THE MARCH 1975
VOL. 54 NO. 3
BELL SYSTEM
TECHNICAL JOURNAL

THE BELL SYSTEM TECHNICAL JOURNAL

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

THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 54 March 1975 Number 3

Preparation of Optical-Fiber Ends for Low-Loss Tape Splices

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

(Manuscript received September 18, 1974)

We describe a reliable method of preparing planar fiber tape ends by fiber fracture. Using this technique, with suitable precautions to preserve cleanliness during splice preparation, we have measured a splice loss of less than 0.25 dB in 99 percent of all attempts.

The alignment of fibers in prefabricated grooves[1] so far remains the simplest and most reliable method of connecting fibers, even in the case of fiber cable subgroups (tapes).[2,3] Two problems were identified as most serious:

(*i*) The preparation of satisfactory fiber ends was found difficult in the case of tapes and cables, because all ends must be in one cross-sectional plane. Grinding and polishing has proven feasible, but may not be entirely satisfactory, particularly in the preparation of field splices.

(*ii*) The splice losses have been higher than expected on the basis of single-fiber splice tests[4] and have been scattered over a wide range.

Although we are not certain that groove alignment necessarily provides the best splicing technique, we have used an advanced form of this technique developed by Cherin[3] to take a closer look at the problems identified above. Even if other techniques prove more promising later on, the problems mentioned may still be present in

some form or other and seem serious enough to require thorough analysis now.

The preparation of fiber ends discussed here is a modification of fracture techniques reported earlier for single-fiber splices.[4] The device used for this purpose was a compact and simple hand tool that could easily be operated in a cramped and narrow space. The essential element of this tool was a spring-steel strip over which the fibers were stretched. The mechanical characteristics of this strip primarily determined the stress distribution in the fibers, and, thus, by a proper choice of strip thickness, we were able to choose the appropriate ratio of bending to tensile stress for the particular fibers to be fractured.

To prepare the ends of the fiber tape for splicing, one proceeds as follows:

(*i*) The plastic of the tape is removed over a short distance so that the fibers are exposed in the area where the end is to be prepared.

(*ii*) The tape is placed between a spring-steel strip and two friction plates, so that the exposed area is located under a diamond stylus.

(*iii*) The spring-steel strip and the tape are bent. At the same time the friction plates slide a small distance along the spring-steel strip. This sliding action exerts an additional amount of tension on the tape, so that the optimal ratio between longitudinal and bending stress is obtained in the fibers.

(*iv*) The diamond stylus (tip radius 50 μm) is now drawn across the exposed fibers to produce scores. The slight pressure of a few grams imparted by a phosphor bronze spring suffices to produce scores of a few micrometers in depth. As each fiber is scored, a fracture starts at the score and proceeds across the fiber producing a flat surface perpendicular to the fiber axis.

For a more detailed explanation of this process, see Ref. 4. The order of the steps explained above is not imperative. As an alternative, the scoring can be done before tension is applied.

Figure 1 shows a typical array of fiber ends obtained in this way. The fibers used were multimode fibers having a high-silica core with a diameter of about 80 μm and an outer diameter of 120 μm. To make a tape splice, we prepared the ends of two tapes in the way discussed above. The tapes were then positioned slightly above a small grooved chip made from lead, copper, or aluminum (see Fig. 2). The chip was roughly 1 cm long and 3 to 4 mm wide. The grooves were embossed using a stainless-steel head that had six adjacent 90-degree grooves, each 80 μm deep. As shown in Fig. 2, the fibers were lowered into

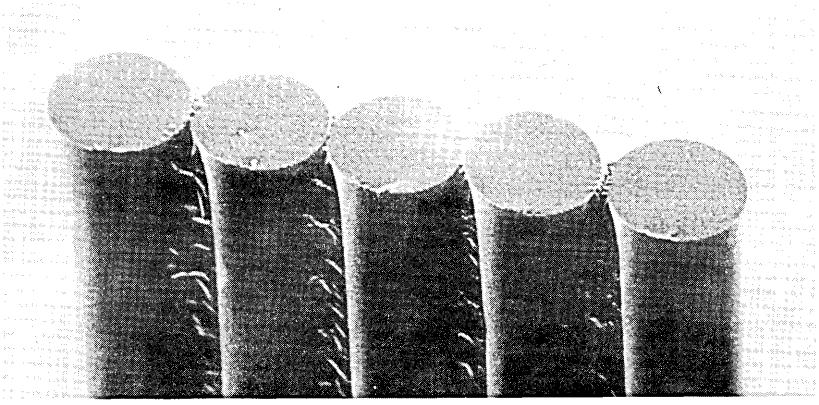

Fig. 1—Tape end prepared by simultaneous fracture of fibers. Fibers were epoxied together after end preparation to keep them aligned for electron micrograph process.

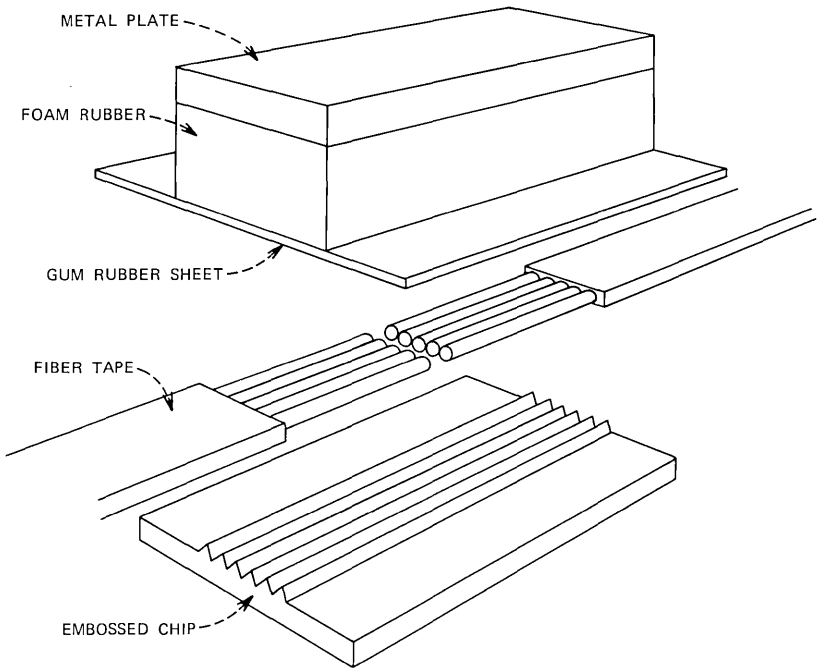

Fig. 2—Sketch of splice arrangement before assembly.

grooves by pressing from the top with a stack made of (from top to bottom) a lead weight, a foam rubber pad, and a sheet of gum rubber about 150 μm thick. After the fibers were lowered into the grooves, the tapes were gently pushed together in axial direction.

Tapes 2 m long were used to measure the splice loss. To simulate longer lengths, we injected light with a power distribution approximating the steady-state distribution of these particular fibers. We first made a large number of loss measurements in unbroken and unspliced tapes, switching back and forth between the five fibers of each tape to determine the measuring uncertainty. We found a distribution that had an rms value of 1.7 percent. The splice loss was then determined by measuring the transmission before and after a small part (a few centimeters) was removed from the middle of each tape and the ends spliced together as explained above, adding a drop of index-matching oil or glycerin before covering the arrangement with the gum rubber sheet. The optical loss of the length of fiber removed was insignificant. Figure 3 shows a histogram of the splice losses measured in 60 attempts. Evidently some loss values were negative as a result of the measuring inaccuracy. Figure 4 shows the (smoothed) cumulative loss distribution as measured and after the 1.7-percent rms measuring uncertainty was discounted. Of all measurements, 99 percent show a loss of less

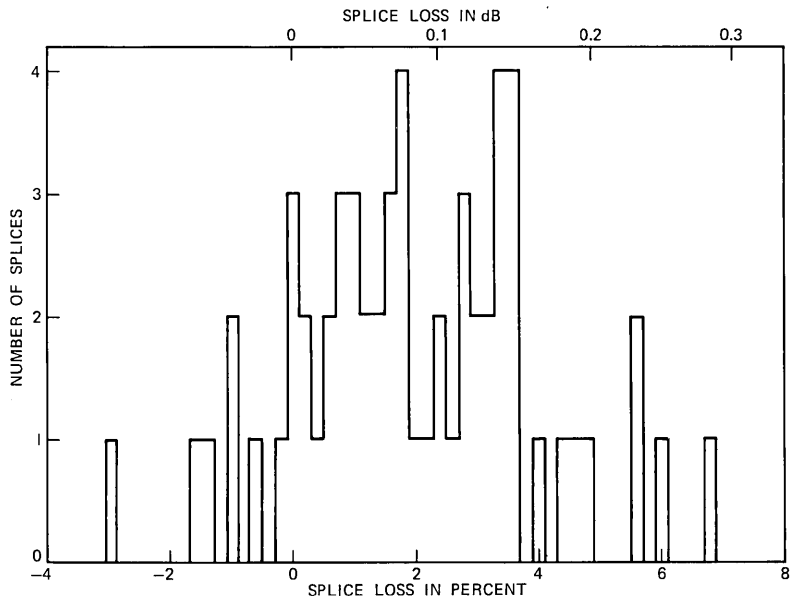

Fig. 3—Histogram of measured splice loss.

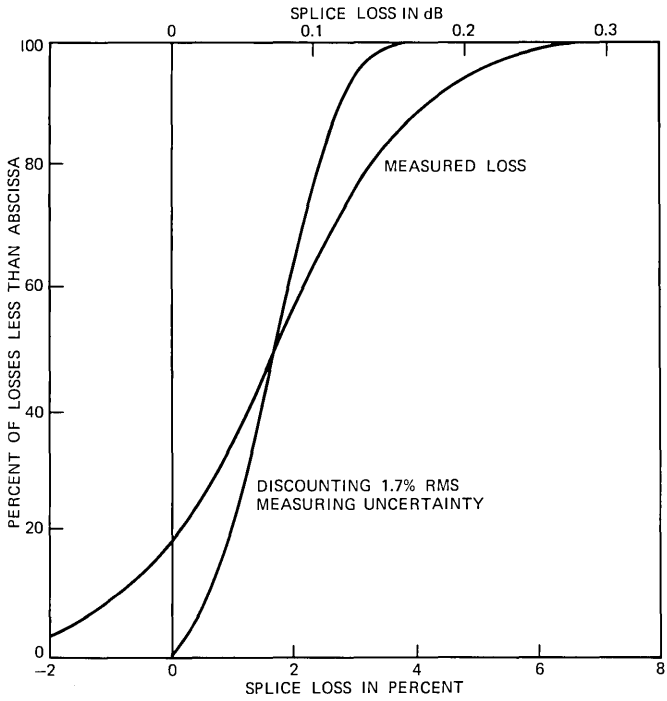

Fig. 4—Cumulative distribution of splice loss as measured and after discounting the 1.7-percent rms measuring uncertainty.

than 0.25 dB. However, a large part of the scatter of these measurements results from the measuring process and the actual cumulative distribution would predict that 99 percent of all splices would have a loss of less than 0.15 dB.

We attribute the low-loss values obtained not only to the quality of the end faces, but also to the extreme care taken during these measurements to keep the splice area clean. Earlier microscopic observations of groove-aligned fiber array splices taught us that, generally, losses in the 10-percent range can be correlated with a contamination in the splice area which reveals itself by large-angle scattering at the joint. In most cases, we were able to identify the contaminating material on the end face of the fiber even after the splice was taken apart; these materials tenaciously adhere to the fiber surface often even after ordinary cleaning procedures. We learned that an extended period of ultrasonic cleaning with isopropyl alcohol of *all* parts involved in the splice was necessary before the splice loss decreased to the levels measured. The contaminant is usually not added during the end prep-

aration; it is not the result of dust accumulation from the surrounding air caused, for example, by electrostatic forces. We believe that the contamination results from a contact of the fiber ends with contaminated surfaces, such as the grooved chip or the rubber sheet. We believe, also, that this sensitivity to contamination is a sufficient reason to consider splicing processes in which the fiber end surfaces are prepared and exposed after alignment has been achieved.[5]

Permanent splices on the basis of the techniques described here were prepared by replacing the index-matching oil with a special epoxy. This epoxy flows down along the grooves and the fibers and permanently attaches the tape ends to the embossed metal chip and the gum rubber sheet. The rubber sheet ends extending beyond the chip (see Fig. 2) are then folded around the chip and attached to its bottom surface. Figure 5 shows a finished five-fiber splice. Splices of this kind were found to have sufficient intrinsic strength to be used as splices of cable subgroups; additional armor would of course in this case be provided around a stack of such subgroup splices in a cable. The loss distribution for these permanent splices showed no deviation from that of splices made using index-matching fluid, and no aging effect was noticed, at least not within the period of a few days. The epoxy has a room-temperature curing time of several hours, but faster-curing epoxies are being studied and should replace the one used without significant alteration of the results.

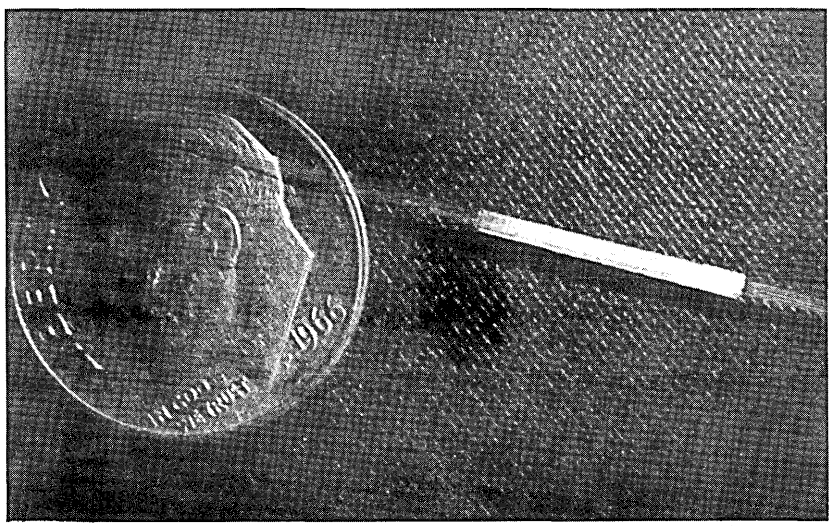

Fig. 5—Finished five-fiber splice.

We are grateful to R. D. Standley who prepared the electron micrograph of Fig. 1.

REFERENCES

1. C. G. Someda, "Simple, Low-Loss Joints Between Single-Mode Optical Fibers," B.S.T.J., *52*, No. 4 (April 1973), pp. 583–596.
2. R. D. Standley, "Fiber Ribbon Optical Transmission Lines," B.S.T.J., *53*, No. 6 (July–August 1974), pp. 1183–1185.
3. A. H. Cherin and P. J. Rich, "A Splice Connector For Forming Linear Arrays of Optical Fibers," Topical Meeting, Optical Fiber Transmission, Williamsburg, Va., January 7–9, 1975.
4. D. Gloge, P. W. Smith, D. L. Bisbee, and E. L. Chinnock, "Optical Fiber End Preparation for Low-Loss Splices," B.S.T.J., *52*, No. 9 (November 1973), pp. 1579–1588.
5. E. L. Chinnock, D. Gloge, P. W. Smith, and D. L. Bisbee, "End Preparation and Splicing of Optical Fiber Ribbons," Topical Meeting, Optical Fiber Transmission, Williamsburg, Va., January 7–9, 1975.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

All-Glass Optical-Fiber Tapes

By D. L. BISBEE and P. W. SMITH

(Manuscript received September 18, 1974)

We propose and demonstrate a new approach to the problem of splicing optical fibers in a fiber cable. The optical fiber cable subgroups (tapes) are made in such a way that the relative positions of the optical fibers are accurately maintained. By using glass as a rigid matrix material in which the optical fibers are held, we demonstrate that a simple scoring and stressing technique can be used to simultaneously prepare all the fiber ends for splicing.

The potential of optical fibers as transmission media for optical communications systems has stimulated much work on the various problems that need to be overcome before a practical system can be built. One of these problems involves the development of techniques for connecting and splicing these fibers and fiber cables. Although several laboratory techniques for splicing fibers and groups of fibers with low splice losses have now been developed, they are all relatively complex techniques that require operations of high precision and are thus difficult to carry out in the field.

In this paper, we propose a different approach to the splicing problem. Linear arrays ("tapes") of optical fibers have been suggested as building blocks for optical fiber cables,[1] and a number of techniques for producing plastic-bonded tapes have been investigated.[2] We propose here a technique for fabricating a precision all-glass optical fiber tape that would have considerable advantages with regard to splicing operations. The precision operations would be performed during the manufacturing of the tapes, and splicing operations in the field would become relatively simple. The basic idea is to make a fiber tape in which the optical fibers are held in a rigid matrix with their relative positions accurately maintained. Further, by using a glass for the rigid matrix material, we can greatly simplify the problem of preparing the optical fiber ends for splicing: All the fibers comprising the tape can be prepared for splicing in a single operation by utilizing the scoring and stressing technique described earlier in Ref. 3.

We made all-glass tapes by fusing conventional clad soda-lime-silicate glass optical fibers together with lower melting point glass fibers in a precision jig. To get a stable bond, the glasses must have very nearly the same thermal expansion coefficient even though their softening temperatures are different. Glasses with these characteristics can be obtained.

As an example, let us consider glass composed of SiO_2, Na_2O, and CaO ("soda-lime-silicate" glass). Morey[4] defines the softening point of glass as that point at which the viscosity becomes $10^{7.6}$ poises. From Ref. 4, we see that by changing the composition of our soda-lime-silicate glass, it is possible to make one composition that has a viscosity of $10^{7.6}$ poises at 650°C and another with a viscosity two orders of magnitude greater than this at the same temperature. The thermal expansion coefficient of the first will be 9.7×10^{-6} per °C, while that of the second will be 11.3×10^{-6} per °C. Thus, we can obtain glasses that have viscosities different by two orders of magnitude at the softening point, while their thermal expansion coefficients differ by only 15 percent. By using such glasses, we could ensure that the deformation during the fusing operation would take place only in the low-melting-point glass, and the optical fibers would be essentially undistorted.

The optical fibers used for these experiments were conventional clad multimode fibers made from soda-lime-silicate glass. The low-melting-point fibers were made from ferrite sealing glass obtained from the Corning Glass Works (No. 8463). The optical fibers had a softening temperature of ≈ 650°C and a thermal expansion coefficient of 9.7×10^{-6} per °C, while the softening temperature of the low-melting-point fibers was ≈ 375°C, and the thermal expansion coefficient was 10.4×10^{-6} per °C.

Figure 1 is a schematic view of the precision guide used to fabricate the all-glass fiber tape. The optical fibers are held accurately in place by the precision guide, while the low-melting-point triangular fibers are introduced in such a way that they press against two adjacent optical fibers. By means of a heating element, the glass is fused at the contact points and then allowed to cool as the completed tape is pulled out of the guide.

The heating element is a 500-μm outer diameter nichrome wire mounted on a manipulator. When the heating wire is brought to within 100 μm of the cold fibers and heated by passing current through it, the triangular fibers melt and form beads of molten glass that touch the heater. This glass is then smoothly spread along the round fibers as the fibers are fed into the guide, giving a uniform bond.

The precision guide is made of boron nitride—a material that resembles soapstone. It is used because it is easily machinable to

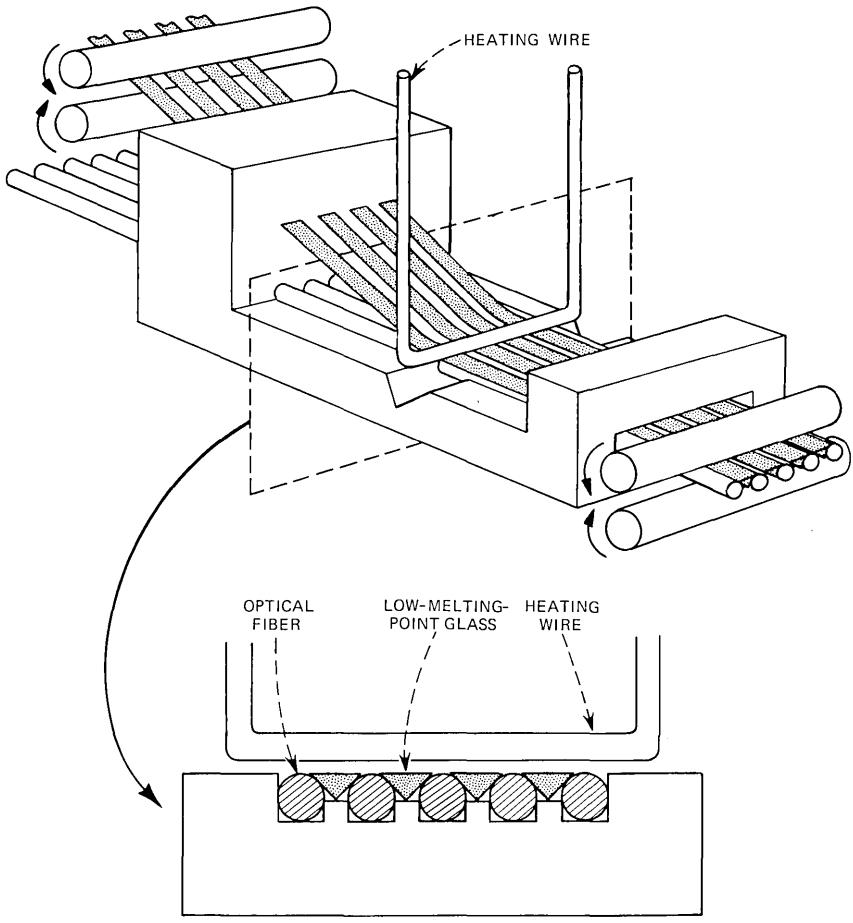

Fig. 1—Schematic diagram of the device used to fuse the all-glass fiber tapes.

close tolerances, has a slick surface, and is not affected by the heat applied.

To feed the fibers through the precision guide, rollers are used as shown in Fig. 1. One pair of rollers is used to push the triangular fibers into contact with the round fibers, and one pair is used to pull the fused tape out of the guide. The lower roller in each pair is a 6.25-mm shaft with a tight-fitting sleeve of vinyl 0.5-mm thick giving an outer diameter of 7.25 mm. The upper roller of each pair is a 6.25-mm shaft with a flexible plastic sleeve giving an outer diameter of 9.38 mm. One of the smaller rollers is driven by a variable speed motor and is connected to the other small roller through an idler gear. Before entering

the guide, the fibers were cleaned by passing through a solvent-soaked wick.

To make the tape, the heater current is turned on while the heating wire is still far from the fibers (more than 250 μm), the driving motor is turned on, and then, while the fibers are passing through the guide, the heating wire is lowered by a manipulator (while the operator watches through a microscope) until the triangular fibers melt and bond smoothly to the round fibers. The tape was formed at about 3 cm per minute. Figure 2 is a photograph of the resultant fiber tape.

It has previously been shown that, by scoring an optical fiber and subjecting it to a properly tailored stress distribution, a smooth fracture perpendicular to the fiber axis can be obtained.[3] It has also been observed that such fracture behavior can be obtained with more complex fiber cross sections. We used a diamond stylus to score the fiber tape and fractured it by subjecting it to bending and tension stress. The fracturing operation was performed with the aid of a device similar to that described in Ref. 5. Internal strains would sometimes cause the tape to fracture in such a way that the break was not perpendicular to the tape axis. Nevertheless, good fiber ends were usually produced. Figure 3 shows a fiber tape end produced in this way.

The solder-glass fibers that were used tended to break very easily

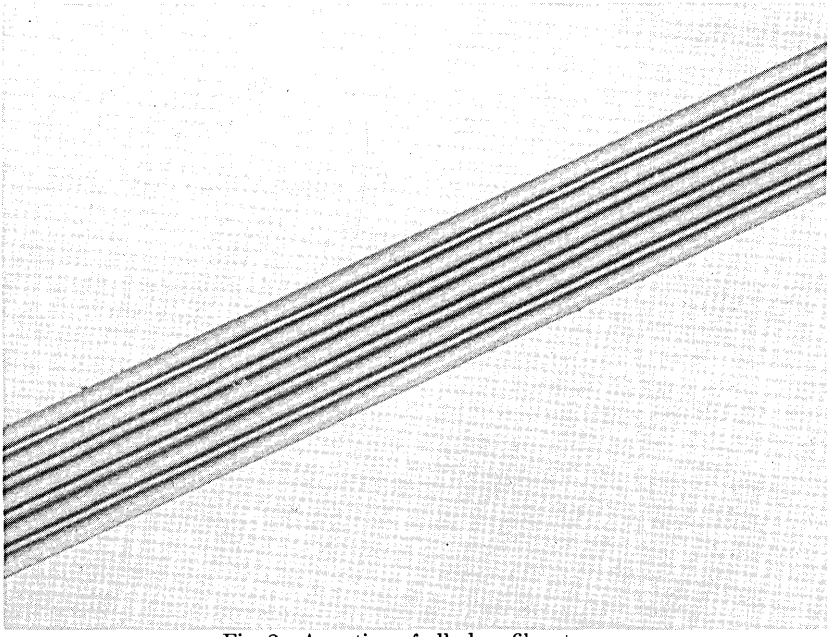

Fig. 2—A section of all-glass fiber tape.

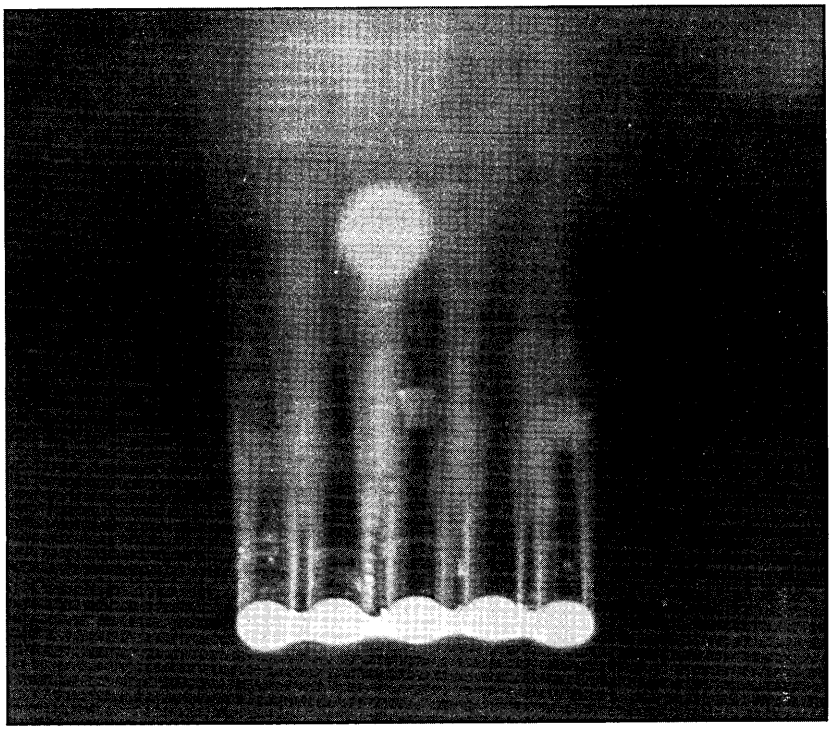

Fig. 3—A fiber tape end prepared for splicing by the scoring and stressing technique.

and had to be handled with great care; thus, the tapes we made were fragile. When fused together, the round and triangular fibers are so intimately bonded together that a crack originating in the glass of a triangular fiber would propagate across the whole tape. If low-melting-point glass with better mechanical properties were used, we would expect the finished tape to be appreciably stronger. As can be seen in Fig. 3, the tape is no thicker than its round fibers, so it is flexible in the direction of its thinner dimension.

Because the optical fibers are accurately positioned during the manufacture of the tape, the most difficult part of the splicing problem has already been solved, and the splicing of these tapes merely involves preparing the tape ends by scoring and bending and placing the prepared ends in a suitable holder with matching fluid and a cover to hold the tape ends in place, as shown in Fig. 4. To make a permanent splice, a transparent index-matching epoxy may be used. Note that at no stage during the entire splicing process does one have to deal with single optical fibers, and that the tapes can be placed by hand in the splicing holder without the need for any precise visual alignment.

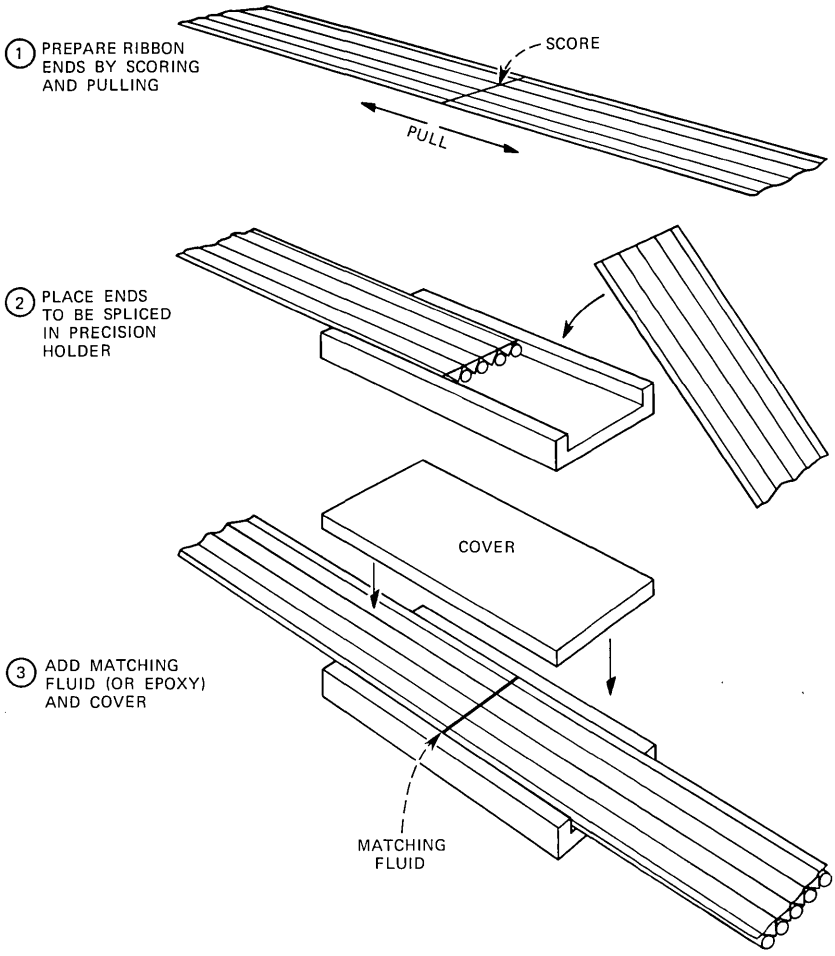

Fig. 4—Schematic representation of tape splicing operations.

We would like to thank A. R. Tynes for pulling the optical fibers and the triangular solder glass fibers used for these experiments.

REFERENCES

1. R. D. Standley, "Fiber Ribbon Optical Transmission Lines," B.S.T.J., *53*, No. 6 (July-August 1974), pp. 1183–1185.
2. B. R. Eichenbaum, unpublished work; C. M. Miller, "A Laminated Fiber Ribbon Concept for Optical Communication Cables," to be published; and M. J. Saunders, "Adhesive Sandwiched Optical Fiber Ribbons and Cables," to be published.
3. D. Gloge, P. W. Smith, D. L. Bisbee, and E. L. Chinnock, "Optical Fiber End Preparation for Low-Loss Splices," B.S.T.J., *52*, No. 9 (November 1973), pp. 1579–1588.
4. G. W. Morey, *The Properties of Glass*, New York: Reinhold, 1954.
5. E. L. Chinnock, D. Gloge, P. W. Smith, and D. L. Bisbee, "Preparation of Optical-Fiber Ends for Low-Loss Tape Splices," B.S.T.J., this issue, pp. 471–477.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

Synthesis of Speech From a Dynamic Model of the Vocal Cords and Vocal Tract

By J. L. FLANAGAN, K. ISHIZAKA, and K. L. SHIPLEY

(Manuscript received July 17, 1974)

We describe a computer model of the human vocal cords and vocal tract that is amenable to dynamic control by parameters directly identified in the human physiology. The control format consequently provides an efficient, parsimonious description of speech information. The control parameters represent subglottal lung pressure, vocal-cord tension and rest opening, vocal-tract shape, and nasal coupling. Using these inputs, we synthesize vowel-consonant-vowel syllables to demonstrate the dynamic behavior of the cord/tract model. We show that inherent properties of the model duplicate phenomena observed in human speech; in particular, cord/tract acoustic interaction, cord vibration, and tract-wall radiation during occlusion, and voicing onset-offset behavior. Finally, we describe an approach to deriving the physiological controls automatically from printed text, and we present sentence-length synthesis obtained from a preliminary system.

I. INTRODUCTION

Speech sounds can be synthesized by a variety of means used to construct signal waveforms. Many ingenious methods have been recorded. But speech synthesis generally has the practical purpose of producing intelligible sounds from control data that are as parsimonious as possible. In other words, the control data should represent an efficient, concise coding of the speech information. This motivation applies as much to analysis/synthesis techniques for speech transmission as to computer voice-response systems which strive for efficient vocabulary storage and high versatility in message fabrication.

Because speech is a human-generated signal, it is unlikely that a synthesis method can achieve the ultimate parsimony of input control without considerable attention to the parameters a human overtly manipulates in speaking. That is, one increases the information "built into" the synthesizer when its design exploits fundamental properties of the human speech mechanism.

We therefore have chosen an approach to synthesis with which we can identify overtly the significant physiological parameters important

in speech production. Major system components obviously are the mechanism of voiced-sound generation and the mechanism for intelligibly modulating sound timbre, that is, the vocal cords and the vocal tract. Our approach, unlike that found conventionally in the speech literature, is not to make a linear separation of the sound source and vocal tract. More than this, we believe that source/tract interaction actually contributes built-in natural behavior that is significant in synthesis. This natural interaction is missing in approaches that assume linear separation of source and tract (unless provided at additional expense and coding effort in the input data).

The initial results stemming from this approach to synthesis are described below.

II. ACOUSTIC MODEL OF VOCAL CORDS AND VOCAL TRACT

We view the acoustic system of the human vocal cords and vocal tract as shown at the top of Fig. 1. The lungs are an air reservoir, maintained at subglottal air pressure P_s by contraction of the rib-cage muscles. The subglottal pressure is applied via the bronchi and trachea passages to the variable-area orifice controlled by the vocal cords.

We model the cords as an acoustic-mechanical oscillator, wherein a single vocal cord is described by two masses, each having an associated stiffness and loss, which are "internally" coupled by a third stiffness. In previous work,[1–4] we established the philosophy leading to this description and gave a quantitative analysis of the vocal cord model.

Oscillation of the vocal cord model results in the glottal volume velocity U_g. This quantity typically has an impulsive waveform and it is the excitatory source for voiced sounds.

The vocal tract proper is a nonuniform tube, about 17 cm long in man, extending from the cords to the mouth. Its cross-sectional area varies from zero to upwards of 20 cm^2. The nasal tract is an ancillary tube about 60 cm^3 in total volume and coupled to the vocal tract by the trap-door action of the velum. Sound is radiated from the system as a result of the volume velocities at the mouth U_m and nostril U_n, and from vibration of the yielding sidewalls of the vocal tract.

Cross-dimensions of the acoustic system are small compared to sound wavelengths of interest, and hence we confine our analysis to plane-wave propagation in the tract. We therefore represent the acoustic system as the bilateral, time-varying transmission line shown in the lower part of Fig. 1. Formulation of this system follows that given by Flanagan.[5]

As illustrated, the lossy lung volume is "charged" to subglottal lung pressure P_s, which is applied via the trachea-bronchi network, to the glottal (vocal-cord opening) impedance Z_g. This nonlinear glottal

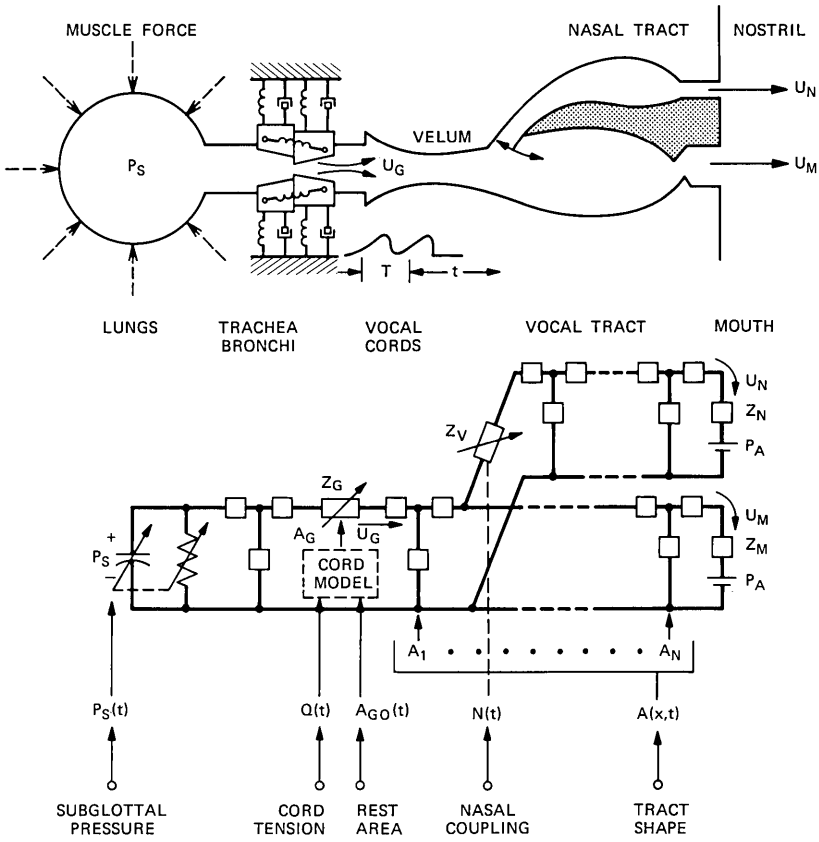

Fig. 1—Schematic diagram of the vocal cord/vocal tract system.

impedance depends upon the glottal flow and area A_g, which in turn depend upon the self-oscillating properties of the vocal cord model described in detail in an earlier paper.[1] The resulting volume flow U_g is the excitation source for the vocal and nasal tracts.

The shape of the vocal tract is defined by its cross-sectional area as a function of distance $A(x)$, and the coupling to the time-invariant nasal cavity is governed by the velar impedance Z_v. Volume velocity at mouth U_m and nostril U_n flow through their respective radiation impedances Z_m and Z_n, both of which are in series with batteries representing the constant atmospheric pressure P_a. (This formulation permits simulation of respiration as well.) The mouth and nostril radiation impedances are those for a circular piston in an infinite baffle.[5]

Parameters of control for the speech synthesis system are the physiologically-based functions shown in Fig. 1. All vary with time.

They are subglottal lung pressure P_s, vocal-cord tension Q, rest (or neutral) area of cord opening A_{go}, nasal coupling N, and cross-sectional area function of the tract shape $A(x)$. We are concerned here only with nonnasal sounds, hence nasal coupling will not figure in the discussion.

Each T-section of the vocal-tract transmission line is represented in Fig. 2.[5] An elemental length Δx of the vocal tube has cross-sectional area A, terminal sound pressures p_1 and p_2, and terminal volume velocities U_1 and U_2. The sidewall has noninfinite mechanical impedance, and vibrates in response to the enclosed sound pressure with displacement ξ. This displacement radiates a per-unit-length sound pressure p_{wall}. Relations between terminal values of pressure and volume

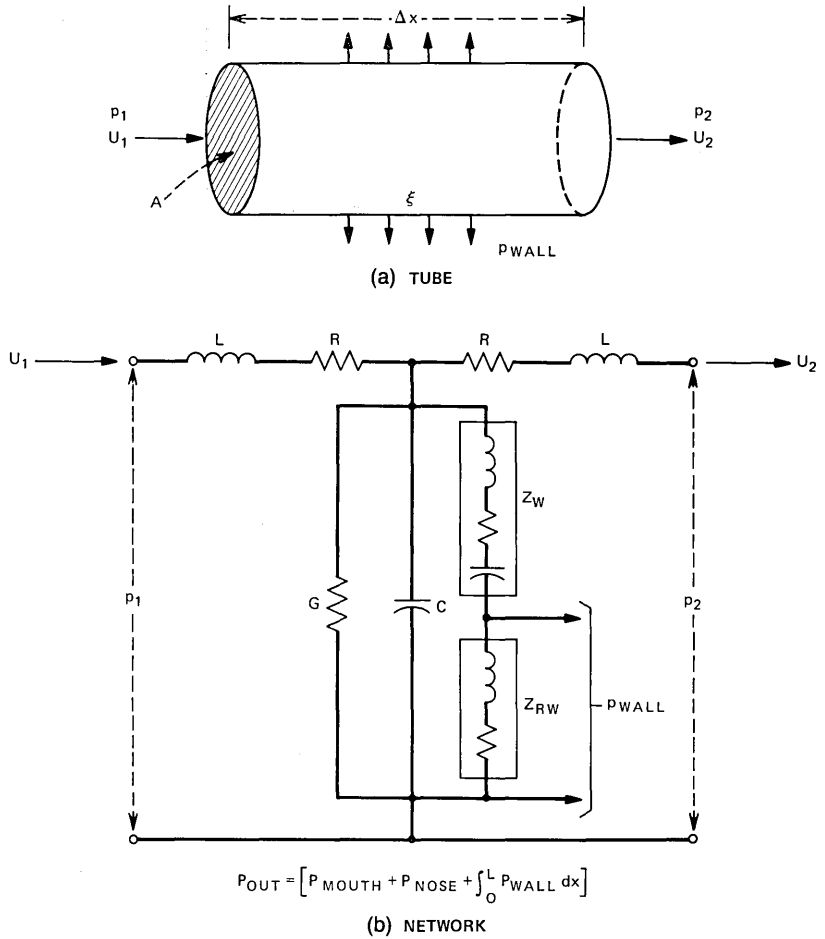

Fig. 2—Circuit representation of plane acoustic wave propagation in an elemental length of tube with yielding sidewalls.

velocity for plane-wave propagation are given in the circuit in Fig. 2b, in which L is the per-unit-length inertance of the air mass of the tube element, R the viscous loss at the sidewall, G the heat-conduction loss, C the acoustic compliance of the contained air volume, Z_w the acoustic equivalent mechanical impedance of the yielding wall, and Z_{rw} the radiation impedance of the wall, assumed to be that for a pulsating right circular cylinder.[6]

The total sound output from the model is, following the long-wave assumptions, the linear superposition of the mouth and nostril radiation plus the spatially summated wall radiation.

In addition, every T-section of the transmission-line network includes a means for introducing turbulent noise excitation. This capability is provided by a series random pressure source P_N with its internal resistance R_N, as shown in Fig. 3. This technique has been given in detail previously.[4] The intensity (or rather mean-square variance) of the random pressure source is controlled by the Reynolds number of the flow at each network section, while the internal resistance is similarly modulated according to the Bernoulli loss in a constriction.[5] In both instances, the specified value of cross-sectional area A and the calculated resulting volume velocity flowing through the serial source completely describe the control functions. That is, no additional input data are required.

More specifically, to simulate the conditions of turbulent-source generation in any section, the amplitude of the noise pressure is made directly proportional to the squared Reynolds number in excess of a

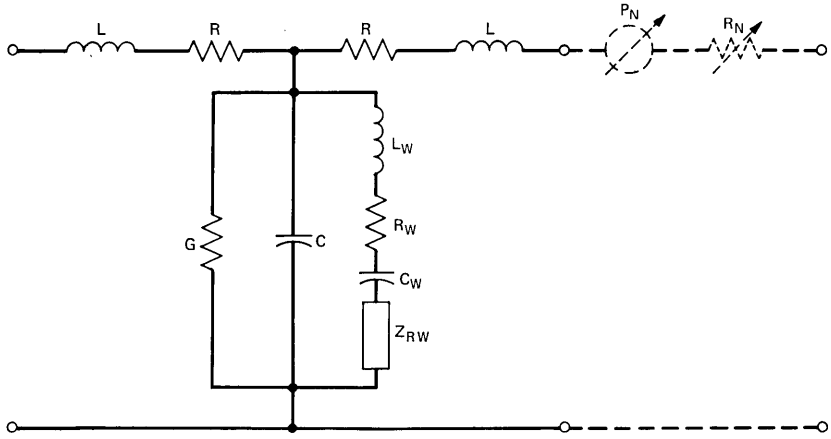

Fig. 3—Circuit representation of the turbulent noise source for each network section. The intensity of the random pressure source, P_N, and its self-resistance, R_N, are controlled by the volume velocity and cross-sectional area at each section.

critical (threshold) Reynolds number for turbulent flow.[4] The squared Reynolds number is proportional to U^2/A, whereas the internal resistance of the turbulent source is, to first order, a flow-dependent loss proportional to $|U|/A^2$. Therefore, as the prescribed section area becomes small in the presence of a large flow-velocity, turbulence conditions are indicated and the intensity of the noise source and the value of its internal resistance are increased. In the simulation, values of every dependent variable are calculated on a sample-by-sample basis to construct the time functions for the output sound pressure and all other pressure and velocity quantities.[4]

As a consequence of continually noting the magnitude of the calculated volume flow in each section and having the tract cross-sectional areas continually prescribed as input data, the synthesizer *automatically* introduces random noise excitation in any section when the Reynolds number is sufficiently high to indicate turbulent flow. The synthesizer, therefore, requires no additional data to produce voiceless sounds, but uses exactly the same control parameters to generate both voiced and voiceless sounds (or combinations of voiced and voiceless sounds). As Fig. 1 has shown, these control parameters are P_s, Q, A_{go}, N, and $A(x)$.

