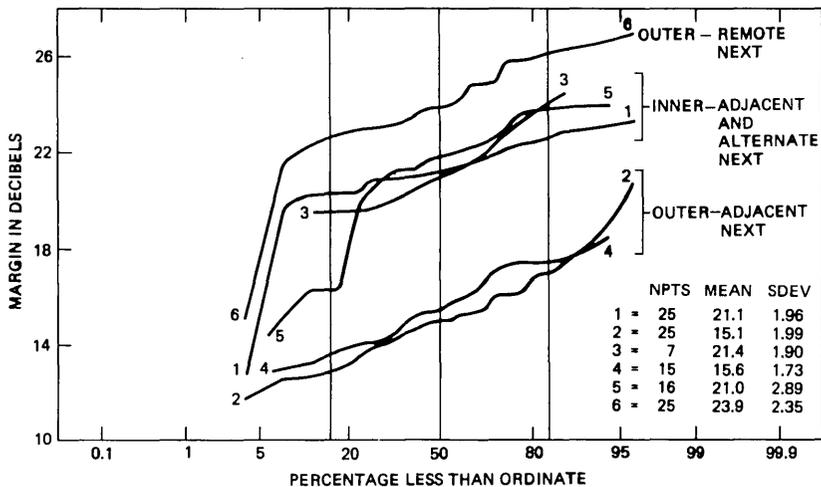


Fig. 27—Test margins for a 23-dB loss cable section with adjacent-unit NEXT. Number of points = 113.

of systems allowed in the cable and repeater section transmission loss. About twice the usual number of systems have been allowed in the cable (thus, raising levels of NEXT), while the section loss has been held to about 23 dB (by using 22-gauge section lengths for 19-gauge cable). Also illustrated is the fact that outer binder groups are more susceptible to adjacent-unit NEXT coupling (especially from other outer binder groups) than are inner binder groups.

VI. SUMMARY

The measurement program described here has provided data on about 2000 T1 intermediate repeater section lines most of which are operating in the 466/468 apparatus-case environment. The principal results of this program are as follows:

(i) The average margin observed for T1 repeater sections (whose average cable insertion loss is 30.7 dB at 772 kHz) is 22.6 dB. The estimated margin at the 0.1-percent point of the distribution is 8.5 dB. These results imply that for properly engineered systems, almost no errors should result from intersystem crosstalk.

(ii) Repeater section margin at 10^{-6} error rate may be accurately predicted if the crosstalk noise power sum at the decision point of the T-carrier repeater is known. Other parameters, such as repeater degradation *A*, or noise peak factor *B*, may be held to constant (average) values in the prediction of margin.

(iii) Apparatus-case crosstalk is the dominant form of crosstalk interference for intermediate sections where 466, 468, and 475 cases are used. 80 percent of lines have more crosstalk energy at 772 kHz

from ACXT than from FEXT; about 5 percent of lines show no FEXT at all ($\alpha = 1$).

(iv) The average repeater performance ($I + A$) observed in the field measurements is about 4 dB better than has been previously assumed in models featuring the 100-percent cosine roll-off transmission channel. The single interferer noise power I is 1.5 to 3.0 dB less for a specific test repeater than for the 100-percent cosine roll-off repeater, and the repeater degradation A for the practical repeater is 1.7 dB less than has been previously assumed in the engineering model.

(v) The mean peak factor determined for intersystem crosstalk noise is about 19.3 dB, very close to the Gaussian value of 19.7. The peak factors for low margin lines are about 0.7 dB less than the average, thereby raising the margins of these lines about 0.7 dB above that expected from knowledge of only crosstalk noise power and the average peak factor. Most end sections exhibited very nearly Gaussian noise; a few did show evidence of noise that was impulsive in nature (high peak factors).

VII. ACKNOWLEDGMENTS

The authors appreciate the help and cooperation given by the management, office craftspeople, and outside plant technicians of the Bell System operating companies in which the field measurements were performed. The contributions to the field-measurement program and subsequent analysis by the members of digital transmission groups at Bell Laboratories in West Long Branch, N. J., as well as discussions with members of interested groups at Bell Laboratories in Holmdel, N. J. and in Merrimack Valley, Mass. are gratefully appreciated.

APPENDIX A

ACXT and FEXT Components of Margin

The predicted margin curves of Section III were calculated from the margin equation (2), assuming constant values for all terms except Q , the power sum of the crosstalk coupling coefficients. The following values were used:

$$S = 13.6 \text{ dBm}$$

$$B = 19.3 \text{ dB}$$

$$A = 4.3 \text{ dB}$$

$$p = 0.69$$

$$\alpha = 0.69$$

$$L = 30.7 \text{ dB}$$

$$I(p, \alpha, L) = 36.53 \text{ (ACXT)}$$

$$I(p, \alpha, L) = 37.04 \text{ (FEXT)}$$

$$I(p, \alpha, L) = 36.45 \text{ (NEXT)}$$

The values for I were obtained by applying eq. (5) to the test repeater used in the margin measurements. Results for ACXT and FEXT are plotted in Fig. 17. For the NEXT value for I , eq. (5) was used except the (f/f_0) dependence was changed from squared to the power 1.5, characteristic of NEXT.

Then the margin equations for ACXT, FEXT, and NEXT become:

$$M_{ACXT} = -46.59 - Q_{ACXT} \quad (7)$$

$$M_{FEXT} = -47.10 - Q_{FEXT} \quad (8)$$

$$M_{NEXT} = -46.51 - Q_{NEXT} \quad (9)$$

Using the values for Q derived from Figs. 2 and 3, the predicted margin curves in Section III for ACXT and FEXT were obtained. Q_{ACXT} and Q_{FEXT} were obtained using eqs. (23) and (24) in Appendix B. The distribution of Q_{NEXT} was obtained by assuming average alternate unit NEXT pair-to-pair coupling averaging -103 dB with standard deviation of 7 dB and assuming 50 alternate-unit disturbers.² The average power sum Q_{NEXT} for this situation is -80 dB with standard deviation 2.0 dB. Applying this to eq. (22) produces the predicted NEXT margin curve of Fig. 8.

APPENDIX B

ACXT and FEXT Components of Noise

Assuming that NEXT is negligible, the noise power density (watts/Hz) at the repeater input (before equalization) because of a single T1 disturber may be broken down into ACXT and FEXT components as defined in the equation:

$$n(p, l, f) = w(p, f)[x_A(f) + x_F(l, f)c(l, f)], \quad (10)$$

where all terms represent single-sided power spectral densities, and it has been assumed that ACXT and FEXT signals are independent so that their powers may be added directly to produce the total noise power.

(i) $w(p, f)$ is the power spectral density of the T1 signal at the disturber source (repeater output) where p is the ones density. It is given by:

$$w(p, f) = \frac{2A^2 T}{R} \frac{p(1-p)}{2} \cdot \left[\frac{\sin \pi f T / 2}{\pi f T / 2} \right]^2 \frac{1 - \cos(2\pi f T)}{1 + 2(2p - 1)\cos(2\pi f T) + (2p - 1)^2}. \quad (11)$$

This represents a bipolar pulse stream with ones density p , rectangular pulses with 50-percent duty cycle, pulse period T , pulse amplitude A , through a reference impedance R . For T1 systems, $1/T = 1.544 \times 10^6$ Hertz, $A = 3$ volts, and $R = 100$ ohms.

(ii) $x_A(f)$ is the crosstalk power transfer function from disturber source to test-repeater input, where the disturber source is located in the same apparatus case as the test repeater.

(iii) $x_F(l, f)c(l, f)$ is the crosstalk power transfer function from disturber source to test repeater input where the disturber source is located at the opposite end of the repeater section from the test repeater and couples through FEXT paths in the main cable. A factor $c(l, f)$, representing the cable (power) transfer function, i.e. the inverse of the cable loss, has been split off to separate crosstalk effects from cable propagation effects. The factor $x_F(l, f)$ represents the equal-level FEXT coupling (power) transfer function for a cable section of length l . The dependence on l is different for different cable types and splicing arrangements. Pulp cable and random splicing result in a tendency for power addition of FEXT noise currents generated along the length of the cable, while PIC cable and color-for-color splicing result in a tendency for voltage addition.

The expression for the noise power density, caused by multiple disturbers, is a sum of terms $n(p, l, f)$ for each disturber, assuming independence of disturber signal sources. If all disturbers have a ones density p , then eq. (10) may be used to describe the total noise at the repeater input, provided that we interpret $x_A(f)$ and $x_F(l, f)$ as *power sums* of the individual pair-to-pair coupling coefficients.

We can simplify the expression (10) by assuming a common dependence on frequency f of the coupling power sums:

$$x_A(f) = a_A(f/f_0)^2 \quad x_F(l, f) = a_F(l)(f/f_0)^2, \quad (12)$$

where f_0 is the one-half baud for T1, 772 kHz. This gives:

$$n(p, l, f) = w(p, f)(f/f_0)^2[a_A + a_F(l)c(l, f)]. \quad (13)$$

Finally, we can parameterize the relationship between ACXT and FEXT by defining:

$$\alpha = \frac{a_A}{a_A + a_F(l)c(l, f_0)}, \quad (14)$$

where α represents the fraction of the noise density at $f = f_0$ that is caused by apparatus-case crosstalk. (Note, this is *not* the same as the fraction of total integrated noise power due to ACXT.) The noise power density at the repeater input may then be expressed:

$$n(p, \alpha, l, f) = w(p, f)(f/f_0)^2 q(l) \left[\alpha + (1 - \alpha) \frac{c(l, f)}{c(l, f_0)} \right], \quad (15)$$

where

$$q(l) = a_A + a_F(l)c(l, f_0). \quad (16)$$

Since, at $f = f_0$, eq. (15) reduces to

$$n(p, \alpha, l, f_0) = w(p, f_0)q(l), \quad (17)$$

then $q(l)$ is the power sum, from all sources, of the crosstalk coupling coefficients at $f = f_0$ to the repeater input of the line being tested.

Equation (15) is a function of the frequency f , and of the parameters p , α , and l . Direct measurements were made of the *relative* value of $n(p, \alpha, l, f)$ for 3 to 5 frequencies for each T1 line. Since the cable loss $1/c(l, f_0)$ was also measured, the parameters p and α could be determined by fitting the equation (15) to the data, where the frequency dependence of the dB cable loss was assumed to be $f^{0.58}$, as was measured for the pulp cables surveyed in these field measurements.

The interpretation of the value obtained for p in the fit is slightly ambiguous, because it is known that all disturbers did not have the same ones density. However, the parameter p , redefined as some sort of *composite ones density* for all the disturbers, was quite useful in characterizing the noise, and behaved in a reasonable manner as discussed in the main body of the text.

The total noise power appearing at the decision point of the repeater (i.e. after equalization) is given by:

$$n_{\text{tot}}(p, \alpha, l) = \int_0^{\infty} n(p, \alpha, l, f)r(f)df, \quad (18)$$

where $r(f)$ is the (power) gain of the repeater input equalization. Note, the repeater input gain normally depends also on the cable loss $1/c(l, f_0)$ through the action of the ALBO in restoring a T1 signal to fixed equalized pulse height. However, for noise measurements, the ALBO was disabled and the repeater gain fixed to equalize a fixed cable loss (31.7 dB) at 772 kHz. For an ideal repeater with 100 percent cosine roll-off, and compensation for 50-percent duty cycle pulses, the gain is given by:

$$[r(l, f)]^{1/2} = \frac{1}{2} \left[1 - \sin\left(\frac{\pi f - f_0}{f_0}\right) \right] \left(\frac{\pi f / 4f_0}{\sin \pi f / 4f_0} \right) \frac{1}{[c(l, f)]^{1/2}} \quad (19)$$

for $|f| \leq 2f_0$ and equals zero otherwise.

Substituting from eq. (15) into eq. (18), the total noise power at the repeater decision point is:

$$n_{\text{tot}}(p, \alpha, l) = q(l) \int_0^{\infty} w(p, f)(f/f_0)^2 \cdot \left[\alpha + (1 - \alpha) \frac{c(l, f)}{c(l, f_0)} \right] r(f)df \quad (20)$$

or

$$n_{\text{tot}}(p, \alpha, l) = q(l)i(p, \alpha, l), \quad (21)$$

where $i(p, \alpha, l)$ is defined as

$$i(p, \alpha, l) = \int_0^{\infty} w(p, f)(f/f_0)^2 \left[\alpha + (1 - \alpha) \frac{c(l, f)}{c(l, f_0)} \right] r(l, f) df. \quad (22)$$

Equation (20) is used to define the terms I and Q appearing in the margin eq. (2) in the main body of the text, where the capital letters refer to dB quantities. In the field measurements, I has been determined by numerical integration for each measured line given the fitted values for p and α , and given the measured frequency dependence of $r(f)$ for the test repeater. The individual contributions of ACXT and FEXT to the power sum Q were then determined by:

$$q_{\text{ACXT}} = a_A = \alpha q(l) = \alpha \frac{n_{\text{tot}}}{i}, \quad (23)$$

where n_{tot} is the measured noise power sum. Also,

$$q_{\text{FEXT}} = a_F(l)c(l, f_0) = (1 - \alpha)q(l) = (1 - \alpha) \frac{n_{\text{tot}}}{i}. \quad (24)$$

The distributions of a_A and $a_F(l)$ obtained by the use of eqs. (23) and (24) with the field measurement data are plotted in Figs. 14 and 15 and discussed in Section 4.4. In the plot of Fig. 15, l is approximately 6000 feet so that $a_F(l)$ represents the equal-level FEXT coupling for a full-length repeater section.

REFERENCES

1. K. E. Fultz and D. B. Penick, "The T1 Carrier System," B.S.T.J. 44, No. 7 (September 1965), pp. 1405-51.
2. H. Cravis and T. V. Crater, "Engineering of T1 Carrier System Repeated Lines," B.S.T.J. 42, No. 1 (March 1963), pp. 431-86.
3. G. F. Erbrecht et al., B.S.T.J., this issue.



Automated Measurement System for T1 Characterization

By G. F. ERBRECHT, J. H. GENTRY, JR., T. C. KAUP, and
R. J. SCHWEIZER

(Manuscript received January 10, 1980)

A computer operated measurement system was developed for the T1 characterization program. Using this automated system, we measured the T1 repeater section crosstalk margin, insertion loss, power sum, line resistance, simplex current, longitudinal currents, crosstalk spectrum, noise voltage distributions, and repeater margin, and made calibration checks. Over 100,000 field measurements have proven this test system to be reliable, versatile, and accurate.

I. INTRODUCTION

As part of the T-Carrier characterization program, T1 repeater section performance measurements were performed in 1977 and 1978. In this paper, we describe the experimental design and equipment implementation for these measurements. Other aspects of this program are described in companion papers.¹ The T1 repeatered line consists of the regenerative sections adjacent to central offices (end sections), and the intermediate sections. Intermediate section measurements performed at manholes are discussed in Section 2 and end-section measurements in the central office are described in Section 3.

The following measurements were performed for each intermediate section:

- (i) Line resistance and simplex power currents
- (ii) Longitudinal ac current
- (iii) Pair insertion loss at various frequencies
- (iv) Crosstalk coupling losses within the apparatus case slot
- (v) Crosstalk power sum
- (vi) Line margin for a test repeater
- (vii) Bit-rate and half-bit-rate output power of transmitting repeater

- (viii) Crosstalk spectrum at repeater input
- (ix) Automatic line buildout current of test repeater
- (x) Voltage amplitude distribution of crosstalk noise
- (xi) Service repeater margin and eye opening
- (xii) Test repeater calibration checks

End sections were characterized by measuring crosstalk power sum, crosstalk margin, and voltage amplitude distribution of crosstalk using separate test equipment designed for single-ended measurements in the central office.

II. INTERMEDIATE SECTION MEASUREMENTS

Analysis of plant records indicates extensive use of 900- and 1100-pair pulp cables for T1 in metropolitan areas.² For one-cable operation, previous engineering rules for these cables allowed 200 and 250 T1 systems, respectively. With these large T1 cross sections, an automated measuring system was deemed essential for fast and reliable measurements and recording of data.

The measurement configuration is outlined in Section 2.1; more detailed descriptions of individual measurements are given in Section 2.2. Use of the minicomputer to control the system is described in Section 2.3 and measurement accuracy is discussed in Section 2.4.

2.1 Automated measuring system

The measuring equipment was housed in two mobile vans stationed at the ends of the repeater section under test. The remote van (Fig. 1) was positioned at the transmitting end and housed signal sources. The control van (Fig. 2) was stationed at the receiving end and contained the bulk of the measuring equipment, the minicomputer, and data-recording equipment.

Access to the pairs to be measured was provided in the apparatus case slots by probes (Fig. 3) permitting connection of side 1 and side 2 inputs and outputs to the control panels (Fig. 4) in the vans by 60-foot shielded patch cables. The specially designed control panels were the heart of the test arrangement. The manhole (service) repeater or a test repeater were inserted in the control panels at each end, which then allowed test equipment to be inserted into the T1 transmission path. Specially packaged, portable, automatic-protection switches were installed in the central offices at the ends of the measured span to allow removal of repeaters on one line at a time without manual patching by central office personnel.

Measurements of power sum and test repeater margin were performed with a special test repeater. The test repeater was a laboratory-modified automatic line build-out (ALBO) repeater with one side (regenerator) modified to provide the following features:

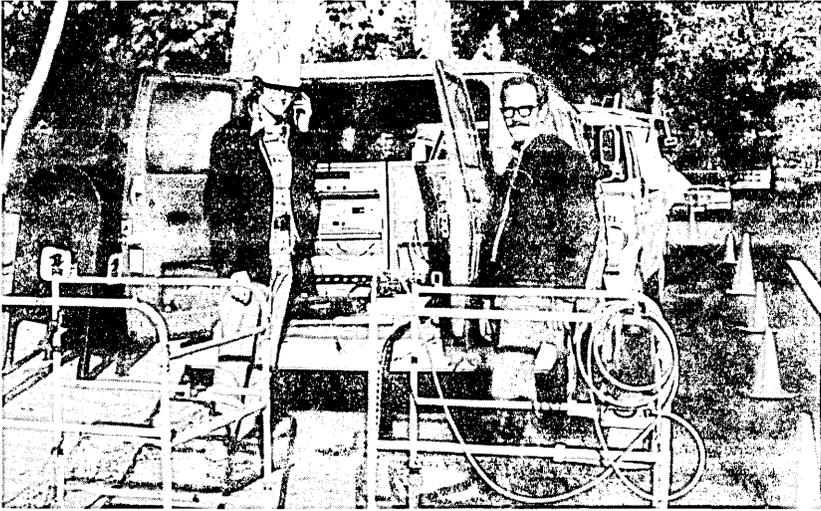


Fig. 1—Remote van containing signal sources.

(i) Access to the regenerator decision point for observation of the eye and for measurement of crosstalk power sum and signal level.

(ii) Option of regulated (normal) or fixed ALBO operation with monitoring of the ALBO current in both cases.

(iii) On/off control of the regenerator clocks.

Measurements were controlled by a minicomputer with 32K of memory (Section 2.3). It performed the following principal functions:

(i) Control of relays to set up the required measurement paths and instrument connections.

(ii) Setting of synthesizer frequencies.



Fig. 2—Control van containing measuring equipment.

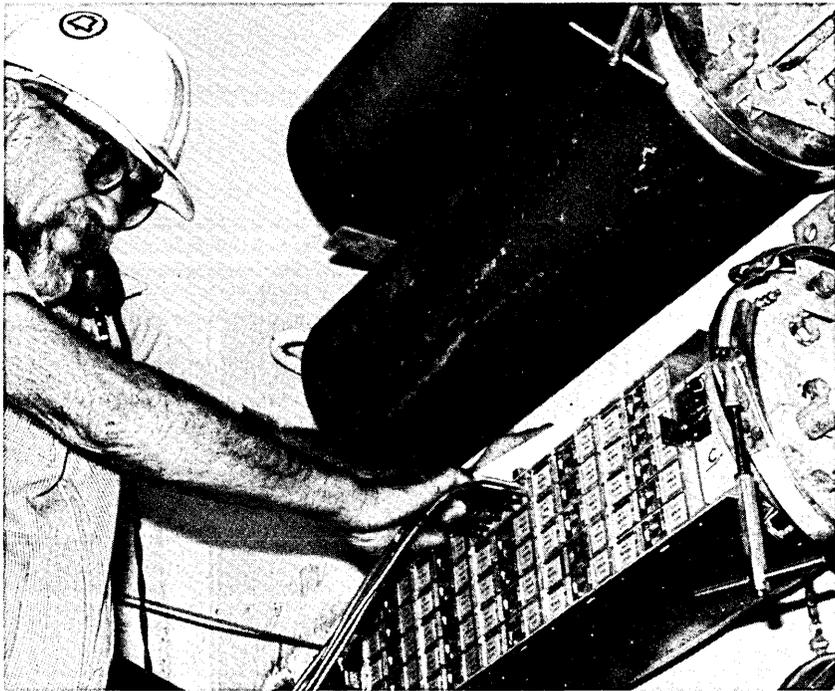


Fig. 3—Apparatus case with access equipment.

(iii) Search routines for attenuator settings to produce voltmeter readings or error counts within a specified range.

(iv) Processing of voltmeter and counter outputs for recording on magnetic tape and for production of hard copy listing by a printing data terminal.

(v) Prompting of the operator for identification of repeater section, case, and slot.

A digital voltmeter and a counter interfaced with the computer via standard interfaces. All other interfaces with the computer were specifically designed for this experiment. A telemetry system, operating over any available pair capable of carrying an audio signal, was employed to transmit control signals from the control van to the remote van. A simplified block diagram of the measurement system is shown in Fig. 5.

2.2 Measurements

This section describes the individual measurements in detail; the relevant portion of the measurement equipment for each of the measurements is shown in Figs. 6 through 13. In these figures, the individ-

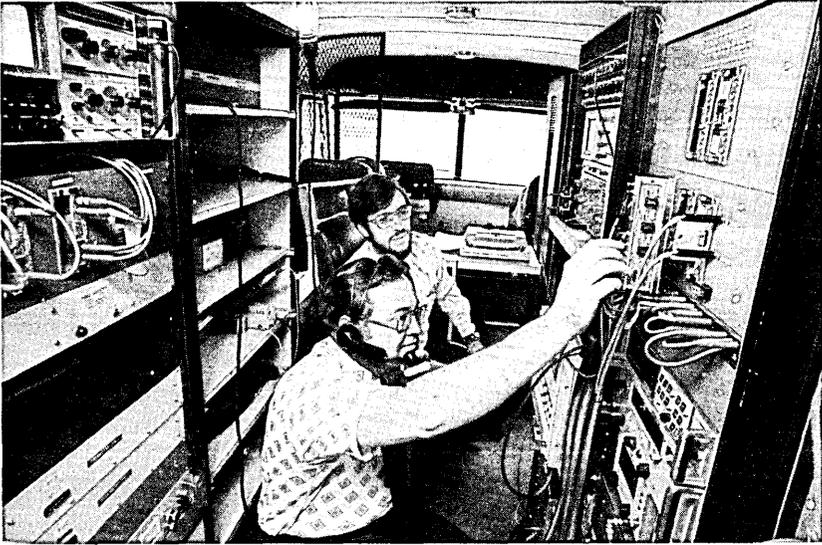


Fig. 4—Interior view of van showing control panels.

ual relays operated from the computer are identified by numbers in circles. The relay contacts are shown by the detached contact method, which indicates the nonenergized relay state by a single straight line through the transmission path for a closed contact and by a cross for an open contact. High isolation contacts are shown as double contacts.

2.2.1 Line resistance and simplex power current

The first measurement, cable resistance, was primarily intended to assure continuity of the transmission path between the control and remote vans once all signal and telemetry connections were made.

Line current was measured under normal operating conditions by measuring the voltage across a known resistor which was connected between the center taps of a 1:1 transformer, as shown in Fig. 6. A 5- $\frac{1}{2}$ digit DVM and a resistor of known value assured an accuracy of ± 0.5 percent for the line-current measurement.

2.2.2 Longitudinal ac current

Low-frequency longitudinal current, band-limited by a filter (20 Hz to 1 kHz), was across a 10-ohm series resistor while a current probe was used to measure the high-frequency currents (1 kHz to 10 MHz), as shown in Fig. 7.