As a practical matter in the computer implementation, we use a P_N source produced from gaussian noise (or, rather, gaussian numbers) bandpass-filtered from 500 to 4,000 Hz. Further, to insure stability, the volume flow which modulates the serial noise source is low-pass filtered to 500 Hz. In other words, the noise source is modulated by low-frequency components of U, including the dc flow.

The transmission line model of Fig. 1 is described by a set of linear and nonlinear differential equations in which all coefficients also vary with time. This set of differential equations is approximated by difference equations, as previously described,[2] and programmed in a laboratory computer for on-line control. Twenty network sections are used to approximate the vocal tract. This formulation has permitted initial experiments with physiologically-based control of the synthesis model.

III. ASSESSMENT OF WALL IMPEDANCE AND EFFECTS ON FORMANT BANDWIDTH

All elements of the transmission line network have been well established in previous work, with the exception of the wall-vibration shunt branch of the circuit in Fig. 2.

Assessment of wall effects in earlier calculations[5] utilized the only available mechanical impedance measurements of human tissue,

namely, chest, stomach, and thigh tissue. These data led to correct order-of-magnitude values for wall-vibration damping of formant resonances, but the values were clearly on the high side.

To better assess the wall impedance, we have done two things. First, we have used data on human formant bandwidths to estimate contributions to losses in our model. And second, we have made direct measurement of the mechanical impedance of the vocal-tract wall.[7]

Formant bandwidths have been measured for the human vocal tract by van den Berg,[8] Bogert,[9] Fujimura and Lindquist,[10] and Dunn.[11] Our programmed model allows us to adjust values of the wall-impedance parameters to effect three consistencies. It permits us to (*i*) adjust the wall-loss component to match glottis-closed formant bandwidths, (*ii*) adjust the inductive reactance of the wall to produce the observed mouth-closed, lowest value of first formant frequency of about 200 Hz, and (*iii*) choose a wall compliance to produce wall resonance substantially below 100 Hz. Small-signal-driven vibration of the cord oscillator in the model permits calculations of model response at any prescribed frequency. Furthermore, formant bandwidths measured on real speech allow additional cross-checks of parameters used in the model formulation, especially for the loss components of the cord-oscillator source. Application of this knowledge in our model yields the formant bandwidth behavior shown in Fig. 4.

In particular, Fig. 4 illustrates how the wall viscous loss parameter can be chosen to match glottis-closed formant bandwidth. This technique has recently been analyzed in extensive quantitative form by Sondhi.[12] The wall loss is selected to match measured formant bandwidths at formant frequencies around 300 to 500 Hz. In this frequency range, the contributions to formant bandwidth are mainly wall loss and glottal source loss. Viscosity, heat conduction at the walls, and mouth radiation resistance represent relatively small values (see Ref. 5, for example, for these calculation techniques). Note, too, that in Fig. 4 the vertically-sloping line of calculated bandwidth indicates the effect of wall impedance on the tuning of formant frequency. In the absence of additional data, we assume a uniform distribution of the per-unit-area wall impedance along the tract. The value we use for the mechanical per-unit-area impedance is $(1600 + j1.5\omega)$g/s/cm^2, where ω is the radian frequency. This value is confirmed well by our direct measurements of wall impedance.[7]

Formant bandwidths measured in real speech[11] permit a cross-check of the glottal oscillator parameters chosen in previous work.[1] Figure 4 shows that the contribution to formant damping of the glottal source falls into the correct range of real speech measurements. This is a

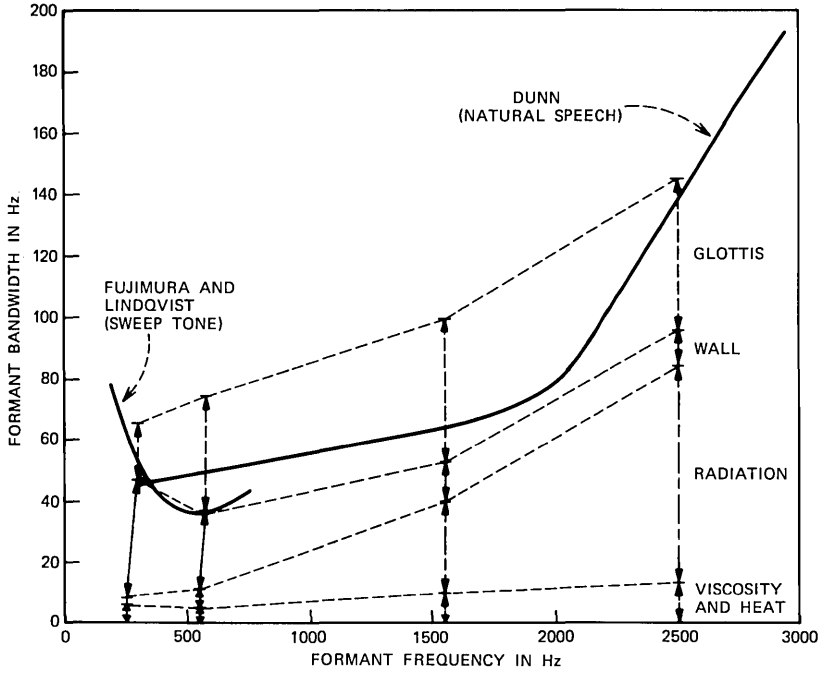

Fig. 4—Variation of formant bandwidth (damping) with formant frequency. The diagram shows the relative contributions of loss to the total formant bandwidth produced by the cord/tract model.

gratifying confirmation of cord parameters chosen strictly on other bases, namely, according to physiological properties and oscillatory behavior.[1]

The loss contributions of the glottal source in Fig. 4 are calculated for a nominal, midrange value of glottal rest area, namely $A_{go} = 0.05$ cm^2. The glottal contribution to formant bandwidth is, of course, a function of A_{go}. Figure 5 shows glottal loss contributions for other values of A_{go}. Note, especially, how the articulatory configuration of the tract influences the contribution of the glottal source to formant damping.

IV. DYNAMIC BEHAVIOR OF THE CORD/TRACT MODEL

How does this physiologically-based model of the vocal cords and vocal tract behave under dynamic control? Time-varying control inputs in the present study are P_s, Q, A_{go}, and $A(x)$. An obvious major problem is the determination of realistic values of these parameters. As a first cut, fairly realistic data can be derived from direct measurements of lung pressure during speech,[13] laryngeal muscle

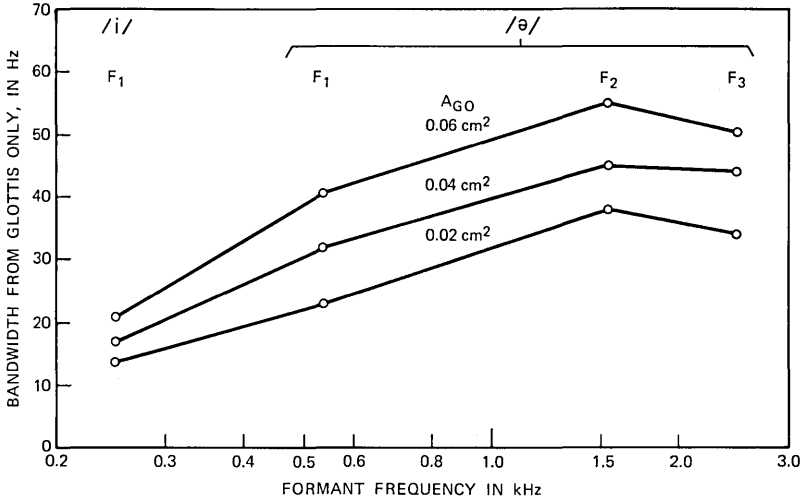

Fig. 5—Variation of the glottal loss contribution to formant bandwidth as a function of formant frequency. The parameter is the neutral glottal area, A_{go}.

electromyography,[13] glottal transillumination,[14] glottal pulses,[15] and cine X-rays of the vocal tract.[16] An important element at present is that all these data are not simultaneously available for a given subject. Experiments now under way aim to provide some simultaneous measurements.[17]

The physiological literature does provide adequate bases for dynamic tests on some simple utterances, using idealized input controls. We have, therefore, made first tests on vowel-consonant-vowel syllables (v-c-v) in which stress may be on either initial or final vowel, and where the intervocalic consonant is a voiced or unvoiced labial stop. These combinations also provide a convenient vehicle for exposing other physiologically realistic properties of the cord/tract model.

Figure 6 shows the synthesis of the syllable /'abə/. Input controls are indicated in the top three traces. Because of the initial stress, P_s falls during the labial closure to a lower value. Because the intervocalic stop is voiced, A_{go} is maintained in a position favorable to cord oscillation throughout. Cord tension, not shown, is also maintained constant. Any pitch changes are effected solely by P_s variation and by the interaction of tract load on the cord oscillator. Articulatory shape $A(x)$ changes from /a → b → ə/. Because of space limitation in the illustration, only the mouth area, A_m, is displayed.

Response behavior of the model to these input controls is shown in the bottom five traces: the sound spectrogram of the total output sound; A_g; U_g; the pressure waveform of the total output sound P;

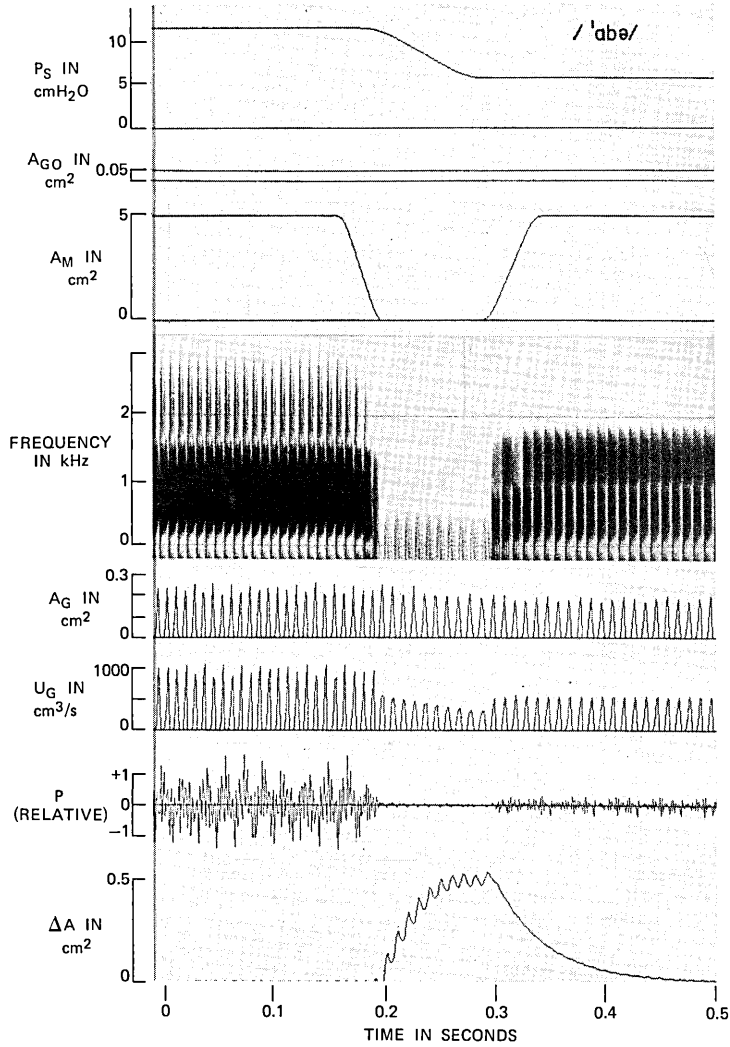

Fig. 6—Control functions and sound output from the cord/tract model synthesizing the syllable /ˈabə/. The effects of sound radiation from the yielding sidewalls is evidenced in synthesized sound, and the vibration of the mouth cavity wall is illustrated by the ΔA trace.

and the incremental change in area ΔA of the oral cavity in response to the contained sound pressure.

Several things are notable. In the sound spectrogram, notice the intense initial vowel /a/ with relatively elevated pitch (about 120 Hz) and with natural formant transition into the stop. Voicing continues

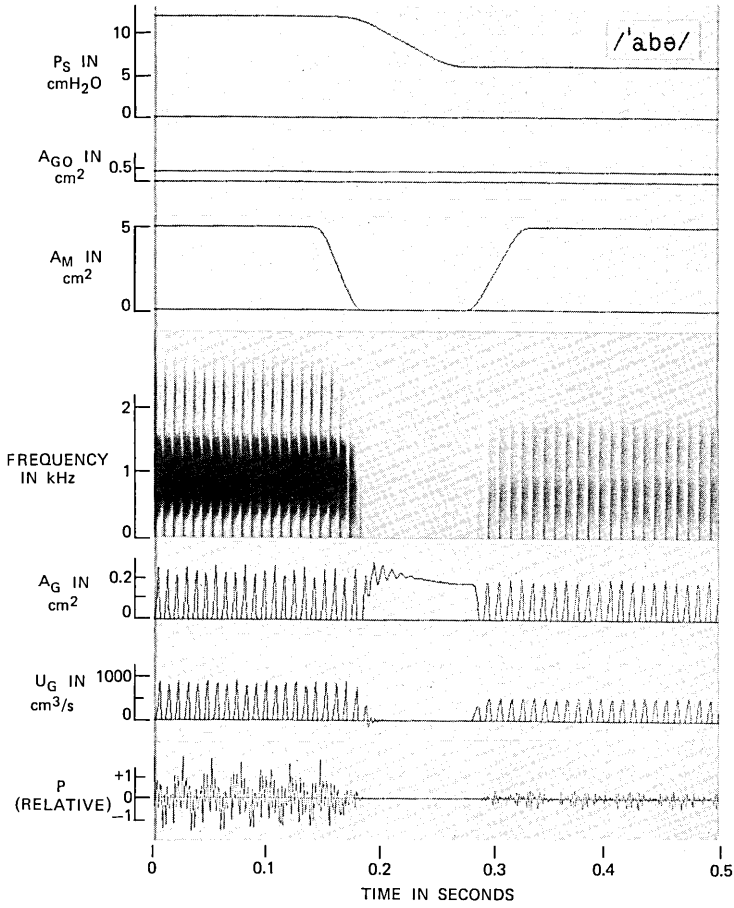

Fig. 7—Behavior of the cord/tract model when the sidewalls are made rigid. The input control functions are identical to those of Fig. 6. The synthesized syllable is /'abə/. Note especially that the cord-oscillator ceases vibration during the mouth closure.

throughout the labial closure, at slightly reduced pitch (about 95 Hz), and with the sound output coming solely from the wall radiation. The sound level during the lip closure is on the order of 20 dB lower than the mouth-radiated vowels. Natural transition into the final vowel follows, with voicing at reduced pitch (104 Hz) and intensity.

The waveforms of the A_g and U_g oscillations confirm the spectrogram display, as does the waveform of output pressure. The wall-radiated sound is dramatized by examining the incremental area change in the yielding-wall oral cavity. The area perturbation is seen to follow pitch-synchronously the glottal pulses of U_g.

It is instructive to contrast this soft-wall behavior with that which obtains when the tract is made hard-walled; i.e., by letting $Z_w \to \infty$. This behavior, for exactly the same input control data, is shown in Fig. 7.

Now, because the tract walls do not yield and permit enlargement, the transglottal pressure is rapidly diminished during the labial closure,

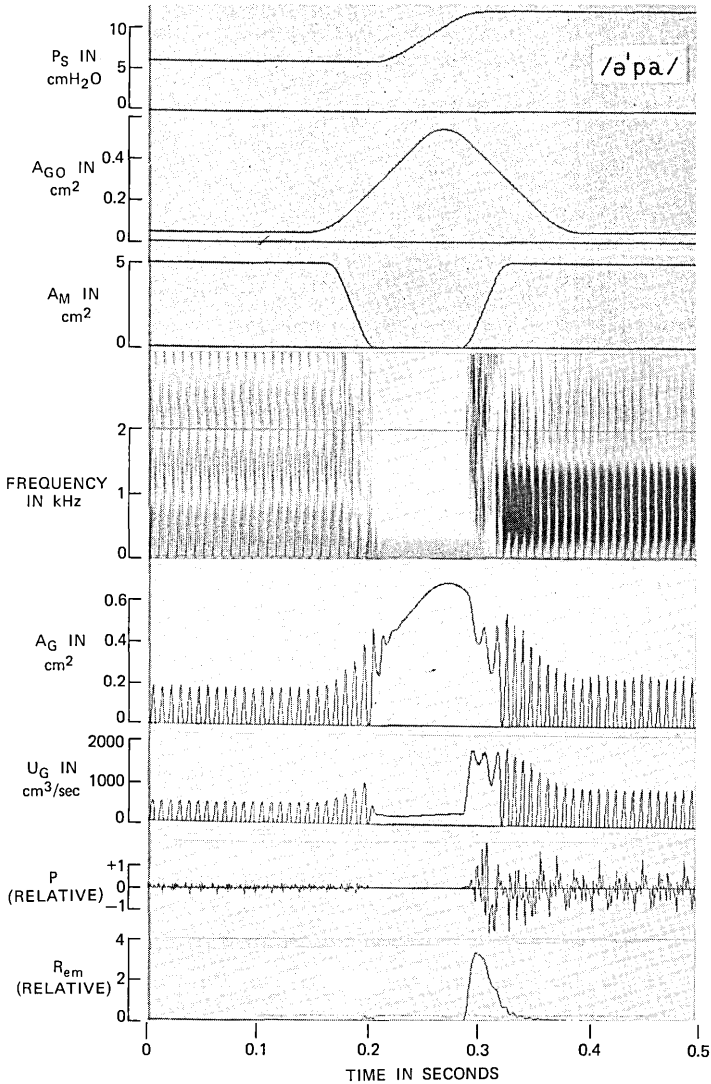

Fig. 8—Control functions and synthesized sound for the syllable /ə'pa/. Sound generation for the voiceless consonant is produced from the distributed random noise sources that are modulated by the Reynolds number for every network section.

and cord oscillation rapidly ceases during the /b/ consonant. Also, no sound is radiated from the tract walls, and only silence prevails during the lip closure. Offset and onset of cord oscillation, with lip closure and release, appears abnormal when compared to transillumination data taken on human vocalization. This latter factor may be more important perceptually than the actual absence of sound during lip closure.

Dynamic behavior for a voiceless intervocalic labial stop is displayed in Fig. 8. The syllable is /ə'pa/, with stressed second vowel. Again, control function input is indicated by the top three traces. Only mouth area A_m is again displayed, and cord tension Q is held constant. Note now, however, the A_{go} control effects voiced-voiceless switching by moving from a value that sustains cord oscillation to one that does not.

The spectrogram of the sound output shows the low-intensity, low-pitch initial vowel with natural formant transitions into the stop. Cord oscillation ceases during the closure because the cords are overtly pulled apart. (The lateral and posterior crico-arytenoid muscles accomplish this in the human larynx.) The cords come back together as the lip closure is released, and oscillation starts with an abrupt bounce that is quite characteristically seen in glottal transillumination data on humans.[18] Natural formant transition is made into the final, high-intensity, relatively-higher-pitched vowel.

The U_g flow continues without cord oscillation through the lip closure, as the tract wall yields and enlarges the volume forward of

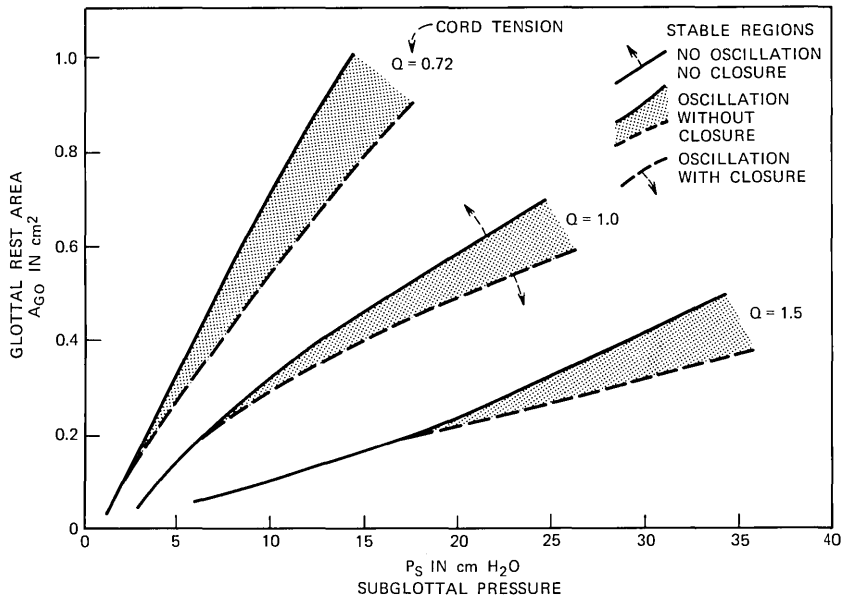

Fig. 9—Regions of stable oscillation for the vocal-cord model.

the cords. As the lips release, the U_g flow reflects a relatively large dc component before oscillation commences. This flow is the source of aspiration in the consonant release.

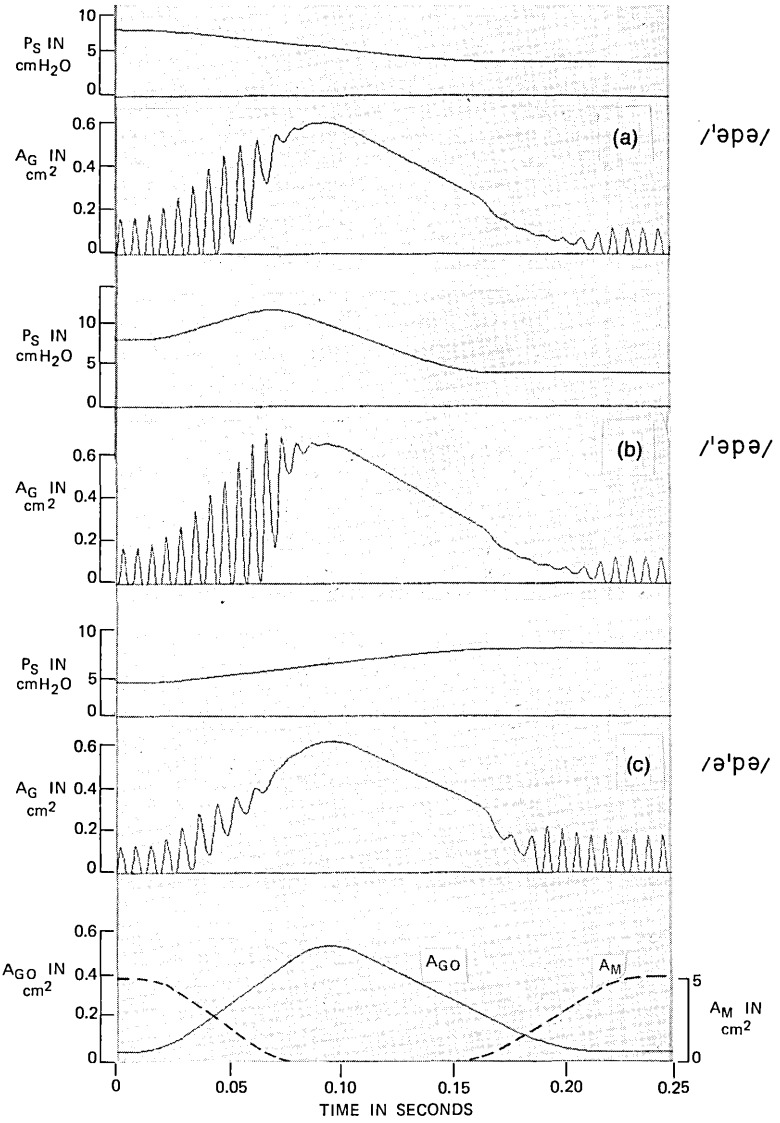

Fig. 10—Behavior of the vocal-cord oscillation as a function of subglottal pressure. The dynamics of tract motion and the control of the vocal cord neutral area are the same for each diagram. Subglottal pressure for conditions (a) and (b) correspond to initial stressed vowel, whereas condition (c) represents final stressed vowel. Note especially the delayed voicing onset in (b).

The automatic turbulence generation is indicated by the lower trace in Fig. 8, which is the squared Reynolds number for the volume flow at the lips. As discussed previously, turbulence (noise) intensity is monotonely related to this function, in excess of a threshold value.[5] The high spike in R_{em} at about $t = 0.3$ s indicates a turbulent burst of noise with approximately this amplitude envelope. The sound output pressure waveform and the spectrogram show the result of this turbulence generation. The result is consistent with aspirated releases seen in the /p/ consonant. Furthermore, auditory assessment of the synthesized sound indicates a natural-sounding syllable.

This synthesis also highlights the importance of the A_{go} control for switching between voiced and unvoiced sounds. A more detailed indication of this behavior is shown in Fig. 9. Three distinct regions of stable cord behavior are indicated. For given cord parameters, stable behavior is determined by the interplay of P_s and A_{go}.

An additional examination of dynamic behavior dramatizes the so-called delayed voicing onset. The syllable /əpə/ is generated with the A_{go} and A_m controls shown at the bottom of Fig. 10. The cord tension, Q, is maintained constant. Lung pressure, P_s, however, is varied to correspond to initially stressed vowels (conditions a and b) and a finally-stressed vowel (condition c). Notice especially in condition b, the initially rising, then abruptly falling P_s conspires with the first opening, and then closing A_{go} control to produce substantial delay in the resumption of voicing. This is found characteristically in human speech.[19]

V. AUTOMATIC GENERATION OF CORD/TRACT CONTROL

Ultimately, we wish to use the cord/tract model as an end-organ for speech synthesis. What are the prospects for obtaining the necessary controls automatically by rule?

In recent work on synthesis-by-rule, Coker and Umeda[20] generated synthetic speech from printed text using programmed algorithms for articulatory dynamics and for speech prosody. Their speaking machine includes a pronouncing dictionary, a syntax and prosody analyzer,

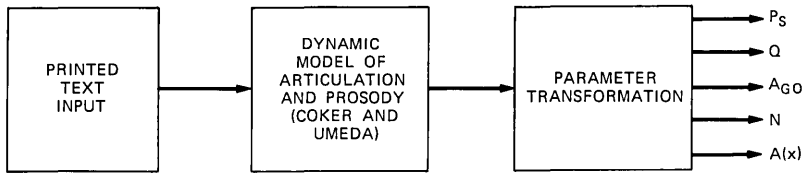

Fig. 11—Automatic generation of control functions for the cord/tract synthesizer from printed text input.

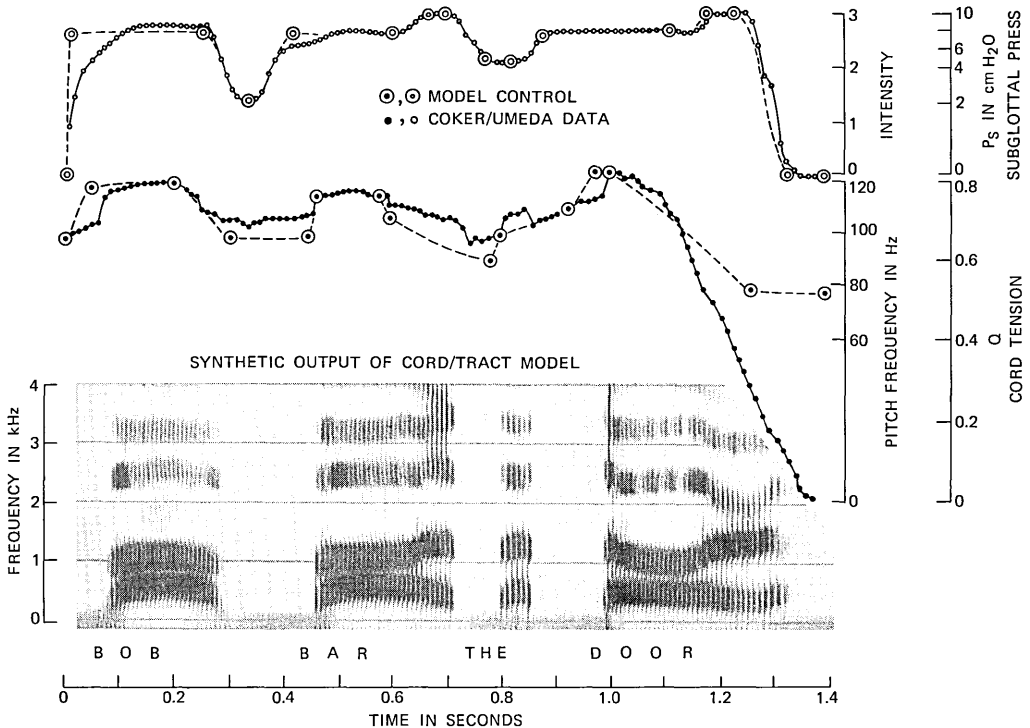

Fig. 12—Example of automatic synthesis of a voiced sentence from printed text using the cord/tract synthesizer.

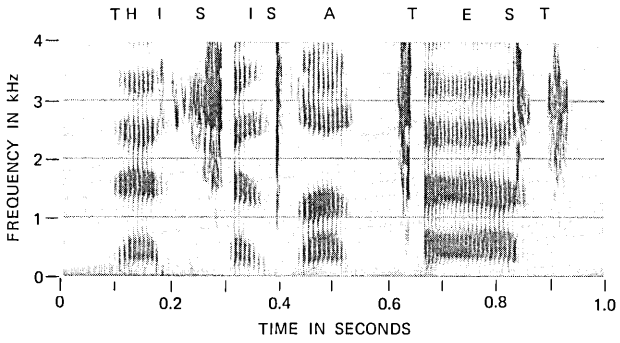

Fig. 13—Example of automatic synthesis from printed text for a sentence containing voiceless consonants.

and a dynamic model of vocal-tract shape. The text synthesis program calculates several functions that can be transformed into the parameters needed for the control of our cord/tract synthesizer. The sequence of conversions is illustrated in Fig. 11. As determined from the Coker-Umeda machine, overall sound intensity can be related to P_s, voice pitch to Q and P_s, voiced-unvoiced switching to A_{go}, and tract shape to N and $A(x)$.

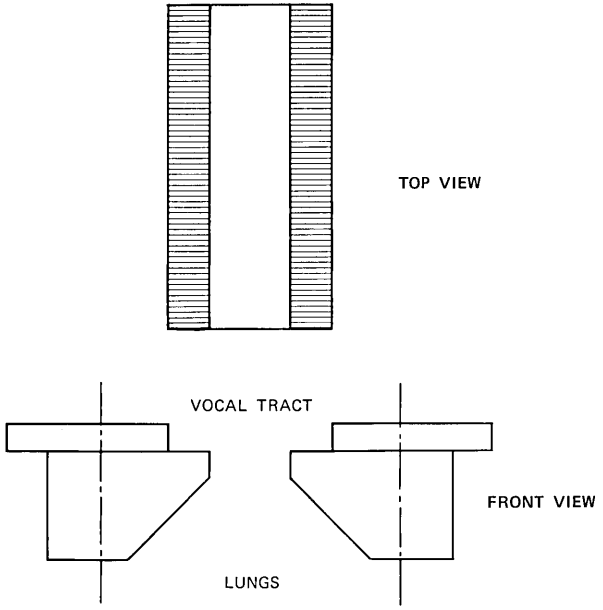

Fig. 14—Format of the computer movie illustrating dynamic behavior of the vocal-cord model.

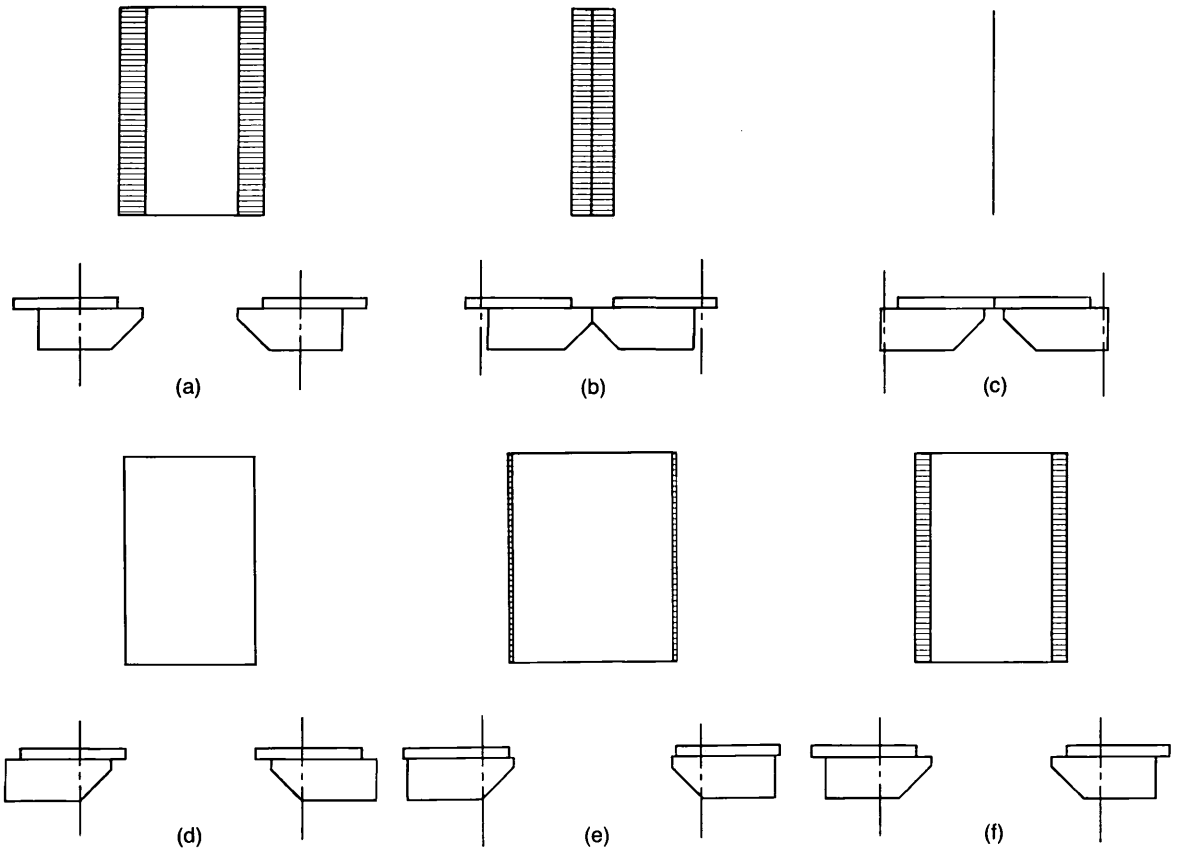

Fig. 15—Frame sequence from the computer movie illustrating vocal-cord vibration.

With the collaboration of Coker and Umeda, we have made an initial trial at synthesis of connected speech by making a transformation of the prosody and area-function output of their text-synthesis machine. An illustration of this first attempt to marry the two systems is shown in Figs. 12 and 13. Figure 12 includes plots to show how the machine-determined values of voice pitch frequency and intensity are transformed into the Q and P_s parameters required by our synthesizer. The spectrogram of Fig. 11 shows completely automatic synthesis of a voiced sentence. Figure 12 shows automatic synthesis of a sentence containing voiceless sounds.

VI. SLOW-MOTION COMPUTER PICTURES OF CORD AND TRACT BEHAVIOR

To aid in visually assessing the complex control and interaction of the model components, we programmed high-speed microfilm displays of the cord and tract motion. The 16-mm movie film, when shown at 24 frames/s, corresponds to a 100:1 slowdown of real time. One can, therefore, examine detailed cord motion and cord/tract interactions.

One display shows details of the two-mass vocal-cord model under dynamic control. The film format is given in Fig. 14 and shows simultaneously a top view of the glottal opening and a front (anterior-posterior) view of the two-mass cord model. Some prints of frame sequences are given in Fig. 15. The time between displayed frames is 20 ms.

A second display, given in Fig. 16, shows a schematized side view of the whole vocal system. The vocal tract is simplified to four cyl-

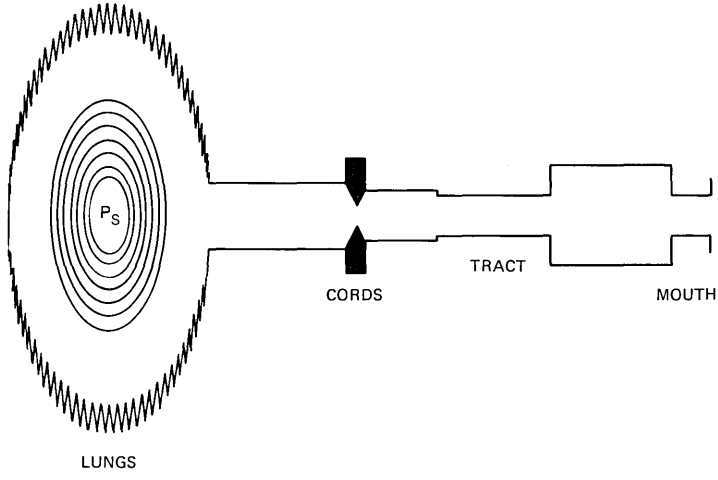

Fig. 16—Format of the computer movie showing dynamic articulatory relations between lung pressure, cord motion, and tract shape.

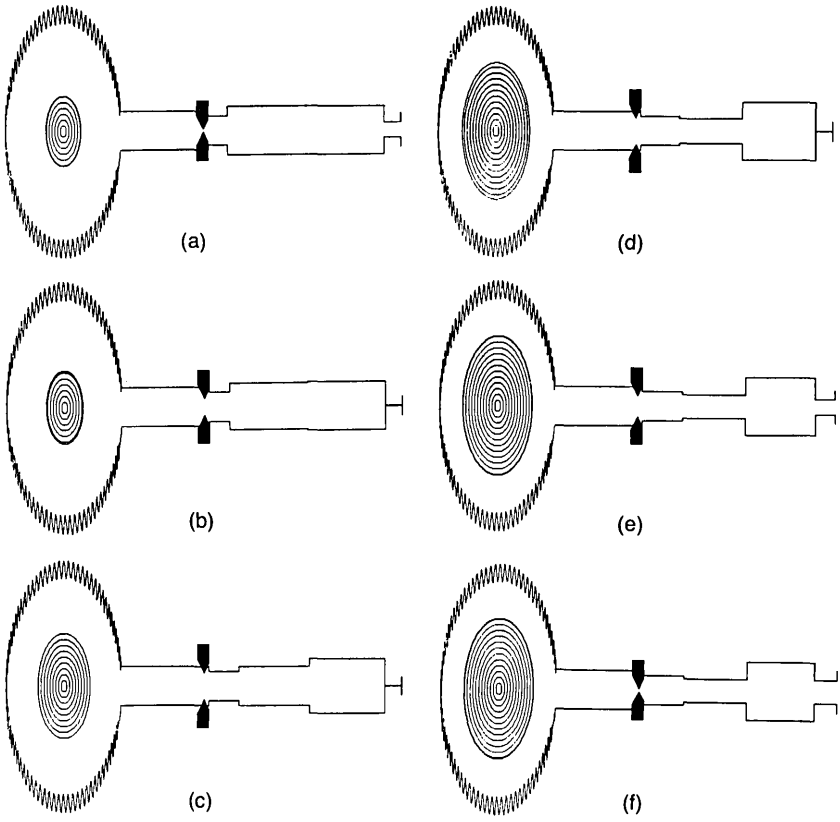

Fig. 17—Frame sequence from the computer movie illustrating dynamic behavior of the cord/tract synthesizer. The sequence is taken from the synthesis of /ə'pa/.

indrical sections only, but with lengths and areas that change with time. The magnitude of subglottal pressure is represented by the elliptical contours in the lung volume which expand or contract with time. Figure 17 shows a sequence of motion frames, spaced by 20 ms, for generation of the syllable /ə'pa/.*

VII. CONCLUSION

Initial experiments with this formulation of cord and tract properties suggest that the physiologically-based control functions have distinct advantages in terms of "built-in" information. That is, much natural behavior—such as vagaries of voicing onset and offset, fine-structure pitch fluctuations occasioned by tract motion, and voicing behavior dur-

* The data of Fig. 17 correspond to those of Fig. 8.

ing occlusion—is produced automatically in the model. In other words, faithful modeling of significant physiological parameters leads to input control data that can be rather parsimonious. It is therefore not necessary to describe input commands with intricate, high-information-rate detail. The model is able to generate many of these intricacies of natural behavior from relatively simple input control.

If continued work proves the cord/tract formulation to indeed possess the desired physiological constraints and attributes, the synthesis approach would also seem promising as a relatively sophisticated end-organ synthesizer which could be driven by models of prosody and articulation, such as provided by the Coker-Umeda text-synthesis system. This is an ultimate long-range goal.

Further than this, however, the model promises some extensive potential for studying the dynamics of real speech. Feasibility is presently being examined for automatically adapting the model's synthetic output to match real speech waveforms (for example, in a least-squares sense). Gradient-climbing adaptive algorithms are being examined for this.[21] Obvious difficulties are model nonlinearities and multiple local-minima traps which may be encountered. Continued work will determine whether these analytical questions can be solved.

Finally, since the present cord/tract synthesis model incorporates the technique for automatic generation of turbulence devised earlier,[4] this feature permits detailed study of the remarkably delicate articulatory timing the human employs in transitions between voiced and voiceless sounds. The cord/tract model therefore fills a critical need for a framework within which to organize and assess articulatory measurements now being accomplished.*

REFERENCES

1. K. Ishizaka and J. L. Flanagan, "Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords," B.S.T.J., *50*, No. 6 (July-August 1972), pp. 1233–1268.
2. J. L. Flanagan and L. L. Landgraf, "Self-Oscillating Source for Vocal-Tract Synthesizers," IEEE Trans. Audio and Electroacoustics, *AU-16* (March 1968), pp. 57–64.
3. K. Ishizaka and M. Matsudaira, "What Makes the Vocal Cords Vibrate," 6th Int. Congr. Acoust., *II* (August 1968), pp. B, 9–12.
4. J. L. Flanagan and L. Cherry, "Excitation of Vocal-Tract Synthesizers," J. Acoust. Soc. Amer., *45*, No. 3 (March 1969), pp. 764–769.
5. J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, Second Edition, New York: Springer Verlag, 1972.
6. S. N. Rshevkin, *A Course of Lectures on the Theory of Sound*, New York: Macmillan, 1963, pp. 400–405.
7. K. Ishizaka, J. C. French, and J. L. Flanagan, "Direct Determination of Vocal-Tract Wall Impedance," J. Acoust. Soc. Amer., *55* (April 1974), p. S79(A).

* In a cooperative study of dynamics of articulation with Dr. T. Shipp, Veterans Hospital, San Francisco, California.

8. J. W. Van der Berg, "Transmission of the Vocal Cavities," J. Acoust. Soc. Amer., *27*, No. 1 (January 1955), pp. 161–168.
9. B. P. Bogert, "On the Band Width of Vowel Formats," J. Acoust. Soc. Amer., *25*, No. 7 (July 1953), pp. 791–792.
10. O. Fujimura and J. Lindqvist, "Sweep-Tone Measurements of Vocal-Tract Characteristics," J. Acoust. Soc. Amer., *49*, No. 2 (February 1971), pp. 541–558.
11. H. K. Dunn, "Methods of Measuring Vowel Formant Bandwidths," J. Acoust. Soc. Amer., *33*, No. 12 (December 1961), pp. 1737–1746.
12. M. M. Sondhi, "A Model for Wave Propagation in a Lossy Vocal Tract," J. Acoust. Soc. Am., *55*, No. 5 (May 1974), pp. 1070–1075.
13. T. Shipp and R. E. McGlone, "Laryngeal Dynamics Associated with Vocal Frequency Change," J. Speech and Hearing Res., *14* (December 1971), pp. 761–768.
14. M. Sawashima, "Movements of the Larynx in Articulation of Japanese Consonants," Annual Bulletin, Research Institute of Logopedics and Phoniatrics, Univ. Tokyo, No. 3 (1968), pp. 11–20.
15. M. M. Sondhi, "Measurement of the Glottal Pulses," 86th Meeting, Acoust. Soc. Amer., *53* (October 1973), G-1.
16. J. S. Perkell, *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*, Boston: M.I.T. Press, 1969.
17. J. French, C. H. Coker, M. M. Sondhi, K. Ishizaka, and J. L. Flanagan, "Measurement of Articulatory Dynamics," Proc. Int. Congr. Acoustics, London (July 1974).
18. C. H. Coker, unpublished laboratory data on transillumination.
19. J. Lindqvist, "Laryngeal Articulation Studied on Swedish Subjects," STL-QPSR, 2–3 (1972), pp. 10–27.
20. C. H. Coker, N. Umeda, and C. P. Browman, "Automatic Synthesis from Ordinary English Text," IEEE Trans. Audio and Electroacoustics, *AU-21*, No. 3 (June 1973), pp. 293–298.
21. E. Hafer, "Speech Analysis by Articulatory Synthesis," M.S. Thesis, Dept. Elec. Eng., Northwestern Univ., June 1974.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

RC Active Filters for the D3 Channel Bank

By R. A. FRIEDENSON, R. W. DANIELS, R. J. DOW, and P. H. McDONALD

(Manuscript received August 1, 1974)

The development of the voice-frequency active filters for the D3 channel bank is described. These filters are the first single-substrate RC active filters using thin-film tantalum RC and silicon integrated-circuit technology to be produced on a large scale by Western Electric. To create complete confidence in both the design and the new technology, an extensive model building and testing program was undertaken. In addition, continuous interaction with manufacturing engineers resulted in a design that facilitated the introduction of this new technology in a large-scale manufacturing environment.

Significant advances were made in reducing the complexity of tuning active filters. In fact, the tuning and testing procedure has been adapted for use with a totally automated computer-controlled test set. Furthermore, a Monte Carlo statistical simulation of the manufacturing process of the filters was developed. This model includes tolerances of the manufactured components, test set errors in measuring gain and component values, resistor adjustment accuracies, and temperature and aging behavior of the components. This computer program has been an invaluable tool in determining the requirements for tuning, testing, and optimizing the final design for minimum manufacturing cost.