2.2.3 Pair loss at various frequencies

Pair-loss measurements at selected frequencies were performed with the following instruments, as shown in Fig. 8:

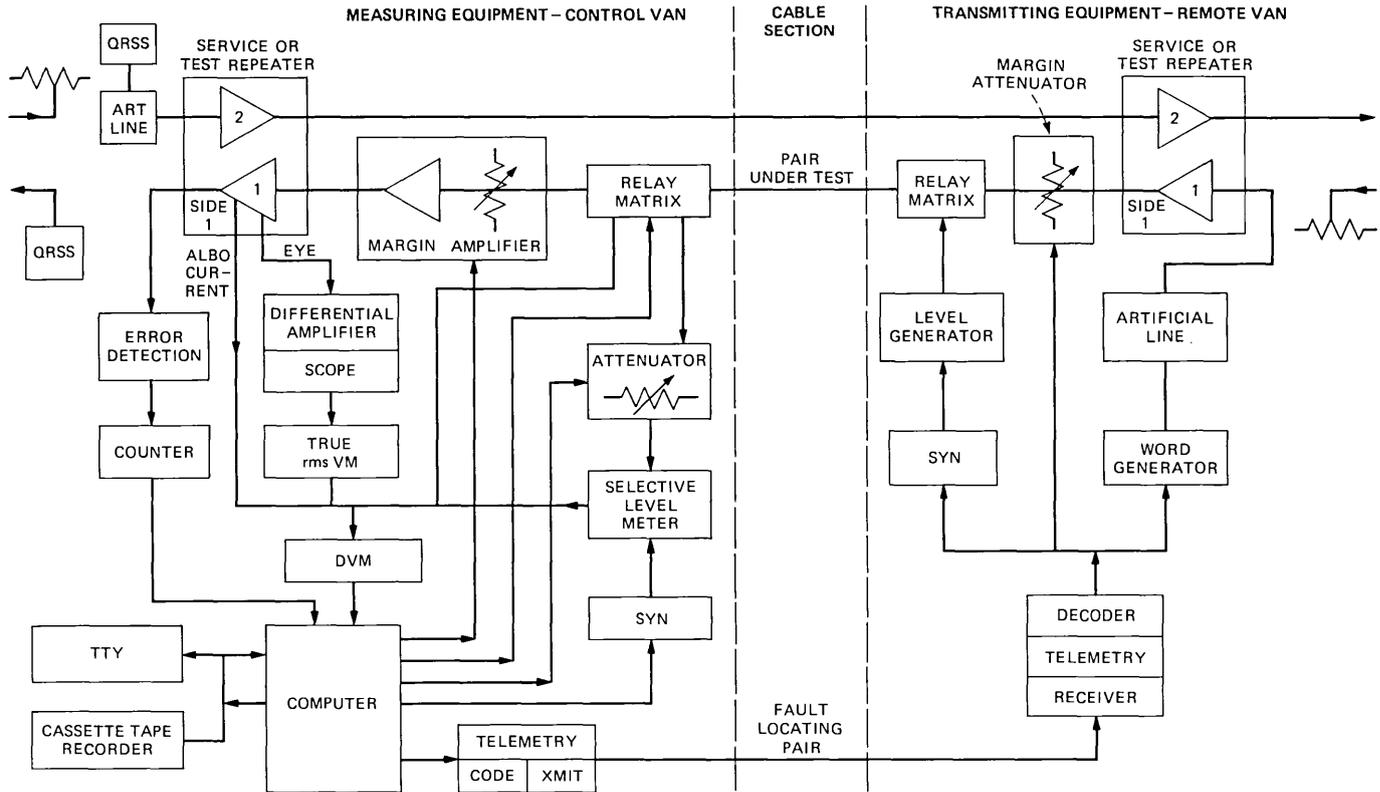


Fig. 5—Block diagram of measurement system.

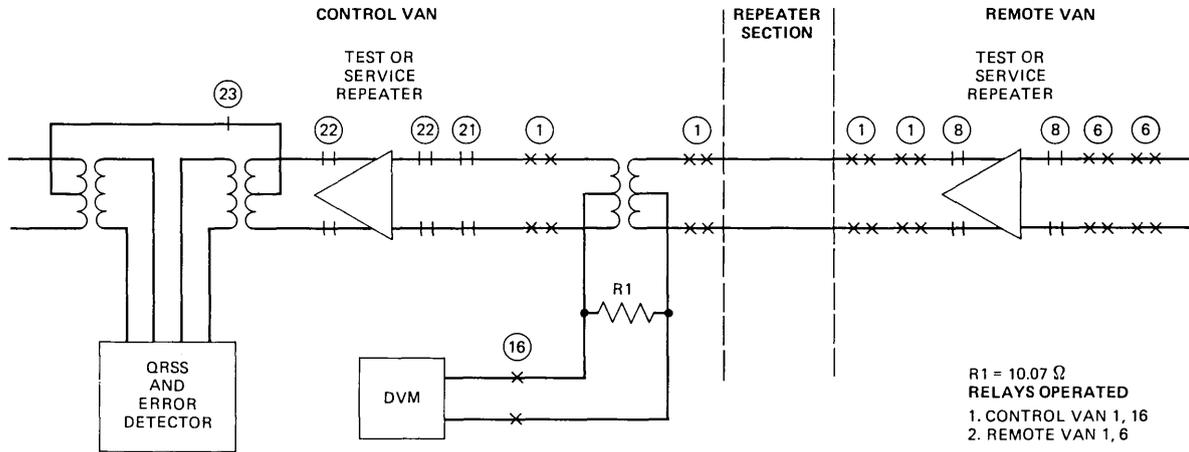


Fig. 6—Line current diagram.

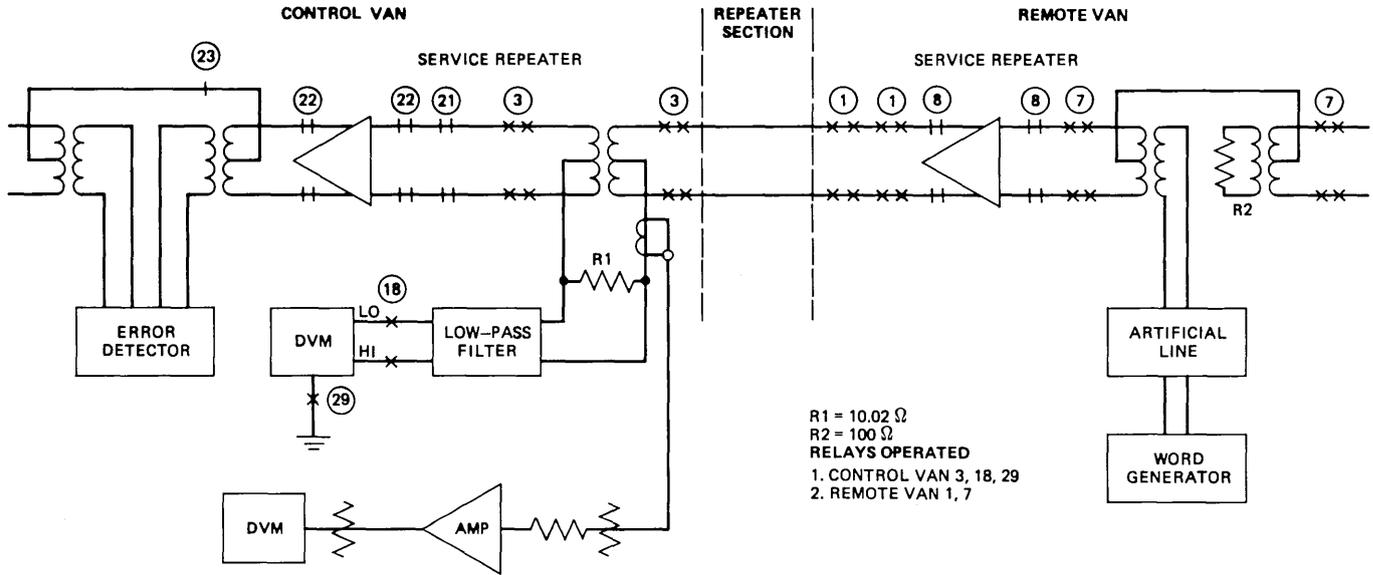


Fig. 7—System for low- and high-frequency longitudinal current measurement.

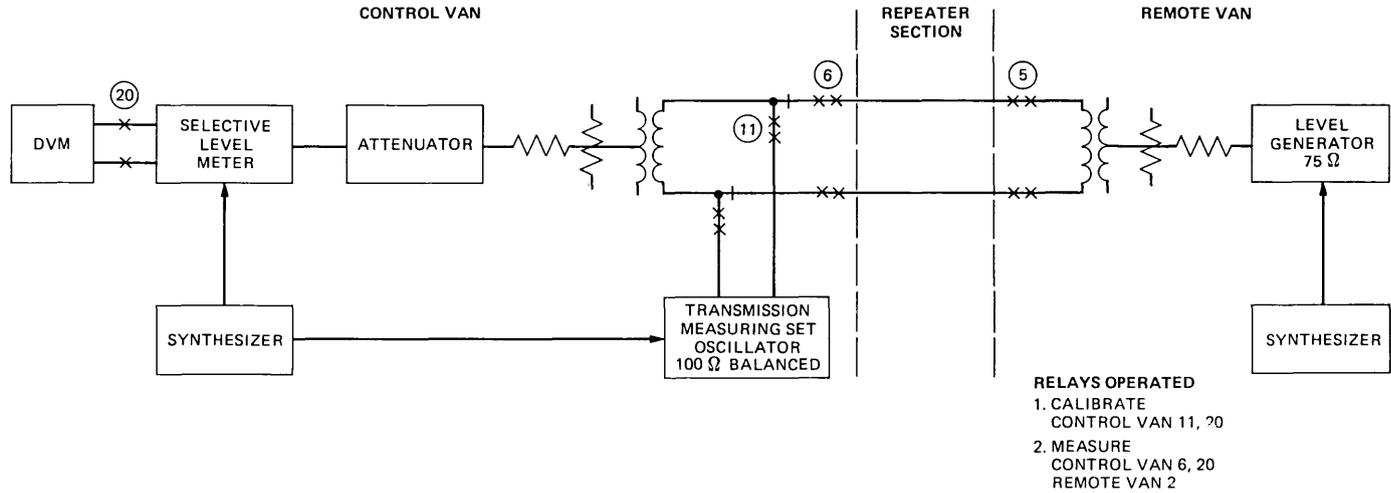


Fig. 8—Insertion-loss measurement system.

(i) A precision level generator was located at the remote end. Its frequency was controlled by a frequency synthesizer set by the computer from the control end via telemetry.

(ii) Located at the control end was a transmission measuring set (TMS) consisting of a stable level generator and a frequency selective level meter. The control frequency was also controlled by a synthesizer set by the computer. Auto-ranging detector operation was obtained by setting the detector to 80 dB of gain and applying signal to it via a programmable precision attenuator (0.1 dB resolution) controlled to produce a small range of dc output from the detector for measurements by the digital multimeter.

Measurement accuracy depended on the absolute level stability (better than 0.1 dB) of the level generators. The control-end detector (subject to drift) was automatically calibrated against the control-end generator (stable) before each measurement.

2.2.4 Crosstalk coupling losses within apparatus case slot

Figure 9 illustrates the measurement approach for within-slot crosstalk measurements. Measurements were performed with the control van's transmission-measuring set (TMS) described in Section 2.2.3. Six selective measurements (40-Hz detector bandwidth) at 772 kHz were made as follows:

(i) Side-1 input noise power with no signal on side-1 or side-2 outputs ("background" noise).

(ii) Side-1 input with side-1 output excited.

(iii) Side-1 input with side-2 output excited.

(iv) Side-2 input noise power with no signal on side-1 or 2 outputs (background noise).

(v) Side-2 input with side-1 output excited.

(vi) Side-2 input with side-2 output excited.

2.2.5 Crosstalk power sum

Measurements of crosstalk power sum at the repeater decision point were performed with the test repeater (Section 2.1). Since there was no T1 signal for the automatic line build-out (ALBO) to regulate properly, the preamplifier gain (and hence equalization) of the test repeater was fixed by forcing the ALBO current to the value it would have if regulating (operating normally) for a particular half-bit-rate pair loss. A convenient choice was 31.7 dB, which was within the loss range of the sections measured.

An oscilloscope with a differential plug-in amplifier monitored the noise signal at the regenerator decision point (repeater eye) relative to circuit ground and provided an unbalanced signal from its vertical output, which was measured by a true rms voltmeter. A dc voltage

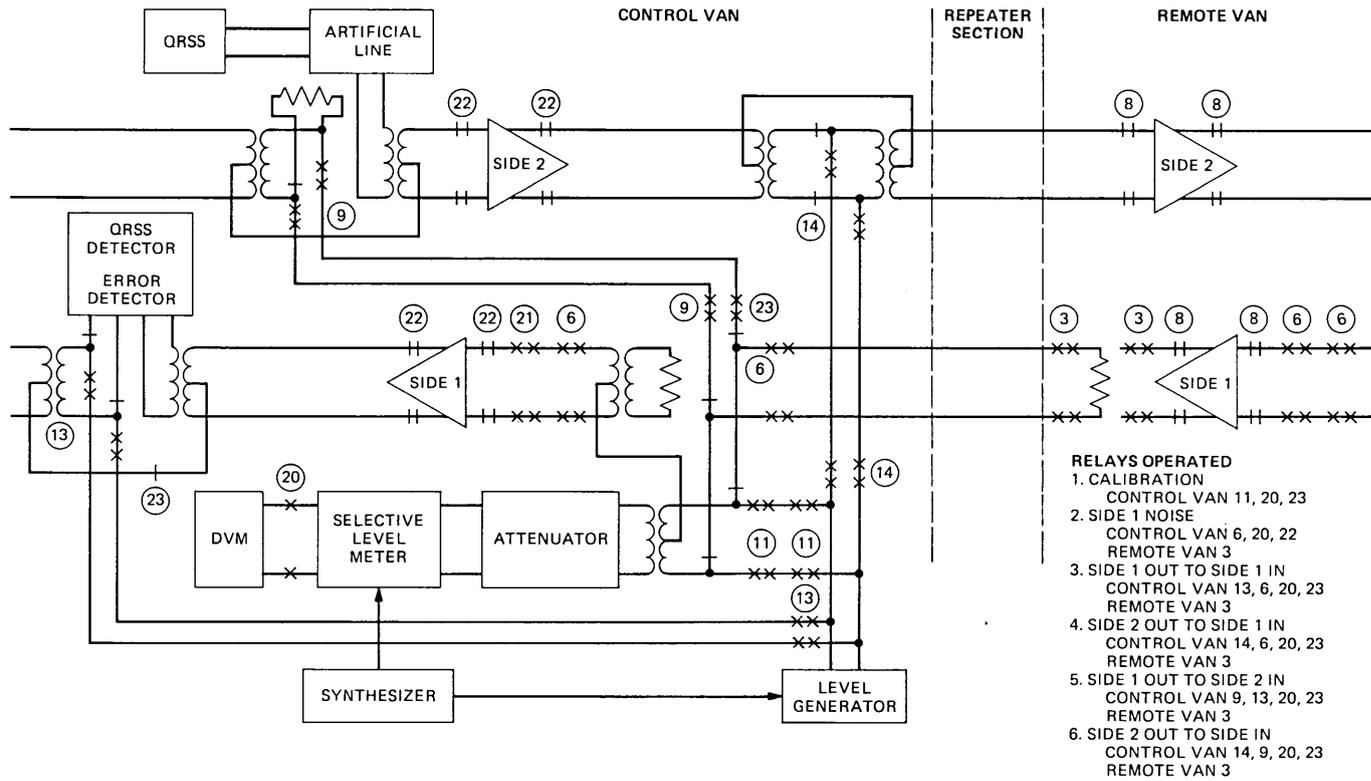


Fig. 9—Within-slot crosstalk measurement system.

proportional to the rms noise voltage was produced by the voltmeter and in turn measured by the digital dc voltmeter, which converted the dc voltage to a digital value suitable for the computer. An amplifier and variable attenuator under control of the computer and ahead of the test repeater brought the repeated-line noise voltage within the measurement range of the true rms voltmeter.

2.2.6 Line margin

Line margin, the principal measure of performance, is defined as the number of dB that the signal-to-noise ratio of a repeater section must be reduced to produce a 10^{-6} error rate. For this measurement program, an amplifier-attenuator technique was used (Fig. 10). A quasi-random signal from a T1 word generator was injected through an artificial line (31.7 dB at 772 kHz) into the repeater input at the remote van. The repeater output was then attenuated by a remotely controlled 75-ohm attenuator, which was matched to the line with a repeating coil and matching pad. The minimum loss through this attenuator-transformer arrangement was 3.2 dB. In the control van, an amplifier-attenuator combination with a total available gain of 41.3 dB was switched in ahead of the repeater under test. The unbalanced-to-balanced conversion for this arrangement was accomplished with 1:1 transformers. A small inductor of 440 μF in series and 0.01- μF capacitor to ground were added to the simplex power loop to prevent amplifier oscillation. The net margin-measuring range for this arrangement was from 3.2 to 38.1 dB with 0.1 dB precision. The computer kept the attenuation at the remote van equal to the amplification at the control end to maintain the correct signal level at the repeater input. The system searched automatically for an attenuator setting that produced an initial error rate close to 10^{-3} . Four additional error-rate measurements were then made with the attenuator reduced in 1-dB steps, and the attenuation for a 10^{-6} error rate was calculated by polynomial interpolation. The functional relationship between attenuation and error rate is a measure of the peak factor of the noise.

The error rate (actually the bipolar-violation rate) was determined at the repeater output by a bipolar-violation detector and a counter. Two separate quasi-random signal sources (QRSS) energized the output pins of the apparatus case connector for the slot being tested to maintain normal within-slot crosstalk.

To obtain a measure of the variation of repeaters in the field, margin was measured for each line with the repeater found in the slot (service repeater) and with the test repeater (ALBO regulating normally).

2.2.7 Bit-rate and half-bit-rate output power of transmitting repeaters

Since the transmitted signal level of the service repeater in the remote van affects the measured service repeater margin, a measure-

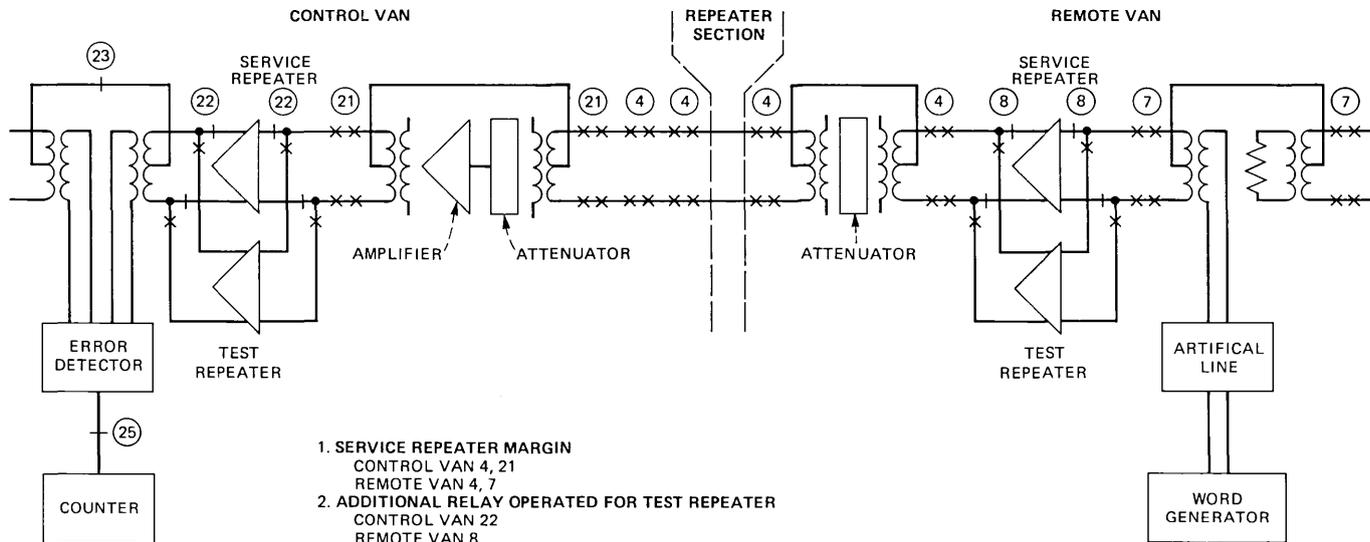


Fig. 10—System for service or test repeater margin measurement.

ment of the remote van's transmitting repeater output was made from the control van. This measurement was made with a 1500-Hz bandwidth at 772 kHz and 1.544 MHz when the remote service repeater was driven with an all ones signal produced by the remote van's word generator under telemetry control. The bit-rate power (1.544 MHz) is a measure of the balance between positive and negative pulses; the half-bit-rate power (772 kHz) is a measure of the output pulse amplitude of the transmitting repeater (after correction for line loss).

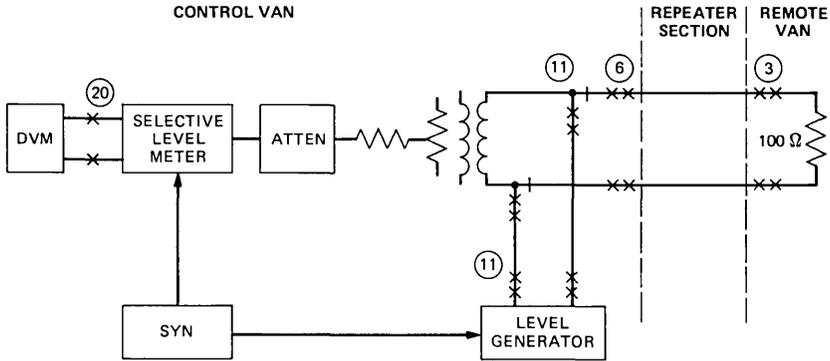
2.2.8 Crosstalk spectrum

The measurement of the crosstalk power spectrum was made at several frequencies with the level detector at the control van's repeater input while the remote end of the repeater section was terminated with a 100-ohm resistor (Fig. 11). The receiver was set for 100-dB gain and 1500-Hz bandwidth to achieve maximum sensitivity.

Measurements at three or more frequencies allowed calculation of the average ones density of the interfering signals and of the partitioning of the total crosstalk into cable far-end, cable near-end, and apparatus case near-end crosstalk (FEXT, NEXT, and ACXT).

2.2.9 ALBO current of test repeater

The test repeater's ALBO current was measured under two operating conditions; first, under the normal regulating condition to determine the ALBO setting as a function of cable loss for a normally regulating repeater, and second, under the fixed ALBO condition where the ALBO current was preset for a specific cable loss. This measurement consisted



- 1. CALIBRATION
CONTROL VAN 11, 20
- 2. MEASUREMENT
CONTROL VAN 6, 20
REMOTE VAN 3

Fig. 11—System for measuring the crosstalk spectrum.

of a voltage measurement across a resistor in series with the repeater's ALBO diodes (Fig. 12).

2.2.10 Voltage amplitude distribution of crosstalk

On selected pairs, especially those with low margin, the probability distribution of the crosstalk voltage amplitude at the decision point of the test repeater was measured to determine possible departure from the Gaussian distribution that is generally assumed for margin analysis (Fig. 13).

The rms noise voltage measured at the repeater decision point (eye) was adjusted by computer control of a variable attenuator to a level V_0 . A comparator, whose thresholds were also set by the computer, produced an output when the noise signal exceeded a selected value. The comparator output was clocked into a counter to determine the percentage of time for which the noise exceeded each level.

The threshold level was incremented from zero in integral multiples of the rms voltage until no counts were obtained in a 1-second interval (typically at $4 V_0$). Then the level was decremented in four steps of one quarter of the rms voltage (e.g., $3.75 V_0$, $3.50 V_0$, $3.25 V_0$, and $3.0 V_0$), and 4-second counts were made to give high resolution and accuracy in the important "tails" of the distribution.