I. INTRODUCTION

Resistance-capacitance active filters are a relatively new addition to the family of frequency selective networks. Although active filters have been in existence for 20 years, they have not been widely used because passive filters have been less expensive. However, the concurrent development of silicon integrated circuits, tantalum thin-film technology, and automated computer-controlled test sets have made RC active filters a practical alternative to passive filters at voice frequencies.

This paper deals with the design and development of the voice-frequency RC active filters used in the D3 channel bank.[1,2] These filters are realized in thin-film tantalum technology with beam-leaded silicon integrated-circuit operational amplifiers on one ceramic substrate.[3]

New analysis[4,5] and modeling techniques, more efficient optimization algorithms,[6] the concept of statistical design,[7] and more flexible machine aids to physical design gave the network designers powerful tools with which to attack the design process.

II. DEVELOPMENT PROCESS

The development process involves a number of interrelated steps. Neglecting any of the steps can result in a design that either does not meet requirements or is unmanufacturable at a reasonable cost. Traditionally, the first step is a careful analysis of the filter requirements. This will determine the order of the filter and, possibly, the technology needed for the physical realization (e.g., LC, RC active, or mechanical technology). Next, the initial design is undertaken. This step requires accurate models for the components and general-purpose analysis and synthesis routines. Unfortunately, when initial designs using new technologies are breadboarded, requirements are very rarely met. In addition, the physical realization is either too large and/or the network is too costly. Thus, we must optimize the network to meet not only electrical requirements but also size, environmental, and cost restrictions. Next, sensitivity studies must be undertaken to determine both the viability of the design and whether or not manufacturing adjustments (tuning) are necessary. Since the designer cannot afford to wait 10 or 20 years to see if his design meets performance objectives in the field, a statistical simulation of the manufacturing process and field performance is useful. This simulation can analyze the effects of component and adjustment tolerances, simulate the tuning procedure and temperature and aging effects, and statistically optimize the design to give both the greatest manufacturing yield and a prediction of the end-of-life performance.

In the design of filters, early attention to the physical realization and tuning and testing for manufacture is important. Lack of attention to these details can result in an unmanufacturable design and/or a design that has excessive cost. In addition, when new technology is introduced, the designer must closely follow the initial manufacture of his design to aid the manufacturer in overcoming initial production hurdles.

III. FILTER ENVIRONMENT AND REQUIREMENTS

The D1 and D3 channel banks[1] are the terminals for the T1 PCM repeatered line.[8,9] The T1 PCM line carries 24 telephone conversations over two cable pairs. A simplified sketch of the voice frequency end of the D3 channel bank is shown in Fig. 1. A voice-frequency analog signal from the switching equipment passes through a hybrid trans-

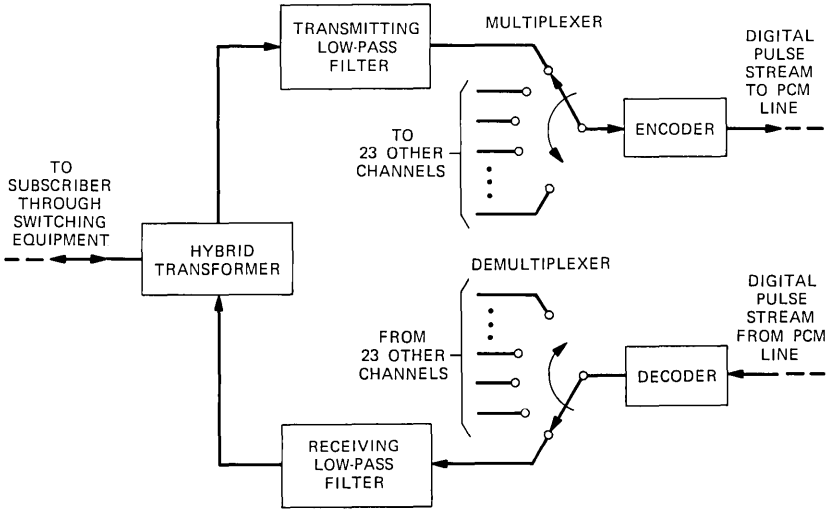

Fig. 1—Simplified model of PCM channel bank.

former. It is then amplified and bandlimited by the transmitting low-pass filter. Subsequently, it is converted into a pulse-amplitude modulated (PAM) signal by a JFET (junction field-effect transistor) switch operating at an 8-kHz rate. The PAM signals from the 24 channels are sequentially encoded into a binary bit stream (1.544 megabits/s) and sent out as PCM signals over a line. On the receiving end, the incoming PCM signals are sequentially decoded and converted to PAM signals. The analog signal for each individual channel is then recovered from the PAM signal by passing it through an interpolating receiving low-pass filter. The analog signal then passes through the hybrid transformer to the switching equipment and subsequently to the subscriber.

The transmitting band-limiting filter must present a good impedance to the hybrid transformer and negate the effect of the switched load impedance. In addition, it must pass frequencies between 200 and 3200 Hz with less than ± 0.09-dB ripple and suppress frequencies above 5 kHz by at least 30 dB. At half the sampling rate of 8 kHz (or 4 kHz), there must be at least 15 dB of suppression. An RC active filter is ideal for this voice-frequency application, since the operational amplifier output will be impervious to the time-varying load, and the passband ripple performance is not degraded by inductor losses. The above requirements can be met with a fifth-order Cauer-Chebyshev filter. To obtain a low-cost, highly reliable filter conducive to high-volume manufacture, thin-film tantalum technology[3] was chosen for the physical realization.

One method of realizing a multiple-order active filter is cascading noninteracting lower-order sections. This realization consists of two stages. The first stage is a modification of a low-pass notch section (Fig. 2) developed by J. J. Friend.[10] It is a differential-input single operational-amplifier section. The second stage is a twin-T notch section[11] that has been modified by the addition of an RC network (Fig. 2) to include the pole on the negative real axis. Their respective transfer functions are

$$T_1(s) = \frac{s^2 + \omega_{z1}^2}{s^2 + \frac{\omega_{p1}}{Q_{p1}} s + \omega_{p1}^2} \tag{1}$$

and

$$T_2(s) = \frac{s^2 + \omega_{z2}^2}{\left(s^2 + \frac{\omega_{p2}}{Q_{p2}} s + \omega_{p2}^2\right)(s + \alpha)}. \tag{2}$$

The nominal performance of each section and the overall design is given in Fig. 3.[12]

The smoothing filter on the receiving side of the channel presents the designer with a difficult problem. It has a time-varying, rather than a time-invariant, generator impedance. This is caused by the JFET switch, which operates at an 8-kHz rate (125 μs) and is "on" for 3.5 μs. The switch working in conjunction with the input impedance of the network contributes an additional frequency-dependent gain characteristic to the filtering function of the network. A detailed description of the design of this filter is given in the next section.

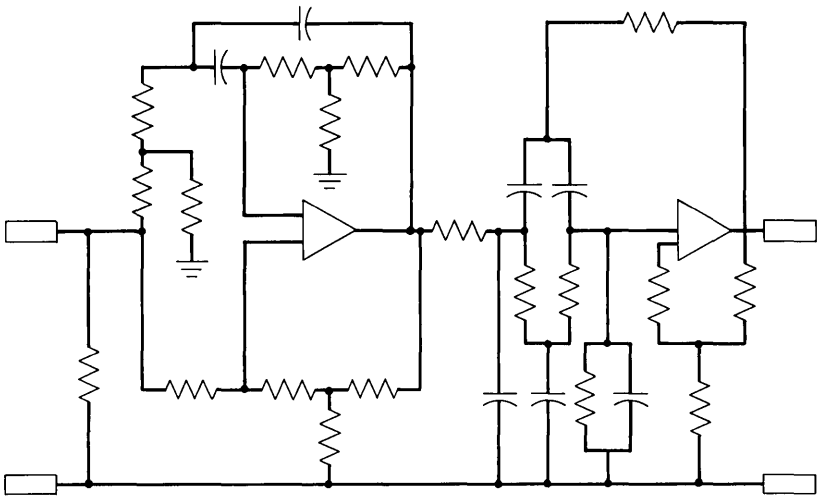

Fig. 2—Configuration of transmitting filter.

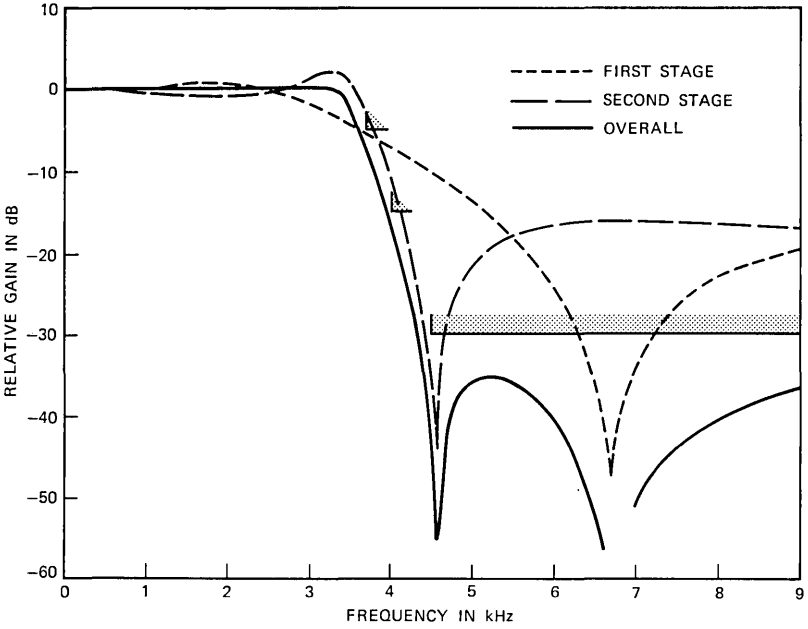

Fig. 3—Nominal performance of transmitting filter.

The design objectives of the receiving filter are as follows:*

(*i*) Passband ripple (200 Hz–3200 Hz) $\leqq \pm 0.16$ dB.
(*ii*) At 3300 Hz: 0.16 dB $\geqq$ gain $\geqq -0.60$ dB.
(*iii*) At 3400 Hz: 0 dB $\geqq$ gain $\geqq -1.20$ dB.
(*iv*) At 4000 Hz: gain < -15 dB.
(*v*) Above 5000 Hz: gain < -30 dB.
(*vi*) Gain at 1000 Hz: 4.75 dB $\pm$ 0.02 dB.

In addition, the filter must absorb the following manufacturing and environmental variations:

(*i*) Switch "on time": 3.1 μs to 3.7 μs.
(*ii*) Switch "on resistance": 50 Ω to 200 Ω.
(*iii*) Temperature: 0°C to 60°C.
(*iv*) Aging: 20-year life.

IV. INITIAL DESIGN OF THE RECEIVING FILTER

The configuration and nominal performance of the receiving filter are given in Figs. 4 and 5, respectively.[13,5] If the switch were not present, the requirements in Section III could be met by a fifth-order

* All gains and losses are relative to 1000 Hz.

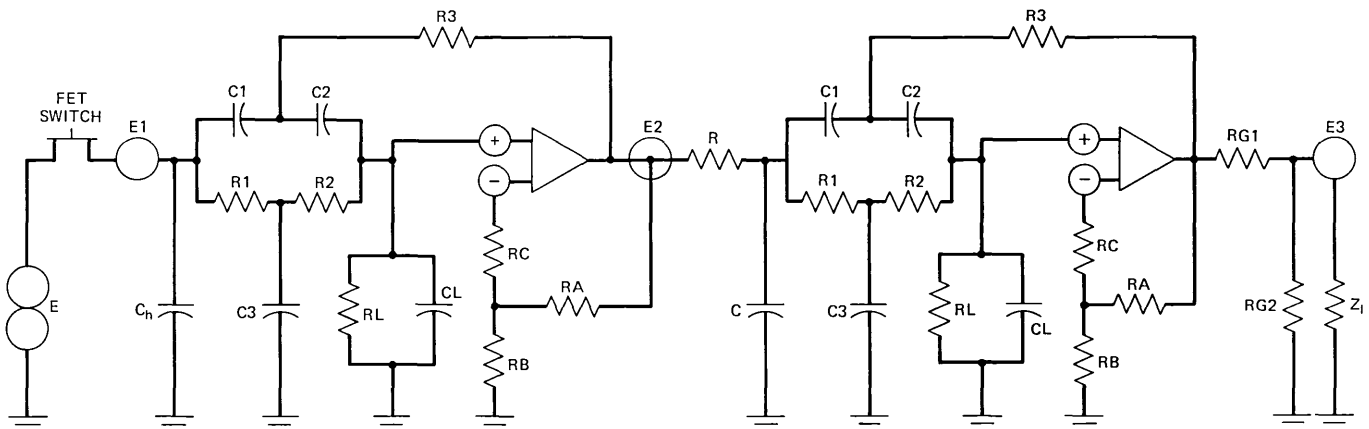

Fig. 4—Configuration of receiving filter.

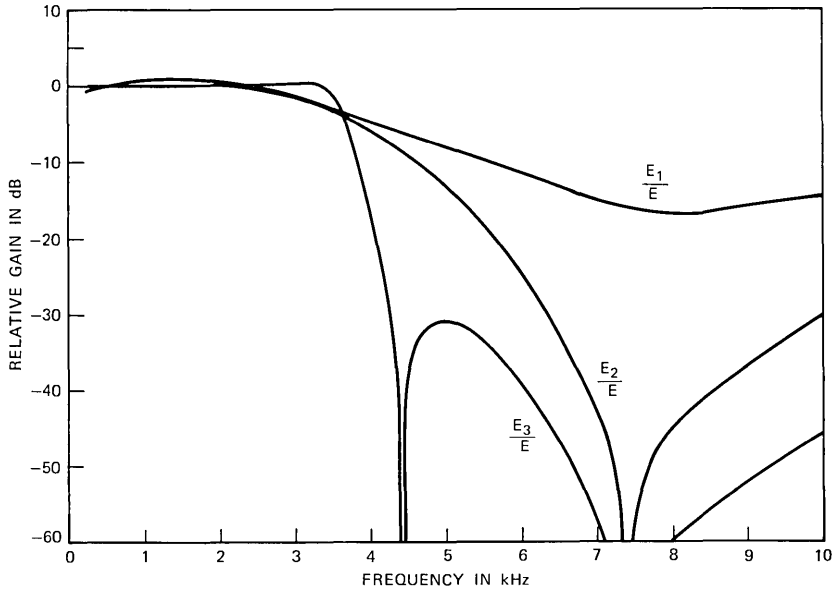

Fig. 5—Nominal performance of receiving filter.

Cauer-Chebyshev transfer function of the form

$$T(s)=\frac{V_{\text{out}}}{V_{\text{in}}}$$

$$=K\frac{(s^2+\omega_{z1}^2)(s^2+\omega_{z2}^2)}{[s^2+(\omega_{p1}/Q_{p1})s+\omega_{p1}^2][s^2+(\omega_{p2}/Q_{p2})s+\omega_{p2}^2](s+\alpha)}. \quad (3)$$

The first pole-zero pair can be realized with a second-order tuned twin-T.[14] The real pole at $s = -\alpha$ could be realized with a first-order section, although an operational amplifer can be saved by combining this pole with the second pole-zero pair to form a third-order twin-T section.

To maintain the gain level and reduce the sensitivity to switch "on time" and "on resistance" variations, the input stage must provide some "holding" function.[13] An ideal sample-and-hold network would accomplish this, but it would require an additional operational amplifier. Therefore, a holding capacitor was added to the input stage. When the switch is on (for a nominal 3.5 μs), the capacitor C_h is charged. During the off time (121.5 μs), the capacitor will slowly discharge through the network if the input impedance is at a high level (>100 kΩ). The switch working in conjunction with the high input impedance of the first stage acts as a lossy sample-and-hold network with a $(\sin x)/x$-type frequency-dependent gain character-

istic. This gain characteristic has minima at multiples of the 8-kHz switch rate (E_1/E in Fig. 5).

To design this filter, we could first assume that no switch or holding capacitor is present. An initial network would then be developed, to which the switch-and-holding capacitor would be added. Because of the roll-off introduced by the lossy sample and hold, further optimization would be required. Using a state-variable switched-filter analysis program in conjunction with a general-purpose optimization program,[4,5] the desired frequency characteristic is obtained (Fig. 6).

The program output gives the element values of the first stage of the filter (including the holding capacitor) and the transfer function coefficients of the second stage of the filter. To realize the second stage, the element values must be described as a function of the transfer function coefficients. The transfer function for a tuned, twin-T, second-order section with a preceding RC stage can be written as

$$T(s) = K \frac{s^2 + \omega_z^2}{[s^2 + (\omega_p/q_p)s + \omega_p^2](s + \alpha)} \tag{4}$$

or

$$T(s) = K \frac{s^2 + \omega_z^2}{s^3 + A_1's^2 + A_2's + A_3'}. \tag{5}$$

If we define

$$\begin{aligned} \beta &= 1 + (RA/RB), \\ c_1 &= (2CL/C1), \\ c_2 &= (CL/C), \\ r_1 &= (2R1/RL), \\ r_2 &= R/RL, \end{aligned} \tag{6}$$

and assume the twin-T is symmetrical and tuned, the coefficients of (5) can be written as

$$K = \frac{\beta}{RC(1 + c_1 + c_2)} \tag{7}$$

$$\omega_z^2 = \frac{1}{(R1\ C1)^2} \tag{8}$$

$$A_1' = \frac{1}{(1 + c_1 + c_2)} \left[\frac{1 + c_1}{RC} + \frac{2}{R1\ C1} \left(2 - \beta + \frac{r_1 + c_1}{2} \right) + \frac{2}{R1\ C} \left(2 - \beta + c_1 + \frac{r_1}{4} \right) \right] = \frac{\omega_p}{q_p} + \alpha \tag{9}$$

$$A_2' = \frac{1}{(1 + c_1 + c_2)} \left[\frac{1 + r_1 + c_2}{(R1\ C1)^2} + \frac{2}{RC\ R1\ R2} \left(2 - \beta + \frac{r_1 + c_1}{2} \right) + \frac{2}{R1^2\ C1\ C} (2 - \beta + r_1) \right] = \omega_p^2 + \frac{\omega_p}{q_p} \alpha \tag{10}$$

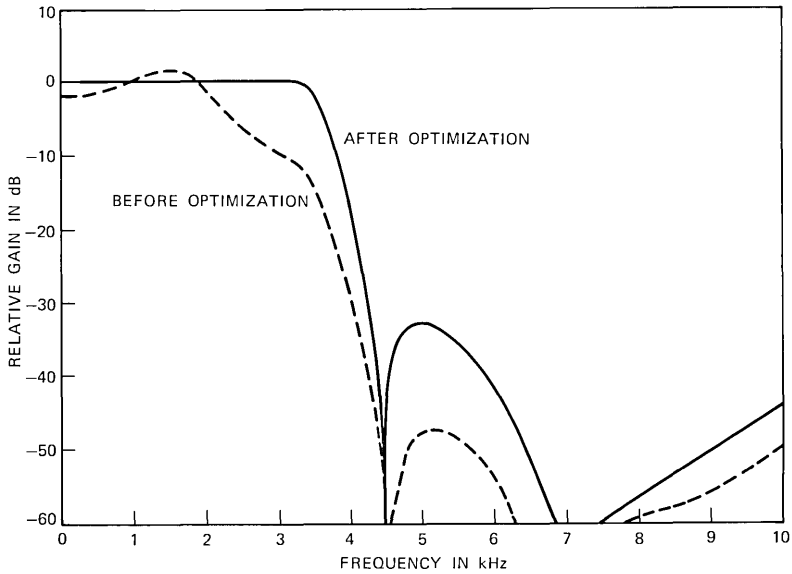

Fig. 6—Receiving filter before and after optimization.

and

$$A_3' = \frac{1}{RC}\left(\frac{1 + r_1 + r_2}{1 + c_1 + c_2}\right)\frac{1}{(R1\ C1)^2} = \omega_p^2\alpha. \tag{11}$$

Since there are more parameters (elements) than constraints (coefficients), the designer can optimize both the absolute values of the elements and the element value ratios, in addition to matching the required coefficients. An interactive computer program[15] was used to obtain the element values of the third-order twin-T.

V. SENSITIVITY CONSIDERATIONS

An important concern in the design of filters is the degradation of the frequency response because of environmental and element value changes. Some portions of this degradation can be controlled by specifying components with tight tolerances, or by carefully tuning the filter. Others can be controlled by matching the temperature coefficients of the components used. However, there are some uncontrollable changes, e.g., aging, adjustment tolerances, and measurement tolerances. The sensitivity of the filter design to these changes will determine whether a particular technology can be used in the realization, the amount of tuning required, the margin requirements of the design, the tolerance of the components, and, subsequently, the cost of the filter.

Many approaches can be taken for analyzing the sensitivity of the filter to element changes. One is the degradation of the frequency

response caused by all elements taken one at a time. This method is very useful in determining which elements require tight tolerances and whether tuning or tracking is required. Another is the variation of the frequency response under true manufacturing and environmental variations. This would involve all components varying according to some statistical description.

As an illustration of the first type of sensitivity, Fig. 7 shows the degradation of the frequency response of the receiving filter when each capacitor is increased in value by 1 percent. Although the response degrades significantly with these capacitor variations, the degradations tend to cancel if the capacitors vary in the same direction (track). In fact, as can be seen from eqs. (5) through (11), if the second-stage capacitors track perfectly, i.e., all capacitors increase by 1 percent, the only effect on the frequency response of that stage will be a shift in frequency.

Also, from eqs. (5) through (11), it can be shown that if, in addition, the resistors vary in the opposite direction from the capacitors, then even this degradation is cancelled. Another point evident from Fig. 7 is that even ± 1-percent capacitor variations would cause the response to miss the frequency requirements. Since thin-film capacitors can only be manufactured to ± 5-percent tolerances, it is evident that initial tuning of the network is necessary.

Fortunately, thin-film resistors and capacitors have opposite temperature and aging characteristics. Thus, once tuned, the frequency response is quite stable. However, since the tracking is not perfect, further analysis must be undertaken.[16] This is explained in Section VII.

VI. TUNING

Many elegant tuning procedures have been developed for hybrid integrated circuits.[17–19] Most rely on gain and/or phase measurements at a number of frequencies (one for each transfer function coefficient) and adjustment of element values in some algorithmic fashion. Usually, some form of descent algorithm, depending upon component sensitivity, with a number of iterations is used to solve this multiparameter optimization problem. Thin-film networks contain critical constraints. The capacitors cannot be adjusted and the resistor values can only be increased either by anodization[20] or laser trimming.[21]

For the above network, nine constraints must be satisfied. They are the zero frequencies of each stage, the frequency of the real pole of the second stage, the frequencies and q's of the two complex poles, and the 1-kHz flat gain. In addition, the manufacturing deviations of the switch "on resistance" and the holding capacitor must be compen-

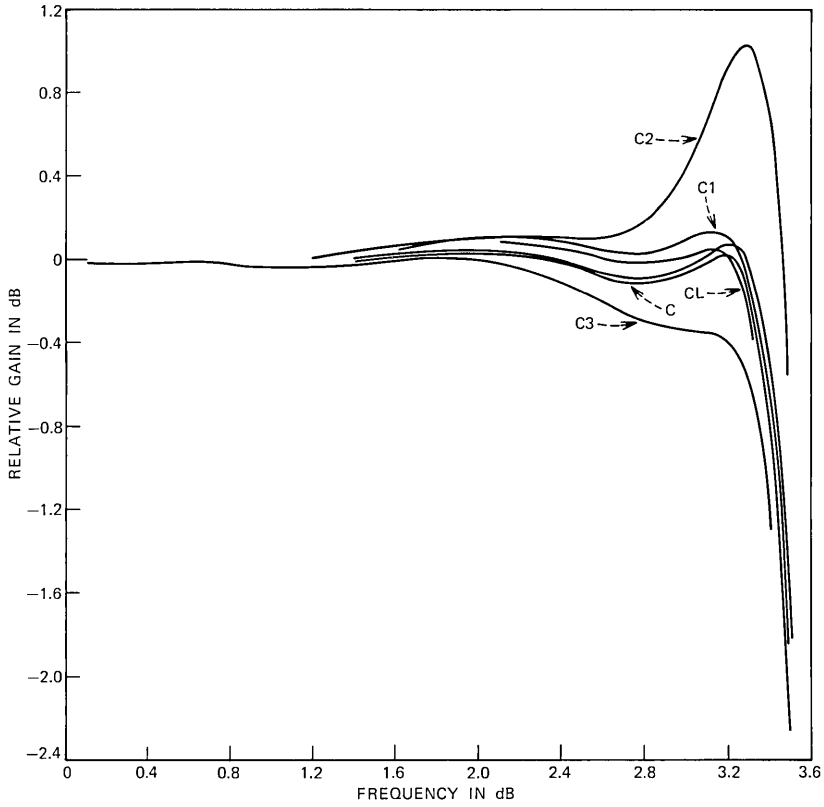

Fig. 7—Sensitivity of frequency response to capacitor changes of +1 percent.

sated for. Consultations with manufacturing engineers indicated that most iterative approaches that involved measurements of the performance at a number of frequencies were impractical from a throughput viewpoint, i.e., they took too much time.

From the manufacturing viewpoint, two points are critical: (*i*) the adjustment sequence must be fast and (*ii*) the adjusted network must meet the given frequency requirements. To accomplish both these aims, the tuning procedure shown in Fig. 8 was developed. It consists of two parts. The first is parametric; that is, it depends only upon component measurements and component adjustments. The second is functional; the gain is measured at *one* frequency for each stage, and an adjustment is made to bring the gain at that frequency to its nominal value. Finally, the voltage divider network at the output is used to trim the absolute through gain at 1 kHz to a predetermined level.

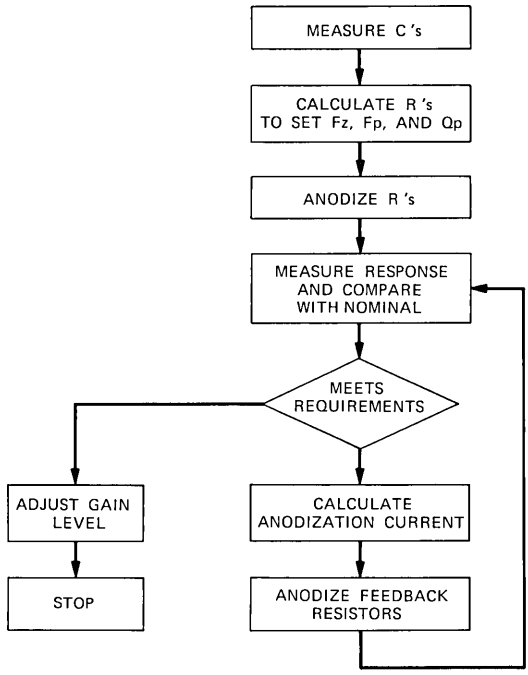

Fig. 8—Tuning procedure.

To tune the first stage of the filter, the switch-and-holding capacitor are initially neglected. The remaining part is a second-order twin-T notch filter whose transfer function is given in the appendix. Since the capacitors $C1$, $C2$, $C3$, and CL cannot be adjusted, their values are measured. To obtain a null at the prescribed frequency, eqs. (22), (25), and (27) dictate that

$$\frac{1}{(C1 + C2)R3} = \frac{R1 + R2}{R1\ R2\ C3}, \tag{12}$$

$$\omega_z^2 = \frac{1}{R1\ R2\ C3\ CS}. \tag{13}$$

Since the capacitor values are known, the constraints given by (12) and (13) can be satisfied if we pick

$$R1 = R2 = \frac{1}{\omega_z(CS\ C3)^{\frac{1}{2}}} \tag{14}$$

and

$$R3 = \frac{R1}{2}\frac{C3}{C1 + C2}. \tag{15}$$

To set the pole frequency, eqs. (22) and (32) can be used. Thus,

$$\omega_p^2 = \frac{1 + (2R1/RL)}{1 + (CL/CS)}\,\omega_z^2. \tag{16}$$

Since all the variables but RL of (16) are known, RL can be used to set ω_p. Rearranging (16),

$$RL = \frac{2R1}{[1 + (CL/CS)](\omega_p/\omega_z)^2 - 1}. \tag{17}$$

At this point, we have calculated the resistor values necessary to set the null, null frequency, and pole frequency of (28). To set the pole q, eq. (30) can be used. It is repeated here for convenience.

$$a_1 = \frac{\omega_p}{q_p} = \left[\frac{1-\beta}{R3\,C1} + \frac{1}{R2\,CS} + \frac{2CL}{R1\,C3\,CS} + \frac{1}{RL\,CS}\right] \Big/ \left(1 + \frac{CL}{CS}\right). \tag{18}$$

At this point, all the capacitors have been measured and $R1$, $R2$, $R3$, and RL have been adjusted. Thus, the only additional adjustment that can be made is on β, which is controlled by negative feedback resistors RA and RB. In fact,

$$\beta = 1 + \frac{RA}{RB}. \tag{19}$$

Solving (18) for β, we have

$$\beta = 1 + R3\,C1\left\{\frac{1}{CS}\left[\frac{1}{R2} + \frac{2CL}{R1\,C3} + \frac{1}{RL}\right] - \frac{\omega_p}{q_p}\left(1 + \frac{CL}{CS}\right)\right\}. \tag{20}$$

To adjust β to coincide with that calculated by (20), RA and RB are adjusted.

Because we have neglected the holding capacitor and switch "on resistance," it is necessary to functionally adjust this stage. This adjustment is done at 2100 Hz, where the desired nominal gain with respect to 1000 Hz is known. Through sensitivity studies, it was found that the gain deviation from nominal at 2100 Hz was linearly dependent upon the ratio of the β resistors, RA and RB. Hence, either RA or RB is adjusted to complete the tuning of the first stage.

Since there is interaction between the real pole, the complex pole, and the transmission zero, the tuning of the second section is more complicated. With $C1$, $C2$, and $C3$ (Fig. 5) measured and the zero frequency ω_z given, the twin-T resistors $R1$, $R2$, and $R3$ can be determined. With C and CL measured and α, ω_p, and q_p given, eqs. (9) through (11) are a nonlinear set of equations in terms of β, R, and RL. Using the nominal values of these parameters as starting guesses, the

new values which set α, ω_p, and q_p to their design values are easily found.

After anodizing the resistors R, $R1$, $R2$, $R3$, RL, RA, and RB to their calculated values, a functional adjustment is performed at 3300 Hz where the nominal gain value is known. This adjustment mops up the previous capacitor measurement and resistor adjustment errors and operational-amplifier variations. As in the first stage, the resistor ratio of the negative feedback network is touched up in a functional adjustment.

Finally, the gain at 1000 Hz must be adjusted to within ± 0.02 dB. To increase the output level, $RG2$ is increased. To decrease the level, $RG1$ is increased.

VII. TOLERANCE ANALYSIS

The network designer's work would be cut significantly if inexpensive components could be manufactured to their nominal values and have no variations with time and temperature. However, manufacturing processes are such that the designer must consider variations in the component values and characteristics from sample to sample. He could do a worst-case analysis in which he would specify tolerances that would ensure that, under all component combinations, the final manufactured network would meet specifications. In general, this would result in specifying tighter tolerances than needed and necessarily increase the cost of the network.

To get a realistic estimate of the performance of the filter as it left the manufacturing facility and a good prediction of its field performance, a statistical simulation of the manufacturing process and environmental behavior was developed. Included in this model were manufactured element variations, measurement errors, adjustment errors, temperature and aging characteristics of the components, and switch timing variations. The variables were the tolerances and distributions associated with the above errors, the adjustment procedure, and the adjustment frequencies. The figure of merit was the highest possible end-of-life yield at the lowest possible cost. Figure 9 is a flowchart for this simulation.

The simulation proceeds in the following manner. First, a set of random numbers is generated for a network. Next, the manufactured capacitor and JFET "on resistance" values are calculated. For the thin-film capacitors, the absolute tolerance and the tracking tolerances are included. The switch "on resistance" has a nominal value of 125 ohms and can vary between 50 and 200 ohms. The next step is a simulation of the tuning procedure, where the capacitors are first measured, the resistor values calculated, and the resistors adjusted.

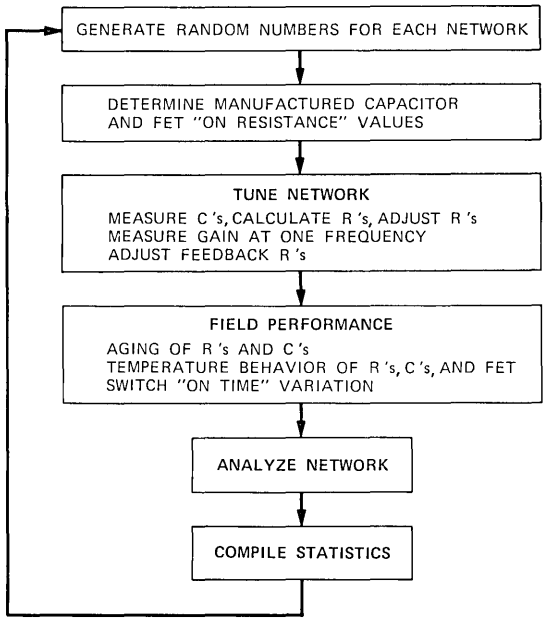

Fig. 9—Flow chart of tolerance analysis program.

For this parametric adjustment, the errors in capacitor measurement and the anodization errors for the resistors are included. During the functional touch-up adjustment, gain measurement errors are simulated. To evaluate the field performance, we use the temperature and aging characteristics of the components.

Finally the network is analyzed for the final element values that are determined by adding the temperature and aging deviations to the manufactured element values. Then the network performance is compared with the specifications at a number of frequencies, and statistics are compiled. If the requirements are not met at any one frequency, the network has failed. Next, a new network is picked, and we repeat the process.

The statistical performance of this filter is shown in Figs. 10 and 11. Only the minimum and maximum deviations at each frequency for 1000 sample networks are shown. At each frequency, there is a normal distribution of the networks' performance between the minimum and maximum deviations. At room temperature, 25°C, 94.5 percent of the samples fell within the design objectives, while at 60°C, 89.3 percent fell within. To ensure that the mathematical modeling was correct, these results were compared with the measured performance of the first 1000 filters manufactured. All those filters fell within the given

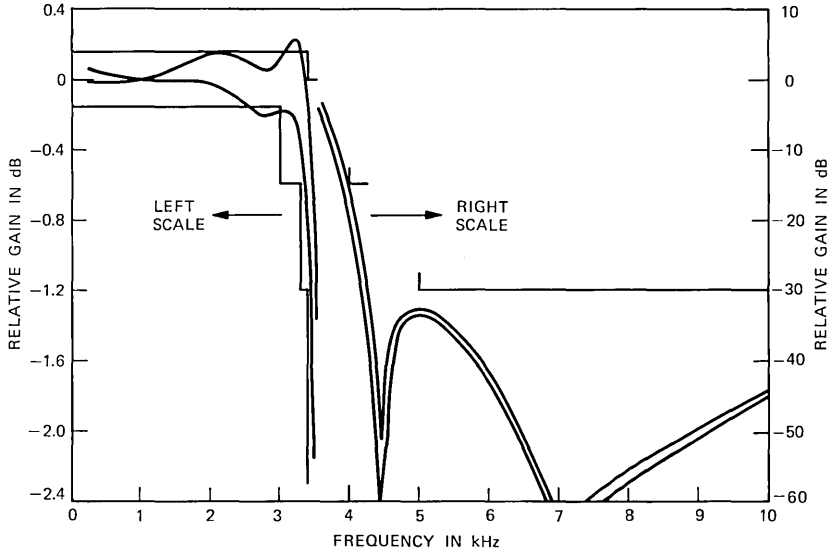

Fig. 10—Overall statistical performance of filter.

statistical bands. This type of extensive simulation is justified when we have a well-characterized technology, automated adjustment and testing, and high-volume manufacturing.

VIII. PRODUCTION EXPERIENCE

As previously mentioned, the tolerance analysis program has aided in obtaining a practical tuning procedure. If temperature and aging variations are ignored, the program predicts the filter yield at the end of the manufacturing process. The assigned component tolerances have a significant effect on manufacturing costs; thus, a trade-off takes place between manufacturing cost and filter yield. Ideally, we would like to find the set of component tolerances that minimizes manufacturing cost. This problem has been attacked by Karafin[22] and Pinel and Roberts.[23] However, their work has been restricted to networks where: (*i*) the performance criteria (i.e., gain or loss constraints at given frequencies) is met 100 percent of the time; (*ii*) tunable elements are banned, and (*iii*) there is no correlation between the elements.

In the case of the D3 filters, it is known that the 100-percent performance criteria restriction will not produce a minimum cost network. In other words, if the yield at final test drops from 100 to 96.7 percent by increasing the tolerance on the resistors during parametric adjustment, the cost drops by more than 3.3 percent. Thus far, cost minimization has been attacked in a heuristic manner. Figure 12 shows the

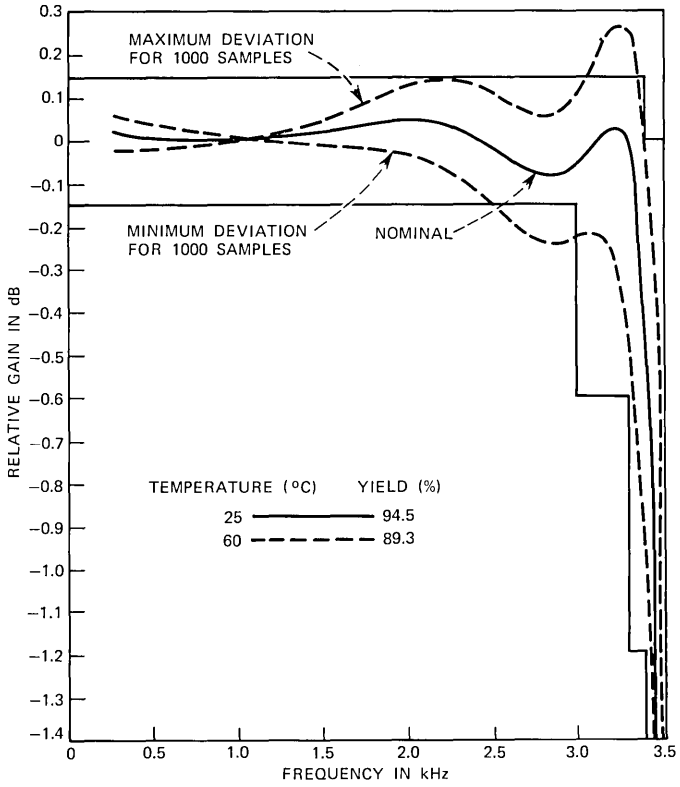

Fig. 11—Statistical performance of passband at temperature extreme.

theoretical manufacturing yields for the transmit filter at 0.1, 0.2, and 0.5-percent resistor tolerances. Current production experience indicates that widening the tolerance on certain resistors to 0.3 percent results in a lower cost, even though the overall yield decreases.

One might ask, Why not remove all tolerance restrictions on the parametric adjustments and accept any filters as long as they are functionally adjusted to meet the frequency requirements at a finite set of test points? The answer is quite simple. If all resistors were adjusted at the parametric step to, say, 0.5 percent tolerance, then only 50 percent of the filters could meet requirements after functional adjustment. Since the silicon is bonded after parametric adjustment, a much higher final yield is needed to justify the additional investment. In addition, as the tolerance is relaxed at the parametric adjustment step, more frequency tests are required at final test.

The tolerance analysis program also helped to provide a temporary solution to a capacitor tracking problem. If the ratio of certain capaci-

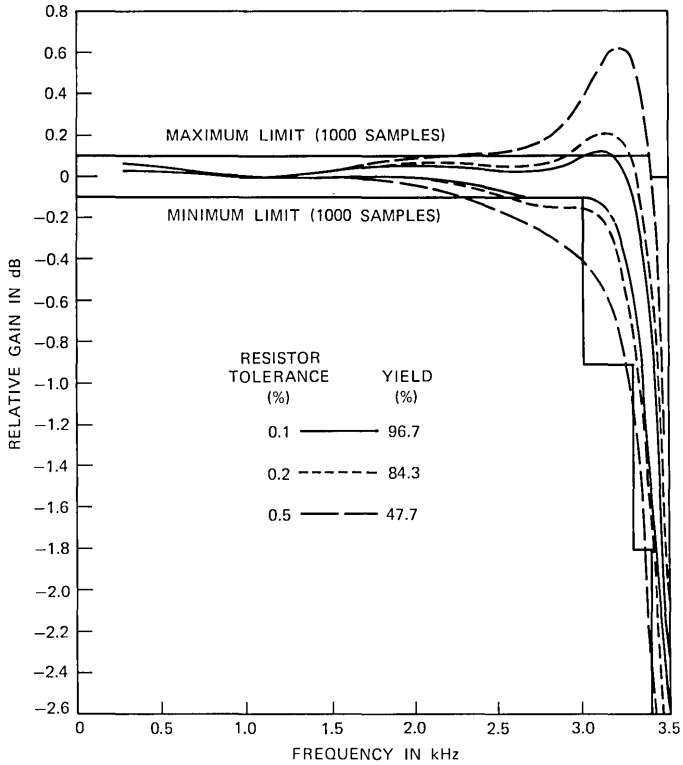

Fig. 12—Statistical performance of transmitting filter with different resistor tolerances.

tors is not within specified limits, then some resistors may have to be anodized excessive amounts—excessive because the resistors will then age poorly. However, the tracking requirement on all sets of capacitors need not be the same; thus, in a particular case, with the tolerance program, it was possible to demonstrate that the tracking requirement on a pair of capacitors could be relaxed and therefore solve a temporary production problem.

A tolerance analysis program can be used only if we have good estimates for the various tolerances that affect the manufacturing process. This was possible for the parameters that influence the frequency performance, but not possible for the dc gain requirement. The 1-kHz frequency gain is required to be adjusted to an accuracy of ± 0.02 dB. This stringent requirement was chosen so that no adjustments would be necessary when D3 channel banks were installed in the field. The no-adjustment philosophy saves on installation and maintenance cost, but it does require critical tuning for the D3 filters.[24]

IX. PHYSICAL REALIZATION

In this case, the physical realization uses single-substrate hybrid-integrated circuit (HIC) technology.[3] The resistors are tantalum nitride thin film, and the capacitors have a base electrode of β-tantalum, a dielectric of tantalum pentoxide, and a counter electrode of nichrome paladium gold. The conductor paths are also nichrome paladium gold, while the operational amplifiers and JFET switch are beam-leaded silicon integrated circuits. They are all placed on a 33-by-20 mm glazed ceramic substrate with a 34-terminal lead frame (Fig. 13).

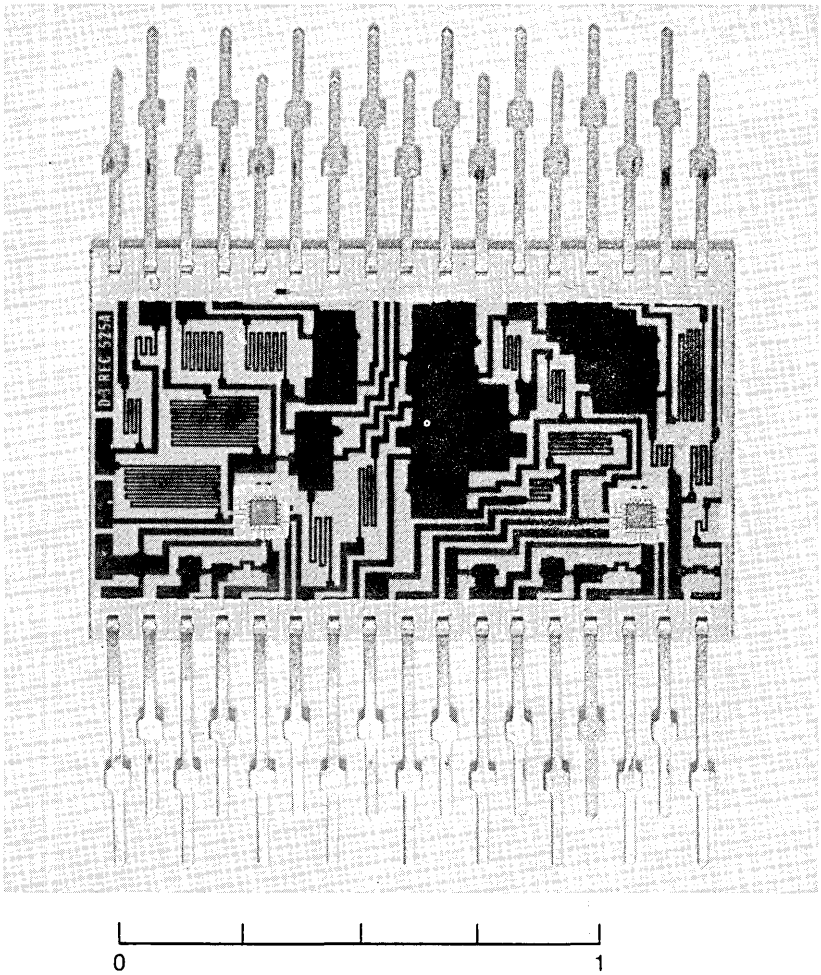

Fig. 13—Hybrid integrated circuit realization of filter.

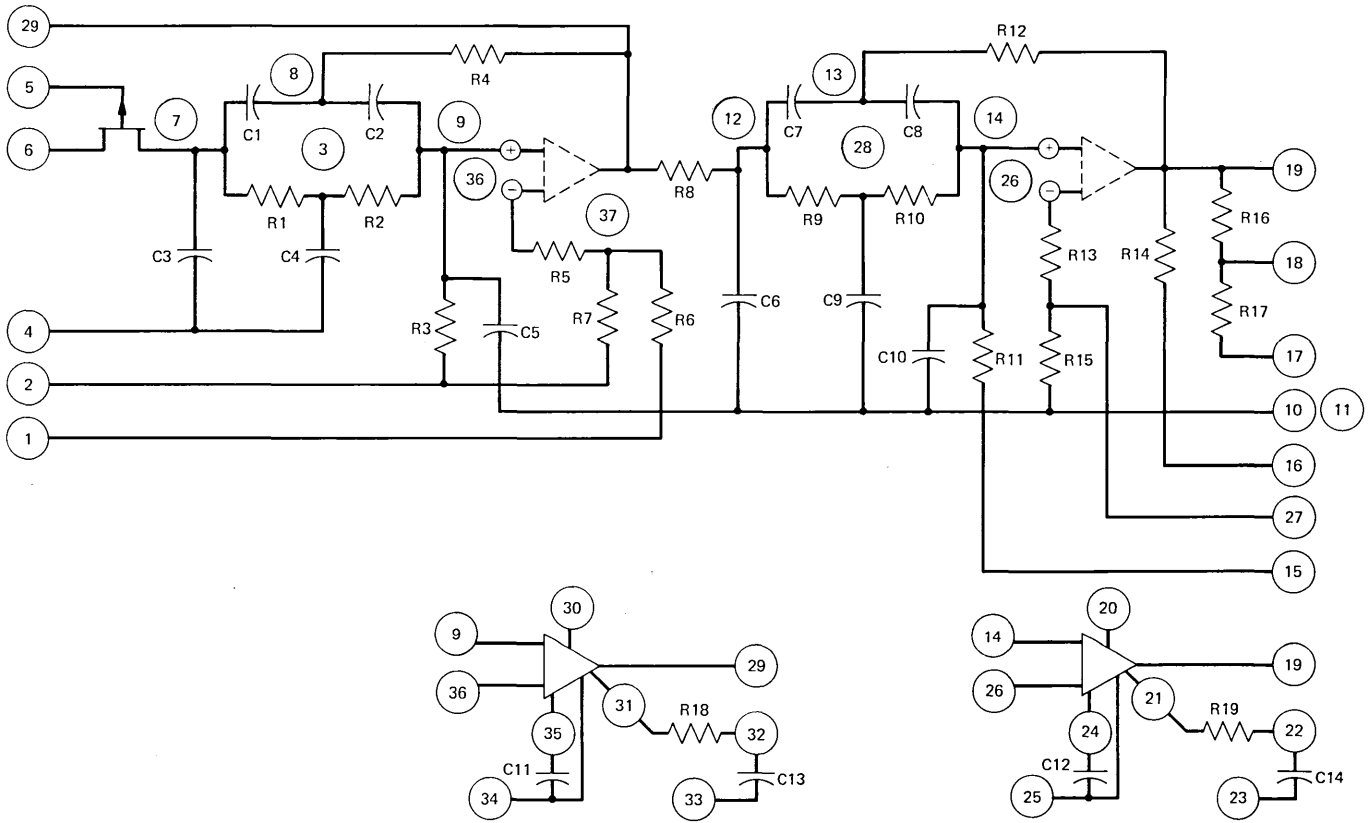

Fig. 14—Electrical layout of HIC realization.

To minimize both the tuning time and the complexity of the test fixtures, the substrate layout was constrained. Thus, all measurements are made from the sides and all adjustments from the top of the substrates. To incorporate the tuning procedure of Section VI, the layout must provide the capability of measuring all element values. Thus, all network nodes must be brought to the edge of the substrate.

Traditionally, layouts follow the flow of a network. Thus, for a network of this complexity, crossovers are generally required, as well as break points, to measure the component values. Crossover and break closures are normally processed following filter tuning. To eliminate breaks and crossovers and separate the measurement and adjustment functions, a unique circuit layout was developed. These aims were accomplished by judiciously separating common grounds and incorporating break points. These redundant points are brought to the substrate edge (Fig. 14) and are subsequently connected when the substrate is inserted into a printed wiring board. Three terminal capacitor measurements and capacitor electrode sharing were necessary to make these measurements. To realize this layout with its subsequent 11 mask levels required considerable reliance on computer-aided graphics.[25,26]

X. ACKNOWLEDGMENTS

The authors are grateful to W. Thelen for his work on the initial design, R. C. MacLean for his unique circuit layout, J. Novak and M. Fiore for building, tuning, and testing many models of the filters, C. F. Kurth, M. L. Liou, and F. J. Witt for many stimulating discussions, and numerous other colleagues who helped to bring this project to fruition.

APPENDIX

Transfer Function of Second-Order Twin-T Section

A.1 Untuned case

The transfer function of the untuned, unsymmetrical second-order twin-T is

$$T(s) = K\frac{(s+\tau_2)s^2 + (s+\tau_1)\omega_z^2}{(s+\tau_1)D(s) + \delta_\tau E(s)}, \tag{21}$$

with the following definitions:

$$\begin{aligned} CP &= C1 + C2, & CS &= (C1\ C2)/CP, \\ RS &= R1 + R2, & RP &= (R1\ R2)/RS, \\ \tau_1 &= 1/(R3\ CP), & \tau_2 &= 1/(RP\ C3), \\ \delta_\tau &= \tau_2 - \tau_1, & \beta &= 1 + (RA/RB), \\ c_1 &= CL/CS, & r_1 &= RS/RL. \end{aligned} \tag{22}$$

The polynomials $D(s)$ and $E(s)$ of (21) are

$$D(s) = s^2 + \left[\frac{1-\beta}{R3\,C1} + \frac{1}{R2\,CS} + c_1\tau_2 + \frac{1}{RL\,CS} \right] \frac{s}{(1+c_1)} + \left(\frac{1+r_1}{1+c_1} \right) \omega_z^2, \quad (23)$$

$$E(s) = \left[\frac{1}{1+c_1} s + \frac{(1-\beta)}{R3\,C1(1+c_1)} \right] s, \quad (24)$$

and

$$\omega_z^2 = \frac{1}{R1\,R2\,C3\,CS}, \quad (25)$$

$$K = \frac{\beta}{1+c_1}. \quad (26)$$

A.2 Tuned case

When the twin-T is tuned,

$$\tau_1 = \tau_2, \qquad \delta_\tau = 0. \quad (27)$$

Thus, (21) reduces to

$$T(s) = K \frac{s^2 + \omega_z^2}{s^2 + a_1 s + a_2} = K \frac{s^2 + \omega_z^2}{s^2 + (\omega_p/q_p)s + \omega_p^2}. \quad (28)$$

If, in addition, a symmetric twin-T is picked, i.e.,

$$\begin{aligned} R1 &= R2 = 2R3 \\ C1 &= C2 = C3/2 \\ CL &= (C1\,c_1)/2 \\ RL &= (2R1)/r_1, \end{aligned} \quad (29)$$

then

$$a_1 = \frac{2}{R1\,C1} \left(2 - \beta + \frac{r_1 + c_1}{2} \right) \frac{1}{1+c_1}, \quad (30)$$

$$a_2 = \left(\frac{1+r_1}{1+c_1} \right) \omega_z^2, \quad (31)$$

and

$$\omega_p = \left(\frac{1+r_1}{1+c_1} \right)^{\frac{1}{2}} \omega_z, \quad (32)$$

$$q_p = \frac{[(1+r_1)(1+c_1)]^{\frac{1}{2}}}{2[2 - \beta + (r_1 + c_1)/2]}. \quad (33)$$

REFERENCES

1. W. B. Gaunt and J. B. Evans, Jr., "The D3 Channel Bank," Bell Laboratories Record, *50*, No. 7 (August 1972), pp. 229–233.
2. R. A. Friedenson, "Active Filters Make it Small in the D3 Channel Bank," Bell Laboratories Record, *51*, No. 4 (April 1973), pp. 104–111.
3. W. H. Orr et al., "Integrated Tantalum Film RC Circuits," Proc. 20th Elec. Components Conf., May 1970, pp. 602–612.
4. C. Pottle, "State-Space Techniques for General Active Network Analysis," in *System Analysis by Digital Computer*, F. F. Kuo and J. F. Kaiser, eds., New York: John Wiley, 1966, Ch. 3.
5. M. L. Liou, "Exact Analysis of Linear Circuits Containing Periodically Operated Switches with Applications," IEEE Trans. Circuit Theory, *CT-19*, No. 2 (March 1972), pp. 146–154.
6. P. E. Fleischer, "Optimization Techniques in System Design," in *System Analysis by Digital Computer*, F. F. Kuo and J. F. Kaiser, eds., New York: John Wiley, 1966, Ch. 6.
7. A. C. Dickieson, J. Chernak, et al., "Statistical Circuit Design," B.S.T.J., *50*, No. 4 (April 1971), pp. 1099–1310.
8. D. F. Hoth, "The T1 Carrier System," Bell Laboratories Record, *40*, No. 10 (November 1962), pp. 358–363.
9. K. E. Fultz and D. R. Penick, "The T1 Carrier System," B.S.T.J., *44*, No. 7 (September 1965), pp. 1405–1452.
10. J. J. Friend, "A Single Op-Amp Biquadratic Filter Section," Proc. 1970 IEEE International Symposium on Circuit Theory, Atlanta, Georgia, December 1970.
11. G. S. Moschytz, "FEN Filter Design Using Tantalum and Silicon Integrated Circuits," Proc. IEEE, *58*, April 1970, pp. 550–556.
12. R. A. Friedenson, "Computer-aided Design . . . A Network Designer's Viewpoint," IEEE Computer-Aided Design Short Course, *72SC-02*, March, April 1972, pp. 18–42.
13. M. L. Liou, P. H. McDonald, and W. Thelen, "Computer Optimization of Active and Passive Switched Low-Pass Filters for PCM Systems." Proc. 20th Electronic Components Conf., May 1970, pp. 301–306.
14. G. S. Moschytz and W. Thelen, "Design of Hybrid Integrated Circuit Building Blocks," IEEE J. Solid-State Circuits, *SC-5*, No. 2 (June 1970), pp. 99–107.
15. R. L. Adams, private communication.
16. D. Hilberman, "An Approach to the Sensitivity and Statistical Variability of Biquadratic Filters," IEEE Trans. Circuit Theory, *CT-20*, No. 4 (July 1973), pp. 382–390.
17. G. S. Moschytz, "Two-step Precision Tuning of Twin-T Notch Filter," Proc. IEEE, *54*, No. 5 (May 1966), pp. 811–812.
18. G. S. Moschytz, "A General Approach to Twin-T Design and Its Application to Hybrid Integrated Linear Active Networks," B.S.T.J., *49*, No. 6 (July-August 1970), pp. 1105–1149.
19. W. H. Orr, "The Computer Design and Precision Tuning of Thin Film Filters," Proc. 3rd Annu. Seminar on Integrated Circuits, New York Chapter IEEE Basic Sciences Division, February 1966, pp. 45–56.
20. D. A. McLean, M. Schwarz, and E. D. Tidd, "Tantalum-film Technology," Proc. IEEE, *52*, No. 12 (December 1964), pp. 1450–1462.
21. M. J. Cohen, B. A. Unger, and J. F. Milkosky, "Laser Machining of Thin Films and Integrated Circuits," B.S.T.J., *47*, No. 3 (March 1968), pp. 305–405.
22. B. J. Karafin, "The Optimum Assignment of Component Tolerances for Electrical Networks," B.S.T.J., *50*, No. 4 (April 1971), pp. 1225–1242.
23. J. F. Pinel and K. A. Roberts, "Tolerance Assignment in Linear Networks Using Nonlinear Programming," IEEE Trans. Circuit Theory, *CT-19*, No. 5 (September 1972), pp. 475–479.
24. J. Dupcak and R. H. DeGroot, "The Manufacture of Thin-Film Active Filters," The Western Electric Engineer, *18*, No. 3 (July 1974), pp. 18–25.
25. Fowler, "XYMASK," Bell Laboratories Record, *47*, No. 7 (June-July 1969), pp. 204–209.
26. J. J. Degan, "Focus on Efficiency: Computer Graphics for HIC Deisgn," Bell Laboratories Record, *52*, No. 9 (October 1974), pp. 286–292.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

FIR Digital Filter Banks for Speech Analysis

By R. W. SCHAFER, L. R. RABINER, and O. HERRMANN

(Manuscript received August 9, 1974)

In using filter banks for processing speech signals, it is often important that the sum of the individual frequency responses of the bandpass filters (composite response) be flat with linear phase. This paper presents a technique for achieving flat composite response using linear-phase FIR digital filters. The design method is based on some special properties of FIR filters designed by the windowing method. Excellent response characteristics can be achieved with complete flexibility in choosing the center frequencies and bandwidths of the individual filters.

I. INTRODUCTION

Filter banks are used to perform short-time spectrum analysis in a variety of speech processing systems.[1–4] Typically, a set of bandpass filters is designed so that a desired portion of the speech band is entirely covered by the combined passbands of the filters composing the filter bank. The outputs of the bandpass filters therefore are considered to be a time-varying spectrum representation of the speech signal. If special care is taken in the design of the bandpass filters, it is possible to reconstruct a very good approximation to the input speech by simply adding together the outputs of the bandpass filters.[5] This is the basic principle of a variety of vocoder systems.

Since the bandpass filters are linear systems, we can characterize the behavior of such filter banks by considering the composite frequency response when all the outputs are added together. Since, ideally, the output should be equal to the input, then we desire that the composite frequency response have constant magnitude and linear phase in the desired band of frequencies. This criterion, together with specifications on the desired bandwidths of the individual frequency channels, forms a meaningful basis for the design of filter banks for speech analysis.

An earlier paper[5] showed that careful attention to the relative phases between channels is important in achieving a flat composite frequency response. That paper, which was concerned primarily with filter banks composed of infinite impulse response (IIR) digital filters, described a method of obtaining flat composite frequency response by a relatively

simple adjustment of the relative phases of the channels. This method was later applied to the design of a speech analysis/synthesis scheme in which finite impulse response (FIR) digital filters were used.[3] Using this method, excellent overall response can be obtained for both IIR and FIR digital filters in filter banks in which the center frequencies are *uniformly* spaced. However, the method is not easily extended to nonuniformly spaced filter banks.

In the present paper, we describe a different approach that is not limited to the design of uniformly spaced filter banks. The method exploits some special properties of linear-phase FIR filters and thus cannot be applied very successfully to the design of IIR filter banks. We first discuss the basic design principles, and then show some design examples. We conclude with a discussion of some computational considerations of FIR digital filter banks.

II. DESIGN METHOD

FIR digital filters are attractive for design of speech filter banks for several reasons. First, such filters can be designed to have precisely linear phase simply by imposing the constraint

$$h(n) = h(N - 1 - n) \qquad 0 \leqq n \leqq N - 1 \tag{1}$$

(on each individual filter band*), where $h(n)$ is the impulse response of the filter and N is its length in samples. This means that the criterion of linear phase for the composite filter bank response is trivially met if the individual filters have identical linear-phase characteristics. Therefore, it is possible to focus attention on achieving arbitrary frequency selective properties for the individual filters and on obtaining the desired flat response for the composite filter bank. The second great advantage of FIR filters is that a variety of design methods exist ranging from the straightforward windowing method[6,7] to iterative approximation methods that allow great flexibility in realizing complicated design specifications.[8]

2.1 FIR bandpass filters

The bandpass filters that we shall consider have impulse responses of the form

$$\begin{aligned} h_k(n) &= h_{lk}(n) \cos(\omega_{ck} nT) \qquad 0 \leqq n \leqq N - 1 \\ &= 0, \qquad\qquad\qquad\qquad \text{otherwise,} \end{aligned} \tag{2}$$

where $h_{lk}(n)$ is the impulse response of the kth linear-phase low-pass

* It is assumed, for simplicity, that the impulse response of each bandpass filter is of duration N samples, although it is trivial to remove this restriction by adding appropriate delays for each channel.

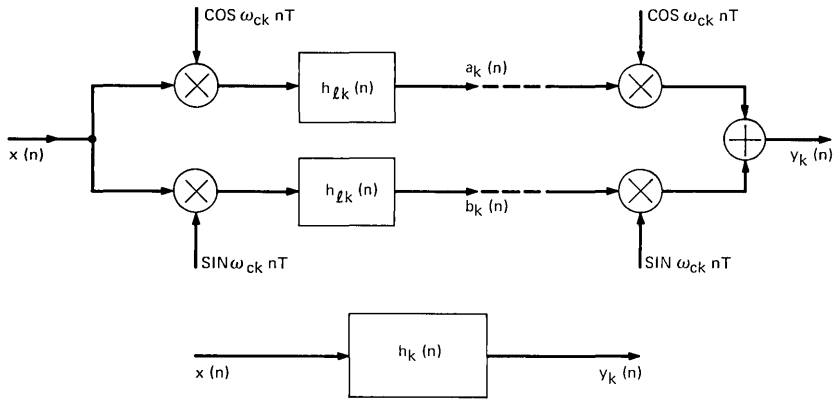

Fig. 1—Implementation of a typical bandpass channel.

filter. This particular form for the impulse response is motivated by the fact that, in some vocoder applications,[2,4] each bandpass channel is implemented as shown in Fig. 1. The overall impulse response of the system of Fig. 1 from input $x(n)$ to output $y(n)$ is easily shown to be given by eq. (2).

The spacing of the individual channels of the filter bank is determined by the choice of the set of center frequencies, ω_{ck}, which is in turn determined by the desired frequency resolution of the filter bank. The frequency selectivity of each channel is determined by the frequency response characteristics of the prototype low-pass filters $h_{lk}(n)$. Since phase considerations can be simply avoided by designing all the bandpass filters to have the same linear phase, we can focus our attention entirely on designing a set of prototype low-pass filters that have the desired individual frequency selective properties and that give the flattest amplitude response for the composite set of bandpass filters.

2.2 Low-pass filter design

The window design method appears to have a number of advantages for design of the prototype low-pass FIR filters. This method is depicted in Fig. 2. First, a desired ideal low-pass filter of the form

$$\begin{aligned} H_{dk}(e^{j\omega T}) &= e^{-j\omega n_0 T} \qquad |\omega| \leqq \omega_{pk} \\ &= 0, \qquad\quad \text{otherwise,} \end{aligned} \tag{3}$$

is defined by choosing the cutoff frequency ω_{pk}. Note that, for simplicity, we have omitted in the figure the linear phase term $\exp(-j\omega n_0 T)$ corresponding to a delay of n_0 samples. The value of n_0 required is $n_0 = (N-1)/2$. This means that, if N is even, the

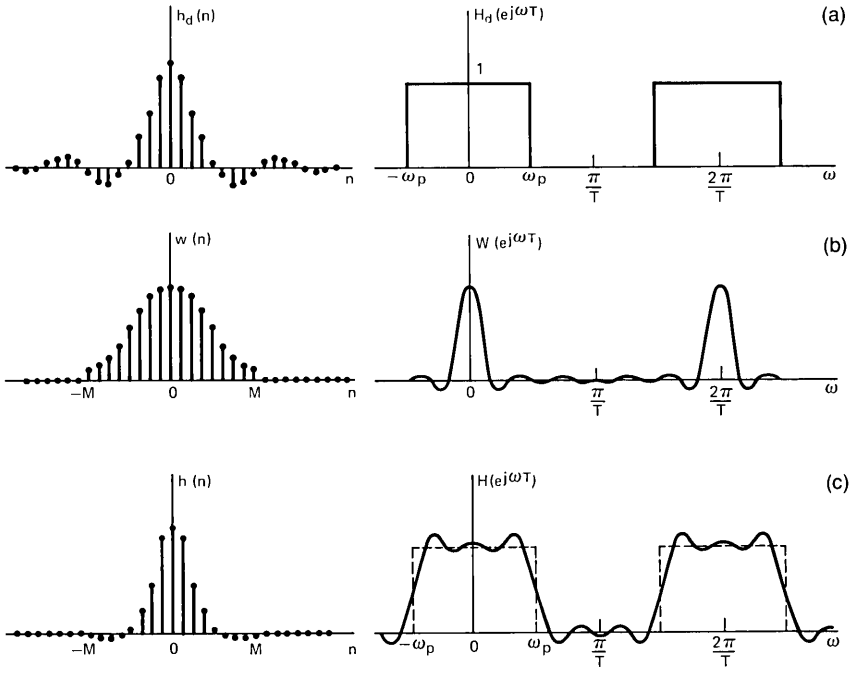

Fig. 2—Windowing technique for a low-pass design.

delay corresponds to a noninteger number of samples. The ideal impulse response for the kth channel is, therefore,

$$h_{dk}(n) = \frac{1}{2\pi}\int_{-\omega_{pk}}^{\omega_{pk}} e^{-j\omega n_0 T} e^{j\omega n T} d\omega = \frac{\sin\left[\omega_{pk}(nT - n_0 T)\right]}{\pi(n - n_0)}. \tag{4}$$

Of course, this impulse response is infinite in extent and must be truncated to obtain an FIR filter. This is done by defining

$$h_{lk}(n) = w(n - n_0)h_{dk}(n), \tag{5}$$

where $w(n)$ is a window function and $h_{lk}(n)$ is the impulse response of the kth prototype low-pass filter. The length of the window, denoted by N, can be either an even integer ($N = 2M$) or an odd integer ($N = 2M + 1$). Figure 2 shows the case when N is odd.

The result of multiplying the ideal low-pass impulse response by the window corresponds to a convolution in the frequency domain of the ideal frequency response and the Fourier transform, $W(e^{j\omega T})$, of the window; i.e.,

$$H_k(e^{j\omega T}) = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} H_{dk}(e^{j\theta T}) W(e^{j(\omega-\theta)T}) d\theta. \tag{6}$$

The result of this convolution is depicted in Fig. 2. It can be seen that

the main effects are the introduction of a smooth transition between the passband and the stopband and the introduction of ripples in the passband and stopband regions. The properties of this approximation are depicted in Fig. 3. If ω_p is larger than the width of the "main lobe" of $W(e^{j\omega T})$, then the following set of properties are generally true:

(*i*) The transition region, $\Delta\omega$, is inversely proportional to N.
(*ii*) The function $H(e^{j\omega T})$ is very nearly antisymmetric about the point $(\omega_p, 0.5)$.
(*iii*) The peak approximation errors in the passband and stopband are very nearly equal.
(*iv*) The approximation error is greatest in the vicinity of ω_p, and it decreases for values of ω away from ω_p.

The above properties of the windowing design method are true of all the commonly used windows. However, Kaiser has proposed a family of window functions that are very flexible and nearly optimum for filter design purposes.[6] Specifically, the Kaiser window is

$$\begin{aligned} w(n) &= \frac{I_0[\alpha\sqrt{1-(n/n_0)^2}]}{I_0(\alpha)} && |n| \leqq n_0 \\ &= 0, && \text{otherwise,} \end{aligned} \tag{7}$$

where $n_0 = (N-1)/2$ and $I_0[\cdot]$ is the modified zeroth-order Bessel function of the first kind. By adjusting the parameter α, one can trade off between transition width and peak approximation error. Furthermore, Kaiser[7] has formalized the window design procedure by giving

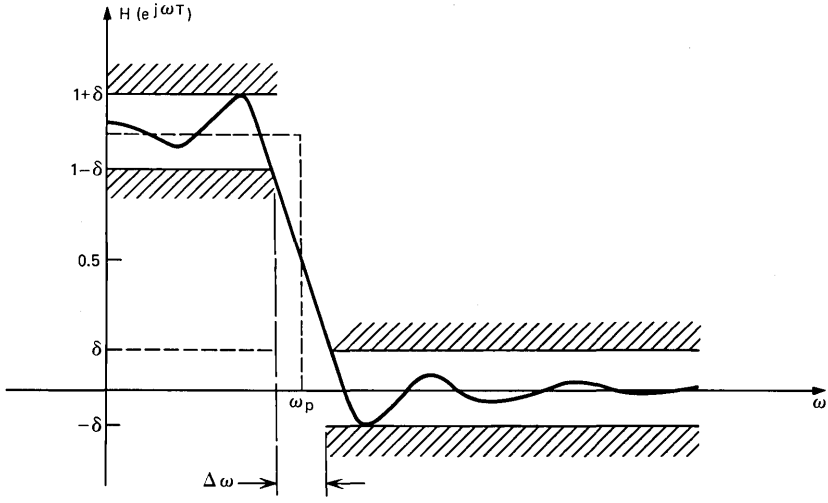

Fig. 3—Resulting low-pass design from windowing.

the empirical design formula

$$N = \frac{-20 \log_{10} \delta - 7.95}{14.36 \Delta f} + 1, \tag{8a}$$

where N is the filter order, δ is the peak approximation error, and Δf is the normalized transition width

$$\Delta f = \frac{\Delta \omega T}{2\pi}. \tag{8b}$$

To use this formula, we fix δ and Δf at values that provide the desired frequency selectivity. Then eq. (8a) can be used to compute N, and the parameter α can be computed from the equation[7]

$$\begin{aligned} \alpha &= 0.1102(-20 \log_{10} \delta - 8.7), && -20 \log_{10} \delta > 50 \\ &= 0.5842(-20 \log_{10} \delta - 21)^{0.4} \\ &\quad + 0.07886(-20 \log_{10} \delta - 21), && 21 < -20 \log_{10} \delta < 50. \end{aligned} \tag{9}$$

In the present application of this design method, the choice of δ and Δf depends upon the specifications of the bandpass filters that constitute the filter bank.

2.3 Filter bank design

To design a filter bank using FIR filters, we must first determine the range of frequencies to be covered by the composite response. Let us assume that these are denoted $\omega_{\min}$ and $\omega_{\max}$, where $\omega_{\max} \leqq \pi/T$. Now, if there are a total of N_f filters, we must choose the bandwidths and center frequencies so that the entire range of frequencies $\omega_{\min} \leqq \omega \leqq \omega_{\max}$ is covered. This is depicted in Fig. 4 for the case $N_f = 3$. This figure shows the ideal responses for each bandpass filter;

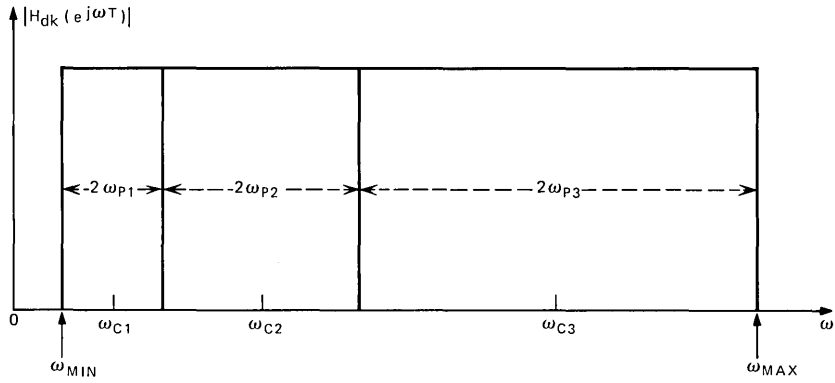

Fig. 4—A typical nonuniform filter bank.

i.e., as would be obtained if windowing were not required. In general, it is clear that

$$\omega_{\max} - \omega_{\min} = \sum_{k=1}^{N_f} 2\omega_{pk} \tag{10}$$

and

$$\begin{aligned} \omega_{ck} &= \omega_{\min} + \sum_{m=1}^{k-1} 2\omega_{pm} + \omega_{pk} \qquad k \geqq 2 \\ &= \omega_{\min} + \omega_{p1} \qquad k = 1. \end{aligned} \tag{11}$$

If all the filters have the same bandwidth, i.e., $\omega_{pk} = \omega_0$, then it is easily seen that

$$\omega_0 = \frac{\omega_{\max} - \omega_{\min}}{2N_f}. \tag{12}$$

Alternatively, if the bandwidths are to increase exponentially; e.g., $\omega_{pk} = 2^{k-1}\omega_0$, then

$$\omega_0 = \frac{\omega_{\max} - \omega_{\min}}{2(2^{N_f} - 1)}. \tag{13}$$

The center frequencies can be found in either case by using eq. (11).

The choice of peak approximation error depends upon how much stopband attenuation is deemed necessary in a given application. Typical values of $-20 \log_{10} \delta$ would most likely be between 40 and 60 dB. Using eq. (9), the appropriate value of α can be computed. Finally, the normalized transition width Δf must be fixed to compute N from eq. (8a). Again, the choice of $\Delta\omega$ (or Δf) is governed by consideration of the desired frequency selectivity for the individual filters. Clearly, the transition width $\Delta\omega_k$ should not be more than $2\omega_{pk}$.

In the filter bank context, we shall require that $\Delta\omega$ be the same for all filters so that we can take advantage of property (*ii*) of Section 2.2. That is, if all the filters have identical transition regions and, furthermore, if these transitions are antisymmetric about the crossover points, then we can expect that the sum of the frequency responses will be very close to unity. This is illustrated in Section III.

III. DESIGN EXAMPLES

In this section, we illustrate the use of the principles established in Section II with examples of both uniform and nonuniform filter banks. For all the examples, the sampling rate is assumed to be 9.6 kHz.

Example 1

Suppose that we wish to design a bank of 15 equally spaced filters that covers the range 200 to 3200 Hz. Then, using eq. (12), we find

that the cutoff frequency for all the low-pass filters* is

$$f_0 = \frac{\omega_0}{2\pi} = 100 \text{ Hz}.$$

Using eq. (11), the center frequencies are

$$f_{ck} = \frac{\omega_{ck}}{2\pi} = 100(2k + 1) \text{ Hz} \qquad k = 1, 2, \cdots, 15.$$

If we assume that 60-dB attenuation is required outside the transition regions of each channel, we find from eq. (9) that $\alpha = 5.65326$. Since the cutoff frequency is 100 Hz for all the prototype low-pass filters, the widest transition band that is reasonable is 200 Hz. Using this value and $-20 \log_{10} \delta = 60$ in eq. (8a), we obtain $N = 175$ as the lowest reasonable value for N. Note that, if lower attenuation is acceptable, then N can be smaller for the same Δf.

The filter bank designed with the above parameters is shown in Fig. 5. Figure 5a shows the individual bandpass filters. Note how the fall-off in the upper transition band of a given filter complements the ascent of the next filter. Also note that adjacent channels cross at an amplitude value of 0.5. Figure 5b shows the composite response of the filter bank. It is clear that the filters merge together very well at the edges of the frequency bands. Indeed, the deviation from unity is less than or equal to the peak approximation error, $\delta = 0.001$, that was used in designing the prototype low-pass filters.

Example 2

A nonuniform spacing of the filters is often used to exploit the ear's decreasing frequency resolution with increasing frequency. Suppose that we wish to cover the same range 200 to 3200 Hz as in Example 1, but we wish to use only four octave band filters. That is, each successive filter will have a bandwidth twice the bandwidth of the previous filter. Using eq. (13), we find that the lowest frequency channel has cutoff frequency

$$f_0 = \frac{\omega_0}{2\pi} = \frac{3200 - 200}{2(2^4 - 1)} = 100 \text{ Hz}.$$

In general, the cutoff frequencies of the prototype low-pass filters are

$$f_{pk} = \frac{\omega_{pk}}{2\pi} = 2^{k-1} f_0 \qquad k = 1, 2, 3, 4,$$

* For the actual low-pass filter, the response will be approximately 0.5 at $\omega = \omega_p$, the cutoff frequency of the ideal low-pass filter.

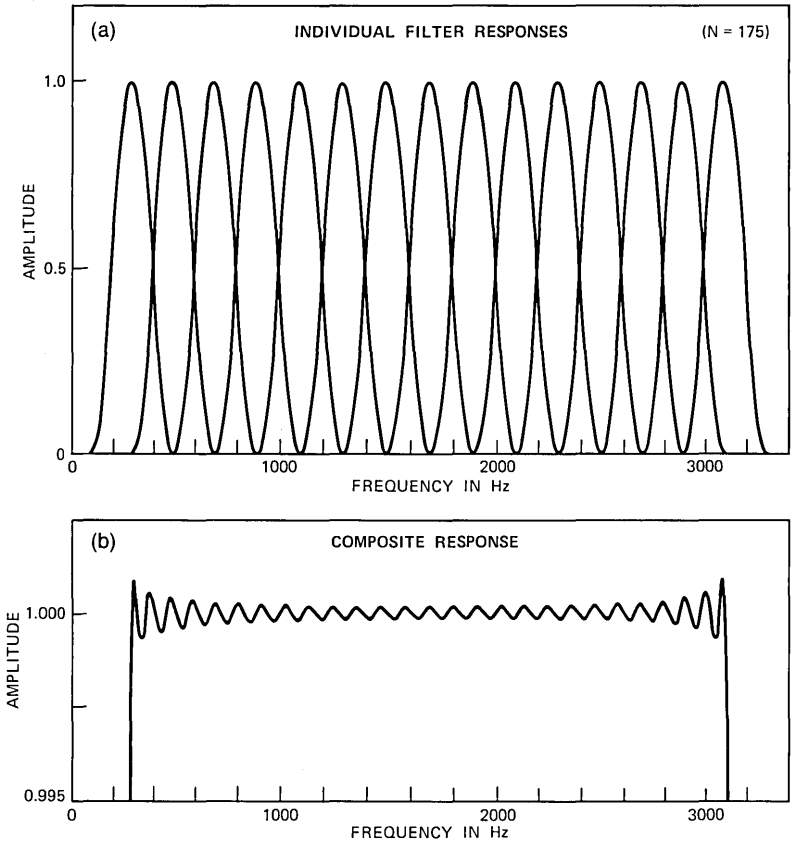

Fig. 5—Individual and composite frequency responses of a bank of 15 uniform bandpass filters for $N = 175$.

or the bandwidths of the bandpass filters are 200, 400, 800, and 1600 Hz, respectively. The center frequencies are found, from eq. (11), to be 300, 600, 1200, and 2400 Hz, respectively. Again requiring 60-dB attenuation, we note that the narrowest bandwidth is 100 Hz, so that the smallest reasonable transition width is 200 Hz. This leads again to a minimum value of $N = 175$. The filter bank corresponding to these design parameters is shown in Fig. 6. In Fig. 6a, again note the relationship between the ascending and descending transitions between adjacent filters. Particularly note that, since N and α are the same for each of the prototype low-pass designs, the shape of the curves in the transition region is independent of the bandwidth. Figure 6b shows the composite response where the deviation from unity is again less than 0.001.

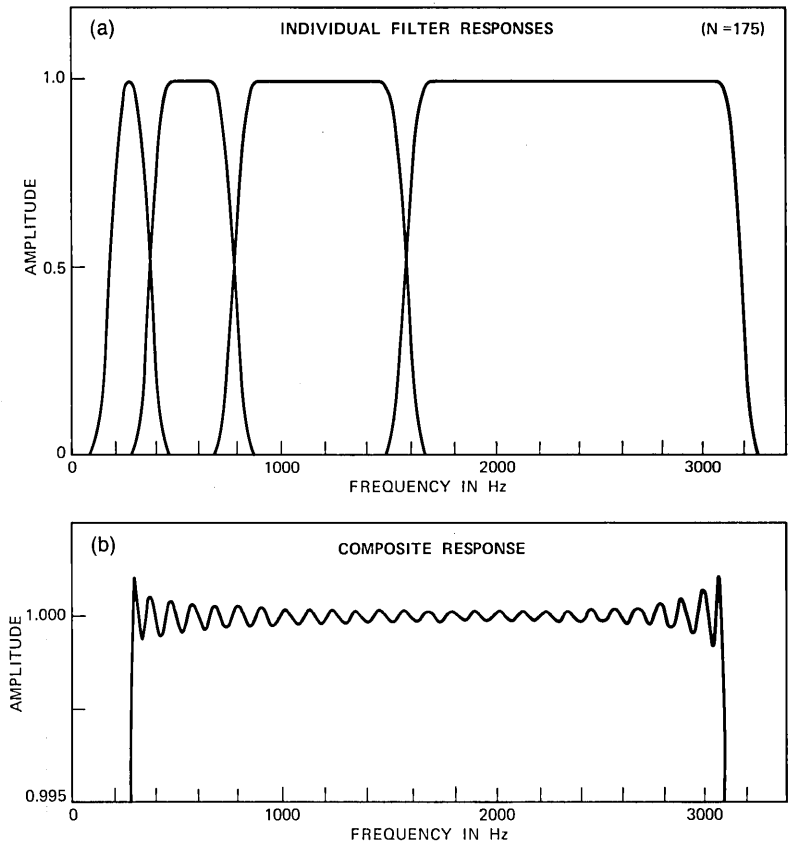

Fig. 6—Individual and composite frequency responses of a bank of 4 nonuniform bandpass filters for $N = 175$.

It is interesting to note that the composite frequency response of the filter bank is *independent* of the number and distribution of the individual filters, so long as the same window is used to design all the individual filters in the bank. This result can be verified by writing the overall frequency response of the filter bank, $H(e^{j\omega T})$, as

$$H(e^{j\omega T}) = \sum_{k=1}^{N_f} H_k(e^{j\omega T}), \tag{14}$$

which, from eq. (6), can be written as

$$H(e^{j\omega T}) = \sum_{k=1}^{N_f} \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} H_{dk}(e^{j\theta T}) W(e^{j(\omega-\theta)T}) d\theta. \tag{15}$$

Interchanging the order of summation and integration, eq. (15)

can be written as:

$$H(e^{j\omega T}) = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T}\left[\sum_{k=1}^{N_f} H_{dk}(e^{j\theta T})\right] W(e^{j(\omega-\theta)T})d\theta \tag{16}$$

$$= W(e^{j\omega T}) \circledast H_T(e^{j\omega T}), \tag{17}$$

where

$$H_T(e^{j\omega T}) = \sum_{k=1}^{N_f} H_{dk}(e^{j\omega T}). \tag{18}$$

Equations (17) and (18) show that the overall frequency response of the filter bank is the circular convolution of the frequency response of the window with the frequency response of the combined *ideal* bandpass filters. As seen in Fig. 4, the combined ideal frequency response of the bandpass filters is an ideal bandpass filter from $\omega = \omega_{\min}$ to $\omega = \omega_{\max}$, independent of the number and distribution of the individual filters. Thus, the composite filter bank frequency responses for the examples in Figs. 5 and 6 are identical because the same window was used in both cases and the filters spanned the identical frequency ranges.

Example 3

Suppose that all the parameters remain the same as in Example 2 except that we require narrower transition regions. This means that a larger value of N is required. In fact, Eq. (8a) shows that N and Δf are roughly inversely proportional. Figure 7 shows the filter bands corresponding to the parameters of Example 2 except that $N = 301$ and $\Delta f = 0.012082$ (transition width is 116 Hz). The sharper transitions are apparent in Fig. 7a, and Fig. 7b shows that the composite response remains very flat.

Example 4

We have assumed throughout that the transition width was less than twice the smallest low-pass cutoff frequency. In our examples, this constraint required that N be at least 175. The result of reducing N below this value is illustrated in Fig. 8. In this case, all the parameters were the same as in Examples 2 and 3, except in the case of $N = 101$ and $\Delta f = 0.0362465$. The transition width is 348 Hz, which is much greater than twice the cutoff frequency of the first low-pass filter. This is clearly in evidence in Fig. 8a. It is clear that reasonable filters are obtained for the wider bandwidth filters; however, the lowest filter does not attain unity response anywhere in its passband.

The preceding examples make it abundantly clear that, for sufficiently long impulse responses, the composite filter-bank response can

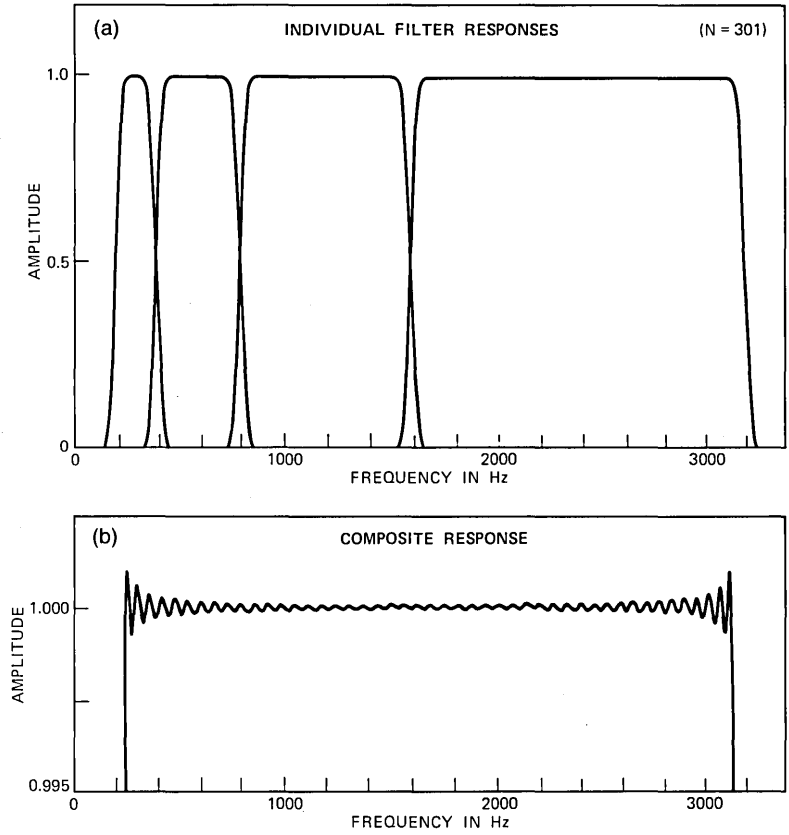

Fig. 7—Individual and composite frequency responses of a bank of 4 nonuniform bandpass filters for $N = 301$.

be very flat. In Ref. 5, where design techniques for IIR filter banks were discussed, the best results achieved for the composite response were approximately 1-dB peak-to-peak ripple for uniform bandwidths and about 2.5-dB peak-to-peak ripple for nonuniform bandwidths. This is in contrast to the results of the examples of this section, where the peak-to-peak ripple in the composite response was about 0.0274 dB for all the filter banks independent of how the bandwidths were chosen. This, together with the precise linear phase that is easily achieved, makes the FIR filter banks superior to what can be achieved for IIR filter banks. The price that is paid for this is that rather large values of N are required to achieve sharp transitions. However, the values of N used in the previous examples are certainly not unreasonable if the filters are implemented by FFT convolution methods or in special-purpose hardware.

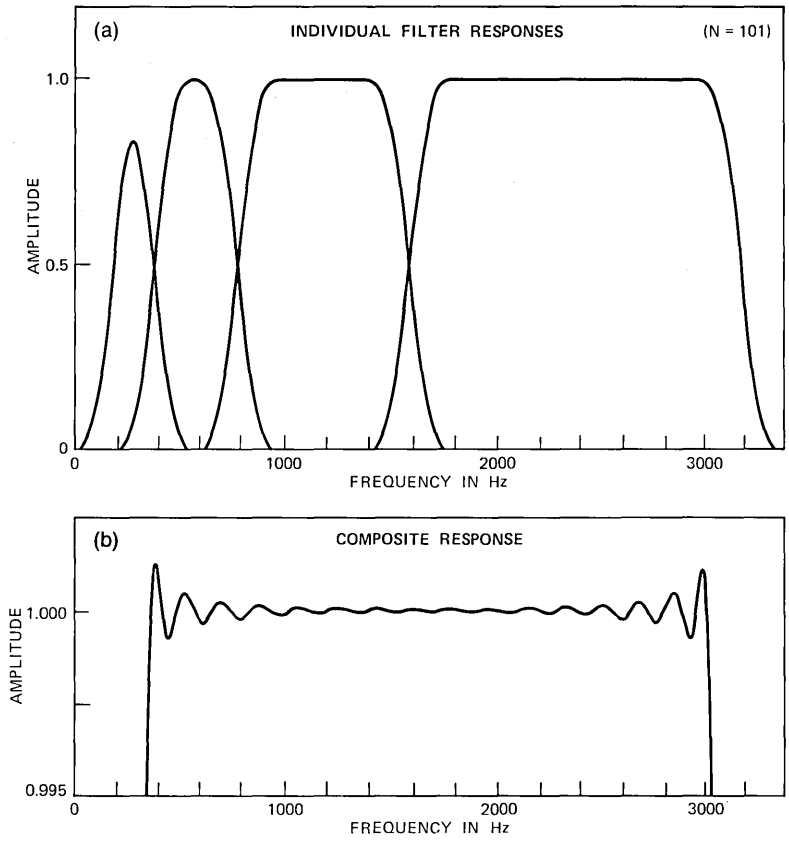

Fig. 8—Individual and composite frequency responses of a bank of 4 nonuniform bandpass filters for $N = 101$.

IV. SUMMARY

We have discussed a design method for filter banks composed of FIR digital filters. The method exploits the linear-phase properties obtainable for such filters, as well as the symmetry of the transition region that results from the windowing method of design. We summarized this method of design for the Kaiser window and illustrated the filter-bank design method with several examples. These examples show that the proposed design method has a great deal of flexibility and that excellent response characteristics can be achieved.

REFERENCES

1. J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, Second Ed., New York: Springer-Verlag, 1972.
2. J. L. Flanagan and R. M. Golden, "Phase Vocoder," B.S.T.J., *45*, No. 9 (November 1966), pp. 1493–1509.