2.2.11 Service repeater margin and eye opening

Both the *repeater*-margin (not to be confused with the *line* margin) and eye-opening tests were designed as independent measurements of service-repeater performance. For these tests, the repeaters are operated on a 31.7-dB artificial line. The eye-opening test employed a quasi-random signal in which the amplitude of a pulse in every 16th time slot could be varied from zero to full amplitude. This measurement was intended to directly measure the service-repeater eye degradation due to the intersymbol interference resulting from imperfect equalization. The eye of the service repeater was closed by reducing the amplitude of a pulse representing a "one" or increasing the amplitude of a small pulse substituted for a "zero" and measuring the error rate.

The service-repeater margin was measured by determining the amount of suitably filtered broadband Gaussian noise necessary at the repeater input in order to cause a given error rate with a quasi-random signal. In addition to the effect of the repeater-eye degradation mentioned above, this test also reflected variation in the noise filtering of the T1 repeater preamplifier.

The intent of these tests was to characterize the performance of the service repeater independently of the T1 plant in which it normally operates.

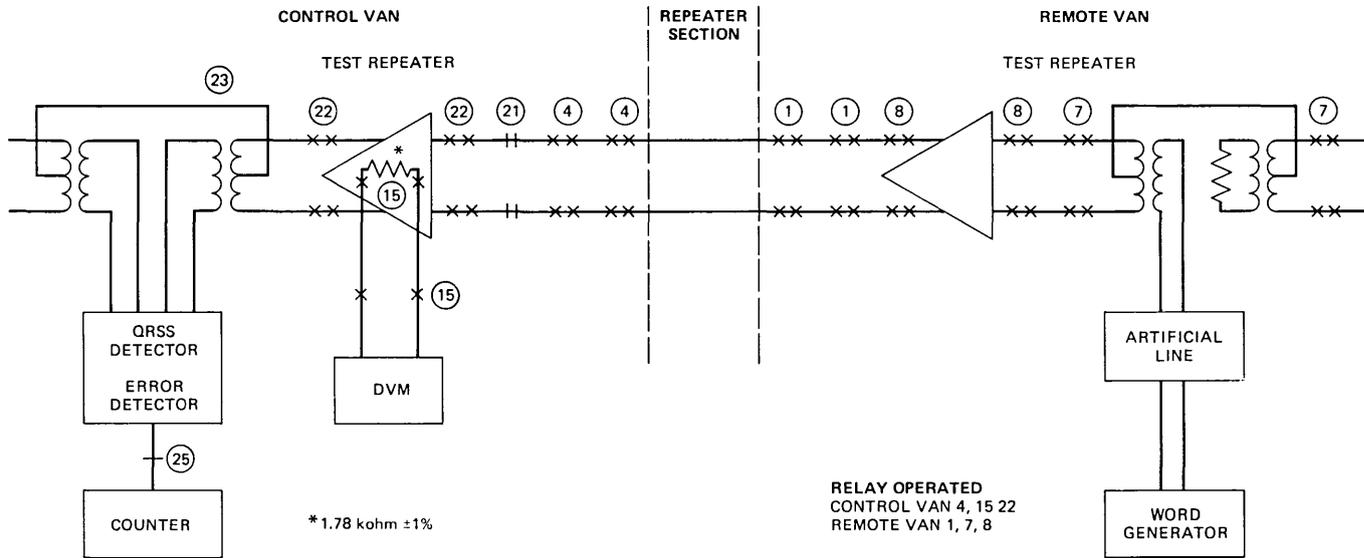


Fig. 12—System for measuring the ALBO current of the test repeater.

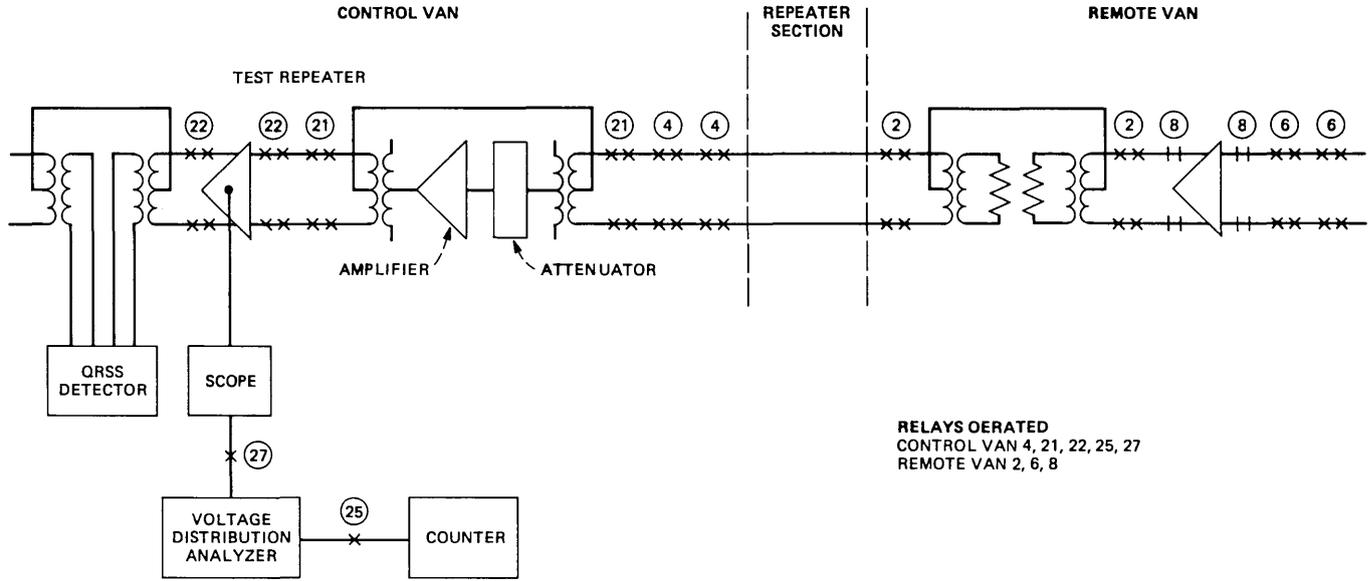


Fig. 13—System for measuring the noise-voltage distribution.

2.2.12 Test-repeater calibration checks

Several checks of the test-repeater performance were provided in this experimental arrangement. For example, the test-repeater-equalizer gain shape and stability were measured under fixed ALBO conditions by injecting a single frequency from the level oscillator and measuring the resulting output at the repeater eye with the rms voltmeter. To test for any drift in the test-repeater ALBO current, a quasi-random signal was directed through an artificial line to the regenerator input while the ALBO current was measured.

2.3 Computer measurement programs

A new operating system for the minicomputer was developed for field measurements to satisfy the following special requirements:

(i) Provide a core-based system for quick start-up and execution of measurement programs as well as decreased dependence on magnetic recording media (tape or disk) during long periods of operation in hostile environments.

(ii) Provide a programming language with a set of basic commands for the sequential control of relays and measurement instruments and with the capability for algebraic calculations and conditional branching.

(iii) Provide for the creation and saving of named measurement sequences (programs) that may thenceforth be called by name in the same manner as the basic commands.

(iv) Eliminate cumbersome compilation and loading procedures when creating, modifying, or executing measurement programs so that the maximum time may be spent on making actual measurements.

The latest version of this system occupied the entire 32,768 words (16 bits) of the minicomputer, where the operating system used about 20,000 words and the measurement programs filled up the rest of the core. Since core storage is nonvolatile, start-up of the system required only power on and a manual jump to a starting address. Only in those cases where previous versions of the operating system or measurement programs were desired (or for certain operator errors or hardware failures) was it necessary to load the computer memory from an external medium (tape or disc).

The structure of the command language designed is very simple. Every command consists of a command name followed by a list of parameters, where all items are separated by spaces. Commands may be preceded by statement numbers, which may be used in control commands (e.g., GOTO) to refer to the given command. Multiple commands may be entered on one line, separated by semicolons. A line entered at the terminal may or may not start with a line number. Command lines entered without a beginning line number are executed

as soon as entered. Command lines entered with a line number are stored in a program area as part of a current program being built up or modified. The current program may be saved under a new name, or may replace a previously saved program, and then becomes available for execution as a new command.

The types of basic commands available in the operating system were as follows:

(i) Instrument commands—e.g., “SYNN 772” sets the control van frequency synthesizer to 772 kHz and “DVM X” causes the digital voltmeter to make a reading and store the results in the variable X.

(ii) Program control commands—e.g., “RANGE X A B 1 2 3” causes program execution to jump to statement 1 if X is less than A, to statement 3 if X is greater than B, and to statement 2 if X is in the range A to B inclusive.

(iii) Algebraic calculations—e.g., “LET A = B + C / D” calculates the quantity $(B + C)/D$ and stores the result in A.

(iv) Trigonometric functions—e.g., “COS A X” calculates the cosine of A and stores the result in X.

(v) Input-output statements—e.g., “WR 2 (F6.2, F8.1) A B” writes the numbers A and B to the terminal in the format indicated.

(vi) System commands—e.g., “SAVE MEASCUR” saves the program residing in the current working area into its own memory area under the name MEASCUR.

Since new programs may be named, saved, and thenceforth called in a manner identical to the basic commands, a hierarchy of programs may be built up, where programs at a given level use previously constructed programs as building blocks. When a standard set of field measurements was being made, the last program was usually just a list of the measurements to be made, in the desired order. For example, a program named MEAS1 (which may be called by simply typing “MEAS1” on the terminal) may look like the following:

```
100 RES; CUR; IL; PSUM; MARGIN
```

and would perform the measurements of line resistance, simplex powering current, pair insertion loss, crosstalk power sum, and line margin in the indicated order.

2.4 Measurement accuracy

Tests were performed to evaluate the equipment accuracy. Measurements of total within-slot crosstalk coupling losses exhibited a Gaussian distribution to about 100 dB (2.8 sigma point) indicating no truncating within this region due to detector sensitivity limitation. Crosstalk coupling losses within the measurement equipment itself were greater than 114 dB for all crosstalk measurement paths. In

addition, short-term measurement stability tests under field conditions revealed only small variations for other measurements. Ten repeated measurements made on three different repeater sections resulted in average standard deviations of 0.03, 0.19, 0.18, and 0.04 dB for power sum, test-repeater margin, service-repeater margin, and 772-kHz insertion loss, respectively.

An upper bound on long-term measurement accuracy for insertion loss can be inferred from field results. A sigma of 0.16 dB has been determined for the measurements of low-frequency insertion loss for each of seven complete cable cross-sections. This total variation includes actual pair-loss variations so that the RMS variation alone was less than 0.16 dB.

Crosstalk power sum was measured twice within less than one minute under similar conditions on each of 503 T1 pairs in one measurement program. The distribution of the difference had a standard deviation of less than 0.1 dB.

III. END-SECTION MEASUREMENTS

The measurements performed on end sections (those repeater sections adjacent to a central office) were carried out in a single-ended mode as opposed to the double-ended master-slave van configuration employed in the T1 intermediate-section measurements. That is, equipment was employed only at the receiving end of the repeater section (the central office) rather than at both ends.

3.1 Test configuration

The measurement procedure called for removal of the office service repeater from its slot, attaching an adapter to its edge connector, and reinserting it into the slot. The adapter allowed the incoming signal pair to be connected to the external measuring equipment while the line-powering current was maintained by the original office repeater reinserted into its slot. The incoming signal was passed through a variable (flat) gain broadband amplifier made up of an attenuator and a video amplifier connected in tandem. This combination provided a range of a available gain of -19.1 dB to $+41.9$ dB, adjusted nominally to 0.0 dB.

Figure 14 represents the test equipment configuration, which employed an office repeater as the test repeater. Modifications were made to permit external monitoring of the test repeater's ALBO current (I_{ALBO}), decision threshold voltage (V_{th}), and the (broadband) rms noise voltage (V_{rms}) present at the eye of the regenerator. In addition, ALBO current and threshold voltage were allowed to operate at their normal levels or fixed to externally provided reference levels.