3. R. W. Schafer and L. R. Rabiner, "Design and Simulation of a Speech Analysis-Synthesis System Based on Short-Time Fourier Analysis," IEEE Trans. Audio and Electroacoustics, *AU-21*, No. 3 (June 1973), pp. 165–174.
4. J. L. Flanagan and R. C. Lummis, "Signal Processing to Reduce Multipath Distortion in Small Rooms," J. Acoust. Soc. Am., *47*, part 1, June 1970, pp. 1475–1481.
5. R. W. Schafer and L. R. Rabiner, "Design of Digital Filter Banks for Speech Analysis," B.S.T.J., *50*, No. 10 (December 1971), pp. 3097–3115.
6. J. F. Kaiser, "Digital Filters," *System Analysis by Digital Computer*, Ch. 7, F. F. Kuo and J. F. Kaiser, eds., New York: John Wiley, 1966, pp. 228–243.
7. J. F. Kaiser, "Nonrecursive Digital Filter Design using the I_0-Sinh Window Function," Proc. 1974 IEEE Int. Symp. on Cir. and Syst., San Francisco, 1974, pp. 20–23.
8. L. R. Rabiner, J. H. McClellan, and T. W. Parks, "FIR Digital Filter Design Techniques using Weighted Chebyshev Approximation," to appear in Proc. IEEE, April 1975.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

Design and Evaluation of Shifted-Companion-Form Active Filters

By J. TOW

(Manuscript received August 12, 1974)

Two techniques for designing a class of low-sensitivity, follow-the-leader, feedback-type active filters have been introduced by Hurtig and Laker-Ghausi. The FLF configuration consists of a cascade of second- and/or first-order sections, with feedback from each section back to the first. This paper presents an approach for designing FLF-type realization for all classes of filter functions. The technique is based on a shifted-companion form of the associated-state equations. Some salient features of Hurtig's primary resonator block, Laker-Ghausi's follow-the-leader feedback, and the shifted-companion-form techniques are presented below.

(*i*) *Hurtig's PRB realizes any all-pole (no finite transmission zeros) filter function. This includes the low-pass, high-pass, and symmetrical bandpass filters without finite zeros. Explicit design equations are available, and the individual sections in the array are identical.*

(*ii*) *Laker-Ghausi's FLF realizes any symmetrical (including finite transmission zeros) bandpass filter function. The sections are not constrained to be identical, which allows optimization using this degree of freedom. Finite zeros are realized by a summation technique.*

(*iii*) *The SCF realizes all types of filter functions, i.e., low-pass, high-pass, bandpass, all-pass, or band-reject filters. Explicit design equations are available. The first section can differ from the rest, thus allowing some optimization with standardization. Feedforward as well as summation techniques can be used to realize the finite zeros.*

Two bandpass design examples using SCF, PRB, and/or Laker-Ghausi FLF techniques are given and compared with the low-sensitivity coupled (leapfrog) biquad, the conventional cascade biquad, and the passive ladder filter designs. The comparison shows that the passive filter gives

the best performance with respect to sensitivity to element deviations. All the coupled designs are significantly better than the cascade design in the passband, with the coupled biquad (leapfrog) design the most significantly better. In the stopband, cascade and coupled designs perform roughly the same.

I. INTRODUCTION

Recently, Hurtig[1,2] introduced a low-sensitivity, multiple-loop-feedback active RC filter configuration for the realization of greater-than-second-order voltage transfer functions. The configuration has been found to exhibit greatly improved stability over cascaded designs. For symmetrical bandpass filters, Hurtig's structure [called the primary resonator block (PRB) configuration] consists of a cascade of identical biquadratic bandpass sections (i.e., same pole-frequency and pole-Q) with feedback from each section (except the first) back to the first section. More recently, Laker and Ghausi[3,4] extended Hurtig's configuration to include symmetrical bandpass filters with finite transmission zeros, e.g., elliptic-type filters. In Laker and Ghausi's approach [called the follow-the-leader feedback (FLF) technique], different pole-Q values can be allowed for the biquadratic bandpass sections.

For the PRB technique, Hurtig has given a set of explicit equations expressing the biquadratic bandpass transfer function and the feedback factors in terms of the coefficients of the all-pole prototype low-pass transfer function.[2] In the FLF approach, Laker and Ghausi used a coefficient-matching technique. Because of the nonuniqueness of solutions in the FLF approach, Laker and Ghausi further proposed a method of choosing the pole-Q values for an optimum design.

In this paper, we present yet another approach based on a shifted-companion form of state variable representation of the voltage transfer function for the design of symmetrical bandpass and band-reject filters with this structure. In the bandpass case, using the proposed method, each biquadratic bandpass section in the cascaded array must be identical, with the possible exception of the first. Hence, it includes the Hurtig PRB configuration as a special case, but does not encompass the Laker-Ghausi cases having three or more different values of pole-Q. As in Laker-Ghausi's approach, the design of symmetrical bandpass filters with finite transmission zeros is included in the discussion of the shifted-companion form. Similarly to Hurtig's approach, the shifted-companion form also gives explicit design formulas as opposed to the coefficient-matching technique used by Laker and Ghausi. Furthermore, in the shifted-companion-form

design, different realizations (with the same configuration) can be obtained by varying the value of a shift parameter. The standard companion-form representation[5] corresponds to the case in which the value of this shift parameter is equal to zero.

In the next section, the shifted-companion-form representation of a voltage transfer function is presented. A brief discussion on the optimal choice of the shift parameter based on our design experience is given in Section III. Two design examples, a three-section Butterworth bandpass filter and a three-section elliptic bandpass filter, are given in Section IV. The section also compares the sensitivity performance, in a Monte-Carlo sense, of the shifted-companion-form designs to the cascade biquad and the coupled biquad[6,7] as well as to the passive designs.

II. SHIFTED-COMPANION-FORM REPRESENTATION OF VOLTAGE TRANSFER FUNCTION

The design technique for the proposed shifted-companion-form representation of a voltage transfer function is obtained as follows. First, a shift is introduced to the complex frequency variable by adding a variable constant α (shift parameter) to the complex frequency variable. Second, the resulting shifted-transfer function is represented by the standard companion form[5] and its corresponding block diagram, which has the desired structure. Third, an inverse shift operation is made on the standard companion form to determine the proper values for the parameters of the structure.

2.1 Representation of voltage transfer function by a shifted-companion form

Let the voltage transfer function be given by

$$\frac{V_{\text{out}}}{V_{\text{in}}}(p) = \frac{n_m p^m + n_{m-1}p^{m-1} + \cdots + n_1 p + n_0}{p^n + d_{n-1}p^{n-1} + \cdots + d_1 p + d_0} + d, \quad \text{for} \quad m < n. \tag{1}$$

Let the following shifting be made in the complex frequency variable of (1):

$$p = s - \alpha, \tag{2}$$

where α, the shift parameter, is a real number. Substitution of (2) into (1) results in the following shifted-transfer function (see Appendix A):

$$\frac{V_{\text{out}}}{V_{\text{in}}}(s) = \frac{\sum_{i=0}^{m} b_{m-i}s^{m-i}}{s^n + \sum_{j=1}^{n} a_{n-j}s^{n-j}} + d, \tag{3}$$

where

$$\left.\begin{aligned} a_{n-j} &= \sum_{k=0}^{j} (-1)^{j-k} \frac{(n-k)!}{(j-k)!(n-j)!} \alpha^{j-k} d_{n-k}, \\ & \qquad\qquad j = 1, 2, \cdots, n \\ b_{m-i} &= \sum_{k=0}^{i} (-1)^{i-k} \frac{(m-k)!}{(i-k)!(m-i)!} \alpha^{i-k} n_{m-k}, \\ & \qquad\qquad i = 0, 1, \cdots, m \end{aligned}\right\}. \tag{3a}$$

Note that $d_n = 1$. Alternatively, the a's and b's can also be obtained by the following implicitly recursive formula:

$$\left.\begin{aligned} a_n &= 1 \\ d_{n-j} &= \sum_{k=0}^{j} \frac{(n-k)!}{(j-k)!(n-j)!} \alpha^{j-k} a_{n-k}, \qquad j = 1, 2, \cdots, n \\ n_{m-i} &= \sum_{k=0}^{i} \frac{(m-k)!}{(i-k)!(m-i)!} \alpha^{i-k} b_{m-k}, \qquad i = 0, 1, \cdots, m \end{aligned}\right\}. \tag{3b}$$

It is well known that a voltage transfer function (with degree n) can be represented by a set of state equations in the (standard) companion form,[5] i.e.,

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{b}v_{\text{in}} \\ v_{\text{out}} &= \mathbf{c}\mathbf{x} + dv_{\text{in}}, \end{aligned} \tag{4}$$

where the state matrix $\mathbf{A}$ is of dimension $n \times n$. In the case of eq. (3), we have

$$\left.\begin{aligned} \mathbf{x} &= (x_1, x_2, \cdots, x_n)^t \\ \mathbf{A} &= \begin{bmatrix} -a_{n-1} & -a_{n-2} & -a_{n-3} & \cdots & -a_1 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \\ \mathbf{c} &= [\gamma_1, \gamma_2, \cdots, \gamma_n]. \end{aligned}\right\} \tag{5}$$

There are two special cases for eq. (5), A and B.

Case A: Transmission-zero forming by an input feed-forward technique:

$$\begin{aligned} \mathbf{c} &= [0\ 0\ \cdots\ 1] \\ \mathbf{b} &= [\beta_1\ \beta_2\ \cdots\ \beta_n]^t \end{aligned}$$

and

$$\beta_{n+1-i} = b_{n-i} - \sum_{j=1}^{i-1} a_{n-i+j} \beta_{n+1-j}, \qquad i = 1, 2, \cdots, n. \tag{5a}$$

Case B: Transmission-zero forming by summation-of-state-variables technique:

$$\begin{aligned} \mathbf{b} &= [1\ 0\ \cdots\ 0]^t \\ \mathbf{c} &= [b_{n-1}\ b_{n-2}\ \cdots\ b_0]. \end{aligned} \tag{5b}$$

To obtain the shifted-companion-form representation of the voltage transfer function of eq. (1), the inverse shift operation, i.e., $s = p + \alpha$ is applied to eq. (4). This is equivalent to the following operation:

$$\begin{array}{c|c|c} \text{time domain} & \text{frequency domain} & \text{time domain} \\ \dot{\mathbf{x}} \Leftrightarrow s\mathbf{X} \xrightarrow[s=p+\alpha]{\text{inverse shift}} (p+\alpha)\mathbf{X} & \Leftrightarrow \dot{\mathbf{x}} + \alpha\mathbf{I}\mathbf{x}, & \end{array} \tag{6}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. Hence, a shifted-companion-form representation of eq. (1) is*

$$\begin{aligned} \dot{\mathbf{y}} &= \mathbf{A}'\mathbf{y} + \mathbf{b}v_{\text{in}} \\ v_{\text{out}} &= \mathbf{c}\mathbf{y} + dv_{\text{in}}, \end{aligned} \tag{7}$$

where

$$\mathbf{A}' = \mathbf{A} - \alpha\mathbf{I} = \begin{bmatrix} -a_{n-1} - \alpha & -a_{n-2} & -a_{n-3} & \cdots & -a_1 & -a_0 \\ 1 & -\alpha & 0 & \cdots & 0 & 0 \\ 0 & 1 & -\alpha & \cdots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 & -\alpha \end{bmatrix}, \tag{7a}$$

and the vectors $\mathbf{b}$ and $\mathbf{c}$ are as given in eqs. (5a) or (5b).

At this point, it is desirable to change the relative level of the state vector $\mathbf{y}$ to obtain more convenient values for the gain (i.e., close to unity) of the individual biquadratic sections. Mathematically, we let

$$\mathbf{y} = \mathbf{K}\mathbf{x}, \tag{8}$$

where $\mathbf{K}$ is a nonsingular diagonal matrix. It has been found convenient to choose $\mathbf{K}$ to have the following form:

$$\mathbf{K} = \text{diag}\,[\alpha^{n-1}\ \ \alpha^{n-2}\ \ \cdots\ \ \alpha\ \ 1]. \tag{9}$$

Substituting (8) and (9) into (7) and (7a), the following shifted-companion-form representation of eq. (1) is obtained:

$$\begin{aligned} \dot{\mathbf{x}} &= \hat{\mathbf{A}}\mathbf{x} + \hat{\mathbf{b}}v_{\text{in}} \\ v_{\text{out}} &= \hat{\mathbf{c}}\mathbf{x} + dv_{\text{in}}, \end{aligned} \tag{10}$$

* The state vector is changed from $\mathbf{x}$ to $\mathbf{y}$.

where

$$\hat{\mathbf{A}} = \mathbf{K}^{-1}\mathbf{A}'\mathbf{K} = \begin{bmatrix} -a_{n-1} - \alpha & -\frac{a_{n-2}}{\alpha} & -\frac{a_{n-3}}{\alpha^2} & \cdots & -\frac{a_1}{\alpha^{n-2}} & -\frac{a_0}{\alpha^{n-1}} \\ \alpha & -\alpha & 0 & \cdots & 0 & 0 \\ 0 & \alpha & -\alpha & \cdots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & \alpha & -\alpha \end{bmatrix} \tag{11}$$

$$\hat{\mathbf{b}} = \mathbf{K}^{-1}\mathbf{b} = \begin{bmatrix} \frac{\beta_1}{\alpha^{n-1}} & \frac{\beta_2}{\alpha^{n-2}} & \cdots & \frac{\beta_{n-1}}{\alpha} & \beta_n \end{bmatrix}^t$$

$$\hat{\mathbf{c}} = \mathbf{c}\mathbf{K} = [0 \quad 0 \quad \cdots \quad 0 \quad 1] \tag{11a}$$

or

$$\hat{\mathbf{b}} = \mathbf{K}^{-1}\mathbf{b} = \begin{bmatrix} \frac{1}{\alpha^{n-1}} & 0 & \cdots & 0 & 0 \end{bmatrix}^t$$

$$\hat{\mathbf{c}} = \mathbf{c}\mathbf{K} = [\alpha^{n-1}b_{n-1} \quad \alpha^{n-2}b_{n-2} \quad \cdots \quad \alpha b_1 \quad b_0]. \tag{11b}$$

Equations (11a) and (11b) correspond to the cases where the transmission zeros are formed by the input feed-forward and the summation techniques, respectively.

2.2 Block diagram representation of the shifted-companion form

Transforming eqs. (10), (11), and (11a) into the frequency domain, the following set of transfer functions representing the shifted-companion form is obtained.

$$\left.\begin{aligned} X_1(p) &= \frac{1}{p + (a_{n-1} + \alpha)}\left[-\sum_{j=2}^{n}\frac{a_{n-j}}{\alpha^{j-1}}X_j(p) + \frac{\beta_1}{\alpha^{n-1}}V_{\text{in}}(p)\right] \\ X_i(p) &= \frac{1}{p + \alpha}\left[\alpha X_{i-1}(p) + \frac{\beta_i}{\alpha^{n-i}}V_{\text{in}}(p)\right] \qquad \text{for} \quad i = 2, 3, \cdots, n \\ V_{\text{out}}(p) &= X_n(p) + dV_{\text{in}}(p) \end{aligned}\right\}. \tag{12}$$

Similarly, by transforming eqs. (10), (11), and (11b), we have

$$\left.\begin{aligned} X_1(p) &= \frac{1}{p + (a_{n-1} + \alpha)}\left[-\sum_{j=2}^{n}\frac{a_{n-j}}{\alpha^{j-1}}X_j(p) + \frac{1}{\alpha^{n-1}}V_{\text{in}}(p)\right] \\ X_i(p) &= \frac{1}{p + \alpha}[\alpha X_{i-1}(p)] \quad \text{for} \quad i = 2, 3, \cdots, n \\ V_{\text{out}}(p) &= \sum_{i=1}^{n}\alpha^{n-i}b_{n-i}X_i(p) + dV_{\text{in}}(p) \end{aligned}\right\}. \tag{13}$$

Equations (12) and (13) are shown in block diagram form in Figs. 1a and 1b, respectively.

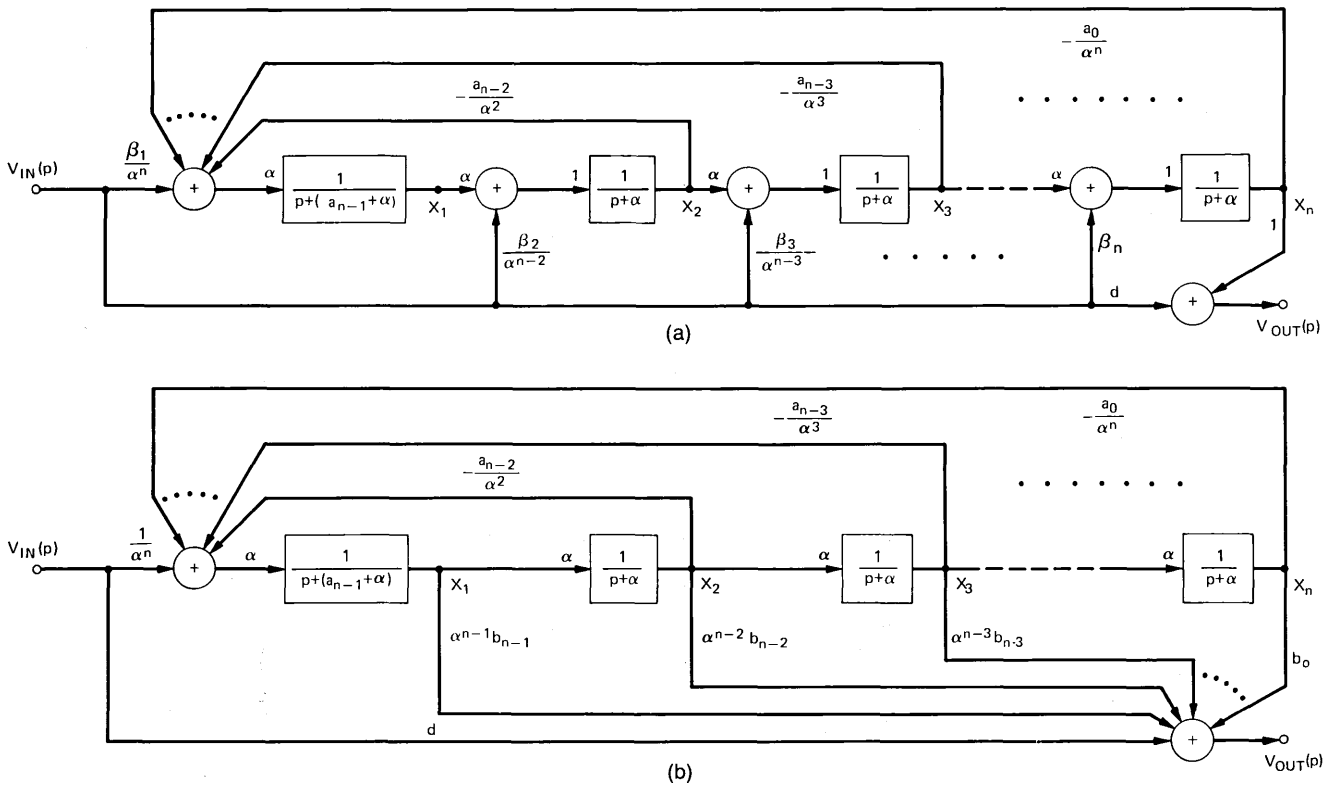

Fig. 1—Shifted-companion form. (a) Feed-forward zero-forming technique. (b) Summation zero-forming technique.

2.3 Block diagram representation of symmetrical bandpass filters via the shifted-companion form

For geometrically symmetrical bandpass filters, eq. (1), eq. (12) or (13), and Fig. 1a or 1b can be taken as the transfer function, the shifted-companion-form representation, and the block-diagram representation of the corresponding low-pass prototype, respectively. To obtain the block diagram representation of the symmetrical bandpass filter transfer function, we can apply the well-known low-pass to bandpass transformation:

$$p = \frac{s^2 + \omega_0^2}{Bs}, \tag{14}$$

where

p = complex frequency for the normalized low-pass function
s = complex frequency for the actual bandpass function
ω_0 = center frequency of the bandpass filter (in radians/s)
B = passband bandwidth of the bandpass filter (in radians/s)

to Figs. 1a and 1b. The resulting block diagram representations are shown in Figs. 2a and 2b.

2.4 Block diagram representation of symmetrical band-reject filters via the shifted-companion form

To obtain the block diagram representation of the symmetrical band-reject filter transfer function, similarly to the development of Section 2.3, we can first apply a low-pass to high-pass transformation, then follow with the usual low-pass to bandpass transformation, eq. (14). Specifically, this results in the following transformation to Figs. 1a and 1b:

$$\frac{1}{p + \alpha} = \frac{1}{\alpha} \cdot \frac{s^2 + \omega_0^2}{s^2 + (B/\alpha)s + \omega_0^2}, \tag{15}$$

where

p = complex frequency for the normalized low-pass function
s = complex frequency for the actual band-reject function
$\omega_0 = \sqrt{\omega_2 \omega_1}$ (in radians/s)
$B = \omega_2 - \omega_1$ (in radians/s)
ω_1/ω_2 = the lower/upper passband edge frequencies of the band-reject filter.

The resulting block diagram representation can also be shown as in Fig. 2, except that

$$T_1(s) = \frac{1}{a_{n-1} + \alpha} \cdot \frac{s^2 + \omega_0^2}{s^2 + [Bs/(a_{n-1} + \alpha)] + \omega_0^2}$$

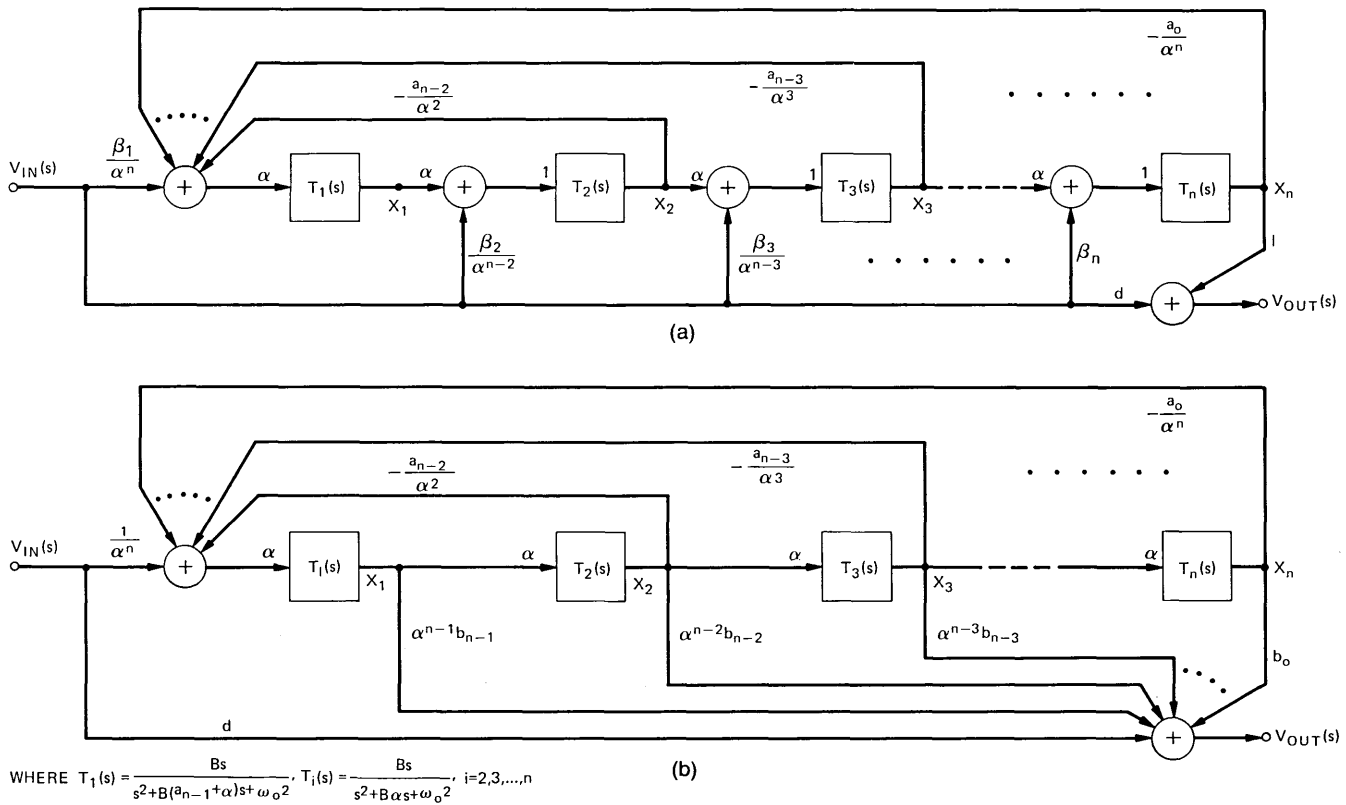

Fig. 2—Symmetrical bandpass filters via the shifted-companion form. (a) Feed-forward zero-forming technique. (b) Summation zero-forming technique.

and

$$T_i(s) = \frac{1}{\alpha} \cdot \frac{s^2 + \omega_0^2}{s^2 + (B/\alpha)s + \omega_0^2}, \qquad i = 2, 3, \cdots, n.$$

In passing, we note that Fig. 2 can also represent the symmetrical band-reject filter provided the parameters in Fig. 2 are determined by making eq. (1) the transfer function of the band-reject's corresponding high-pass prototype.

III. OPTIMAL CHOICE OF THE SHIFT PARAMETER, α

For symmetrical bandpass filters (Fig. 2), it is seen that all the biquadratic sections, with the possible exception of the first section, have a pole-Q value equal to $\omega_0/B\alpha$. The pole-Q value for the first section is

$$\frac{\omega_0}{B(a_{n-1} + \alpha)} \quad \text{or} \quad \frac{\omega_0}{B[d_{n-1} - (n-1)\alpha]}.$$

The value of these Q's versus α is illustrated in Fig. 3.

Before we proceed with a discussion on the optimal choice of α, two special cases are pointed out. The first is the standard companion form which corresponds to the case where $\alpha = 0$. From Eq. (3a),

$$a_{n-1} = d_{n-1} - n\alpha. \tag{16}$$

Letting $\alpha = d_{n-1}/n$, $a_{n-1} = 0$. With this value, (d_{n-1}/n) for α, a second special case is obtained where all the biquadratic sections (including the first) will have a pole-Q value equal to $\omega_0/B \cdot n/(d_{n-1})$. For simple symmetrical bandpass filters, this special case reduces to Hurtig's PRB configuration, and Hurtig's design formula[2] is identical to that given by eq. (3b).

Since an infinite number of realizations, depending upon the choice of α, can be obtained for the shifted-companion-form representation, is there an optimal choice of α? This optimal choice would, perhaps, depend also upon the performance criterion chosen. Laker and Ghausi have proposed an optimization scheme for their configuration based on a minimization of a certain statistical multiparameter sensitivity measure.[3,4] Their scheme can also be used here for the determination of an optimal α with respect to their performance criterion. In the following, we present two observations based on our limited design experience with bandpass filters using the proposed shifted-companion form where minimizing the filter's passband sensitivity is of primary concern. In our discussion here, the filter designs are subjected to a computer-simulated Monte-Carlo analysis and sensitivity is examined from the standpoint of standard deviation (dB) vs frequency.

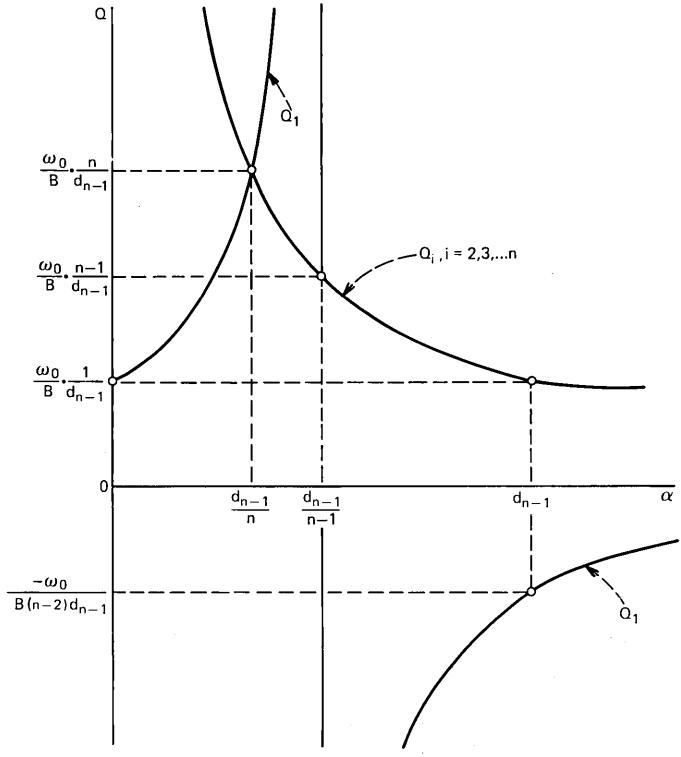

Fig. 3—Biquadratic sections Q vs shift parameter α.

(*i*) It appears that a broad range of values exists for α where the improvements* over the cascade biquad design are relatively constant. This range includes Hurtig's design, i.e., $\alpha = d_{n-1}/n$.

(*ii*) Performance of the standard companion-form (i.e., $\alpha = 0$) design is about the same as that of the cascade biquad design.

The above empirical rule (*i*) is observed in the design examples to follow.

IV. DESIGN EXAMPLES

Two examples given here illustrate the shifted-companion-form design technique as well as demonstrate its performance relative to that of the cascade biquad, coupled (leapfrog) biquad as well as to passive ladder designs. Comparisons among these designs are based on a Monte-Carlo analysis of the filters with passive components selected randomly from a uniform distribution within a given tolerance interval.

* Improvement is to be broadly interpreted as less sensitive or having a smaller standard deviation.

4.1 Example 1—A three-section Butterworth bandpass filter*

The normalized transfer function of a third-order low-pass Butterworth filter is given by

$$\frac{V_{\text{out}}}{V_{\text{in}}}(p) = \frac{1}{p^3 + 2p^2 + 2p + 1}. \tag{17}$$

Let the desired bandpass filter have center frequency (f_0) of 1 Hz and 3-dB bandwidth ($B/2\pi$) of 0.04 Hz. The PRB version of the shifted-companion form is designed here. Hence,

$$\alpha = \frac{d_{n-1}}{n} = \frac{2}{3}.$$

From eq. (3a), we obtain

$$a_2 = 0, \quad a_1 = 0.66666667, \quad a_0 = 0.25925926$$
$$b_2 = b_1 = 0, \quad b_0 = 1.$$

And from eq. (5a),

$$\beta_1 = 1, \quad \beta_2 = 0, \quad \beta_3 = 0.$$

For this simple bandpass filter, the output summing amplifier (Fig. 2) is not needed. Furthermore,

$$\alpha T_i(s) = \frac{0.08\pi(\frac{2}{3})s}{s^2 + 0.08\pi(\frac{2}{3})s + (2\pi)^2} \qquad i = 1, 2, 3.$$

Note that

$$Q_i = \frac{2\pi}{0.08\pi} \cdot \frac{3}{2} = 37.5 \qquad \text{for} \qquad i = 1, 2, 3.$$

For this example, each of the $T_i(s)$ is chosen to be realized by the single-amplifier biquad (SAB) configuration of Ref. 8. The complete configuration† is shown in Fig. 4, with the element values tabulated in Appendix B. The element values as well as circuit topologies for the cascade SAB, coupled SAB (or leapfrog SAB),‡ and the optimized Laker-Ghausi design[3] are also given in Appendix B. Each of the three biquadratic bandpass sections in the shifted-companion-form, Laker-Ghausi, and coupled-biquad designs has a pole frequency of 1 Hz; whereas for the cascade design, the pole frequencies are 1, 1.01747, and 0.982828 Hz. The pole-Q values for these four designs are tabu-

---

* This example can also be found in Ref. 3.

† The inverting amplifier A_2 can be eliminated by feeding the output of section 3 to the positive input terminal of the summing amplifier A_1. This has not been done in the example.

‡ For symmetrical bandpass filters derived from an all-pole low-pass prototype, the coupled biquad and the leapfrog designs can be made identical.

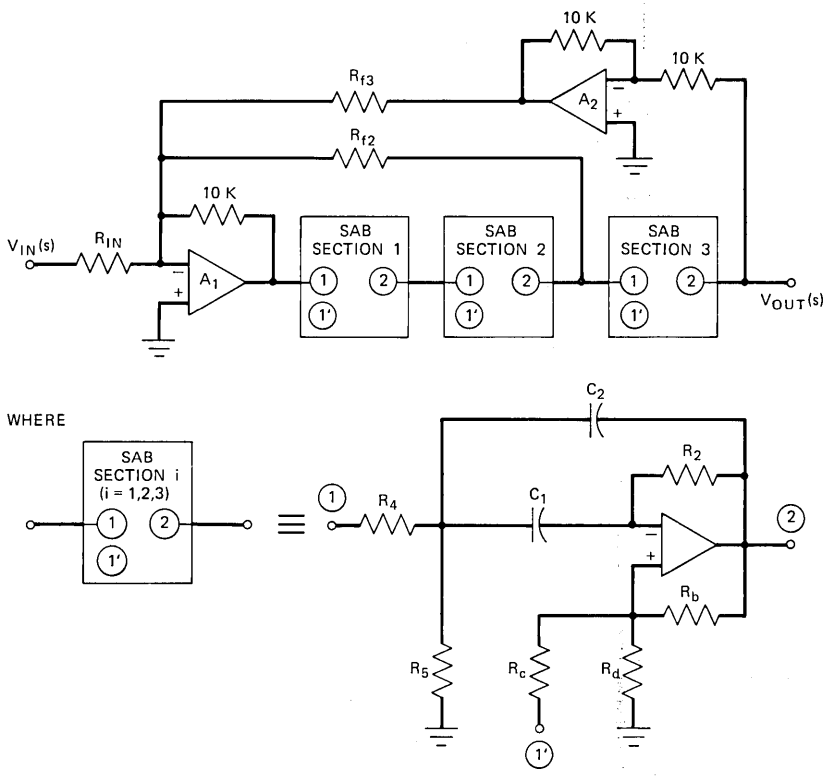

Fig. 4—Configuration for the three-section Butterworth bandpass filter.

lated in Table I. These four realizations of the Butterworth filter as well as the passive ladder realization were compared by a Monte-Carlo study (with 200 trials) using the computer program BELTAP.[9] The following assumptions are made:

(*i*) The operational amplifiers are ideal.
(*ii*) All passive components have the same tolerance with a uniform distribution.

Table I — Pole-Q values for the three-section Butterworth filter

Filter Type \ Pole-Q	Section		
	1	2	3
Shifted-companion form (PRB)	37.5	37.5	37.5
Laker-Ghausi (FLF)	44.2	44.2	28.8
Cascade biquad	25.0	50.0075	50.0075
Coupled biquad/leapfrog	25.0	∞	25.0

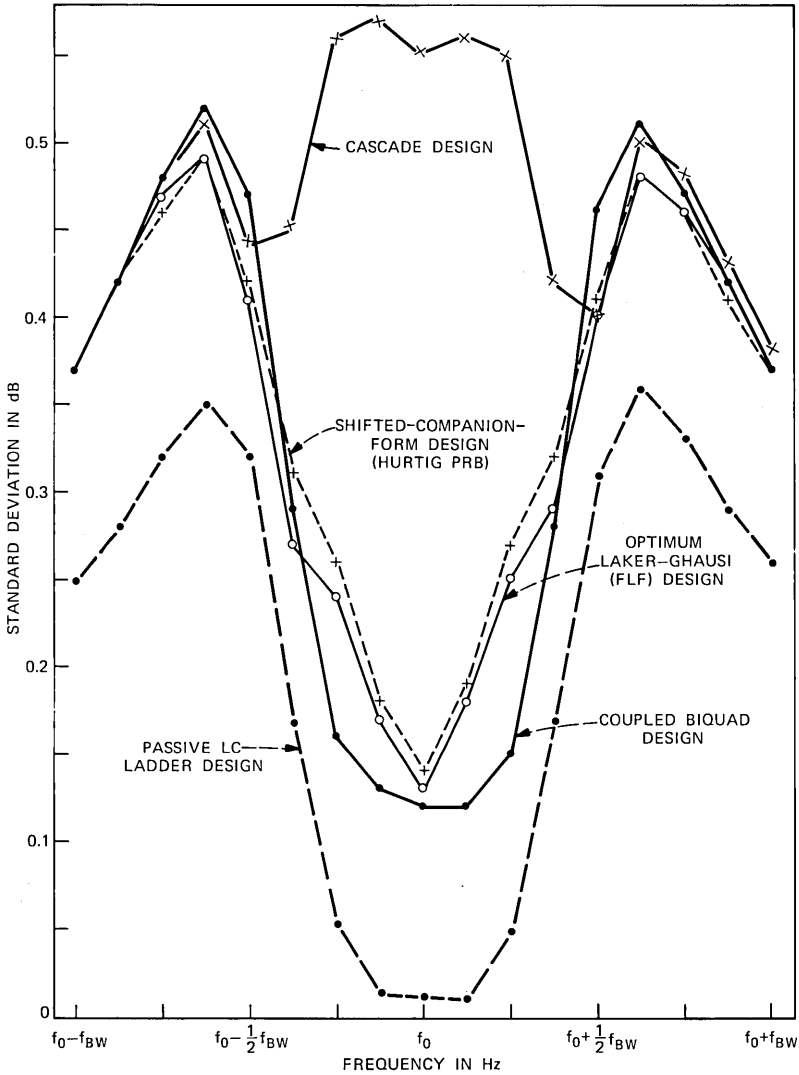

Fig. 5—Simulated variations of the three-section Butterworth bandpass filter (0.1732-percent passive components tolerance).

Two different tolerances were simulated, the first having a realistic tolerance of ± 0.1732 percent and the second a large tolerance of ± 1.732 percent.* The resulting comparisons based on the standard

* Realistic in the sense that the statistical variation of the filter response is within a reasonable bound from the nominal. The large tolerance corresponds to a component standard deviation of 1 percent, which was used in the example of Ref. 3.

deviations of the transfer function (dB) of the various designs plotted vs frequency (Hz) are shown in Figs. 5 and 6.

It is observed that, over most of the passband (between the 3-dB points), the coupled biquad/leapfrog, Hurtig's PRB/shifted-companion form, and Laker-Ghausi FLF designs show roughly the same improvement (3-to-1 reduction in standard deviations) over the cascade biquad design. The passive filter is, however, seen to be the least sensitive.

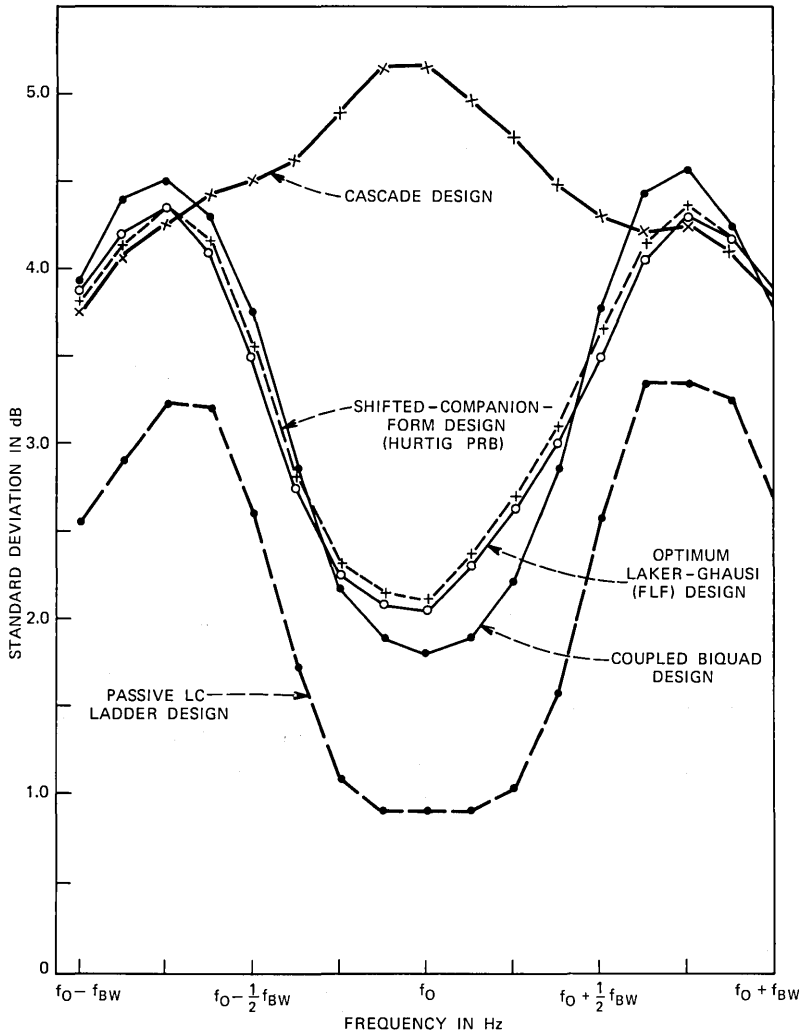

Fig. 6—Simulated variations of the three-section Butterworth bandpass filter (1.732-percent passive components tolerance).

4.2 Example 2—A three-section elliptic bandpass filter*

A sixth-order elliptic bandpass filter with center frequency (f_0) at 2805 Hz, 0.1-dB passband ripple with bandwidth ($B/2\pi$) of 90 Hz, and a minimum 30-dB loss below 2694.8 Hz and above 2919.8 Hz is desired. The corresponding third-order normalized low-pass prototype transfer function is given by

$$\frac{V_{\text{out}}}{V_{\text{in}}}(p) = \frac{N(p)}{D(p)},$$

where

$$N(p) = 0.214115(p^2 + 8.158500)$$
$$D(p) = p^3 + 1.897376p^2 + 2.543168p + 1.746858.$$

Once again, the Hurtig criterion is used for the shifted-companion-form design. Hence,

$$\alpha = \frac{d_2}{3} = 0.6324587.$$

From eq. (3a),

$$a_2 = 0, \quad a_1 = 1.3431556, \quad a_0 = 0.64438116$$
$$b_2 = 0.214115, \quad b_1 = -0.2708378, \quad b_0 = 1.8325041.$$

And from eq. (5a),

$$\beta_1 = 1.5449143, \quad \beta_2 = -0.2708378, \quad \beta_3 = 0.214115.$$

The feed-forward zero-forming configuration (Fig. 2a) is used for the realization† where

$$T_i(s) = \frac{180\pi s}{s^2 + 180\pi(0.6324587)s + (2\pi \cdot 2805)^2}.$$

For this example, each of the $T_i(s)$ is chosen to be realized by the three-amplifier biquad configuration.[10] The complete configuration‡ is shown in Fig. 7 with the element values tabulated in Appendix C. The element values for the cascade biquad and coupled biquad designs are also given in Appendix C.§ A leapfrog design is also available,[11] the performance of which is similar to but slightly inferior to the coupled biquad design. Once again, a Monte-Carlo study was made on these

---

* This example can also be found in Ref. 6.

† It was found, for this example, that the design with the feed-forward zero-forming technique outperforms the summation zero-forming technique design.

‡ With the three-amplifier biquad sections, it is possible to eliminate the input summing amplifier A_1 by using node 1 of section 1 as the summing point. This has not been done in the example.

§ Without the availability of a computer program to choose an optimized Laker-Ghausi circuit, no FLF design is included in this example.

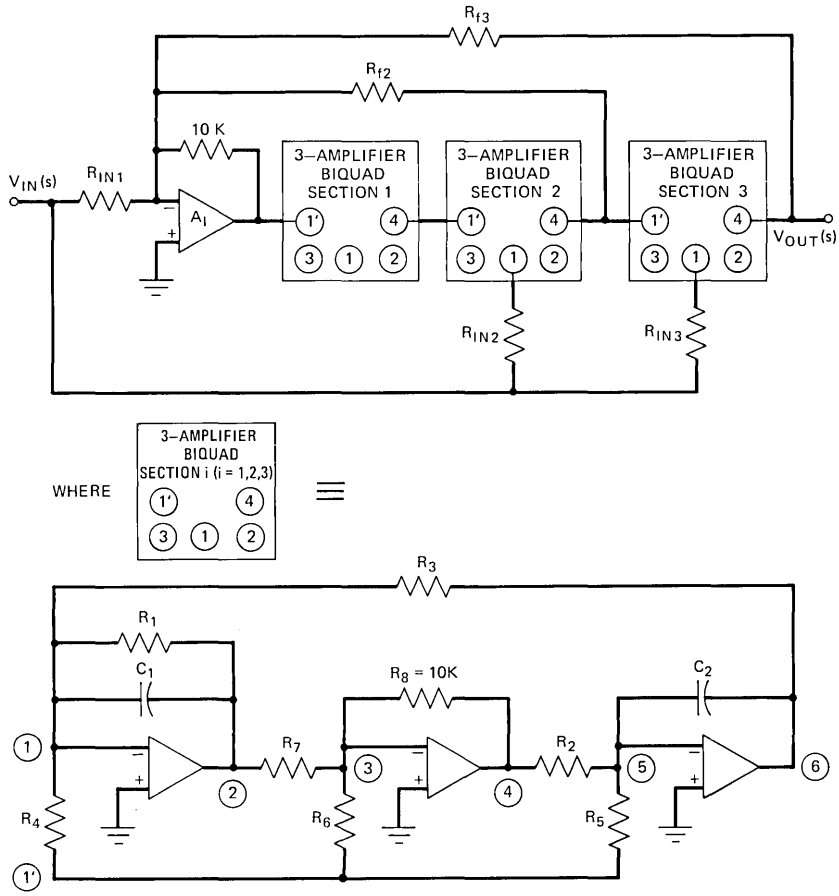

Fig. 7—Configuration for the three-section elliptic bandpass filter.

designs as well as a passive ladder design. The resulting comparisons are shown in Figs. 8 and 9, where a 0.25-percent tolerance (with uniform distribution) is assumed for all passive components.

It is observed that, in terms of standard deviation, the passband improvements over the cascade design are noticeably less for the shifted-companion-form (α chosen by Hurtig's PBR criterion) design than the roughly 4-to-1 improvement of the coupled biquad design. Once again, the passive filter outperforms its active counterparts.