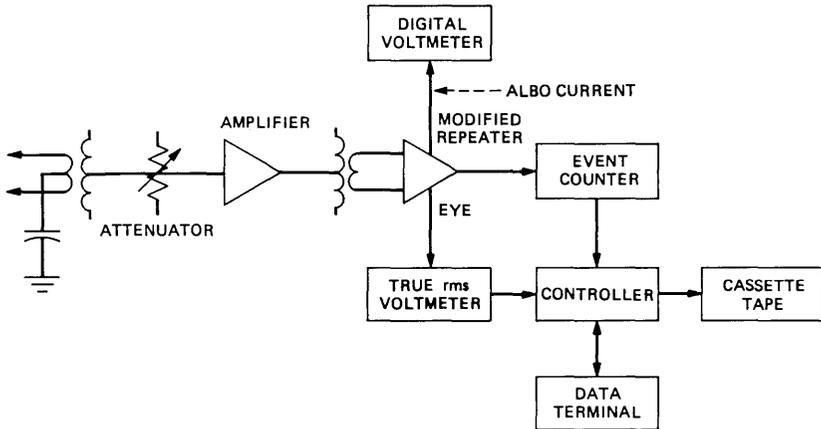


Fig. 14—End-section test configuration.

The regenerator output pulse stream was routed (with bipolar to unipolar conversion) to a digital counter, and the regenerator V_{rms} to an rms voltmeter. Outputs of both these instruments were interfaced to a microprocessor-driven measurement controller that performed operator prompting and interaction, measurement-algorithm control, and data-logging functions.

3.2 End-section measurements

Line margin at the 10^{-6} error rate was determined for end sections by amplifying crosstalk noise without signal on the line and counting output pulses. This was done by disabling the internal clock-extraction circuitry of the test repeater and inserting a local 1.544-MHz clock signal, thus enabling continuous sampling of the noise at the repeater's decision point, independent of input-signal timing energy. The test repeater was set up by adjusting and maintaining its threshold at one-half of the peak signal, ($V_{P/2}$), and its ALBO current at a level corresponding to the section length being measured. Then, with the test repeater operating essentially as a strobed comparator, flat gain was added at its front end until noise peaks present on each incoming line exceeded the threshold voltage ($V_{P/2}$), and began producing "ones" at the repeater's output. The margin was then taken as the amount of added gain required to produce pulses at a 10^{-6} rate. Note that this margin was in effect measured with a perfect eye; that is, the noise occupied fully 50 percent of the peak signal before producing an error. Thus, margins under operating conditions will be on the order of 4 dB lower to account for intersymbol interference under normal operating conditions.

In addition to margin, two parameters were measured to characterize the amplitude distribution of noise on end section lines. While the test repeater was producing pulses at a 10^{-6} rate, the rms value of the noise present at the repeater eye was measured and recorded. This value was used to determine the "noise peak factor," defined as the ratio "peak" to rms noise voltage in dB. In this instance, the peak voltage was taken to be the regenerator threshold voltage (V_{th}) exceeded 0.0001 percent of the time, corresponding to a 10^{-6} error rate, and was maintained in the test repeater at $V_{P/2}$.

The second noise-distribution indicator was provided by degrading the signal-to-noise ratio in steps of 1.0 dB and recording the repeater's error rate at each step. The set of points obtained may then be compared to the theoretical probability of error for Gaussian noise.

IV. CONCLUSION

The measurement equipment for the T1 characterization program has proven reliable, accurate, and versatile under actual field conditions. In fact, over 100,000 data items or a total of about 3,000 repeater section lines were measured in 1977 and 1978 with no more than 10 hours of down time due to computer and transmission measuring set failures. High accuracy was maintained throughout the experiment by the elimination of human recording errors and the high inherent accuracy of the automated test procedures. Program modifications were readily accomplished throughout the experiment, permitting modification of existing programs for increased operational efficiency and the inclusion of new test programs.

REFERENCES

1. T. C. Kaup et al., "T1 Carrier Characterization—Field Measurement Results," B.S.T.J., this issue.
2. J. R. Davis, L. E. Forman, and L. T. Nguyen, "The Metropolitan Digital Trunk Plant," B.S.T.J., this issue.

CONTRIBUTORS TO THIS ISSUE

Anthony S. Acampora, B.S.E.E., 1968, M.S.E.E., 1970, Ph.D., 1973, Polytechnic Institute of Brooklyn; Bell Laboratories, 1968—. At Bell Laboratories, Mr. Acampora initially worked in the fields of high-power microwave transmitters, radar systems studies, and signal processing. Since 1974, he has been studying high-capacity digital satellite systems. His current research interests are modulation and coding theory, time division multiple access methods, and efficient frequency re-use techniques. Member, Eta Kappa Nu, Sigma Xi, IEEE.

John G. Ackenhusen, B.S. (Physics), B.S.E. (Nuclear Engineering), M.S. (Physics), M.S.E. (Nuclear Engineering), 1976; Ph.D. (Nuclear Engineering), 1977, University of Michigan; Bell Laboratories, 1978—. Mr. Ackenhusen is developing practical implementations of digital speech processing techniques using microprocessors and custom signal processing architectures to perform speech recognition, synthesis, and analysis. He is supervisor of the Speech Processing Group in the Integrated Circuits Subsystems Department. Member, IEEE, American Physical Society, Optical Society of America, Sigma Xi.

Darryl L. Brownlow, B.S. (Physics), 1971, Tougaloo College; M.S. (Physics), 1974, University of Maryland; Bell Laboratories, 1974—. Mr. Brownlow joined Bell Labs to work on fracture problems associated with synthetic α -quartz seed plates. He later worked on the electronic properties of rapidly quenched metallic glasses, mechanical, microstructural, and magnetic properties of Fe-Co-based alloys, and devitrification tendencies of fused silica crucibles used in the Czochralski growth of silicon crystals. Presently Mr. Brownlow is a member of the Crystal Growth and Glass Research Development Department where his projects have included laser and furnace drawing of optical fibers, characterization of substrate tubes used in MCVD technology, and research and development in the MCVD method of fabricating optical fibers. Member, American Association of Crystal Growth—New Jersey Section.

Paul E. Butzien, B.S. (Electrical Engineering), 1961, Newark College of Engineering; M.E.E., 1963, New York University; Pacific Telephone and Telegraph Co., Portland, Oregon, 1948–1953; Bell Laboratories, 1953—. Mr. Butzien has been involved with research in electron tubes, superconductivity, cryogenic and high-vacuum devices, solid state microwave amplifiers, and phased-array antennas. His

current interests are in exploratory antennas and characterization and improvements to existing Bell System antennas.

Xian Mo Chang, B.S., 1967, Nankai University, Tienjin, China; Sunyang 21st High School, mathematics teacher, 1969–1977; Huangu Normal School, mathematics teacher, 1977–1978. A graduate student in mathematics at Nankai University since 1978, Mr. Chang's interests include graph theory, combinatorial algebra, and combinatorial optimization. Mr. Chang worked with F. K. Hwang of Bell Laboratories in 1980, while Mr. Hwang taught at the Institute of Applied Mathematics, Beijing, China.

John R. Davis, B.A., 1964, Allegheny College; B.S.E.E., 1964, M.S.E.E., 1965, Columbia University; Bell Laboratories, 1965—. Mr. Davis was initially engaged in the circuit development of the repeater and terminal electronics for the T2 digital line. Recently he has worked on the performance characterization of paired-cable digital systems. He is presently supervisor of the Digital Wire Line Performance Studies Group. Member, IEEE, Tau Beta Pi, Eta Kappa Nu.

Frank V. DiMarcello, B.S. (Geochemistry), 1960, Pennsylvania State University; M.S. (Ceramics), 1966, Rutgers University; Bell Laboratories, 1960—. Mr. DiMarcello has worked on various glass and ceramic materials problems including substrates for thin-film circuitry, microwave windows for hardened antennae, and glass and ceramic-to-metal seals. He is currently a member of the technical staff involved in materials and processing aspects of optical waveguides.

Ding Z. Du, 1967, Middle School, Qiqihaer, China. Mr. Du pursued independent studies of mathematics, and became a graduate student at the Institute of Applied Mathematics, Academia Sinica, Beijing, China, in 1978. He is currently interested in mathematical programming and combinatorics. Mr. Du worked with F. K. Hwang of Bell Laboratories in 1980, while Mr. Hwang taught at the Institute of Applied Mathematics, Beijing, China.

Gunter F. Erbrecht, B.S., 1960, Columbia University; M.S.E.E., 1962, New York University; Bell Laboratories, 1960—. Mr. Erbrecht was initially concerned with circuit development for a high-speed digital repeater. He was subsequently involved in the development of the regenerator for the T2 digital transmission system. From 1976 to

1980 he participated in the T-carrier characterization program. He is now involved with digital multiplex development for lightwave systems. Member, IEEE, Tau Beta Pi, Eta Kappa Nu.

Leonard E. Forman, B.E.E., 1959, Polytechnic Institute of Brooklyn; M.E.E., 1961, New York University; Bell Laboratories, 1959—. Upon joining Bell Laboratories, Mr. Forman designed hybrids for the millimeter waveguide transmission system, varactor frequency multipliers for *TELSTARR*, and tunnel diode regenerators for high-speed digital repeaters. He then worked on the timing circuits for the T2 digital repeater, and subsequently on characterizing the T1 digital network. He is presently concerned with system applications for the FT3 lightwave system. Member, Tau Beta Pi, Eta Kappa Nu; Associate Member, Sigma Xi.

William B. Gardner, B.S. (Physics), 1961, University of Alabama; Ph.D. (Physics), 1968, Johns Hopkins University; Bell Laboratories, 1968—. Since 1972, Mr. Gardner has worked on optical-fiber properties and measurement techniques. Member, Optical Society of America, IEEE.

James H. Gentry, Jr., A.S.E.E., 1961, Temple University Community College; Bell Laboratories, 1961—. Mr. Gentry has worked on system and maintenance aspects of video communications and digital transmission outside the metropolitan area. He is presently engaged in digital wire line performance studies.

Frank K. Hwang, B.A., 1960, National Taiwan University; M.B.A., 1964, City University of New York; M.E.S., 1966, Ph.D., 1968, North Carolina State University; Bell Laboratories, 1967—. Mr. Hwang has been engaged in research in discrete mathematics, computing algorithms, statistical theory, and switching networks. Member, SIAM.

Thomas C. Kaup, U.S. Naval Electronics Technician School, 1965–1966, Nassau Community College, 1974–1977, Polytechnic Institute of New York, 1977—. Prior to joining Bell Laboratories in 1977, Mr. Kaup held a number of positions in the electronics industry, including several years as a microprocessor systems design consultant. He is presently involved in the characterization of digital transmission facilities, and his current interests include the design of multimicroprocessor-based systems.

David G. Leeper, B.S.E.E., 1969, Washington University; M.E. (Electrical Engineering), 1970, Cornell University; Ph.D. (Electrical Engineering), 1977, University of Pennsylvania; Bell Laboratories, 1969—. Mr. Leeper has worked on waveguide, coaxial, and wire-pair digital transmission systems. He currently supervises a group concerned with characterization of end-to-end performance of the telephone network. Member, Tau Beta Pi, Eta Kappa Nu, IEEE.

Charles R. Lovelace, Electronic Technician Diploma, 1976, DeVry Institute of Technology; Bell Laboratories, 1978—. Mr. Lovelace is engaged in measurements of the transmission properties of optical fibers.

Suzanne R. Nagel, B.S. (Ceramic Engineering), 1968, Rutgers University; M.S., 1970, Ph.D., 1972 (Ceramic Engineering), University of Illinois; Bell Laboratories, 1972—. Ms. Nagel has worked in the area of materials and methods of glass-fiber-optical waveguides. She is presently supervisor of the Fiber Materials and Process Technology Group. Member, American Ceramic Society, Society of Women Engineers, National Institute of Ceramic Engineers, Tau Beta Pi, Keramos.

Liem T. Nguyen, B.S.E.E., 1975, M.E.E.E., 1976, University of Virginia; Bell Laboratories, 1977—. Upon joining Bell Laboratories, Mr. Nguyen was first engaged in studies of applications of digital wire line systems. He subsequently worked on the crosstalk model and performance analysis of digital subscriber loop systems. He is currently with the Microwave Radio Characterization Studies Group working on systems engineering aspects of *PICTUREPHONE*® Meeting Service. Member, Tau Beta Pi, Eta Kappa Nu.

Kinichiro Ogawa, B.S.E.E., 1966, M.S.E.E., 1968, University of Tokyo, D.Sc. (Electrical Engineering), Washington University, Nippon Telegraph & Telephone Public Corporation, 1968-1976; Bell Laboratories, 1976—. At N.T.T., Mr. Ogawa worked on long-haul analog coaxial cable systems and on video transmission. He also worked on the development of pulse-code modulated/frequency division multiplex digital systems. At Bell Laboratories, he has been involved in research on optical data links and digital lightwave systems. Member, IEEE, Optical Society of America, IECE of Japan.

Lawrence R. Rabiner, S.B. and S.M., 1964, Ph.D. (Electrical Engineering), The Massachusetts Institute of Technology; Bell Labo-

ratories, 1962—. From 1962 through 1964, Mr. Rabiner participated in the cooperative plan in electrical engineering at Bell Laboratories. He worked on digital circuitry, military communications problems, and problems in binaural hearing. Presently, he is engaged in research on speech communications and digital signal processing techniques. He is coauthor of *Theory and Application of Digital Signal Processing* (Prentice-Hall, 1975) and *Digital Processing of Speech Signals* (Prentice-Hall, 1978). Former President, IEEE, ASSP Society; former Associate Editor, ASSP Transactions; former member, Technical Committee on Speech Communication of the Acoustical Society. Member, ASSP Technical Committee of the Acoustical Society, ASSP Technical Committee on Speech Communication, IEEE Proceedings Editorial Board, Eta Kappa Nu, Sigma Xi, Tau Beta Pi. Fellow, Acoustical Society of America, IEEE.

Arthur K. Reilly, B.S.E.E., 1968, Virginia Polytechnic Institute and State University; National Aeronautics and Space Administration, 1964–1968; M.E. 1969, Cornell University; Bell Laboratories, 1968—. Mr. Reilly's early work involved studies to determine transmission characteristics for broadband analog channels. He has also worked on system engineering projects to characterize digital paired-cable systems and estimate the application of various transmission products. At present, he supervises a group conducting studies of microwave radio multipath fading and digital radio technology performance. Member, IEEE, Tau Beta Pi, Eta Kappa Nu, Phi Kappa Phi, Omicron Delta Kappa.

Manuel R. Santana, B.S. (Electrical Engineering), 1970, University of Hartford; M.S. (Electrical Engineering), 1971, Georgia Institute of Technology; Bell Laboratories, 1971—. Mr. Santana is supervisor of the Lightguide Cable Group and is engaged in optical-fiber cable design, analysis, and testing. Member, IEEE.

Paul E. Scheffler, S.B. (E.E.), S.B. (Physics), 1967, Massachusetts Institute of Technology; Ph.D. (Physics), 1972, California Institute of Technology; Bell Laboratories, 1972—. Mr. Scheffler is currently with the Digital Wire Line Performance Studies Group of the Transmission Facilities Engineering Laboratory. He is working on the modelling and analysis of the performance of digital wire lines and their interactions with terminals and multiplexers. Results of this work are used to predict performance of new digital services currently under development and to provide guidelines for the maintenance and administration of digital lines in the emerging nationwide digital network.

James A. Schiavone, B.S. (Engineering Physics), 1970, Lehigh University; M.S. (Atmospheric Physics), 1972, Drexel University; Bell Laboratories, 1972—. Mr. Schiavone has done work in laboratory spectroscopy of small molecules and in modeling the photochemistry of the lower troposphere. He is currently exploring the meteorological effects on line-of-sight microwave radio systems. Member, American Physical Society, American Meteorological Society, New York Academy of Sciences.

Morton I. Schwartz, B.E.E., 1956, City College of New York; M.E.E., 1959, Eng. Sc.D. 1964, New York University; ITT Laboratories, 1956-1961; Bell Laboratories, 1961—. At ITT, Mr. Schwartz's principal efforts were in the field of radar systems studies and design. At Bell Laboratories, he has been engaged in theoretical and experimental work in radar, sonar, and communications. Since 1972, he has worked on the development of optical-fiber communication media. Senior Member, IEEE.

Richard J. Schweizer, A.A.S., 1968, Mercer County Community College; B.S.E.E., 1975, Monmouth College; M.S.E.E., 1979, Polytechnic Institute of New York; Bell Laboratories, 1968—. Mr. Schweizer was initially involved with system studies for *PICTUREPHONE*® service. More recently he has been involved in performance studies and characterization of cable and central office environments for short-haul digital transmission on paired wire lines. Member, Sigma Pi Sigma, Eta Kappa Nu.

Edward A. Sigety, Eastern Michigan University, Ypsilanti, Michigan, 1944; Cornell University, Ithaca, 1946; Bell Laboratories, 1952—. At Bell Labs, Mr. Sigety was formerly active in research and development on both glasses and polycrystalline ceramic materials. He is currently involved in measuring attenuation and characterizing optical-fiber waveguides.

Arvids Vigants, B.E.E., 1956, City College of New York; M.S. (E.E.), 1957, Eng. Sc.D. (E.E.), 1962, Columbia University; Bell Laboratories, 1962—. Mr. Vigants has worked on various electromagnetic wave propagation topics and is currently working on problems in line-of-sight microwave propagation and microwave radio systems. Member, Eta Kappa Nu, Tau Beta Pi, Sigma Xi, URSI/USNC Commission F, IEEE.

Neil Weste, B.Sc., 1973, B.E. (Electrical Engineering), 1974, Ph.D., 1978, University of Adelaide, South Australia; Bell Laboratories, 1977—. Since joining Bell Laboratories, Mr. Weste has been involved in the areas of high-performance color graphics displays, integrated design automation tools, special processor architectures, and VLSI circuit design.

Jay G. Wilpon, B.S. (Mathematics), A.B. (Economics), cum laude, 1977, Lafayette College; Bell Laboratories, 1977—. At Bell Laboratories, Mr. Wilpon has been engaged in speech communications research and is presently concentrating on problems of speech recognition.

PAPERS BY BELL LABORATORIES AUTHORS

PHYSICAL SCIENCES

Absolute Sense and Model for the Piezoelectric and Pyroelectric Coefficients in $\text{Ba}(\text{NO}_2)_2 \cdot \text{H}_2\text{O}$ and $\text{Cs}_2\text{S}_2\text{O}_6$. R. Liminga, S. C. Abrahams, and J. L. Bernstein, *J Appl Crystallogr*, 13, No. 6 (December 1980), pp 516-26.

On the Anomalies in Density, Young's Modulus, and Glass Temperature of Pd-Si Glasses. H. S. Chen, J. T. Krause, K. Shirakawa, and T. Masumoto, *J Noncrystal Sol*, 41 (1980), pp 79-88.

Conductivity Profiling of GaAs/(AlGa)As Multilayer Structures. J. L. Zilko and P. J. Anthony, *J Electrochem Soc*, 128, No. 4 (April 1981), pp 871-4.

Directional Memory Effects in Inelastic Photoelectron Scattering. D. P. Woodruff et al., *Surf Sci*, 95 (1980), pp 403-10.

The Disappearance of High Temperature Superconductivity in $\text{Ba Pb}_{1-x}\text{Bi}_x\text{O}_3$ Perovskite. C. E. Methfessel, G. R. Stewart, and B. T. Matthias, *Proc Natl Acad Sci*, 77, No. 11 (1980), pp 6307-8.

Effect of Metalloid Elements on Strength and Thermal Stability of Iron Based Glasses. M. Naka, T. Masumoto, and H. S. Chen, *J Phys (Paris)*, C8, No. 41 (1980), pp 839-42.

Efficient Transverse Deflection of Neutral Atomic Beams Using Spontaneous Resonance—Radiation Pressure. J. E. Bjorkhelm, R. R. Freeman, and D. B. Pearson, *Phys Rev A*, 23 (February 1981), pp 491-7.

Electric Field Depinning of Charge-Density Waves in NbSe_3 . R. M. Fleming, *Phys Rev B*, 22, No. 12 (December 15, 1980), pp 5606-12.

EXAFS Studies of Glassy Metallic Alloys. H. S. Chen, B. K. Teo, and R. Wang, *J Phys (Paris)*, C8, No. 41 (1980), pp 254-62.

Neutron Activation Study of a Gettering Treatment for Czochralski Silicon Substrates. L. E. Katz, P. F. Schmidt, and C. W. Pearce, *J Electrochem Soc*, 128, No. 3 (March 1981), pp 620-4.

Neutron Activation Analysis Study of the Sources of Transition Group Metal Contamination in the Silicon Device Manufacturing Process. P. F. Schmidt and C. W. Pearce, *J Electrochem Soc*, 128, No. 3 (March 1981), pp 630-7.

Opto-Acoustic Study of Weak Optical Absorption of Liquid Methane. C. K. N. Patel, E. T. Nelson, and R. J. Kerl, *Nature*, 286 (July 1980), pp 368-70.

Order-Disorder Phase Transitions in 9-Hydroxyphenalene. S. C. Abrahams et al., *J. Chem Phys*, 74, No. 1 (January 1, 1981), pp 644-6.

Picosecond Photoconductivity in Radiation Damaged Silicon-on-Sapphire Films. P. R. Smith et al., *Appl Phys Lett*, 38, No. 1 (January 1, 1981), pp 47-50.

Pressure Induced Structural Transitions in Partially Ionic Semiconductors: A Self-Consistent Pseudopotential Approach to ZnSe. W. Andreoni and K. Maschke, *Phys Rev B*, 22 (November 15, 1980), pp 4816-24.

Reactive Ion Etching of GaAs and InP Using $\text{CCl}_2\text{F}_2/\text{Ar}/\text{O}_2$. E. L. Hu and R. E. Howard, *Appl Phys Lett*, 37 (December 1, 1980), pp 1022-4.

Relativistic Tight-Binding Calculation of Core-Valence Transitions in Pt and Au. L. F. Mattheiss and R. E. Dietz, *Phys Rev B*, 22 (August 15, 1980), pp 1663-76.

Relaxation of Optically Pumped Electron Spins Through a Virtual Photon: Experimental Evidence in Heavily Zn-Doped GaAs. D. A. Kleinman and R. C. Miller, *Phys Rev Lett*, 46, No. 1 (January 1981), pp 68-71.

Silver-Molecule Separation Dependence of Surface Enhanced Raman Scattering. C. A. Murray, D. L. Allara, and M. Rhinewine, *Phys Rev Lett*, 46, No. 1 (January 1981), pp 57-61.

Spin-Glass-Paramagnetic Phase Boundary in Amorphous Magnetic Alloys. K. Moorjani et al., *J Phys (Paris)*, C8, No. 41 (1980), pp 718-22.

Strain-Induced Softening of Certain Eutectics. H. Suhl, B. T. Matthias, and S. Hecker, *Phys Rev Lett*, 45 (1980), pp 1707-8.

Structural Phase Transitions: Defects and Dynamics. P. A. Fleury, *J Phys (Paris)*, C6, No. 7 (July 1980), pp C6-419-C6-24.

Superconductivity and Magnetic Ordering in the Pseudoternary $\text{HO}(\text{Ir}_x\text{Rh}_{1-x})_4\text{B}_4$. B. T. Matthias, H. C. Ku, and F. Acker, *Phys Lett*, 76A (1980), pp 399-402.

MATHEMATICS

Continuity of Generalized Semi-Markov Processes. W. Whitt, *Math Op Res*, 5 (November 1980), pp 494-501.

The Effect of Variability in the GI/G/s Queue. W. Whitt, *J Appl Probab*, 17 (December 1980), pp 1062-71.

Error-Correcting Codes and Cryptography. N. J. A. Sloane, in *The Mathematical Gardner*, edited by D. A. Klarner, Boston: Prindle, Weber, and Schmidt, 1981, pp 346-82.

Imbedding Smooth Involutions In Trivial Bundles. K. J. Prevot, *Pacific J. Math*, 89, No. 1 (July 1980), pp 163-8.

The Infinitesimal Group of the Navier-Stokes Equations. S. P. Lloyd, *Acta Mechanica*, 38, Nos. 1, 2 (1981), pp 85-98.

Statistical Computing: History and Trends. J. M. Chambers, *Am Statist*, 34 (November 1980), pp 238-43.