V. CONCLUSIONS

The FLF/PRB multiple-loop feedback active filter structure is known to have better sensitivity performance than the popular cascade approach. This sensitivity improvement is particularly acute in high-Q

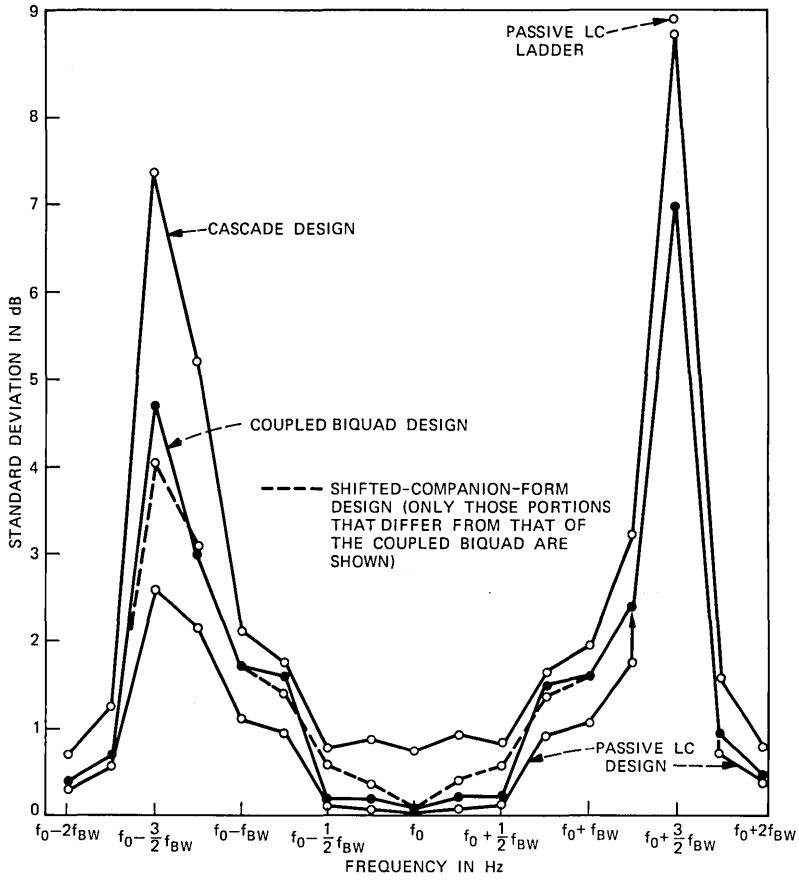

Fig. 8—Simulated variations of the elliptic bandpass filter.

bandpass filter designs, as exemplified in the two examples given. With the described shifted-companion-form representation of the filter transfer function, it is straightforward to obtain explicit design formulas for this feedback structure as contrasted to the coefficient matching technique used by Laker and Ghausi. In practice, the shift parameter can be chosen such that identical biquadratic blocks (i.e., the extended PRB version)* are used. The proposed shifted-companion-form design has the following advantages over the optimized FLF design. First, no matrix inversion and involved sensitivity minimization routine are needed. Second, all sections are identical, and the sections' pole-Q can be much lower than the highest pole-Q required for the FLF design. Furthermore, little difference is usually observed for this shifted-

* Hurtig's PRB does not treat the cases with finite transmission zeros.

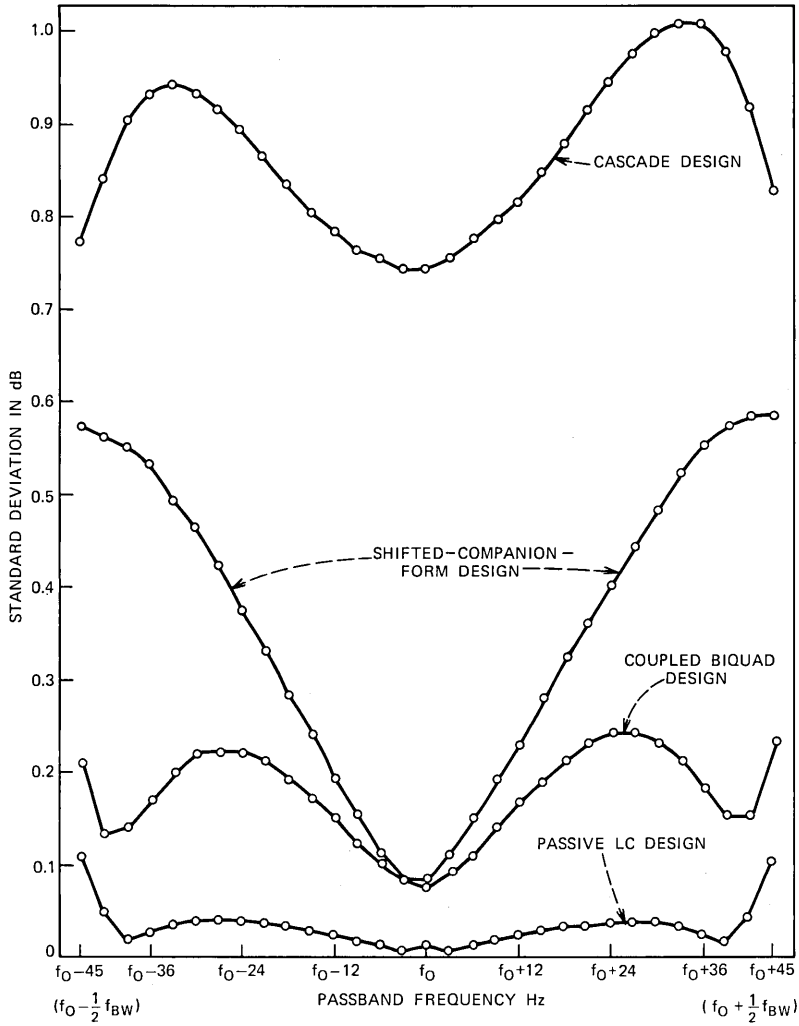

Fig. 9—Simulated variations (passband) of the elliptic bandpass filters.

companion-form design and the optimized FLF design. On the other hand, the two examples also show that the coupled biquad and leapfrog designs may have better sensitivity performance than the corresponding shifted-companion-form designs. However, for those types of filter functions having finite transmission zeros, the designs of coupled biquad and/or leapfrog require extensive computer aids that are not yet generally available. Hence, the proposed shifted-companion-form technique may provide an alternative to the coupled biquad/leapfrog active filter design techniques.

VI. ACKNOWLEDGMENT

The author wishes to thank R. C. Drechsler for helpful discussions.

APPENDIX A

Derivation of the Shifted Transfer Function

Let the polynomial $P(p)$ be

$$P(p) = \sum_{i=0}^{n} d_{n-i} p^{n-i}, \tag{18}$$

and let

$$p = s - \alpha. \tag{19}$$

Then

$$\begin{aligned} P(s) &= \sum_{i=0}^{n} d_{n-i}(s-\alpha)^{n-i} \\ &= \sum_{i=0}^{n} d_{n-i} \sum_{r=0}^{n-i} \binom{n-i}{r} s^{n-i-r}(-\alpha)^r, \end{aligned} \tag{20}$$

where

$$\binom{n-i}{r} = \frac{(n-i)!}{r!(n-i-r)!}. \tag{21}$$

Equation (20) can be rearranged in decreasing power of s as follows:

$$\begin{aligned} P(s) = d_n s^n &+ s^{n-1}\left[\sum_{r=0}^{1}\binom{n-r}{1-r}(-\alpha)^{1-r}d_{n-r}\right] \\ &+ s^{n-2}\left[\sum_{r=0}^{2}\binom{n-r}{2-r}(-\alpha)^{2-r}d_{n-r}\right] \\ &+ \cdots + s^{n-i}\sum_{r=0}^{i}\binom{n-r}{i-r}(-\alpha)^{i-r}d_{n-r} \\ &+ \cdots + \sum_{r=0}^{n}\binom{n-r}{n-r}(-\alpha)^{n-r}d_{n-r}. \end{aligned}$$

Or

$$\begin{aligned} P(s) &= \sum_{i=0}^{n} s^{n-i}\left[\sum_{r=0}^{i}(-1)^{i-r}\binom{n-r}{i-r}\alpha^{i-r}d_{n-r}\right] \\ &= \sum_{i=0}^{n} a_{n-i}s^{n-i}, \end{aligned} \tag{22}$$

where

$$a_{n-i} = \sum_{r=0}^{i}(-1)^{i-r}\binom{n-r}{i-r}\alpha^{i-r}d_{n-r},$$

or

$$= \sum_{r=0}^{i}(-1)^{i-r}\frac{(n-r)!}{(i-r)!(n-i)!}\alpha^{i-r}d_{n-r}. \tag{23}$$

Equation (23) can be used to obtain the coefficients of the shifted polynomial $P(s)$ from the coefficients of the polynomial $P(p)$.

Similarly, if we start with the polynomial

$$P(s) = \sum_{i=0}^{n} a_{n-i}s^{n-i} \tag{24}$$

and let

$$s = p + \alpha, \tag{25}$$

then we obtain

$$P(p) = \sum_{i=0}^{n} d_{n-i}p^{n-i},$$

where

$$d_{n-i} = \sum_{r=0}^{i} \frac{(n-r)!}{(i-r)!(n-i)!} \alpha^{i-r}a_{n-r}. \tag{26}$$

APPENDIX B

Element Values for Example 1 (Fig. 4)

For the various realizations, resistors are in kilohms and all capacitor values are 10 μF.

B.1 Shifted-companion-form (Hurtig's) realization

Element \ Section	1	2	3
R_2	128.5	128.5	128.5
R_4	613.2	613.2	613.2
R_5	1.977	1.977	1.977
R_b	73.07	73.07	73.07
R_d	2.0	2.0	2.0
R_c	∞	∞	∞

and $R_{\text{in}} = 2.963$, $R_{f2} = 6.667$, $R_{f3} = 11.43$.

B.2 Laker-Ghausi realization

Element \ Section	1	2	3
R_2	128.5	128.5	127.7
R_4	722.3	722.3	470.9
R_5	1.976	1.976	1.993
R_b	71.77	71.78	74.73
R_d	2.0	2.0	2.0
R_c	∞	∞	∞

and $R_{\text{in}} = 2.781$, $R_{f2} = 4.600$, $R_{f3} = 23.73$.

B.3 Cascade realization

The summing amplifier A_1 and inverter A_2 are not needed, and the input goes directly to node 1 of section 1.

Element \ Section	1	2	3
R_2	128.5	126.3	130.8
R_4	408.1	409.1	409.1
R_5	1.981	1.946	2.015
R_b	77.76	70.93	70.93
R_d	2.0	2.0	2.0
R_c	∞	∞	∞

B.4 Coupled biquad realization

The summing amplifier A_1 and inverter A_2 are not needed, and the input goes directly to node 1 of section 1.

Element \ Section	1	2	3
R_2	128.5	128.4	128.5
R_4	204.1	820.2	408.1
R_5	1.990	1.977	1.981
R_b	77.39	64.98	77.76
R_d	2.0	2.0	2.0
R_c	414.1	829.8	∞

In addition, node 1′ of section 1 is connected to node 2 of section 2 and node 1′ of section 2 is connected to node 2 of section 3.

APPENDIX C

Element Values for Example 2 (Fig. 7)

For the various realizations, resistors are in kilohms and all capacitor values are 0.01 μF.

C.1 Shifted-companion-form (Hurtig's criterion) realization

Element \ Section	1	2	3
R_1	279.6	279.6	279.6
R_2	5.674	5.674	5.674
R_3	5.674	5.674	5.674
R_4	279.6	279.6	279.6
R_5	∞	∞	∞
R_6	∞	∞	∞
R_7	10.0	10.0	10.0

and

$$R_{f2} = 2.978, \qquad R_{f3} = 3.926$$
$$R_{in1} = 1.638, \qquad R_{in2} = 413.0, \qquad R_{in3} = -825.9.^*$$

C.2 Cascade realization

The summing amplifier A_1 and all input feed-forward paths (R_{in}'s) are not needed. The input goes directly to node 1′ of section 1.

Element \ Section	1	2	3
R_1	167.5	428.6	412.2
R_2	5.674	5.786	5.564
R_3	5.674	5.786	5.564
R_4	167.5	1190.0	732.2
R_5	∞	16.93	9.378
R_6	∞	27.75	17.76
R_7	10.0	10.0	10.0

C.3 Coupled biquad realization

The summing amplifier A_1 is not needed. The input, V_{in}, goes directly into node 1′ of section 1 and nodes 1 of sections 2 and 3 through the feed-forward resistors R_{in2} and R_{in3}, respectively.

Element \ Section	1	2	3
R_1	311.5	∞	133.0
R_2	5.674	5.674	5.674
R_3	5.674	5.674	5.674
R_4	97.92	∞	∞
R_5	∞	∞	∞
R_6	∞	∞	∞
R_7	10.0	10.0	10.0

and $R_{in2} = 1324.0$, $R_{in3} = 825.9$. In addition, the following resistors are needed with value and connections noted.

(*i*) 180.5 kΩ, node 2 of section 1 and node 1 of section 2.
(*ii*) 180.5 kΩ, node 1 of section 1 and node 4 of section 2.
(*iii*) 194.3 kΩ, node 2 of section 2 and node 1 of section 3.
(*iv*) 194.3 kΩ, node 1 of section 2 and node 4 of section 3.

---

*In practice, with the following modifications of Fig. 7, $R_{in3} = 825.9$ is used. Change the connection of sections 2 and 3 to between node 2 (section 2) and node 1′ (section 3); the connection of R_{f2} remains unchanged. Change the connection of R_{f3} to between node 2 (section 3) and the summing amplifier A_1.

REFERENCES

1. G. Hurtig, III, "The Primary Resonator Block Technique of Filter Synthesis," International Filter Symposium, Santa Monica, California, April 15–18, 1972, p. 84.
2. G. Hurtig, III, "Filter Network Having Negative Feedback Loops," United States Patent 3,720,881, March 13, 1973.
3. K. R. Laker and M. S. Ghausi, "Synthesis of a Low-Sensitivity Multiloop Feedback Active RC Filter," IEEE Trans. on Circuits and Systems, *CAS-21*, No. 2 (March 1974), pp. 252–259.
4. K. R. Laker, "On Multiparameter Sensitivity in Active RC Networks," Ph.D. Thesis, New York University, May 1973.
5. L. A. Zadeh and C. A. Desoer, *Linear System Theory: The State Space Approach*, New York: McGraw-Hill, 1963, pp. 282–283 and 357–358.
6. J. Tow and Y. L. Kuo, "Coupled Biquad Active Filters," IEEE Proc. International Symposium on Circuit Theory, N. Hollywood, California, April 18–21, 1972, pp. 164–168.
7. G. Szentirmai, "Synthesis of Multiple-Feedback Active Filters," B.S.T.J., *52*, No. 4 (April 1973), pp. 527–555.
8. J. J. Friend, "A Single Operational-Amplifier Biquadratic Filter Section," IEEE Proc. International Symposium on Circuit Theory, Digest of Technical Papers, December 1970, pp. 179–180.
9. C. L. Semmelman, E. D. Walsh, and G. T. Daryanani, "Statistical Circuit Design: Linear Circuits and Statistical Design," B.S.T.J., *50*, No. 4 (April 1971), pp. 1149–1171.
10. P. E. Fleischer and J. Tow, "Design Formulas for Biquad Active Filters Using Three Operational Amplifiers," Proc. Letters, IEEE, *61*, May 1973, pp. 662–663.
11. G. Szentirmai, private communication.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

Timing Recovery and Scramblers in Data Transmission

By R. D. GITLIN and J. F. HAYES

(Manuscript received August 28, 1974)

This paper considers several problems associated with envelope-derived timing recovery, equalization, and scrambling in synchronous data transmission. Particular attention is focused on the time intervals in which periodic data sequences are transmitted, such as during start-up or when an idle code is being transmitted. It is shown that the standard envelope-derived timing-recovery system may be significantly improved by zonal filtering of the received passband signal prior to forming the envelope. For phase-modulated systems, we discuss the limitations of the "precession" technique employed for the purpose of providing a periodic timing wave when there is an input of short period. The advantages of using a phase-locked loop to filter the envelope instead of a narrow-band filter are also described. A study of scrambler operation has provided an extension of previous results concerning the relationship between the input and output period. It is shown that the output period of several scramblers connected in tandem does not necessarily double with the addition of a stage, and that if a particular stage does not lock up then no succeeding stage can lock up.

I. INTRODUCTION

Recovery and tracking of the symbol rate, or timing frequency, is one of the most critical functions performed by a synchronous modem. Most modems are "self-timed" in that they derive their timing frequency and phase directly from the information-bearing signal, instead of using a separate subchannel to send synchronization information. A technique that is commonly used to acquire the symbol rate* (which is the receiver's basic sampling rate) is to filter the envelope of the modulated data signal. Our investigation will consider several problems related to this method of timing recovery which arise in high-speed modems incorporating both an adaptive equalizer and a scrambler.

* This technique is also used to provide the sampling epoch, or phase, within a symbol interval.

The envelope-derived timing recovery system is a well-studied topic.[1,2] However, as the degree of excess bandwidth decreases, the ease with which timing can be recovered using this approach rapidly diminishes. We focus our attention on periodic input sequences. These sequences are used to train (or adapt) the data receiver during start-up and during the idle period between blocks of random data. To provide a densely spaced line spectrum of uniform amplitude (which is necessary if the equalizer coefficients are to remain properly adjusted for random input data), high-speed modems use a scrambler to "randomize" the short periodic inputs commonly used during the idle period. We investigate the effect of the scrambler on both the line spectrum and the strength of the timing tone. It is observed that zonal filtering of the received data signal prior to taking the envelope can significantly improve the relative strength of the timing tone by suppressing the jitter component.

Using transform theory, a discussion is presented on the relationship between the scrambler input and output periods. We refine Savage's[3] well-known results for periodic inputs; these refinements are applied to the study of the tandem and parallel scrambler configurations.

Sections II to IV review the basic envelope-derived timing system and give expressions for the power in the timing and interfering tones. The role of the phase-locked loop in the timing recovery system is described in Section V. Section VI considers the effect of precessing* the data symbols on timing recovery. The necessary background material on self-synchronizing scramblers is presented in Section VII. The transform approach is used in Section VIII to determine the scrambler output period. In Section IX the performance of a cascaded scrambler configuration is contrasted with the conventional serial arrangement. The parallel scrambler configuration is discussed in Section X.

II. BASIC TIMING RECOVERY SYSTEM

In this section we describe the commonly used technique of acquiring the timing frequency and phase by processing the envelope of the received signal. The object is to extract a tone, located at the symbol rate, which is then used in the sampled-data receiver. Figure 1 shows a simplified receiver structure of an in-phase and quadrature (e.g., QAM) data-transmission system, where we have focused attention on the timing recovery and equalization functions of the receiver. For

* The advancing of the transmitted angle by a fixed phase (in a differential phase-modulated modem), independently of the input, is known as precession.

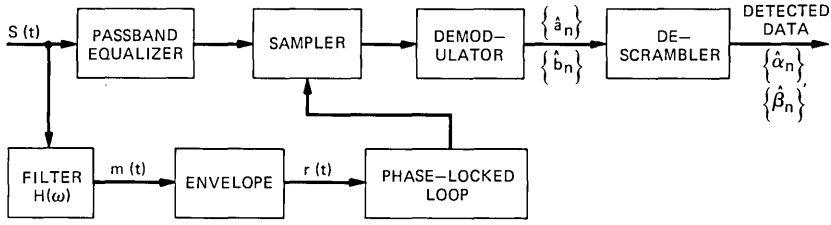

Fig. 1—Simplified QAM receiver.

our purposes it will be convenient to ignore both the additive noise and the quadrature component of channel distortion. Using the notation of Figs. 1 and 2, the received signal $s(t)$ is given by

$$s(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT) \cos \omega_c t - \sum_{n=-\infty}^{\infty} b_n g(t - nT) \sin \omega_c t, \quad (1)$$

where a_n and b_n are respectively the discrete-valued in-phase and quadrature data sequences obtained from the binary sequences α_n and β_n, $g(t)$ is the spectral-shaping pulse, ω_c is the carrier frequency, and $1/T$ is the symbol rate or timing frequency. The envelope of a filtered version of the received signal is tracked by a phase-locked loop tuned to the receiver's best *a priori* knowledge of the timing frequency. The output of a properly designed phase-locked loop will be a tone with frequency equal to the symbol rate and whose zero crossings may be used to derive a sampling wave. Once the timing frequency is acquired, the estimated and unscrambled data sequences $\{\hat{\alpha}_n\}$ and $\{\hat{\beta}_n\}$ are available to the user. The decoder maps the sequence of multilevel two-tuples $(\hat{a}_n, \hat{b}_n)$ into a binary sequence which serves as the input to the inverse scrambler.

The data sequences $\{a_n\}$ and $\{b_n\}$ may be thought of as random (when user data are being sent) or as periodic (during start-up when the equalizer and timing parameters are being acquired, and during an

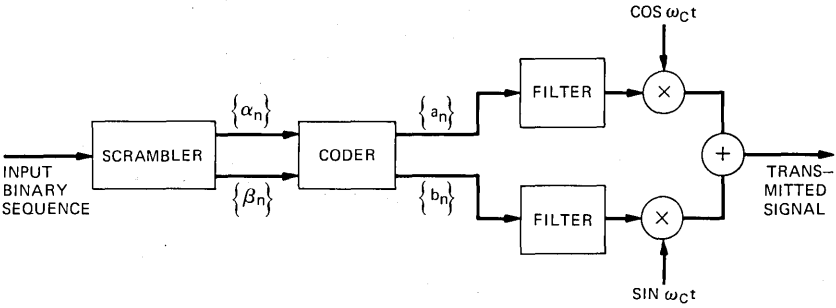

Fig. 2—Simplified QAM transmitter employing a scrambler.

idle period between random data transmissions). As we shall shortly see, the presence of a short periodic input can play havoc with the equalizer tap settings; hence, a scrambler is generally used at the transmitter to "randomize" these periodic sequences. As depicted in Fig. 2, the coder maps the binary stream of scrambled 0s and 1s into the channel pulse levels (e.g., 0 and 1 may be mapped into -1 and 1 respectively).* The effects of choosing a particular scrambler structure (e.g., serial vs parallel or serial vs cascade) on the timing recovery system will be treated in Sections VII and VIII.

III. SPECTRUM OF THE RECEIVED SIGNAL

We confine our attention to periodic inputs, beginning with a calculation of the Fourier transform of the received signal. Rewriting (1) in complex notation, we have

$$s(t) = \text{Re}\left\{\sum_{n=-\infty}^{\infty} c_n g(t - nT)e^{j\omega_c t}\right\}, \tag{2}$$

where $c_n = a_n + jb_n$ and Re denotes the real part of a complex number. With a periodic data sequence, c_n, the signal $s(t)$ is periodic. This latter periodicity is best exhibited via the discrete Fourier transform[4] (DFT) of the periodic sequence. With the period of c_n denoted by N, the DFT of c_n is defined by

$$C(k\Omega) = \sum_{n=0}^{N-1} c_n e^{-jkn\Omega T} \qquad k = 0, 1, \cdots, N-1 \tag{3a}$$

and the inverse relation is

$$c_n = \frac{1}{N}\sum_{k=0}^{N-1} C(k\Omega)e^{jkn\Omega T} \qquad n = 0, 1, \cdots, N-1, \tag{3b}$$

where $\Omega = (1/N)(2\pi/T) = (1/N)\cdot(\text{symbol frequency})$. Hence, the DFT has N components uniformly spaced $1/NT$ Hz apart and the spectrum repeats every $2\pi/T$ Hz. Denoting the Fourier transform of $s(t)$ by $S(\omega)$ and convolution by $\circledast$, we have

$$S(\omega) = \tfrac{1}{2}\left[\sum_n c_n e^{-j\omega nT}G(\omega)\right] \circledast \delta(\omega - \omega_c), \qquad \omega > 0, \tag{4}$$

and using (3b) in (4) gives†

$$S(\omega) = \frac{1}{2}\sum_{k=0}^{N-1} C(k\Omega)\left[\sum_n G\left(k\Omega + \frac{2\pi n}{T}\right)\delta\left(\omega - \omega_c - k\Omega - \frac{2\pi n}{T}\right)\right]; \qquad \omega > 0. \tag{5}$$

Letting the timing frequency be denoted by $\omega_s = 2\pi/T = N\Omega$, it is

* In practice, the data would also be differentially and Gray encoded.
† We use the identity $\sum_n e^{-jn\omega T} = (2\pi/T)\sum_n \delta[\omega - (2\pi n/T)]$.

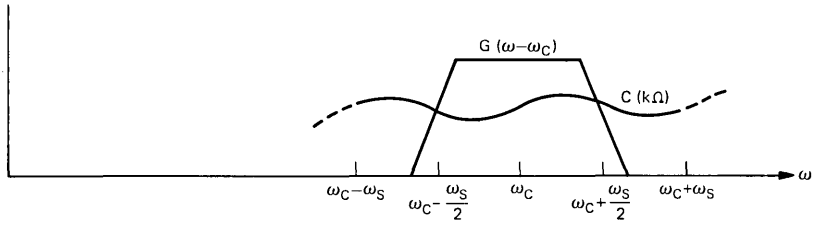

Fig. 3—The spectrum of the line signal for a periodic input is $S(w) = \mathrm{Re}\ \{\sum_{k=0}^{N-1} C(k\Omega)[G(k\Omega)\delta(\omega-\omega_c-k\Omega)+G(k\Omega-\omega_s)\delta(\omega-\omega_c-k\Omega+\omega_s)]\}$ where the input period is NT seconds.

clear that $S(\omega)$ has discrete tones at $\omega_c + k\Omega + n\omega_s$. In practical data-transmission systems with pulses of less than 100-percent excess bandwidth, $G(k\Omega + 2\pi n/T)$ will be zero if $n \neq 0$ or -1, hence,

$$S(\omega) = \frac{\pi}{NT} \sum_{k=0}^{N-1} C(k\Omega)[G(k\Omega)\delta(\omega - \omega_c - k\Omega) + G(k\Omega - \omega_s)\delta(\omega - \omega_c - k\Omega - \omega_s)]; \quad \omega > 0. \quad (6)$$

This spectrum is illustrated in Fig. 3, where it is seen that the envelope $C(k\Omega)$ modulates the baseband pulse shape, $G(\omega - \omega_c)$, in the range $\omega_c - \omega_s$ to $\omega_c + \omega_s$.

Since the signal $s(t)$ is used by the equalizer to adjust the tap weights, ideally the spectrum $C(k\Omega)$ should approximate that of random data, i.e., be constant. Of course, it is more critical that the equalizer be presented with a closely spaced line spectrum; for example, if the input period were two symbol intervals, it is clear that the equalizer can only compensate for distortion at two frequencies in the Nyquist band. Consequently, at the instant when the data return from the periodic to the random mode, the equalizer tap settings will be far from their optimum (for random data) values, and the distortion at the equalizer output could be much larger than the channel distortion. This situation generally causes the receiver to make so many errors that it is necessary to retrain the equalizer. As we shall see, the role of the scrambler is to lengthen the period of the transmitted sequence, thereby keeping the equalizer trained. Hence, for the rest of our discussion, we will assume that the scrambler is such that the periodic spectrum is (essentially) flat and densely spaced. Section VIII deals specifically with the factors that determine the period of the scrambler output.

IV. SPECTRUM OF THE ENVELOPE

The timing frequency is to be acquired from the envelope of the filtered line signal. Let the filtered line signal $m(t)$ be

$$m(t) = \sum_n a_n f(t - nT) \cos \omega_c t - \sum_n b_n f(t - nT) \sin \omega_c t, \quad (7)$$

where $f(t)$ is the (equivalent baseband) pulse shape after filtering at the receiver. The (squared) envelope of $m(t)$ is defined as

$$r(t) = [\sum_n a_n f(t - nT)]^2 + [\sum_n b_n f(t - nT)]^2. \tag{8}$$

As before, we introduce complex-signal notation by letting

$$d(t) = \sum_n c_n f(t - nT)$$

$$d^*(t) = \sum_n c_n^* f(t - nT), \tag{9}$$

so that we can write

$$r(t) = d(t) \cdot d^*(t), \tag{10}$$

where * stands for the complex conjugate. Thus, the Fourier transform of $r(t)$ is given by

$$R(\omega) = D(\omega) \circledast D_*(\omega) = [\sum_n c_n e^{-j\omega nT} F(\omega)] \circledast [\sum c_m^* e^{-j\omega mT} F(\omega)], \tag{11}$$

where $D_*(\omega)$ is the Fourier transform of $d^*(t)$, and $F(\omega)$ is the transform of $f(t)$. Using (3b), we have that

$$D(\omega) = \sum_n \left[\sum_{k=0}^{N-1} C(k\Omega) e^{-j(2\pi/N)kn} \right] e^{-j\omega nT} F(\omega)$$

$$= \sum_{k=0}^{N-1} C(k\Omega) F(\omega) \sum_n \delta(\omega - k\Omega - n\omega_s). \tag{12}$$

Substituting (12) into (11) and performing the convolution gives

$$R(\omega) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} C(k\Omega) C^*(l\Omega) \sum_n \sum_m F(k\Omega + n\omega_s) F(m\omega_s - l\Omega) \times \delta[\omega - (k - l)\Omega - (n + m)\omega_s]. \tag{13}$$

Evidently there are tones at $p\Omega + q\omega_s$ (where $p = k - l$ and $q = n + m$); the desired tone is at ω_s (i.e., $p = 0$, $q = 1$) and all other tones may be regarded as interferers. Again, practical bandlimiting of $F(\omega)$ and filtering of $R(\omega)$ will eliminate all terms where $q \neq 0$ or 1. The power in the desired tone is

$$R(\omega_s) = \sum_{k=0}^{N-1} |C(k\Omega)|^2 F(k\Omega) F(\omega_s - k\Omega), \tag{14}$$

while the power in an interfering tone (or sidelobe) is

$$R(\omega_s + p\Omega) = \sum_{k=0}^{N-1} C(k\Omega) C^*[(k - p)\Omega] F(k\Omega) F[(k - p)\Omega - \omega_s] \quad p = 1, 2, 3, \cdots. \tag{15}$$

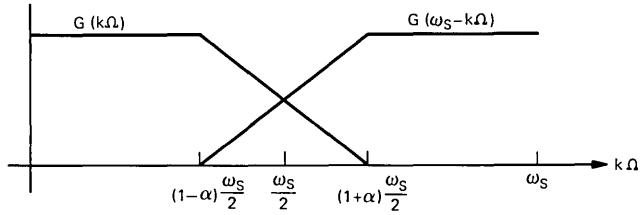

Fig. 4a—Strength of timing tone without loop filtering is $\sum_{k=0}^{N-1} |C(k\Omega)|^2 G(k\Omega) G(\omega_s - k\Omega)$.

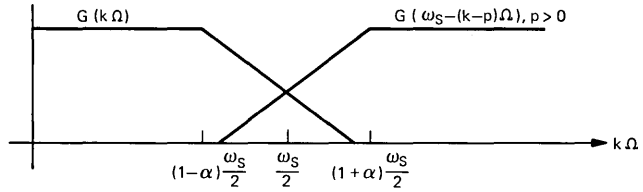

Fig. 4b—Strength of sidelobe interference at $\omega_s + p\Omega$ is $\sum_{k=0}^{N-1} C(k\Omega) C^*[(k-p)\Omega] G(k\Omega) G[(k-p)\Omega - \omega_s]$.

Conventionally, the signal $r(t)$ is fed to a phase-locked loop (PLL), which acts like a narrow-band filter in accepting tones in some region about ω_s (e.g., from $\omega_s - B\Omega$ to $\omega_s + B\Omega$, where $2B\Omega$ is the effective bandwidth of the loop) and produces an output that is dominated by a tone at ω_s. We first consider the above spectra in the absence of any timing loop filtering [i.e., $F(\omega) = G(\omega)$] as shown in Fig. 4. It is clear from (14) that the problem of timing-frequency recovery becomes more difficult as the amount of excess bandwidth (as measured by the parameter α) decreases—for zero excess bandwidth, this timing recovery technique clearly fails since the pulses $G(k\Omega)$ and $G(\omega_s - k\Omega)$ are disjoint. Figures 4a and 4b show how to compute the power of the tones at ω_s and at $\omega_s + p\Omega$ respectively. We note that, in general, $R(\omega_s + p\Omega) \neq R(\omega_s - p\Omega)$, and moreover, for the particular spectral shaping shown in the figure, it is clear that $R(\omega_s + p_1\Omega) > R(\omega_s + p_2\Omega)$ for $-B < p_2 < p_1 < B$; thus, half of the sidelobe tones are greater in magnitude than the desired tone. Thus, without any prefiltering in the timing loop, the desired tone is rather weak in comparison to the interfering tones. As we have already mentioned, this problem has a direct solution*: choose the loop filter $F(\omega - \omega_c)$ to be a narrow zonal filter around $\omega_c + (\omega_s/2)$ and $\omega_c = (\omega_s/2)$ as shown in Fig. 5. The resulting signal $m(t)$ has its energy concentrated at $\omega_c - (\omega_s/2)$ and $\omega_c + (\omega_s/2)$, and the relative strength of the timing tone is illustrated

* A filter in the timing loop has also been proposed by Franks and Bubrowski.[5]

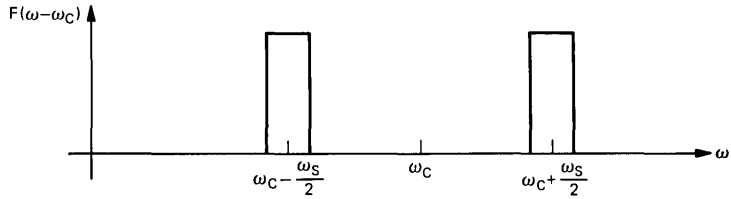

Fig. 5—Timing-loop filter shape which improves the tone-to-interference ratio.

in Figs. 6a, 6b, and 6c. The attenuation of the interferers is aided further by the fact that the magnitude of $\sum_{k=0}^{N-1} C_k C_{k-p}^*$ is a maximum for $p = 0$ (this follows from the Schwarz inequality). Clearly, by making $F(k\Omega) = \delta(k\Omega - \omega_s)$, we can make $R(\omega_s + p\Omega) = 0$ for all $p \neq 0$; however, *any narrow-zonal prefilter* of the type shown in Fig. 5 should significantly improve the relative strength of the timing tone. Since we merely require the filter to be narrow-band, any reasonable

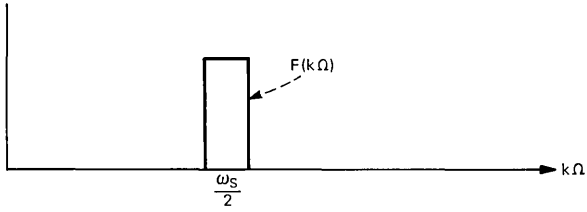

Fig. 6a—Strength of tone at $\omega_s \sim \sum_{k=0}^{N-1} |C(k\Omega)|^2 F(k\Omega)F(\omega_s - k\Omega)$.

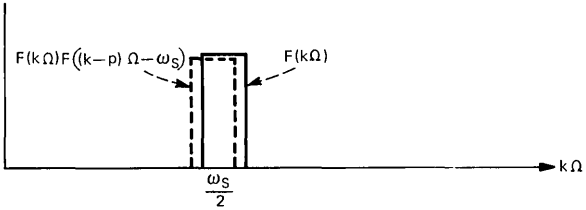

Fig. 6b—Strength of tone at $\omega_s + p\Omega \sim \sum_{k=0}^{N-1} C_k C_{k-p}^* F(k\Omega)F[(k-p)\Omega - \omega_s]$.

Fig. 6c—Spectrum of envelope.

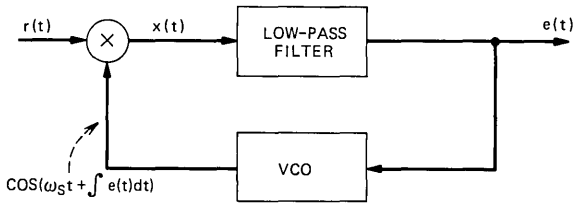

Fig. 7—First-order phase-locked loop.

choice wide enough to accommodate the uncertainty in ω_s would be adequate over a wide range of channel characteristics.

V. THE ROLE OF THE PHASE-LOCKED LOOP

As we have discussed, the signal $r(t)$ contains a desired tone as well as interfering tones. In this section, we wish to demonstrate that a phase-locked loop (PLL) can provide extremely narrow-band filtering even to the extent of extracting a single tone from a spectrum of adjacent interferers. Consider the standard first-order PLL[6] shown in Fig. 7. Let us assume that the input is the desired tone plus two interfering sidelobes, i.e.,

$$r(t) = A \sin \omega_s t + B \sin [(\omega_s + \Delta)t + \gamma] + B \sin [(\omega_s - \Delta)t - \gamma], \quad (16)$$

where Δ is the frequency displacement of the sidelobe from the desired tone and γ is the corresponding phase shift. Note that we have specialized the situation to the case where both interferers have the same amplitude and opposite phase angles (i.e., the distortion in the timing recovery system is symmetric about ω_s radians). We also assume that a perfectly tuned loop (i.e., the free-running frequency of the voltage-controlled oscillator (VCO) is ω_s) is employed.* From Fig. 7 the loop error signal is given by

$$e(t) = A \sin \left(\int e(t)dt + \alpha \right) + B \sin \left(\int e(t)dt + \alpha - \Delta t - \gamma \right) + B \sin \left(\int e(t)dt + \alpha + \Delta t + \gamma \right). \quad (17)$$

If we define $\phi(t)$ as the phase difference between VCO output phase and the PLL input phase corresponding to the desired tone, i.e.,

$$\phi(t) \triangleq \omega_s t - \left(\omega_s t + \int e(t)dt + \alpha \right), \quad (18)$$

* A perfectly tuned loop could arise by varying the free-running VCO frequency. As we show, via (22), when this condition is achieved the output will be a single tone. This observation suggests a feedback or error-sensing procedure for varying the nominal VCO frequency.

it is necessary that $\phi(t) \to 0$, as $t \to \infty$, since this implies successful tracking of the tone. Using (18) in (17) gives

$$e(t) = -[A + 2B \cos(\Delta t + \gamma)] \sin \phi(t), \tag{19}$$

and since from (18) $\dot{\phi}(t) = -e(t)$, the PLL is governed by the first-order differential equation

$$\frac{d\phi(t)}{dt} = -[A + 2B \cos(\Delta t + \gamma)] \sin \phi(t). \tag{20}$$

To solve (20), we first separate the variables and write

$$\frac{d\phi}{\sin \phi} = A + B \cos(\Delta t + \gamma) dt, \tag{21}$$

and, by direct integration, we obtain the solution

$$\phi(t) = 2 \tan^{-1} \{e^{-At} \exp(2B/\Delta)[\sin(\Delta t + \gamma) - \sin \gamma]\}. \tag{22}$$

We then have $\phi(t) \to 0$ as $t \to \infty$, i.e., the loop locks on the desired tone for any strength of the interference tone. Clearly, the same would be true for a collection of interferers provided they met the assumed symmetry conditions on their amplitude and phase. This example illustrates the power of a PLL to capture a desired tone in the presence of considerable interference.

VI. EFFECT OF PRECESSION ON THE RECOVERY OF A TIMING TONE

In modems not employing adaptive equalization, the question arises as to whether or not a scrambler is needed to generate a timing tone during the idle period. Since there is no equalizer in the system, we are not concerned with having a densely spaced line spectrum but only that there be at least two spectral lines, spaced ω_s apart, in the pass-band signal. Using the framework we have developed in the preceding sections, we investigate the effect of "precessing" the data symbols. Let us consider the phase-modulated signal

$$s(t) = \sum_{n=-\infty}^{\infty} g(t - nT) \cos(\omega_c t + \theta_n), \tag{23}$$

whose idle code is $\theta_n = 0$ for all n. The spectrum of $s(t)$ is, by using (5) with $C(k\Omega) = \delta_{k0}$ and $N = 1$,

$$S(\omega) = \sum_{n=-\infty}^{\infty} G(n\omega_s)\delta(\omega - \omega_c - n\omega_s), \tag{24}$$

and for an excess bandwidth of less than 100 percent,

$$S(\omega) = G(0)\delta(\omega - \omega_c); \tag{25}$$

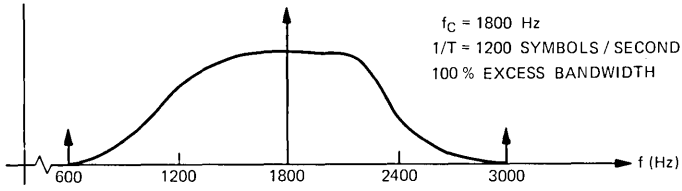

Fig. 8a—Spectrum without precession.

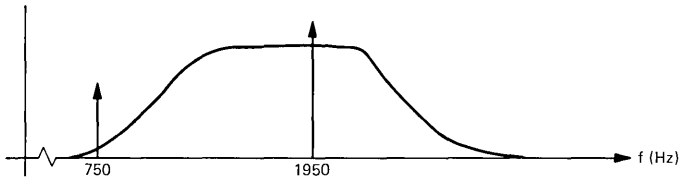

Fig. 8b—Spectrum with precession.

i.e., the spectrum consists of a single tone, which is obviously not sufficient to provide a timing tone. This situation is illustrated in Fig. 8a for a pulse $g(t)$ with 100-percent excess bandwidth and with $f_c = 1800$ Hz and $1/T = 1200$. To avoid the above situation, let $\theta_n = 2n\pi/M$, where $M/2$ is the number of points in the signal constellation and, thus, θ_n has period M. The advancing of the data symbol by $2\pi/M$ degrees, independently of any change in the input data, is known as *precession.** Using the notation in Section I, we have

$$c_n = e^{j\theta_n} = e^{j(2n\pi/M)},$$

$$C(k\Omega) = \sum_{n=0}^{M-1} c_n e^{-jnk(2\pi/M)} = \sum_{n=0}^{M-1} e^{-j(2\pi/M)n(k-1)} = \delta_{k-1}, \qquad (26)$$

which from (5) gives

$$S(\omega) = \sum_n G\left(n\omega_s + \frac{\omega_s}{M}\right)\delta\left(\omega - \omega_c - \frac{\omega_c}{M} - n\omega_s\right). \qquad (27)$$

Thus, the effect of precession is to offset the tones by ω_s/M, producing the spectrum shown in Fig. 8b. Clearly, when squared, this signal provides a tone at $\omega_s = 1200$.

The situation is different, however, for the spectrum shown in Fig. 9a with $\theta_n = 0$. The spectrum with precession shifts the tone by 100 Hz, and clearly no pair of in-band tones is present. Thus, for spectra that

* Differential phase modulation with precession would generate a data sequence $\theta_n = \theta_{n-1} + \phi_n + 2n\pi/M$, where ϕ_n is one of $M/2$ equally spaced angles.

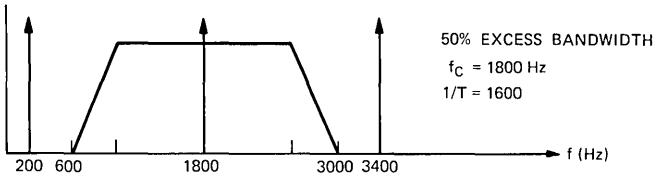

Fig. 9a—Signal spectrum without precession.

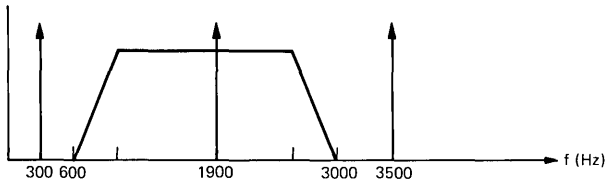

Fig. 9b—Signal spectrum with precession.

use small amounts of excess bandwidth, the tones necessary to provide timing would not be present with or without precession. Also, precession has little or no bearing on equalizer training since it simply shifts the spectrum and does not enrich it. For high-speed transmission in which the amount of excess bandwidth is small, spectral enrichment is provided by the scrambler and insures the proper operation of the equalizer and timing recovery system.

VII. SCRAMBLERS: BACKGROUND MATERIAL

We have shown in the preceding sections how the transmission of short repetitive patterns can play havoc with both the equalizer and timing recovery systems. As the name suggests, scramblers serve to "randomize" deterministic data sequences. The effect of this randomization on periodic sequences is to lengthen the period of the input sequence to the scrambler. Strictly speaking, the periodic output of the scrambler is not random. However, the scrambled data stream results in a line signal that has many more spectral components than the input data stream, and, thus, it looks more like the continuous spectrum that results when purely random data are encoded.

We confine our attention to the so-called self-synchronizing scrambler.[3] The generic forms for the self-synchronizing scrambler and the descrambler are shown in Figs. 10 and 11 and consist of, respectively, feedback and feedforward shift registers. Data symbols are fed into the scrambler every T seconds. These symbols are added (modulo p)*

* In practice, $p = 2$.

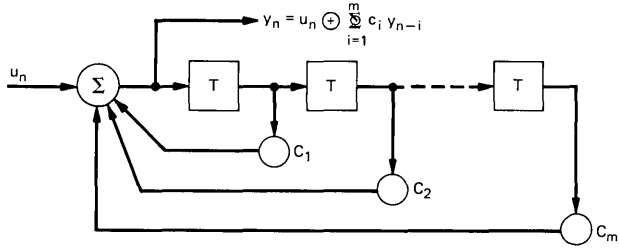

Fig. 10—Self-synchronizing scrambler.