COMPUTING

Application of Computers to Maintenance of Telecommunication Networks. F. B. Six, *Seminar on Maintenance of Telecommunication Networks, October 6-15, 1980*, Geneva, Switzerland, Union Int. Telecommunications, 1981, pp 285-320.

Consistency of Redundant Data Bases in a Weakly Coupled Distributed Computer Conferencing System. B. I. Strom, *Proc Fifth Berkeley Workshop on Distributed Data Management and Computer Networks*, February 3, 1981, pp 143-53.

Constructing Call-by-Value Continuation Semantics. R. Sethi and A. Tang, *J ACM*, 27, No. 3 (July 1980), pp 580-97.

The Control of Store and Forward Congestion in Packet Switched Networks. E. F. Wunderlich, L. Kaufman, and B. Gopinath, *Proceedings of the Fifth International Conference on Computer Simulation*, Amsterdam: North-Holland, 1980, pp 851-6.

The Edge-Flag Algorithm—A Fill Method for Raster Scan Displays. B. D. Ackland and N. Weste, *IEEE Trans Comput*, C-30, No. 1 (January 1981), pp 41-7.

Efficient Computation of Expressions With Common Subexpression. B. Prabhala and R. Sethi, *J ACM*, 27, No. 1 (January 1980), pp 146-63.

Variations on the Common Subexpression Problem. P. J. Downey, R. Sethi, and R. E. Tarjan, *J ACM*, 27, No. 4 (October 1980), pp 758-71.

ENGINEERING

Bell's Echo-Killer Chip. D. L. Duttweiler, *IEEE Spectrum*, 17 (October 1980), pp 34-7.

Coupling Efficiency of Optics in Single-Mode Fiber Devices. R. E. Wagner and W. J. Tomlinson, *J Opt Soc Am*, 70 (December 1980), p 1596.

The Determination of Ruthinium in Plating Solutions and Electroplated Deposits. T. Y. Kometani and L. D. Blitzer, *Plat Surf Finish* (October 1980), pp 57-60.

Development of a Flame Resistant Noncontaminating PVC Jacket for Coaxial Cable. J. T. Loadholt and S. Kaufman, *29th Intl Wire and Cable Symp Proc 1980*, Cherry Hill, New Jersey, November 18-20, 1980, pp 245-52.

The Effect of Frictional Polymer on the Performance of Palladium Connector Contacts. E. S. Sproles, Jr., *Proc IEEE Electron Compon Conf*, 1980, pp 317-23.

Effects of the Lower Atmosphere on Radio Prolongation at Frequencies Above 1 GHz. T. S. Chu, *URSI Symp Proc*, Lennorville, Canada, May 26-30, 1980, pp 6.8.1-6.8.7.

The Effects of Multipath Fading on 16 QAM Digital Radio. W. W. Toy, *ICC Conf Rec*, 3 (June 1980), pp 52.1.1-52.1.5.

Experimental and Theoretical Classification of Submicron MOSFETs. W. Fichtner et al., IEDM Tech Dig, 1 (1980), pp 24-7.

Generalized Box Medal for Electron Irradiated Teflon Foils. B. Gruss et al., 1980 Ann Rep Conf Electrical Insulation and Dielectric Phenomena, Boston, Mass., October 1980, pp 313-20.

High Resistivity Interfacial Layers Due to Compensation by Sulfur in (Al, Ga)As Devices. P. J. Anthony, N. E. Schumaker, and J. L. Zilko, IEEE Electron Device Lett, *EDL-2*, No. 2 (February 1981), pp 46-9.

Joint Carrier Phase and Symbol Timing Recovery for PAM Systems. M. H. Meyers and L. E. Franks, IEEE Trans Commun, *COM-28*, No. 8 (August 1980), pp 1121-9.

Lightwave Communications. J. Simpson, Proc Intl Laser/Electro-Optics Training Conf, Wausau, Wisconsin, June 4-5, 1980, pp 105-23.

Magnetic and Transport Properties of Transition-Metal Based Glassy Ribbons. K. V. Roa et al., IEEE Trans Magn, *MAG-16* (1980), pp 896-901.

Magnetic Phase Diagram of Amorphous $(\text{Fe}_x\text{Ni}_{1-x})_{-15}\text{P}_{16}\text{B}_6\text{Al}_3$ Alloys. S. M. Bhagat, J. A. Geohogen, and H. S. Chen, Solid State Commun, *36* (1980), pp 1-5.

A New Multifunctional Mass Termination Connector. A. W. Carlisle and D. R. Frey, IEEE Trans Components, Hybrids Manuf Technol, *CHMT-3*, No. 4 (December 1980), pp 610-6.

New Routing Capability—Direct Signaling. M. A. McGrew and R. F. Frerking, Proc Natl Electron Conf, Chicago, Illinois, October 27-29, 1980, pp 159-63.

Optical Modulators. I. P. Kaminow, in *Encyclopedia of Science and Technology*, New York: McGraw-Hill, 1981, 5th ed.

Overview of Telecommunications via Optical Fibers. S. E. Miller, Proc IEEE, special issue on optical fibers, *68* (October 1980), pp 1173-4.

Preliminary Study of a Two Wire Split Beam Connection. E. Guancial and S. Das Gupta, IEEE Trans Compon, Hybrids Manuf Technol, *3* (December 1980), pp 654-8.

Preparation of Glassy Metals. H. S. Chen, H. J. Leamy, and C. E. Miller, Annu Rev Mater Sci, *10* (1980), pp 363-91.

Quality Audit Using Adaptive Kalman Filtering. M. S. Phadke, Intl Conf Am Soc Quality Control, San Francisco, California, May 27-29, 1981.

Reliability of 0.87 μm (Al,Ga)As Double Heterostructure Lasers with Ga(As,Sb) Active Layers. P. J. Anthony et al., IEEE Electron Device Lett, *EDL-2*, No. 2 (February 1981), pp 50-2.

Scanning the Issue. S. E. Miller, Proc IEEE, special issue on optical communications, *68* (October 1980), p 1172.

Switching Performance of a 12-GHz Pin Phase Shifter/Driver Module for Satellite Communication Phased Array. N. Amitay and B. Glance, IEEE Trans Commun, *COM-29*, No. 1 (January 1981), pp 46-50.

Thin Film Photoconductor Mounting Schemes for Picosecond Optical Detectors, P. R. Smith, D. H. Auston, and A. M. Johnson, Rev Sci Instrum, *52*, No. 1 (January 1981), pp 138-40.

Threshold-Voltage Instability in MOSFETs Due to Channel Hot-Hole Emission, R. B. Fan and R. C. Sun, IEEE Trans Electron Devices, *ED-28* (January 1981), pp 83-94.

Visibility of Borders: Separate and Combined Effects of Color Differences, Luminance Contrast, and Luminance Level. F. S. Frome, S. L. Buck, and R. M. Boynton, J Opt Soc Am, *71*, No. 2 (February 1981), pp 145-50.

SOCIAL AND LIFE SCIENCES

On the Depth of the Cyclopean Retina. C. W. Tyler and B. Julesz, Exp Brain Res, *40* (1980), pp 196-202.

Task—Oriented Approaches to Automatic Speech Recognition. A. E. Rosenberg et al., Proc 1980 SAE Congress (February 1980), pp 1-9.

Telecommunications in Earthquakes. J. W. Foss and J. D. Scobee, Soc Econ Impact Earthquake Util Lifeline Conf, San Francisco, California, May 20, 1980. New York: Am. Soc. Civil Engineers, 1981, pp 134-59.

MANAGEMENT AND LIFE SCIENCES

Approximating the Admissible Set in Stochastic Dominance. D. Goroff and W. Whitt, *J Econ Theor*, 23 (October 1980), pp 218-35.

A Control Theory Approach for Modeling and Analysis of Economical Systems. F. Aminzadeh, *Proc IEEE Control System Soc, NJ Section, Murray Hill, New Jersey*.

Errata

I. W. Sandberg, "On Newton-Direction Algorithms and Diffeomorphisms," *B.S.T.J.*, 60, No. 3 (March 1981), pp. 339-346.

On page 345, replace the first sentence of Theorem 3 with: Let $B = R^n$, and let f be a continuously differentiable map of R^n into R^n such that f' is Lipschitz on bounded subsets of R^n .

K. Nassau, "The Material Dispersion Zero in Infrared Optical Waveguide Materials," *B.S.T.J.*, 60, No. 3 (March 1981), pp. 327-337.

In eq. 2 (page 328), for d^2 read d^3 .

CONTENTS, SEPTEMBER 1981

Part 1

Sampling From Structured Populations: Some Issues and Answers
V. N. Nair and T. E. Dalenius

A First-Come-First-Serve Bus-Allocation Scheme Using Ticket Assignments
D. K. Sharma and S. R. Ahuja

Accurate Simulation Models for TTL Totem Pole and MOS Gates and Tristate Devices
Y. L. Levendel, P. R. Menon, and C. E. Miller

Special-Information-Tone Frequency Detection
A. Feuer

Effect of Echo Cancelor on Common Channel Interoffice Signaling Continuity Check
G. S. Fang

Profile Characterization of Optical Fibers—A Comparative Study
H. M. Presby

Statistical Behavior of Crosstalk Power Sum With Dominant Components
S. H. Lin

Current-Carrying Capacity of Fine-Line Printed Circuits
A. J. Rainal

A Comparative Study of Several Dynamic Time-Warping Algorithms for Connected Word Recognition
C. S. Myers and L. R. Rabiner

Part 2

DIGITAL SIGNAL PROCESSOR

Overview: The Device, Support Facilities, and Applications
J. R. Boddie

An Overview of the Silicon Very-Large-Scale-Integration Implementation
F. E. Barber, T. J. Bartoli, R. L. Freyman, J. A. Grant, J. Kane, and R. N. Kershaw

Architecture and Performance

J. R. Boddie, G. T. Daryanani, I. I. Eldumiati, R. N. Gadenz,
J. S. Thompson, and S. M. Walters

Logic and Fault Simulations

I. I. Eldumiati and R. N. Gadenz

Software Simulator

J. Aagesen

Design of the Assembler

C. L. Semmelman

A Tutorial Introduction to Digital Filtering

E. J. Angelo, Jr.

Adaptive Differential Pulse-Code-Modulation Coding

J. R. Boddie, J. D. Johnston, C. A. McGonegal, J. W. Upton,
D. A. Berkley, R. E. Crochiere, and J. L. Flanagan

Private Communications

C. A. McGonegal, D. A. Berkley, and N. S. Jayant

Receiver for *TOUCH-TONE*[®] Service

J. R. Boddie, N. Sachs, and J. Tow

Voice Frequency Transmission for Special-Service Telephone Circuits

R. B. Blake, A. C. Bolling, and R. L. Farah

Speech Synthesis

M. R. Buric, J. Kohut, and J. P. Olive

Subband Coding

R. E. Crochiere

Tone Generation

D. L. Favin

Power Measurements

D. L. Favin, D. Yorkgitis, and S. P. Cordray

Tone Generation

R. N. Gadenz

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, Current Contents/Engineering, Technology & Applied Sciences, Current Index to Statistics, Current Papers in Electrical & Electronic Engineering, Current Papers on Computers & Control, Electronics & Communications Abstracts Journal, The Engineering Index, International Aerospace Abstracts, Journal of Current Laser Abstracts, Language and Language Behavior Abstracts, Mathematical Reviews, Science Abstracts (Series A, Physics Abstracts; Series B, Electrical and Electronic Abstracts; and Series C, Computer & Control Abstracts), Science Citation Index, Sociological Abstracts, Social Welfare, Social Planning and Social Development, 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.

CONTENTS (continued)

| | | |
|---|--|-------------|
| T. C. Kaup, D. G. Leeper, A. K. Reilly, and P. E. Scheffler | T1 Carrier Characterization—Field Measurement Results | 965 |
| G. F. Erbrecht, J. H. Gentry, Jr., T. C. Kaup, and R. J. Schweizer | Automated Measurement System for T1 Characterization | 1005 |
| | CONTRIBUTORS TO THIS ISSUE | 1027 |
| | PAPERS BY BELL LABORATORIES AUTHORS | 1035 |
| | CONTENTS, SEPTEMBER ISSUE | 1039 |
| | ERRATA | 1038 |