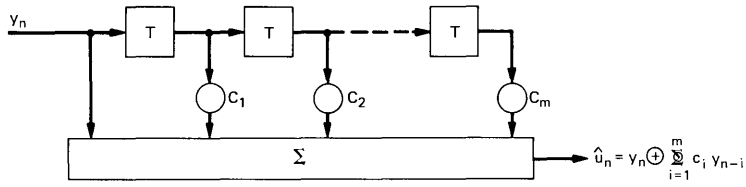

Fig. 11—Inverse scrambler.

to past outputs to produce the current output. The inputs to the delay elements shown in Fig. 10 are delayed by T seconds. The output of the scrambler is then encoded for transmission over the channel. After decoding at the receiver, the resulting sequence is put through a descrambler, shown in Fig. 11, where the original sequence is recovered. The inverse scrambler is self-synchronizing, and it will eventually cleanse itself of a transmission error once the error has propagated through the shift register. The number of errors in the descrambler output sequence is the number of channel errors multiplied by the number of nonzero tap weights in the shift register.

We shall study the input-output relationships of scramblers using d-transforms. Using this tool we are able, quite simply, to extend and clarify Savage's theorem[3] on scramblers with periodic inputs. Before getting into details on scrambler input-output relationships, a summary of some necessary background material on polynomials over Galois fields is in order.*

With p a prime number, we speak of a polynomial $Q(d)$ over $GF(p)$, where the coefficients of $Q(d)$ are modulo-p numbers. Multiplication, addition, and division of such polynomials are carried out in the usual fashion using modulo-p arithmetic on the coefficients. The *degree* of a polynomial $Q(d)$ is the highest power of d appearing in $Q(d)$. A polynomial of degree m is *irreducible* if it cannot be factored into poly-

* Much of the background material is taken from Ref. 7.

nomials of lower order. Two polynomials are relatively prime if they have no common factors. A crucial concept in our study of the scrambler is the *exponent* of a polynomial. The exponent of the polynomial $Q(d)$ is the minimum value of l such that $Q(d)$ divides $1 - d^l$, i.e., $(1 - d^l)/Q(d)$ is a polynomial of finite degree. For example, the exponent of the polynomial $1 + d^2 + d^3$ in $GF(2)$ is 7 since it divides $1 + d^7$, yielding $1 + d^2 + d^3 + d^4$, but it does not divide $1 + d^i$, $i < 7$. If the polynomials $P(d)$ and $Q(d)$ are relatively prime with exponents l_1 and l_2 respectively, it can be shown that the exponent of $P(d)Q(d)$ is the least common multiple (lcm) of l_1 and l_2. The exponent of $[Q(d)]^j$, where $Q(d)$ is over $GF(p)$, is $p^r l$, where l is the exponent of $Q(d)$ and r is such that $p^{r-1} < j \leqq p^r$. An irreducible polynomial of degree m is *primative* or of maximum exponent if its exponent is $p^m - 1$. Given a polynomial $Q(d)$ of order m, its reciprocal polynomial is $d^m Q(1/d)$, and it is known that reciprocal polynomials of irreducible polynomials are themselves irreducible, and that reciprocal polynomials of primative polynomials are themselves primative.

This theory of polynomials over a Galois field is applicable to the d-transforms[7] of the input and output sequences of a scrambler. Consider a time series $x_0, x_1, x_2, \cdots$, such that the x_i, $i = 0, 1, \cdots$ are elements from a Galois field, e.g., $01101\cdots$. The d-transform of this series is defined as

$$X(d) \equiv \sum_{i=0}^{\infty} x_i d^i, \tag{28}$$

and inversion is accomplished by "reading" the coefficients of $X(d)$. The d-transform of a periodic sequence is of the form $R(d)/(1 - d^\lambda)$, where λ is the period of the sequence and $R(d)$ is a polynomial, of degree less than λ, in d over $GF(P)$. To illustrate, suppose we have a series of elements in $GF(3)$, 1021, 1021, $\cdots$. Using the relationship for a geometric progression we find that the d-transform of this series is $(1 + 2d^2 + d^3)/(1 - d^4)$. In general, it can be shown that the d-transform of a periodic time series is of the form $P(d)/Q(d)$, where $P(d)$ and $Q(d)$ are polynomials over a Galois field. If $P(d)$ and $Q(d)$ are relatively prime, the period of the time series represented by $P(d)/Q(d)$ is the exponent of $Q(d)$.

Linear sequential machines over $GF(p)$ are composed of modulo-p adders, multipliers, and delay elements connected according to a few elementary rules. As the name implies, such circuits are *linear* over modulo-p arithmetic. The laws of commutativity, associativity, and superposition apply. For example, the response of a circuit to the sum of two inputs is the sum of the responses to each input separately. The summations are carried out term-by-term modulo p on the input and output sequences.

The response of such circuits to inputs can be found using the classical techniques of linear system theory. The output consists of the sum of the *free response* and the *forced response.* The free response is due solely to initial conditions within the circuit with no input. If the circuit is in the quiescent state, i.e., it has zero initial conditions, there is no output without an input. The forced response is the output when an input is applied to a circuit in the quiescent state. As in the case of conventional linear circuits, the forced response can be found by convolving the impulse* response with the input sequence. Thus, if h_n, $n = 0, 1, \cdots$ is the impulse response of a circuit and u_n is its input at time nT, $n = 0, 1, \cdots$, then the output at time nT is

$$y_n = \sum_{k=0}^{n} h_k u_{n-k}, \tag{29}$$

where $\sum$ indicates summation modulo p. If we take the d-transform of both sides of (29), we find

$$Y(d) = U(d)H(d), \tag{30}$$

where $U(d)$ and $H(d)$ are the d-transforms of u_n and h_n respectively, and where $n = 0, 1, \cdots$.

VIII. SCRAMBLER INPUT-OUTPUT RELATIONSHIPS

Scramblers are linear sequential circuits and their input-output relationships can be found using linear system theory. In this section, we wish to demonstrate the utility of the d-transform approach in characterizing the nature of the output sequence for a given input sequence. Consider the m-stage scrambler shown in Fig. 10 with feedback coefficients $c_1, \cdots, c_m$. The output at time nT, y_n, is given by

$$\begin{aligned} y_n &= c_1 s_{1n} \oplus c_2 s_{2n} \oplus \cdots \oplus c_m s_{mn} \oplus u_n \\ s_{1n} &= y_{n-1} \\ s_{in} &= s_{i-1,n-1} \qquad i \geqq 2, \end{aligned} \tag{31}$$

where u_n is the input at time nT and s_{kl} is the output of the kth delay element at time l.† Now we find the impulse response. Let

$$u_n = \begin{cases} 1 & n = 0 \\ 0 & n > 0 \end{cases}$$

and

$$s_{i0} = 0, \forall i.$$

---

* By an impulse we, of course, mean a time series which is unity at the time origin and is zero elsewhere.

† The output may be rewritten as $y_n = \sum_{i=1}^{m} c_i y_{n-i} \oplus u_n$. At the descrambler we form $z_n = y_n \oplus \sum_{i=1}^{m} c_i y_{n-i}$, which recovers the input when there are no channel errors.

The output sequence can be written

$$y_n = \begin{cases} 1 & n = 0 \\ \sum_{i=1}^{n} c_i y_{n-1} & n \leqq m \\ \sum_{i=1}^{m} c_i y_{n-i} & n > m. \end{cases} \tag{32}$$

If we take d-transforms of both sides of (32) we find after some manipulation that the transform of the impulse response, i.e., the transfer function, is given by

$$H(d) \triangleq \sum_{k=0}^{\infty} y_k d^k = 1 \Big/ \left(1 - \sum_{i=1}^{m} c_i d^i\right). \tag{33}$$

The d-transform of the forced response of the scrambler can be found from eqs. (30) and (33).

The free response of the scrambler can also be found from (31) when $u_n = 0$, for all n. We begin by assuming a particular initial state vector. Assume that the output of all of the delay elements but one are zero. Let the nonzero output be that of the ith delay element and denote this output by s_{i0}. It can be shown that y_n^i, the output of the scrambler due solely to state s_{i0}, is

$$y_n^i = \begin{cases} c_i s_{i0} & n = 0 \\ \sum_{j=1}^{n} c_j y_{n-j}^i + c_{i+n} s_{i0} & 0 < n \leqq m - i \\ \sum_{j=1}^{m} c_j y_{n-j}^i & n > m - i. \end{cases} \tag{34}$$

If we take the d-transform of both sides of eq. (34), we find that the d-transform of the response to initial condition s_{i0} is

$$Y^i(d) = \left(s_{i0} \sum_{k=0}^{m-i} c_{i+k} d^k\right) \Big/ \left(1 - \sum_{j=1}^{m} c_j d^j\right). \tag{35}$$

Now, to find the response to any initial condition $s_{10}, s_{20}, \cdots, s_{m0}$, we sum over i. Thus, the d-transform of the free response is

$$Y_{\text{free}}(d) = S(d) \Big/ \left(1 - \sum_{j=1}^{m} c_j d^j\right) = S(d)H(d), \tag{36}$$

where $S(d) \triangleq \sum_{i=1}^{m} s_{i0} \sum_{k=0}^{m-i} c_{i+k} d^k$.

A fact that is crucial to our analysis in the sequel is that the polynomial $S(d)$ spans the space of polynomials of degree $m - 1$ in $GF(p)$. By choosing the initial conditions s_{i0}, $i = 1, 2, \cdots, m$, $S(d)$ can be any

polynomial of degree $m - 1$. To show this, suppose we have the polynomial $T(d) = t_0 + t_1 d + \cdots + t_{m-1} d^{m-1}$. Equating $T(d)$ and $S(d)$ term by term, we have m equations in m unknowns, $s_{10}, s_{20}, \cdots, s_{m0}$. The equations can be represented in the form

$$\begin{bmatrix} c_m & 0 & \cdots & 0 \\ c_{m-1} & c_m & \cdots & 0 \\ \vdots & \vdots & & \\ c_1 & c_2 & \cdots & c_m \end{bmatrix} \begin{bmatrix} s_{10} \\ s_{20} \\ \vdots \\ s_{m0} \end{bmatrix} = \begin{bmatrix} t_{m-1} \\ t_{m-2} \\ \vdots \\ t_0 \end{bmatrix}. \tag{37}$$

The $m \times m$ lower triangular matrix in (37) is nonsingular (since $c_m \neq 0$); therefore the m simultaneous equations have a unique solution.

From the foregoing, we see that the total response of a scrambler to an input, with transform $U(d)$, is

$$Y(d) = [U(d) + S(d)]/\Phi(d), \tag{38}$$

where $\Phi(d) \triangleq 1 - \sum_{i=1}^{m} c_i d^m$ is the transform of the feedback coefficients. The above equation completely describes the behavior of the scrambler to any input for any given initial state.

Now we consider the input-output relationships for the scrambler based on eq. (38). Throughout our discussion we shall assume that $\Phi(d)$ is a primitive polynomial implying that it has exponent $\phi = p^m - 1$, and thus can be written as $(1 - d^\phi)/\Phi'(d)$, where $\Phi'(d)$ is a finite degree (remainder) polynomial of degree $\phi - m$. Note that $\Phi'(d)$ is one "cycle" of the periodic polynomial $1/\Phi(d)$. Suppose that the input is zero, the transform of the output is simply $S(d)/\Phi(d)$. Since the degree of $S(d)$ is one less than that of $\Phi(d)$, $S(d)$ and $\Phi(d)$ are relatively prime,* and the output transform is $S(d)\Phi'(d)/(1 - d^\phi)$; hence, the output is periodic with period $= p^m - 1$. If the input is a sequence of finite duration j, then $U(d)$ is a polynomial of degree $j - 1$. If $j \leqq m$, then the above output transform is $U(d)\Phi'(d)/(1 - d^\phi)$, and since the degree of the numerator is less than ϕ, it is clear that the output is purely periodic with period ϕ. Note that there is no output transient. If $j \geqq m$, and if $U(d) + S(d)$ and $\Phi(d)$ are relatively prime, then it is easy to show that the output consists of a transient component $(j + 1 - m)$ long† and a periodic component with period ϕ. For any input $U(d)$ of finite duration, there are a unique set of initial conditions

---

* Since $\Phi(d)$ is a primative polynomial, $S(d)$ cannot be a factor of $\Phi(d)$; and since the degree of $S(d)$ is less than a $\Phi(d)$, $\Phi(d)$ cannot be a factor of $S(d)$. Thus, $S(d)$ and $\Phi(d)$ are relatively prime.

† This should be intuitively clear, since once $j - (m - 1)$ bits are accepted in the scrambler, the situation is one where the (remaining) input sequence is of a length less than m.

$S(d)$ such that

$$U(d) + S(d) = T(d)\Phi(d),$$

where $T(d)$ is some polynomial. In this situation the output has finite duration given by $T(d)$, i.e., the periodic component of the solution has been annihilated. To show this, we cite the following theorem.[8] Let $U(d)$ and Φ (d) be polynomials in $GF(p)$. Then there are unique polynomials $T(d)$ and $S(d)$ in $GF(p)$ such that $U(d) + S(d) = T(d)\Phi(d)$, where $S(d) \equiv 0$ or $S(d)$ is of lower degree then $U(d)$. Recall that by suitably choosing initial conditions, $S(d)$ can be any polynomial of degree $m - 1$ over $GF(P)$.

We turn to the important case of periodic inputs. The input sequence $U(d)$ can always be written in the form $U(d) = P(d)/Q(d)$, where $P(d)$ and $Q(d)$ are relatively prime. Let the exponent of $Q(d)$ be ℓ, i.e., the period of the input is ℓ. The d-transform of the output becomes

$$Y(d) = [S(d)Q(d) + P(d)]/\Phi(d)Q(d). \tag{39}$$

We consider first the case where $\Phi(d)$ and $Q(d)$ are relatively prime. If the numerator and denominator of eq. (39) are relatively prime, then, using the background material presented in Section VII, it is clear that the output is periodic with period N, where $N = \text{lcm } (l, p^m - 1)$. However, we will show that given $P(d)$ and $Q(d)$, there is a set of initial conditions for which

$$S(d)Q(d) + P(d) = T(d)\Phi(d), \tag{40a}$$

where $T(d)$ has degree $l - 1$. When (40) holds, the output period is l. Thus, assuming that all initial states are equiprobable, with probability p^{-m} the initial state will be such that the scrambler "locks up" and the output period equals the input period. (As we have previously mentioned, this is a very undesirable situation.) To support (40) we cite the following theorem.[8] There exist (unique) polynomials $T'(d)$ and $S'(d)$ such that

$$S'(d)Q(d) + T'(d)\Phi(d) = 1 \tag{40b}$$

only if $Q(d)$ and $\Phi(d)$ are nonzero relatively prime polynomials over $GF(p)$. Now multiply both sides of the above equation by $-P(d)$ and let $S(d) = -P(d)S'(d)$ and $T(d) = P(d)T'(d)$. Again we make use of the fact that $S(d)$ spans the space of polynomials of degree $m - 1$ to guarantee that for every $S'(d)$ there corresponds a $S(d)$. We now summarize the above.

Let the scrambler be defined by the primative polynomial $1 - \sum_{i=1}^{m} c_i d^i$, and also suppose that the transform of the input to the

scrambler is $P(d)/Q(d)$, where $P(d)$ and $Q(d)$ are relatively prime. It is also assumed that $\Phi(d)$ and $Q(d)$ are relatively prime. For a particular set of initial conditions, the output period of the scrambler is the input period, l, where l is the exponent of $Q(d)$. For all other initial conditions the output period is the least common multiple of l and $p^m - 1$.

Our description is the same as Savage's Theorem 1 with two differences—one superficial, the other crucial. Savage requires the polynomial $h(d) = d^m - \sum_{i=1}^{m} c_i d^{m-i}$ to be primative. However, $1 - \sum_{i=1}^{m} c_i d^i$ and $h(d)$ are reciprocal polynomials and, as we have seen, the reciprocals of primative polynomials are themselves primative with the same exponent. The second requirement is that $\Phi(d)$ and $Q(d)$ be relatively prime. This requirement, which is not part of Savage's theorem, is essential for a complete description of scrambler behavior.*

We will now show that the requirement that $\Phi(d)$ and $Q(d)$ be relatively prime is satisfied whenever the exponent of $Q(d)$ is not a multiple of $p^m - 1$. The proof is by contradiction. Suppose $\Phi(d)$ and $Q(d)$ are not relatively prime, then it is possible to write†

$$Q(d) = R(d)\Phi^j(d) \qquad j = 1, 2, \cdots, \tag{41}$$

where $R(d)$ is a polynomial, with exponent r, which is relatively prime to $\Phi(d)$. The exponent of $Q(d)$ is lcm $[r, p^k(p^{m-1})]$, where k is such that $p^{k-1} < j \leqq p^k$. Clearly, the exponent of $Q(d)$ is a multiple of $p^m - 1$, thus proving the desired result. Thus, when the input period is less than p^{m-1} (the practical case), then $\Phi(d)$ and $Q(d)$ are relatively prime. It is interesting to note that even if the input to the scrambler has an exponent which is a multiple of $p^m - 1$, it may be that $Q(d)$ and $\Phi(d)$ are still relatively prime. For example, $Q(d)$ can be the reciprocal polynomial to $\Phi(d)$.

Consider now the situation when $Q(d)$ and $\Phi(d)$ are not relatively prime. As above, we can then factor $Q(d)$ in the form $Q(d) = \Phi^j(d)R(d)$, $j \geqq 1$, where $R(d)$ is either 1 or a polynomial relatively prime to $\Phi(d)$. From (39) and (41) the d-transform of the output is

$$Y(d) = [S(d)\Phi^j(d)R(d) + P(d)]/\Phi^{j+1}(d)R(d). \tag{42}$$

Since by assumption $P(d)$ is relatively prime with $Q(d) = \Phi^j(d)R(d)$, the numerator and denominator of (42) are relatively prime. Since $\Phi^{j+1}(d)$ and $R(d)$ are relatively prime, the output period is then the least common multiple of $p^k(p^m - 1)$ and r, with k given by $p^{k-1} < j$

---

* In other words, Savage states that, apart from the special case when the output period equals the input period, the output period is the lcm(l, p^{m-1}). This is not strictly true since, as we shall show, if $\Phi(d)$ and $Q(d)$ are not relatively prime, the output period is not necessarily the lcm(l, p^{m-1}).

† Since $\Phi(d)$ is irreducible, we could not write $Q(d)$ as a factor of $\Phi(d)$.

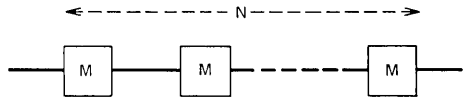

Fig. 12—Cascade of N M-bit scramblers.

$+ 1 \leqq p^k$. Note that this result holds independently of initial conditions.*

The above discussion provides a refinement of Savage's basic result† by indicating that the output period is contingent on whether or not $\Phi(d)$ and $Q(d)$ are relatively prime. It was shown that if the exponent of $Q(d)$ is not a multiple of $p^m - 1$, then $\Phi(d)$ and $Q(d)$ are relatively prime. However, when $Q(d)$ and $\Phi(d)$ are not relatively prime the output period must be determined from (42).

IX. CASCADED SCRAMBLERS

The cascade of identical scramblers provides an interesting example of when $Q(d)$ and $\Phi(d)$ are not relatively prime. Suppose, as in Fig. 12, we have n identical m-stage scramblers in tandem so that the output of the first is the input to the second and so on. Thus, assuming no lockup, the input to the second stage will have the same period as the free-running period of the second stage. With $S(d) = 0$ for all the scramblers, the output transform of the nth scrambler is

$$Y(d) = U(d)/\Phi^n(d), \tag{43}$$

where $U(d)$ is the transform of the input. Consider an example where $U(d) = 1/(1 + d) = 1/Q(d)$ with $p = 2$. Note that the exponent of $Q(d)$ is unity. The transform of the second output is $1/Q(d)\Phi^2(d)$, and we apply the results of the previous section to show that the output period is $2(2^m - 1)$, i.e., $k = 1$. Applying the above result to successive stages produces the data in Table I, which shows the period of the output as a function of n for a binary scrambler ($p = 2$). Table I points out that adding a stage in cascade does not always double the output period.

By considering each scrambler successively, we can comment on the output period of the cascade scrambler for arbitrary initial conditions and input, assuming $Q(d)$ and $\Phi(d)$ are relatively prime. Let the input to the first scrambler have exponent $l < p^m - 1$. From Section VIII

* If cancellation between the numerator and denominator in (42) were to occur, (40) would imply that $P = \Phi(1 + SR\Phi^{j-1})$. Now since (41) states that $Q = R\Phi^j$, it is clear that P and Q have the common factor Φ. This contradicts the assumption that P and Q are relatively prime. Thus, under the above conditions (i.e., Q and Φ are not relatively prime), the initial condition cannot force the output period to equal the input period.

† Results similar to ours were stated without proof in Ref. 9.

Table I

n	Output Period
1	$(2^m - 1)$
2	$2\ (2^m - 1)$
3	$4\ (2^m - 1)$
4	$4\ (2^m - 1)$
5	$8\ (2^m - 1)$
6	$8\ (2^m - 1)$
7	$8\ (2^m - 1)$
8	$8\ (2^m - 1)$
9	$16\ (2^m - 1)$

we know that the probability of the output having period l is p^{-m}. Otherwise, the output has period $\text{lcm}(l, p^m - 1)$. If the input to the second scrambler has period l we have the same situation as the first scrambler. However, if the output of the first scrambler has period lcm $(l, p^m - 1)$, the input polynomial to the second scrambler has denominator $Q(d)\Phi(d)$. By an argument analogous to that surrounding (42), it is clear that (40) cannot be satisfied, since $Q(d)\Phi(d)$ and $\Phi(d)$ are not relatively prime; thus, the scrambler cannot "lock up," and applying the results in Table I indicates that the output of the second scrambler has period $\text{lcm}[l, p(p^m - 1)]$. Thus, if a particular scrambler does not lock up, then no succeeding scrambler can lock up. The situation is summarized in Table II for four binary scramblers in tandem. We assume in Table II that all initial states are equiprobable.

We compare Table II to the serial scrambler in which all delay elements are combined into a scrambler that has $4m$ elements. With input period l, the output period is l with probability 2^{-4m} and $\text{lcm}(l, 2^{4m} - 1)$ with probability $1 - 2^{-4m}$. Both the cascade and serial scramblers lock up and have period l with the same probability (2^{-4m}); however, since

(*i*) the longest period of the cascade scrambler, $\text{lcm}[l, 4(2^m - 1)]$, is less than the largest period of the serial scrambler, $[\text{lcm}(l, 2^{4m} - 1)]$, and

Table II

Output Period	Probability
l	2^{-4m}
$\text{lcm}(l, 2^m - 1)$	$2^{-3m}(1 - 2^{-m})$
$\text{lcm}[l, 2(2^m - 1)]$	$2^{-2m}(1 - 2^{-m})$
$\text{lcm}[l, 4(2^m - 1)]$	$1 - 2^{-2m}$

(*ii*) the probability of the cascade scrambler attaining its largest period $(1 - 2^{-2m})$ is less than the probability of the serial scrambler attaining its largest period $(1 - 2^{-4m})$,

the superiority of the serial over the cascaded scrambler in terms of spectral density is clear.

X. PARALLEL SCRAMBLERS

Serial scramblers have the property that if a single bit error is made in demodulation, then M errors will appear in the unscrambled output sequence, where M is the number of nonzero coefficients in the scrambler primative polynomial. A parallel scrambler configuration has been proposed to ameliorate this error multiplication. In this section, we shall illustrate a spectral cancellation effect that can take place with parallel scrambling. For simplicity, we shall consider two parallel data streams. Suppose that the binary data, as in Fig. 13, are split into two data streams a_n' and b_n', where a_n is the scrambled* version of a_n', while $b_n = a_{n-m} \oplus b_n'$. The a_n and the b_n streams are then encoded for transmission over the channel. At the receiver, inverse operations recover the a_n' and the b_n' streams. A channel error in the a_n stream will cause M errors in the a_n' stream and one error† in the b_n' stream. A channel error in the b_n stream will cause a single error in the b_n' stream. Now suppose that a_n and b_n are Gray encoded so that a_n is the least significant bit. The result is that the probability of error in the a_n stream is much less than the probability of error in the b_n stream; thus, the average number of errors in the a_n' and the b_n' streams will be decreased compared with serial scrambling.

Now we wish to examine the effect of "slaving" the b_n stream to the a_n stream. For our purposes it will be sufficient to code the scrambled output sequences into 1 and -1, i.e., the transmitted data sequence is given by‡ (recall the notation of Section III)

$$\begin{aligned} c_n &= (2a_n - 1) + j(2b_n - 1) \\ &= 2a_n - 1 + j[2(a_{n-m} \oplus b_n') - 1]. \end{aligned} \tag{44}$$

As we have shown in Sections III and IV, *both* the line and envelope spectra critically depend on the discrete Fourier transform (DFT) of the c_n sequence, $C(k\Omega)$. Unfortunately, it is not possible to express

* At the inverse scrambler, a_n' is recovered as in the standard configuration, while b_n' is estimated as $b_n + a_{n-m}$.

† Since the estimated b_n' is formed as the "mod 2" sum of b_n and a_{n-m}, a single channel error will only affect the b_n' output once; however, the a_n' output will see the propagation of this error through the shift register.

‡ The function $(2a_n - 1)$ maps "0" into "-1" and "1" into "1" and thus serves as a mapping from the scrambler output sequence to the transmitted (line) sequence.

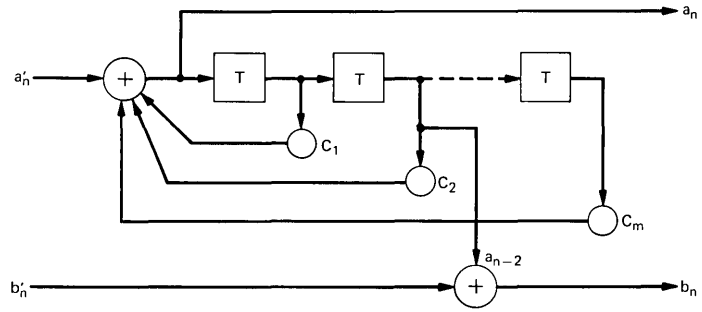

Fig. 13—Parallel scrambler configuration.

$C(k\Omega)$ directly in terms of the scrambler primative polynomial, which is a most difficult problem since, in the space of *real*-valued sequences, the DFT is a linear operation and the scrambler output is a nonlinear function of the input [of course, the scrambler input and output are linearly related in $GF(p)$]. However, in the present situation we are able to proceed since we are only interested in illustrating the possibility of spectral cancelling due to the parallel structure. We first indicate two simple relations between mod 2 operations and the corresponding real variable operations: with $a, b \in GF(2)$,

$$a \oplus b = (a - b)^2 \tag{45a}$$

$$a^2 = a. \tag{45b}$$

Using (45) we write (44) as

$$\begin{aligned} c_n &= (2a_n - 1) + j[2(a_{n-m} - 2b'_n a_{n-m} + b_n) - 1] \\ &= 2(a_n + ja_{n-m}) + 2jb'_n - 4jb'_n a_{n-m} - (1 + j). \end{aligned} \tag{46a}$$

Let us consider the effect of the first term on the spectrum of c_n. Now with L and $C(k\Omega)$ denoting respectively the period of a_n (and b_n) and the DFT of $a_n + ja_{n-m}$, we have*

$$\begin{aligned} C(k\Omega) &= [1 + je^{-j(2\pi/L)mk}]A(k\Omega) \\ &= 1 + e^{j[(\pi/2)-(2\pi/L)mk]}A(k\Omega). \end{aligned} \tag{46b}$$

Suppose that the scrambler produces a flat output spectrum (i.e., it would be a satisfactory scrambler if used solely in the serial mode), it is clear that $C(k\Omega)$ will have periodically spaced nulls. Since the energy in the timing tone is given by

$$R(\omega_s) = \sum_{k=0}^{N-1} |C(k\Omega)|^2 F(k\Omega)F(\omega_s - k\Omega), \tag{14}$$

* For the purposes of our discussion, we, in effect, assume that $b_n = 0$, i.e., $c_n = a_n + ja_{n-m}$.

any amplitude tapering provided by $|C(k\Omega)|^2$ could impair the timing recovery system. From (46) we have

$$\begin{aligned} |C(k\Omega)|^2 &= [1 + \sin(2\pi/L)mk]^2 + [\cos(2\pi/L)mk]^2 \\ &= 2[1 + \sin(2\pi/L)mk], \end{aligned} \tag{47}$$

where $k = 1, 2, \cdots, L$, and $m = 1, 2, \cdots, M$ with M being the number of stages in the scrambler. The strength of the tone, described by (14), will be particularly attenuated when $|C(k\Omega)|$ has a null at $k = L/2$, which corresponds to a frequency of $\omega_s/2$. It is easy to see that for some values of "m," the attenuation of the tone can be quite severe near $\omega_s/2$ (i.e., $k = L/2$).

In practice, the remedy is to change the value of m so that the null occurs as far away from $\omega_s/2$ as is possible. Of course, this cannot be done prior to transmission since the exact value of ω_s is unknown.

In this section, we have described a possible pitfall associated with the use of a parallel scrambler configuration. In practice, whether or not there is severe attenuation of the timing tone depends on the details of the pulse shaping and the operation of the phase-locked loop.

XI. CONCLUSIONS

We have examined several problems occurring in data-transmission systems that employ envelope-derived timing recovery, adaptive equalization, and self-synchronizing scramblers. Several conclusions have been reached regarding both the individual and joint action of these subsystems.

(*i*) The performance of the envelope-derived timing recovery system can be significantly improved by narrow-zonal prefiltering of the received signal prior to extracting the envelope.

(*ii*) The technique of "precessing" the data symbols in a phase-modulated modem is sufficient to provide a timing tone in a large excess-bandwidth system, but does not provide a tone in a small excess-bandwidth system.

(*iii*) A complete description was given of the output period of a cascaded scrambler as a function of the number of stages. Of interest are the facts that the output period does not necessarily double with the addition of a stage, and that if a particular scrambler stage does not lock up, then no succeeding stage can lock up.

(*iv*) It was demonstrated that the parallel scrambler configuration can, via spectral cancellation, cause the strength of the timing tone to be attenuated.

XII. ACKNOWLEDGMENT

The authors would like to thank T. M. Dennis for apprising them of the interaction between the parallel scrambler configuration and the timing-recovery system.

REFERENCES

1. W. R. Bennett, "Statistics of Regenerative Digital Transmission," B.S.T.J., *37*, No. 6 (November 1958), pp. 1501–1542.
2. Y. Takasaki, "Timing Extraction in Baseband Pulse Transmission," IEEE Trans. Commun., *COM-20*, No. 5 (October 1972), pp. 877–884.
3. J. E. Savage, "Some Simple Self-Synchronizing Digital Data Scramblers," B.S.T.J., *46*, No. 2 (February 1967), pp. 449–487.
4. B. Gold and C. M. Rader, *Digital Processing of Signals*, New York: McGraw-Hill, 1969.
5. L. E. Franks and J. P. Bubrowski, "Statistical Properties of Timing Jitter in a PAM Timing Recovery Scheme," IEEE Trans. Commun., *COM-22*, No. 7 (July 1974), pp. 913–920.
6. A. J. Viterbi, *Principles of Coherent Communication*, New York: McGraw-Hill, 1966.
7. A. Gill, *Linear Sequential Circuits*, New York: McGraw-Hill, 1967.
8. H. Pollard, *Theory of Algebraic Numbers*, New York: John Wiley, 1950.
9. K. Nakamura and Y. Iwaderl, "Data Scramblers for Multilevel Pulse Sequences," NEC Research and Development, No. 26 (July 1972), pp. 53–63.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

Analysis and Optimal Design of a Multiserver, Multiqueue System With Finite Waiting Space in Each Queue

By B. A. WHITAKER

(Manuscript received July 29, 1974)

An analysis of a particular type of multiserver, multiqueue system is presented in which each queue has a finite number of waiting positions and the waiting positions are not *vacated until service is completed. Thus, several customers in one queue can be served simultaneously. The steady-state distribution of states is derived and is used to obtain the probability of loss for each queue and the average delay of the system. This analysis is then used in the development of a design procedure to determine the minimum-cost configuration of waiting positions and servers to meet specified single-hour grade-of-service constraints. The results are applicable to the design of systems that utilize automatic call distributors. While this model does not include such effects as day-to-day variation and noncoincidence of peak loads among trunk groups, nevertheless the results properly reflect for the first time the interactions among the trunk groups terminated on the automatic call distributor and the attendants at the automatic call distributor.*

I. INTRODUCTION

The purpose of this paper is to present an analysis of a particular type of multiserver, multiqueue system in which each queue has a finite number of waiting positions and the waiting positions are *not* vacated until service is completed. In particular, the steady-state distribution of states are derived, and expressions for the probability of loss for each queue and the average delay are given. It is shown that the queuing model described is representative of systems characterized by a finite number of trunk groups that carry calls to a group of attendants who then perform some service for the caller. (One such system is used in the directory assistance service provided by the telephone company.) Results are given to illustrate the effects of varying the number of servers and number of positions in each queue. Finally, a design procedure to determine the minimum-cost configura-

tions for such systems under various grade-of-service constraints is developed. This procedure ignores such effects as day-to-day variation, noncoincidence of peak loads among incoming trunk groups, and retrials of blocked calls, which should be investigated in the development of procedures for traffic engineering and administration. However, as in most cases, it is difficult, if not impossible, to obtain analytical results with these effects included. This paper should provide useful insight that can later be incorporated in a complete traffic engineering procedure.

The system analyzed consists of l input queues each with a finite number of waiting positions, N_i ($i = 1, \cdots, l$), which have full access to M servers. When an arrival seizes one of the M servers, it does *not* vacate a waiting position, but remains in the position until its service has been completed. This characteristic, which allows calls that are not at the head of a queue to be in service, distinguishes this system from the usual queuing system. In particular, the system is no longer completely described by the number of calls in each queue since a record of the number of calls from each queue that are in service must be kept. An arrival that finds all the waiting positions for its queue occupied is cleared or lost from the system. An arrival finding no idle servers but at least one vacant waiting position in its queue enters the queue and is delayed until its service begins. In the context of directory assistance systems, the input queues are the trunk groups and the waiting positions are represented by the trunks. The information operators are the servers.

In telephone traffic theory, the described system has been referred to as a combined loss-and-delay system. Previous work in this subject can be segmented into three parts:

(*i*) One input flow—one queue.[1–8]
(*ii*) Several input flows—one queue.[9–11]
(*iii*) Several input flows—several queues.[12–14]

The last segment, of which this analysis is a part, has been investigated by Kühn. He analyzed systems with $g > 1$ queues, each with a finite number of waiting positions s_i ($i = 1, 2, \cdots, g$). Associated with each queue is a Poisson arrival process with mean rate λ_i, which is assumed to be independent of the others. An arrival that finds all waiting positions in its queue occupied is lost. Arrivals that are not cleared from the system are served by one of n servers. The service time distribution for the ith server is exponential with a mean rate ϵ_i. When a server becomes idle, queue i is chosen to receive service with probability p_i. Within a queue, calls can be selected randomly, first-come, first-served, or according to a priority scheme.

As Kühn indicates, analytical solutions to these systems exist only for special cases. In general, the linear equations representing the equilibrium conditions must be solved numerically for the particular values of the parameters. However, Kühn gives a solution to one particular system, which will now be discussed. In this system, it is assumed that the service rate for all servers is identical ($\epsilon_i = \epsilon$) and the interqueue discipline is defined by the p_j's that are

$$p_j = \frac{Z_j}{\sum_{k=1}^{g} Z_k} \qquad (j = 1, \cdots, g),$$

where Z_j is the number of waiting positions occupied in the jth queue.

The system analyzed in this paper is an extension of the one examined by Kühn, since an arrival does not release a waiting position until his service has been completed. This complicates the state analysis, since it is now not sufficient to know the number of servers that are busy to determine the equilibrium equations; information as to the number of calls from each queue that are in service must be included.

Kühn indicates also that, for the above interqueue discipline, the waiting time distribution can be found numerically only for small systems. The calculation of the waiting time is complicated by the fact that an arrival's waiting time is influenced by the number of arrivals that occur *after* it has entered the system. More is said about this difficulty later.

II. MATHEMATICAL FORMULATION

In this section, a mathematical model of the queuing system is formulated. Equilibrium equations are given and their solutions derived.

2.1 Queuing model

The queuing system consists of l input queues, each with a finite length denoted by N_i, $i = 1, 2, \cdots, l$. Requests for service arrive at queue i according to a Poisson distribution with mean rate, λ_i. If we let $A_i(t)$ denote the number of arrivals at queue i in $(0, t)$, then

$$P[A_i(t) = k] = \frac{(\lambda_i t)^k}{k!} e^{-\lambda_i t} \qquad (k = 0, 1, 2, \cdots). \tag{1}$$

The arrival process at queue i is assumed to be independent of the arrival process at each of the other queues. Arrivals from each queue have full access to a group of M servers. The service time distributions

of the servers, denoted by $H(t)$, are independent and identical exponential distributions with mean service rate μ, i.e.,

$$H(t) = \begin{cases} 1 - e^{-\mu t} & (t \geqq 0) \\ 0 & (t < 0). \end{cases} \tag{2}$$

Since the arrival process is Poisson and the service process is exponential, the queuing model is a multidimensional birth-death process.

Arrivals to queue i that find N_i waiting positions in queue i occupied are not allowed to enter the system. If an arrival to queue i finds at least one unoccupied position in queue i but all M servers busy, it enters the queue and waits as long as necessary for service. Within queue i, arrivals enter service on a first-in, first-out (FIFO) basis. An arrival that finds at least one unoccupied position in its queue and at least one free server immediately enters service. When an arrival enters service, it does not release a waiting position but remains in the queue until its service has been completed. Hence, the word "queue" is being used in a nonstandard manner and refers to the number of calls waiting for service *and* in service. As discussed earlier, an example of such a queue is a group of trunks that carry calls into a switchboard.

The interqueue service discipline—the order in which the queues receive service—is characterized by the number of calls waiting for service. When a server becomes free, queue i receives service with a probability, p_i, which is the proportion of queue-i calls waiting for service. If we denote the number of calls in queue i by n_i and the number of calls in queue i that are in service by m_i, then this probability, which is dependent on $(n_1, n_2, \cdots, n_l, m_1, \cdots, m_l) \equiv (\mathbf{n}, \mathbf{m})$, can be expressed as

$$p_i(\mathbf{n}, \mathbf{m}) = \frac{n_i - m_i}{\sum_{j=1}^{l} (n_j - m_j)} = \frac{n_i - m_i}{\sum_{j=1}^{l} n_j - M} \qquad \left(\sum_{j=1}^{l} n_j \geqq M + 1 \right)$$
$$(i = 1, 2, \cdots, l). \tag{3}$$

The effects of other interqueue service disciplines have been investigated, but will not be discussed here.

The fact that arrivals remain in the queue during service distinguishes this queuing system from the standard system, since the amount of information required to fully describe a state of the system is increased. The system is also complicated by the fact that the interqueue service discipline is state-dependent. However, as is shown in later sections, this "complication" leads to a closed-form solution of the

equilibrium equations, which is generally not the case for such systems. The equations of equilibrium are given in the following section.

2.2 Equilibrium equations

The system described in the previous section is characterized by a finite number of states that indicate the number of calls in each queue and, of these calls, the number receiving service. We denote by $(n_1, n_2, \cdots, n_l, m_1, \cdots, m_l)$ the state in which there are n_i calls in queue i with m_i of these calls in service. For notational simplification, we also refer to the state in vector notation as $(\mathbf{n}, \mathbf{m})$. In this notation, $(\mathbf{n}_{i+}, \mathbf{m})$ represents the state $(n_1, n_2, \cdots, n_i + 1, \cdots, n_l, m_1, \cdots, m_l)$ and similarly $(\mathbf{n}_{i-}, \mathbf{m}) \equiv (n_i, \cdots, n_i - 1, \cdots, m_l)$. It should be clear that, if the total number of calls in the system is less than or equal to M, then $n_i = m_i$ for all i.

Assuming stationarity, let $P(\mathbf{n}, \mathbf{m})$ be the probability that at an arbitrary instant of time the system is in state $(\mathbf{n}, \mathbf{m})$. Moreover, if the arrival processes to the system are Poisson, the equilibrium-state distribution $\{P(\mathbf{n}, \mathbf{m})\}$ at an arbitrary instant is equal to the equilibrium-state distribution at the instant of an arrival. By equating the rate into a state to the rate out of a state, we can write the equilibrium-state equations where we have introduced the function

$$u(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leqq 0 \end{cases}$$

to include the boundary conditions and $a_i = \lambda_i/u$:

$$\left\{ \sum_{i=1}^{l} [a_i u(N_i - n_i) + n_i] \right\} P(\mathbf{n}, \mathbf{m}) = \sum_{i=1}^{l} a_i u(n_i) P(\mathbf{n}_{i-}, \mathbf{m}_{i-}) + \sum_{i=1}^{l} (n_i + 1) u(N_i - n_i) P(\mathbf{n}_{i+}, \mathbf{m}_{i+})$$
$$\left(\sum_{i=1}^{l} n_i < M \right) \qquad (0 \leqq n_i \leqq N_i) \qquad i = 1, 2, \cdots, l \quad (4)$$

$$\left\{ \sum_{i=1}^{l} [a_i u(N_i - n_i) + m_i] \right\} P(\mathbf{n}, \mathbf{m}) = \sum_{i=1}^{l} a_i u(n_i) P(\mathbf{n}_{i-}, \mathbf{m}_{i-}) + \sum_{i=1}^{l} m_i u(N_i - n_i) P(\mathbf{n}_{i+}, \mathbf{m})$$
$$+ \sum_{i=1}^{l} \sum_{\substack{j=1 \\ j \neq i}}^{l} (m_i + 1) u(N_i - n_i) u(m_j) P(\mathbf{n}_{i+}, \mathbf{m}_{i+,j-})$$
$$\left(\sum_{i=1}^{l} n_i = M \right) \qquad (0 \leqq n_i \leqq N_i) \qquad i = 1, 2, \cdots, l \quad (5)$$

$$\left\{\sum_{i=1}^{l}\left[a_i u(N_i - n_i) + m_i\right]\right\} P(\mathbf{n}, \mathbf{m}) = \sum_{i=1}^{l} a_i u(n_i) P(\mathbf{n}_{i-}, \mathbf{m})$$
$$+ \sum_{i=1}^{l} \frac{m_i(n_i + 1 - m_i)}{\left(\sum_{k=1}^{l} n_k + 1 - M\right)} u(N_i - n_i) P(\mathbf{n}_{i+}, \mathbf{m})$$
$$+ \sum_{i=1}^{l} \sum_{\substack{j=1 \\ j \neq i}}^{l} \frac{(m_i + 1)(n_j + 1 - m_j)}{\left(\sum_{k=1}^{l} n_k + 1 - M\right)} u(N_i - n_i) u(m_j) P(\mathbf{n}_{i+}, \mathbf{m}_{i+,j-})$$
$$\left(\sum_{i=1}^{l} n_i > M\right) \qquad (0 \leqq n_i \leqq N_i) \qquad i = 1, 2, \cdots, l. \quad (6)$$

Equations (4) to (6), together with the normalization condition

$$\sum_{n_1=0}^{N_1} \cdots \sum_{n_l=0}^{N_l} \sum_{m_1=0}^{\min(M,N_1)} \cdots \sum_{m_l=0}^{\min(M,N_l)} P(\mathbf{n}, \mathbf{m}) = 1, \quad (7)$$

where all nonexistent states, such as states in which both $n_i = 0$ and $m_i > 0$, in the sum are assumed to have probability zero, determine the equilibrium-state distribution.

2.3 Steady-state solution

Since the process described by the equilibrium equations is a finite-state birth-death process in which the arrival rate *into* the system is always less than the service rate of the system as a result of overflow from the finite queue, a unique solution to eqs. (4) to (7) exists, and the solution is a genuine probability distribution.[15] This solution is given in terms of $P(\mathbf{0}, \mathbf{0})$ by

$$P(\mathbf{n}, \mathbf{m}) = \begin{cases} \prod_{i=1}^{l} \frac{a_i^{n_i}}{n_i!} P(\mathbf{0}, \mathbf{0}) & \left(\sum_{i=1}^{l} n_i \leqq M\right) \quad (8) \\ \dfrac{\begin{bmatrix} \sum_{i=1}^{l} n_i - M \\ n_1 - m_1, \cdots, n_l - m_l \end{bmatrix} \begin{bmatrix} M \\ m_1, m_2, \cdots, m_l \end{bmatrix} \prod_{i=1}^{l} a_i^{n_i}}{M! M^{\sum n_i - M}} P(\mathbf{0}, \mathbf{0}) & \left(\sum_{i=1}^{l} n_i > M\right), \quad (9) \end{cases}$$

where $P(\mathbf{0}, \mathbf{0})$ is determined from (7). The general solution was determined from examination of various small systems. By substitution, it can be shown that this solution in fact satisfies the equilibrium-state equations (4) to (6).

When the number of calls in the system is less than or equal to the number of servers, no one is waiting for service and the queues have no interaction. This fact is shown in (8) by the product form of the solu-

tion. However, as the number of calls increases to the point where calls are waiting in more than one queue, the queues no longer are acting independently.

Since the queues behave independently as long as there are free servers, we would expect that, as the number of servers was increased, the system would approach l independent loss systems. By examining the marginal probabilities, it can be shown that this is, in fact, true when $M \geqq \sum_{i=1}^{l} N_i$. That is, the marginal state probability of n_1 calls in queue 1 is

$$P(n_1) = \sum_{n_2} \cdots \sum_{n_l} P(\mathbf{n}, \mathbf{m}) = \frac{a_1^{n_1}/n_1!}{\sum_{k=1}^{N_1} a_1^k/k!}, \tag{10}$$

which is identical to the state probability for a pure loss system.

If no calls were blocked from the system, then the system would act as a pure delay system with the offered load $a = \sum_{i=1}^{l} a_i$. This is easily shown by taking the limit of (8) and (9) as $N_i \to \infty$ for all $i = 1, 2, \cdots, l$.

We first consider the case in which $\sum_{i=1}^{l} n_i \leqq M$. Since the number of calls in each queue is unrestricted, it is easily seen that the multinominal expansion of $(a_1 + \cdots + a_l)^{\Sigma n_i}$ divided by $(\sum_{i=1}^{l} n_i)!$ can be obtained from (8). That is,

$$\lim_{\substack{N_i \to \infty \\ i=1,2,\cdots,l}} \sum_{\Sigma n_i = k} P(\mathbf{n}, \mathbf{m}) = \frac{(a_1 + \cdots + a_l)^k}{k!} P(\mathbf{0}, \mathbf{0}) \qquad \left(\sum_{i=1}^{l} n_i < M \right).$$

Hence,

$$P\left(\sum_{i=1}^{l} n_i = k \right) = \frac{a^k}{k!} P(\mathbf{0}, \mathbf{0}) \qquad (k < M). \tag{11}$$

For the case in which $\sum_{i=1}^{l} n_i \geqq M$, first note that, by the Vandermonde convolution of multinomial coefficients,

$$\sum_{m_1} \cdots \sum_{m_l} \begin{bmatrix} \sum_{i=1}^{l} n_i - M \\ n_1 - m_1, \cdots, n_l - m_l \end{bmatrix} \begin{bmatrix} M \\ m_1, \cdots, m_l \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{l} n_i \\ n_1, n_2, \cdots, n_l \end{bmatrix}. \tag{12}$$

Therefore,

$$P(\mathbf{n}) = \sum_{m} P(\mathbf{n}, \mathbf{m}) = \frac{\left(\sum_{i=1}^{l} n_i \right)!}{M!\, M^{\Sigma n_i - M}} \prod_{i=1}^{l} \frac{a_i^{n_i}}{n_i!} P(\mathbf{0}, \mathbf{0}) \qquad \left(\sum_{i=1}^{l} n_i \geqq M \right), \tag{13}$$

and thus we can denote this state probability by

$$P\left(\sum_{i=1}^{l} n_i = k\right) = \sum_{\substack{l \\ \sum_{i=1} n_i = k}} \cdots \sum \frac{k!}{M!\,M^{k-M}} \prod_{i=1}^{l} \frac{a_i^{n_i}}{n_i!} P(\mathbf{0}, \mathbf{0}) \qquad (k \geqq M). \quad (14)$$

Consequently,

$$\lim_{\substack{N_i \to \infty \\ i=1,\cdots,l}} P\left(\sum_{i=1}^{l} n_i = k\right) = \frac{a^k}{M!\,M^{k-M}} P(\mathbf{0}, \mathbf{0}) \qquad (k \geqq M). \quad (15)$$

The normalization constant is then expressed as

$$P(\mathbf{0}) = \left[\sum_{k=0}^{M-1} \frac{a^k}{k!} + \sum_{k=M}^{\infty} \frac{a^k}{M!\,M^{k-M}}\right]^{-1} \qquad (0 \leqq a < M). \quad (16)$$

Hence, comparison of (11), (15), and (16) with the pure delay system completes the proof.

2.4 Blocking and delay probabilities

In the analysis and design of queuing systems, performance measures for each configuration must be calculated. In telephone traffic theory, these performance measures are generally referred to as "grades of service." Two such measures of the grade of service are:

(*i*) For loss systems, the blocking probability or probability of loss.
(*ii*) For delay systems, the average delay experienced by calls that enter the system. The average delay $\bar{W}(s, a)$ for a pure delay system is expressed in terms of the Erlang delay formula as

$$\bar{W}(s, a) = \frac{C(s, a)}{(s - a)\mu}. \quad (17)$$

In the system described in Section 2.1, the blocking probabilities for each input queue, the average delay experienced by calls that enter the system, and the average delay of only those customers who experience a positive delay are important characteristics to be examined. The latter is not used in the remaining analysis.

The blocking probability for queue i is defined as the probability that an arrival to queue i finds N_i calls in the queue. This probability, which is denoted by $B_i(\mathbf{N}, M, \mathbf{a})$, is a function not only of the number of positions in queue i and the offered load to the queue, but also of the traffic load offered to each of the other queues, the number of positions in each of these queues, and the number of servers. Recalling that, for systems with Poisson input, the state probabilities at an arbitrary instant are equal to the state probabilities at arrival times,

$B_i(\mathbf{N}, M, \mathbf{a})$ is calculated from the marginal distribution for queue i as

$$B_i(\mathbf{N}, M, \mathbf{a}) = \sum_{n_1} \cdots \sum_{n_{i-1}} \sum_{n_{i+1}} \cdots \sum_{m_1} \cdots \sum_{m_l} P(n_1, \cdots, N_i, \cdots, n_l, \mathbf{m}). \quad (18)$$

The average delay experienced by calls that enter the system (successfully occupy a waiting position in their queue) is denoted by $\bar{D}(\mathbf{N}, M, \mathbf{a})$. The delay calculated for this system is the overall mean waiting time measured from an arrival's entry into its queue until its service begins. Hence, it does not include service time of the call. It should also be noted that arrivals into the system that find at least one free server experience no delay.

Since the average number of calls waiting for service must be finite and the mean waiting time is finite as a result of the loss structure of the system, the well-known equation of Little,[16] $L = \lambda W$, can be used to calculate $\bar{D}(\mathbf{N}, M, \mathbf{a})$. In particular, we must define our "queue length" as the total number of calls waiting for service and "λ" is defined as the effective arrival rate *into* the system. Hence,

$$\bar{D}(\mathbf{N}, M, \mathbf{a}) = \frac{\sum_{l} \cdots \sum \sum_{m_1} \cdots \sum_{m_l} \left[\left(\sum_{i=1}^{l} n_i - M \right) P(\mathbf{n}, \mathbf{m}) \right]}{\sum_{i=1}^{l} \lambda_i [1 - B_i(\mathbf{N}, M, \mathbf{a})]} \cdot \quad (19)$$

Calls that are blocked from the system do not enter the queue and hence do not affect the average queue length. Consequently, they are not included in the arrival rate *into* the system. The numerator of (19) is the average number of waiting calls. It should be apparent that the average delay for any particular queue can easily be obtained by using the appropriate marginal probabilities. Also, the conditional average delay, the average delay experienced by only those that must wait, is found by dividing (19) by the probability of being delayed.

For some design purposes, it might be deemed necessary to constrain the probability of waiting longer than some time, t_0, to be less than a specified value. In this case, the waiting time distribution for each queue must be obtained. For the interqueue discipline examined for this system, the calculation of the waiting-time distribution is extremely difficult. (Kühn[13] mentions that, for his problem, numerical techniques can be used for very small systems, after which approximate methods must be formulated.) The difficulty in determining the waiting-time distribution lies in the fact that the time a particular call must wait for service is not just a function of the number of calls in

the system when it arrives, as is usually the case. In particular, the waiting time of a call is related to the number of calls that arrive after the particular call enters the system, since queues are chosen for service according to their queue lengths at a service completion, so that later calls can "pass" earlier ones. For this reason, the waiting time distributions are not calculated in this study.

2.5 Macrostate analysis

As discussed in the previous section, blocking probabilities and average delay values are of interest to system designers. Using the state probabilities given in (8) and (9), it is possible to obtain not only the overall average delay, $\bar{D}$, as shown in (19), but also the average delay, $\bar{D}_i$, for queue i. In certain types of design, we might want to engineer the system so that the average delay in *every* queue is less than a specified level. In such cases, $\bar{D}_i$ would be needed. However, for this study, we consider only the overall average delay.

Therefore, it is apparent from (19) that, for computational purposes, we only need to know the number of calls in each queue without distinguishing between those in service and those waiting. If we denote by $\mathbf{n}$ the state $(n_1, n_2, \cdots, n_l)$, we can find the steady-state probabilities $P(\mathbf{n})$ from (8) and (9). Of course, we could have written the state equations directly and solved this easier set of equations.[17] However, for further studies, it is essential to know the probabilities $P(\mathbf{n}, \mathbf{m})$.

Since the state probabilities for $(\mathbf{n}, \mathbf{m})$ in which $\sum_{i=1}^{l} n_i \leqq M$ are independent of $\mathbf{m}$, $P(\mathbf{n}) = P(\mathbf{n}, \mathbf{m})$. To obtain $P(\mathbf{n})$ for $\sum_{i=1}^{l} n_i > M$, we sum $P(\mathbf{n}, \mathbf{m})$ given by (9) over all possible values of $\mathbf{m}$. Using the Vandermonde multinomial convolution, we find that

$$P(\mathbf{n}) = \sum_{m_1} \cdots \sum_{m_l} P(\mathbf{n}, \mathbf{m}) = \frac{\left(\sum_{i=1}^{l} n_i\right)!}{M!\, M^{\sum n_i - M}} \prod_{i=1}^{l} \frac{a_i^{n_i}}{n_i!} P(\mathbf{0}), \qquad (20)$$

where $P(\mathbf{0})$ is the normalization constant. We can calculate the blocking probabilities and the average delay as before. The number of states is $\prod_{i=1}^{l} (N_i + 1)$.

It should be repeated that the macrostate probabilities are of use *only* if one is interested in the overall mean delay. To calculate the individual average delays, one must use the microstate probabilities.

III. RESULTS

In this section, we investigate the effects of varying N_i and M on the blocking probabilities for each queue and the overall average delay

of a call. In particular, these effects are illustrated by comparing the results obtained from the analysis presented in this paper, for a particular example, with the results obtained if independence is assumed. This "independence assumption" is often used in practice to determine the number of positions for each queue and the number of servers needed. In essence, this assumption permits a designer to design each queue (trunk group) independently of the other queues and the number of servers, and the number of servers is determined assuming that no calls are blocked in the queues. It is shown that this assumption is generally not even a good approximation. Finally, four properties that are used in a design procedure established in the next section are postulated.

3.1 Comparison of results with independence assumption

In the engineering of automatic call distributor systems, a traffic engineer generally dimensions each trunk group using the Erlang loss formula (assuming that it is independent of the other groups and the number of attendants) and often determines the number of attendants required from the Erlang delay formulas using the total offered load to the queues (assuming no blocking in the queues). This procedure is clearly invalid, but up to now an exact procedure has not been available. As a means of illustrating the significance of the results presented in this paper, we now compare, for a particular system configuration, the system characteristics that a traffic engineer would expect to obtain using the independence assumption and what he really will find. Of course, the interqueue service discipline will affect these results in a way that will be described in later work. The actual operation of such systems is quite complicated, and is not readily characterized by any of the disciplines usually used such as first-in, first-out, random, and last-in, first-out. However, the results of simulations indicate that the discipline presented here is a good approximation of the actual method of operation.

For purposes of this comparison, we assume a simple system with only two queues: the first with an offered busy-hour load of 10 erlangs; the second, 5 erlangs. It is further assumed that the queues have coincident busy hours and that the average holding time per call is 30 seconds. Assuming that a P.01 grade-of-service constraint has been placed on each group, the number of trunks required, if independence is assumed, would be $N_1 = 18$ and $N_2 = 11$. (These numbers can be found from tables of the Erlang B formula.) If it is then required that, on the average, no call must wait longer than 3 seconds for an answer, we find, from the Erlang delay formula, that $M = 19$ (assuming no

blocking in the groups). The traffic engineer would expect this system to have the following characteristics:

Blocking probability for group 1 = 0.0071.
Blocking probability for group 2 = 0.0083.
Overall average delay = 1.83 seconds.

However, analyzing this system configuration with the results of this paper, we find that the service levels (which stated above are approximations to the actual levels) would be

Blocking probability for group 1 = 0.014.
Blocking probability for group 2 = 0.013.
Overall average delay = 0.949 second.

The directions of the changes are intuitively obvious, since longer holding times in the queues result in more blocked calls and the higher blocking levels decrease the load to the servers, which in turn results in a lower average delay. It should be noted that the trunk groups are performing at unsatisfactory levels, but the overall average delay has been decreased and is considerably under the required level. Often, customers who have such systems measure only the delay or speed of answer and periodically remove attendants if they feel that the speed of answer is not above the required level. Unfortunately, such a customer generally does not realize the effect of removing attendants on the blocking probabilities on his incoming trunks and consequently on other network customers.

As an example of customer behavior, consider the system discussed above. The customer, having measured the average delay and finding it to be considerably under his required level, would most likely remove two attendants. The average delay would then become 3.01 seconds, but the blocking probabilities increase to 0.028 for group 1 and 0.023 for group 2. Hence, even though the delay constraint is essentially satisfied, the probabilities of loss are more than double their desired levels.

If, instead of the P.01 service level, P.05 or P.10 had been chosen for the above delay constraint, the independence assumption would generally give a configuration that would satisfy all service constraints. The reason for this is that the higher blocking levels decrease the offered load to the attendants, and consequently a very small delay results. This delay is small enough that it has little effect on the holding time of the calls and, hence, the offered load to queue i is approximately a_i. Therefore, the resulting blocking probability, although larger than the Erlang loss probability, is generally in the acceptable

range. However, the configuration would not be optimal because too many attendants are provided. In summary, as the blocking probabilities increase, the discrepancies between the Erlang loss formula and the formula given by (18) decrease, but the discrepancies between average delays increase.

Several variations in the procedures to engineer these systems exist in practice. Generally, in determining the number of attendants, the *measured* offered load into the attendants is used. This load accounts for the blocking in each trunk group. For the above example, the measured offered load would be 14.79 erlangs (if we assume that the assumptions made in this analysis are valid) and, from the Erlang delay formula, an average delay of 1.58 seconds results, which is still higher than the actual delay. The reason for the discrepancies is that the offered load to the attendants is no longer Poisson as a result of the blocking in the groups. In fact, the variance of this offered load will be lower than that of the Poisson load, since the peakedness of the traffic has been decreased by clipping. Hence, since the actual average offered load and variance of the load are lower, this leads us to postulate the following property:

Property 1:

$$\bar{W}(M, a) \geqq \bar{D}(\mathbf{N}, M, \mathbf{a}) \quad \text{where} \quad a = \sum_{i=1}^{l} a_i.$$

This property is illustrated in Fig. 1. The equality holds in the limit as $N_i \to \infty$ for all $i = 1, \cdots, l$, as shown previously. The significance of this property is that we now have a method of obtaining an upper bound on the average delay for the combined system.

Another variation that is sometimes used is to add the speed of answer into the offered load to each group. If we add the average delay of 0.949 second to each call and use this new offered load in the Erlang loss formula, the blocking probabilities that result are 0.0091 and 0.01, for groups 1 and 2, which are still lower than the actual blocking.

The discrepancies result because the Erlang loss formula assumes exponential holding times on the trunks but, in fact, the holding time for the combined system is the sum of an exponential distribution and the delay distribution. Also, the holding times of calls in the system are no longer independent (unless $\sum_{i=1}^{l} n_i < M$). The variance of this new service time distribution is higher than that of the exponential service time distribution and, of course, the mean is larger. Therefore, one would expect the average queue length to be larger which, in turn, implies an increase in blocking.

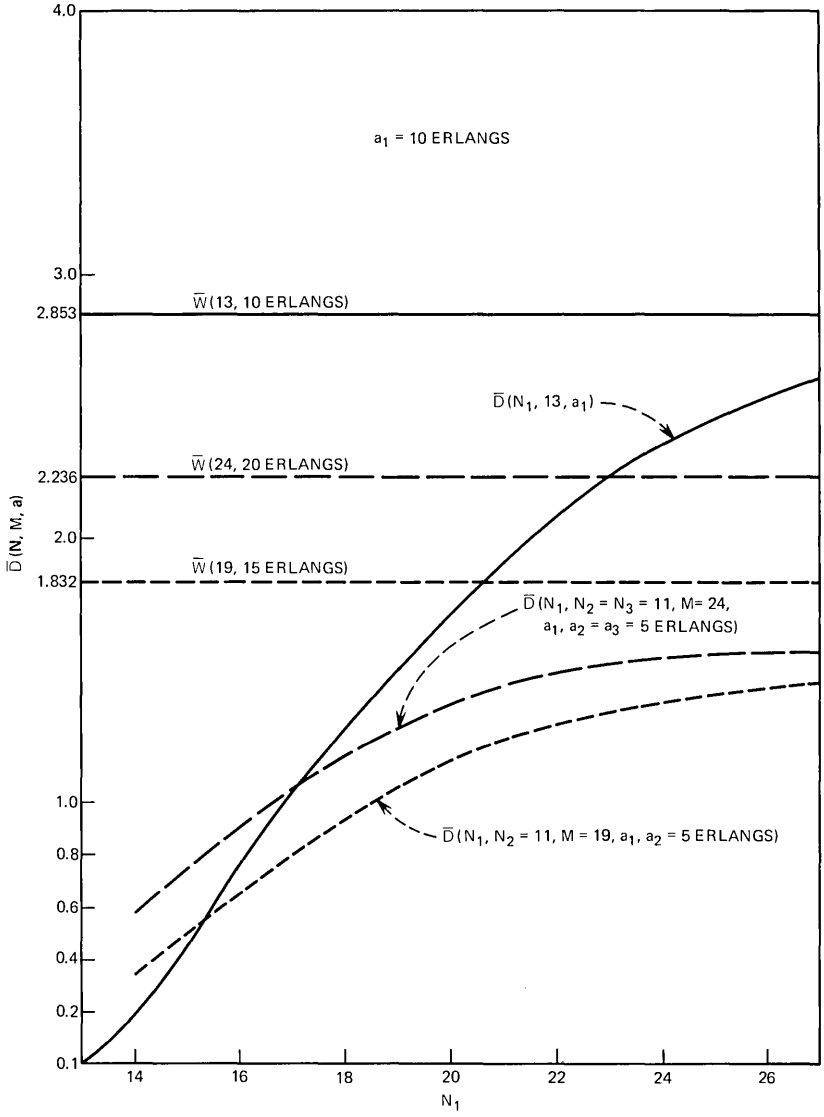

Fig. 1—Average delay as a function of number of positions in one queue.

With these facts in mind and as a result of empirical evidence, we postulate a second useful property:

Property 2:

$$B(N_i, a_i) \leqq B_i(\mathbf{N}, M, \mathbf{a}) \qquad (i = 1, \cdots, l),$$

where $B(N_i, a_i)$ is the Erlang loss formula for N_i servers and an offered

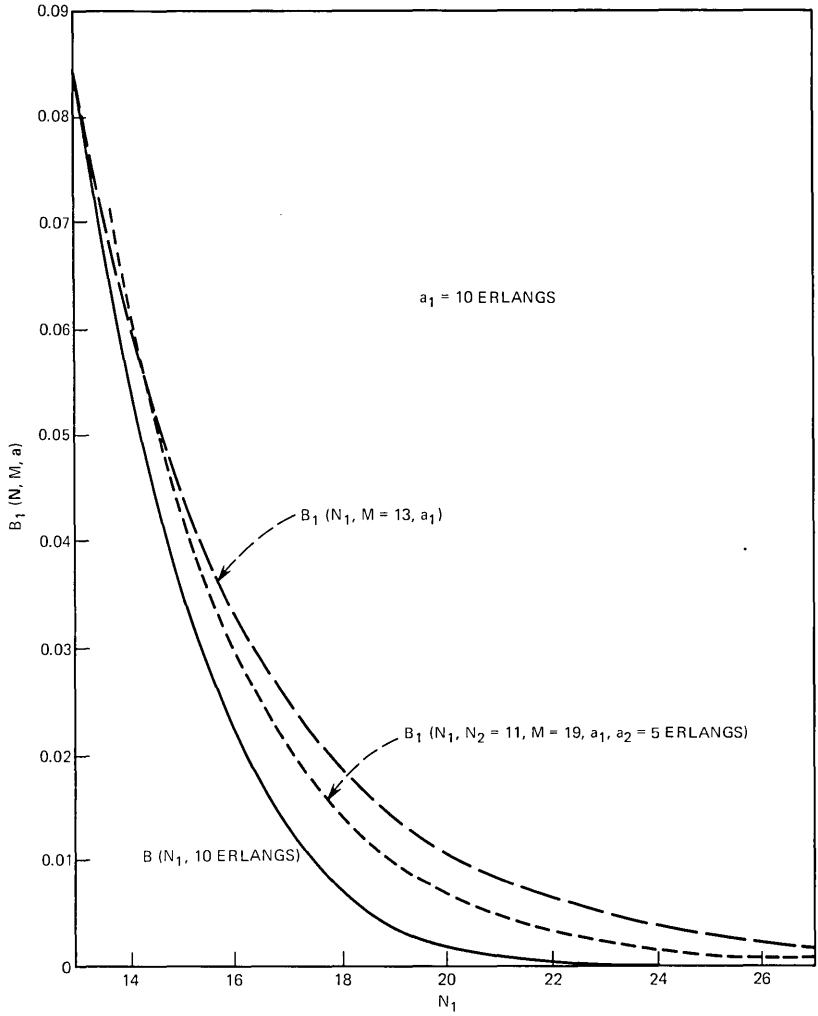

Fig. 2—Blocking probability as a function of number of positions in the queue.

load, a_i. The equality holds when $M \geqq \Sigma N_i$, since the queues then behave independently (there is no delay, so in fact the holding time per call is exponential). This property is illustrated in Fig. 2. Intuitively, one would expect $B_i(\mathbf{N}, M, \mathbf{a})$ to be larger since, as a result of a positive delay added to each call, calls hold the trunks longer, therefore increasing the probability of an arriving call finding all trunks busy. The significance of Property 2 is that a lower bound on the number of trunks required for a given service level can be found using the Erlang loss formula.

3.2 Effects of varying N_i and M

For a given input process and service time distribution, the designer can affect the blocking probabilities for a queue or the average delay by changing the number of positions, N_i, in a queue and/or changing the number of servers, M. We first investigate the effects of varying the number of servers, M. As noted earlier, when M has been increased to the point where $\sum_{i=1}^{l} N_i = M$, the average delay becomes zero and the queues behave independently. The blocking probability for each queue is then given by $B(N_i, a_i)$ which, by Property 2, is a lower bound on the blocking for any value of M. What is of importance, however, is: Do the blocking probabilities and the average delay monotonically decrease to their lower bounds as we increase M to the value $\sum_{i=1}^{l} N_i$? Empirical evidence, such as shown in Figs. 3 and 4, indicates that the answer to this question is *yes*. We postulate this property as:

Property 3:

$$B_i(\mathbf{N}, M + 1, \mathbf{a}) \leqq B_i(\mathbf{N}, M, \mathbf{a}) \qquad (i = 1, \cdots, l)$$
$$\bar{D}(\mathbf{N}, M + 1, \mathbf{a}) \leqq \bar{D}(\mathbf{N}, M, \mathbf{a}).$$

Intuitively, one would not expect that increasing the number of servers in a system would increase the average delay. Moreover, a decrease in the holding time of calls would imply that the average queue lengths would decrease, which would result in a decrease in the blocking probability for that queue. However, this decrease in blocking results in an increase in offered load to the servers but, as we postulate, this increase is less than the marginal carrying capacity of the added server. The significance of this property is that, with added servers, not only is the average delay decreased but also the blocking probabilities are decreased; that is, adding servers improves the service performance of the servers *and* of the queues (trunk groups).

The other system parameters that may be varied to improve system performance are the numbers of positions in each queue. Supported by quantitative evidence, such as given in Figs. 1 and 5, and by intuition we postulate the following:

Property 4:

$$B_i(\mathbf{N}_{i+}, M, \mathbf{a}) \leqq B_i(\mathbf{N}, M, \mathbf{a})$$
$$B_j(\mathbf{N}_{i+}, M, \mathbf{a}) \geqq B_j(\mathbf{N}, M, \mathbf{a}) \qquad j \neq i.$$
$$\bar{D}(\mathbf{N}_{i+}, M, \mathbf{a}) \geqq \bar{D}(\mathbf{N}, M, \mathbf{a})$$

The first part of this property states that, if we increase the number of positions for calls to occupy in a given queue, then the probability of

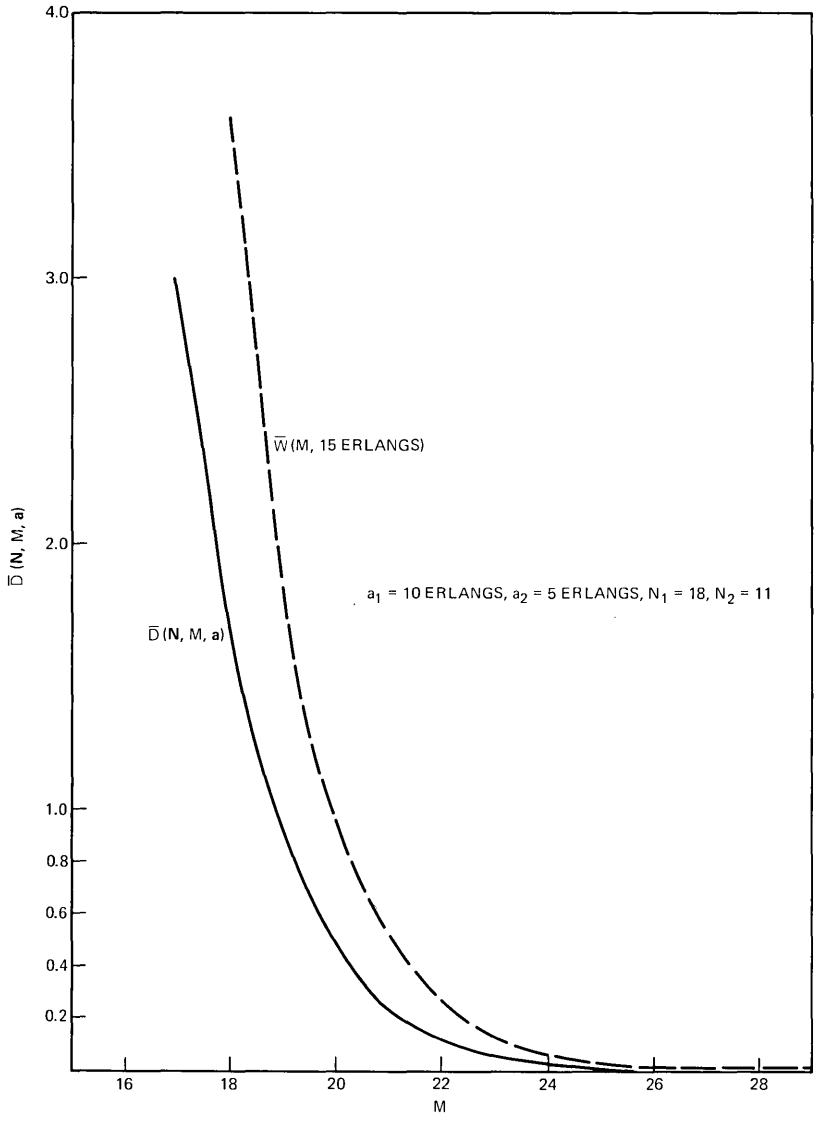

Fig. 3—Average delay as a function of number of servers.

loss for that queue is decreased. (This, of course, is true in pure loss systems.) The intuitive argument is that calls that previously found N_i calls in queue i were blocked, but now are not. Therefore, the number of calls blocked is decreased. However, as the third part of this property implies, the average delay of calls is increased as a result of this increase in calls from queue i. The intuitive counter-

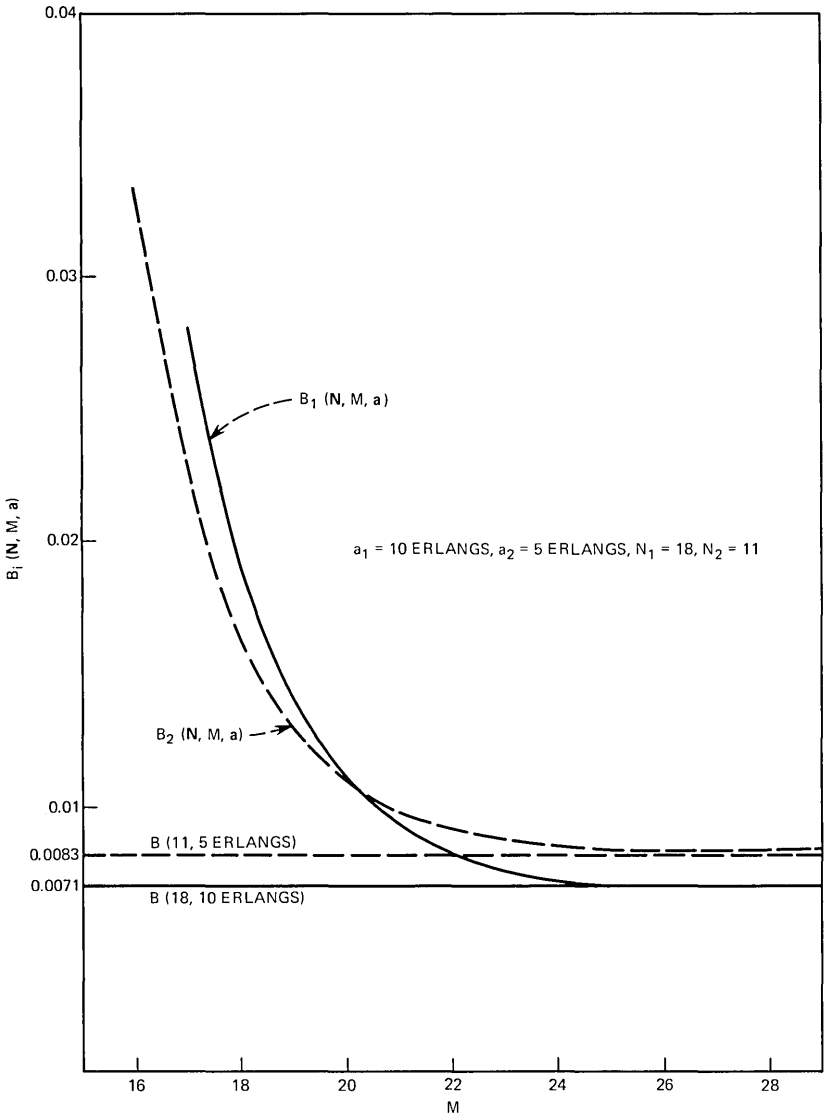

Fig. 4—Blocking probability as a function of number of servers.

argument is that this increase in holding time per call might result in an increase in blocking for queue i. But we postulate that the increase in delay is not substantial enough to eliminate the increased efficiency obtained in queue i by the addition of a position.

However, for the other queues, the number of positions remains fixed, and this increase in average delay results in a larger traffic level

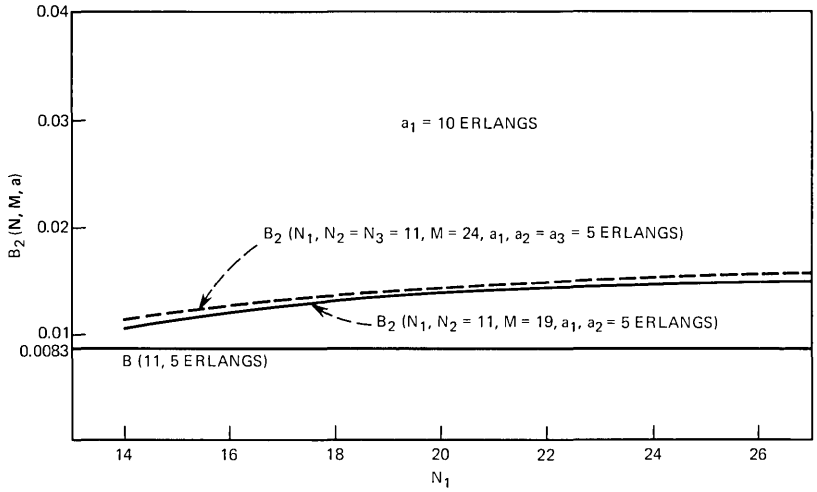

Fig. 5—Blocking probability in one queue as a function of number of positions in another.

per position being placed on the queues. Hence, this increase in load causes the average queue length to increase and, consequently, causes the blocking probabilities to increase. This fact is stated in the second part of Property 4 and is illustrated by the example given in Fig. 5. Equality in the three statements holds only in the limit as N_i goes to ∞ for all $i = 1, 2, \cdots, l$.

The significance of Property 4 is that the loss probability for a given queue cannot be reduced by adding positions to any other queue. Therefore, by Properties 3 and 4, we see that the loss probability for a given queue can be reduced only by increasing the number of servers or by increasing the number of positions in that queue.

The four properties postulated indicate relationships between the system parameters and the system characteristics. Proofs for the simplest cases (i.e., $l = 1$ for all the properties except for the second part of Property 4 for which $l = 2$) are given in Ref. 18. The proofs for the general cases have not been constructed because of the difficulties involved (e.g., the proof of the second part of Property 4 required 17 pages for $l = 2$). However, based on the intuitive explanations given, empirical evidence, and these proofs, I feel that the properties are valid in the general case of l queues. To obtain an optimal configuration (minimum cost), one must balance the cost of servers against the cost of positions for the queues in such a way that all required service levels are met. These properties are used in the next section in the development of a procedure to determine this optimal

configuration. The algorithm and proof of convergence given below assume the validity of these properties.

IV. SYSTEM DESIGN WITH GRADE-OF-SERVICE CONSTRAINTS

4.1 The design problem

In determining the optimal design of a queuing system, one is generally interested in minimizing the operating costs in such a way that specified grade-of-service constraints are met. These constraints are a function of the particular queuing system under study. In the pure loss system, the grade of service is measured by the probability that a call is lost or blocked. Hence, the optimal system configuration is the minimum number of servers that satisfies the constraint, $B(s, a) \leqq b$.

In delay systems, at least three measures of service are useful. The first is the average delay experienced by a call in obtaining service. The second measure is the "extremal" delay—the probability that the delay for any call exceeds a specified limit. Finally, the average delay of only those customers who experience some delay is a useful measure.

However, in the combined loss and delay systems described in Section II, the determination of an optimal configuration is not as straightforward. We will measure the grade of service of the system by

(*i*) The blocking probability for each queue.
(*ii*) The average delay of all calls that enter the system.

The blocking probability for a particular queue is dependent on the number of positions in each queue and the number of servers, and the average delay depends on these same variables. A procedure must be developed to balance these measures of congestion in such a manner that the costs of the system are minimized.

More formally, the problem can be expressed as the following nonlinear problem in integer programming. We denote the monthly cost of a waiting position in queue i by C_i and the monthly cost of each server by C. It is assumed that C_i and C are positive, finite numbers. The blocking objective for queue i will be denoted by b_i and the average delay objective by d. The following assumptions have been made: The system is engineered for the system busy-hour traffic load and the busy hours for the queues are coincident. The optimization problem is then expressed as:

Minimize the cost

$$Z = \sum_{i=1}^{I} C_i N_i + CM$$

subject to

$$\text{(I)} \qquad B_i(\mathbf{N}, M, \mathbf{a}) \leqq b_i \qquad (i = 1, 2, \cdots, l) \tag{21}$$

$$\bar{D}(\mathbf{N}, M, \mathbf{a}) \leqq d \tag{22}$$

$$N_i (i = 1, 2, \cdots, l), \qquad M \geqq 0 \text{ integers.}$$

Since no efficient algorithm to solve a general nonlinear integer-programming problem exists, a procedure was developed that utilizes the four properties stated in Section III.

4.2 Optimization procedure

In this section, a procedure is developed that determines an optimal solution to the nonlinear integer-programming problem expressed by (I). The procedure is a direct-search routine based primarily on the properties presented in the previous section. A description of the procedure is now given and is followed by a concise summary of the algorithm. Figure 6 is a flowchart of the algorithm.

The first step in the algorithm, as in most mathematical programming algorithms, is the determination of a feasible solution to the problem. To obtain an initial feasible solution to (I), we utilize Properties 2 and 3 of the previous section. In particular, by Property 2, we know that the minimum number of waiting positions for queue i can be determined from the Erlang loss formula, which is easily computed from a recurrence relationship.[19] We begin the search for an initial feasible solution with

$$N_i^{(0)} = \min \{n \mid B(n, a_i) \leqq b_i\},$$

since it has been shown that, in fact, a feasible solution to (I) exists. That is, $(\mathbf{N}^{(0)}, \tilde{M})$, where $\tilde{M} = \sum_{i=1}^{l} N_i^{(0)}$, is a feasible solution since $\bar{D}(\mathbf{N}^{(0)}, \tilde{M}, \mathbf{a}) = 0$ and $B_i(\mathbf{N}^{(0)}, \tilde{M}, \mathbf{a}) = B(N_i^{(0)}, a_i)$ for $i = 1, 2, \cdots, l$. However, since this solution will generally not be near the optimal solution, the search will not begin at $\tilde{M}$ but instead with $M_{(0)}$, which is the minimum value of M that satisfies the constraint:

$$\bar{W}(M, a) \leqq d. \tag{23}$$

$M_{(0)}$ is the number of servers that would be selected if the Erlang delay formula with an offered load $a = \sum_{i=1}^{l} a_i$ were used. By Property 1, it is seen that if (23) is satisfied, then (22) will also be satisfied.

Using the set of parameters $(\mathbf{N}^{(0)}, M_{(0)})$, the system characteristics, $\{B_i(\mathbf{N}^{(0)}, M_{(0)}, \mathbf{a})\}$ and $\bar{D}(\mathbf{N}^{(0)}, M_{(0)}, \mathbf{a})$ are determined. One of two results occurs:

(*i*) The parameters ($\mathbf{N}^{(0)}$, $M_{(0)}$) satisfy the constraints of (I), in which case an initial feasible solution has been determined. We then proceed to find the feasible solution of minimum cost.

(*ii*) At least one of the blocking constraints (21) is violated. (As noted above, the delay constraint will be satisfied.) By Property 3, we know that the addition of a server will reduce the blocking probability for each queue and that, by the addition of enough servers, a feasible solution can be obtained. We denote this initial feasible solution by ($\mathbf{N}^{(0)}$, $M^{(0)}$) and note two interesting properties of this solution.

(*a*) By Property 2, $N_i^{(0)}$ is the *minimum* number of positions that must be considered for queue i.

(*b*) $M^{(0)}$ is the *maximum* number of servers that must be considered. This is true since a further increase in the number of servers can be justified only if some waiting positions can be eliminated, and the positions are already at their minimum levels, $\mathbf{N}^{(0)}$.

The next step in the algorithm is to attempt to improve the initial feasible solution. Since by construction we are initially at the maximum number of servers, we attempt to decrease the costs by decreasing the number of servers while maintaining feasiblity. To maintain feasibility, it may be necessary to add waiting positions to certain queues. If a feasible solution is found, then it will be an improvement only if the accumulated cost of those servers removed (since the last feasible solution) is greater than the accumulated cost of all waiting positions that have been added to maintain feasibility. Hence, as we remove servers, one of three things results.

(*i*) All the constraints of (I) are satisfied. If the accumulated cost of removed servers is greater than the cost of all waiting positions that have been added, this new feasible solution represents a cost improvement and should be stored as the tentative "optimal" solution. The accumulated costs are set to zero, a server is removed, and the search continues. If the cost of servers is less than the cost of positions, then we reduce the number of servers by one and continue the search for a solution with lower cost.

(*ii*) The delay constraint (22) is violated. In this case, we stop the search and the tentative optimal solution is the global* optimum. (A justification for stopping the procedure is given later.)

(*iii*) At least one of the blocking constraints is violated. In this case, we add one position to each queue in which the corresponding block-

---

*I have taken the liberty of using "global" since, in fact, the procedure does produce the global optimum *if* the four properties are true in the general case.

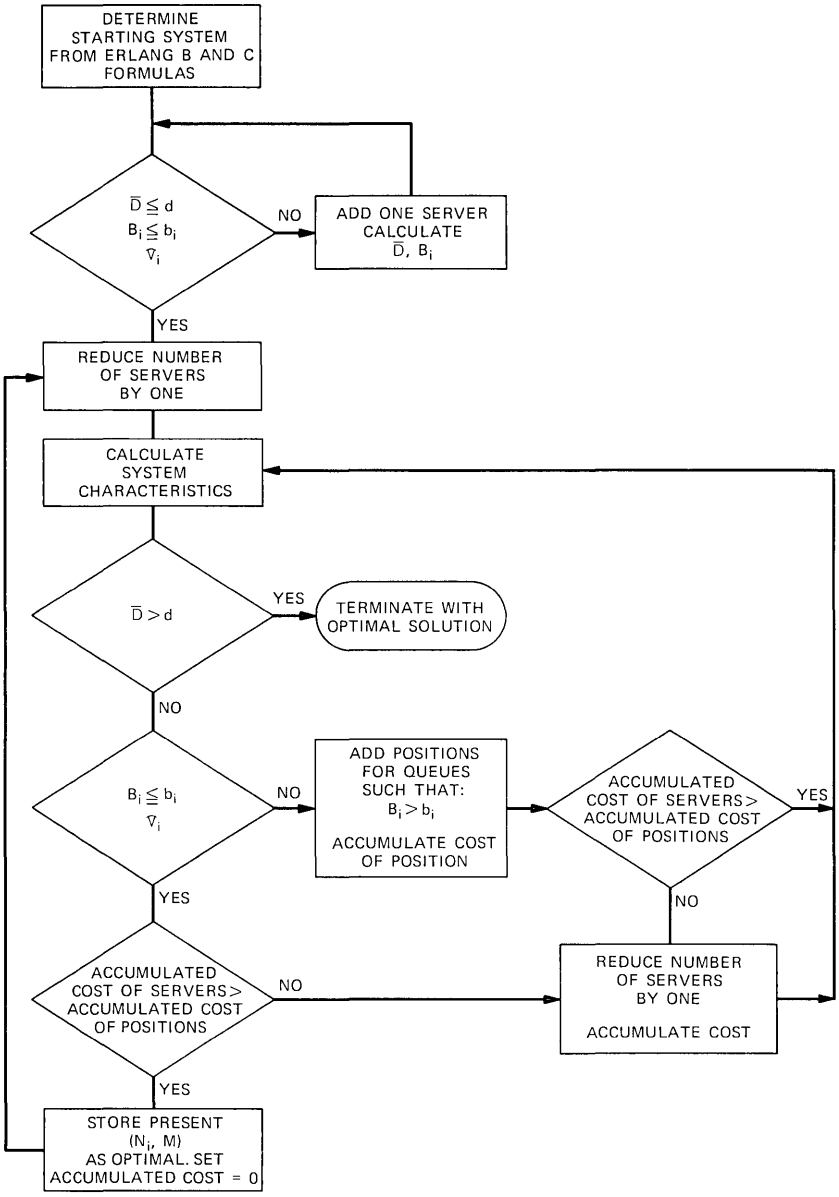

Fig. 6—Design procedure.

ing constraint is violated and increase the accumulated cost of added positions appropriately. If the cost of the additional positions is less than the accumulated cost of the servers that have been removed, we determine if this solution is feasible. If it is feasible, we proceed as in

(*i*). If it is not feasible, either (*ii*) or (*iii*) must be true. If the cost of the positions is more than the cost of servers, then no feasible solution with lower cost can be obtained with this number of servers and, hence, we reduce the number of servers by one and calculate the system characteristics with the new set of system parameters. Then either (*i*), (*ii*), or (*iii*) must be true, and the appropriate action is then taken.

In summary, the procedure removes servers until either the delay constraint is violated, in which case it terminates, or a blocking constraint is violated. If a blocking constraint is violated, waiting positions are added, and an additional server *may* be removed, depending on the incremental costs and on the feasibility of the tentative solution.

A justification of the procedure is in order. We first discuss the case in which at least one blocking constraint has been violated. To reduce the blocking probability for each queue whose constraint has been violated, a position must be added to this queue (by Property 4). The only other way to reduce the blocking probability is to add a server, but this branch has already been terminated. Assume that NS servers have been removed since the last "optimal" feasible solution and that the cost of all the positions added since the last "optimal" feasible solution is G. If $G > NS \times C$, then the new parameters cannot give a lower cost solution and, by Property 4, no lower cost solution for this M exists. Therefore, we terminate the branch with M servers and begin the search of the branch with $M - 1$ servers (increment NS by 1) with the present number of positions. At least this number of positions must be considered, since, by Property 3, the reduction of M results in an increase in the blocking probabilities. If $G \leqq NS \times C$, this set of parameters may be a lower-cost solution. Therefore, we determine the system characteristics and see if the solution is feasible.

We now prove that the procedure converges in a finite number of steps to a global optimum.

Lemma 1: *If a feasible solution for a given M has been found, the branch corresponding to that M can be terminated.*

Proof: Trivially, any further feasible solutions with that M must be more expensive, since these solutions must have more waiting positions. Q.E.D.

Theorem 1: *The algorithm terminates in a finite number of steps.*

Proof: First, we know that an initial feasible solution can be obtained in at most

$$\left[\sum_{i=1}^{l} N_i^{(0)} - M_{(0)} \right]^+$$

steps, as discussed earlier, where $[x]^+ = \max(0, x)$. Second, for any given M, the corresponding branch of the solution tree will be terminated in a finite number of steps. That is, since C is finite and C_i is nonzero for all i, then in a finite number of steps we will either find a feasible solution for this M, reach the point where the cost of the positions added for this M exceeds C, or violate the delay constraint. In the first case, by Lemma 1, we know that we can terminate this branch. In the second case, since none of the positions can be removed and feasibility be maintained, no feasible solution of lower cost exists for this M. In the latter case, the procedure terminates. The number of iterations performed for a given M is bounded by the number of times positions are added before the cost of these additions exceeds C.

Finally, since the maximum number of servers that need be considered is finite, we reach the case in which the delay constraint is violated in a finite number of steps (at most, $M^{(0)}$ values of M).

Since there are only a finite number of values of M to be considered and since, for each M, only a finite number of steps are performed, the algorithm terminates in a finite number of iterations. Q.E.D.

Theorem 2: *The solution* $(\mathbf{N}^*, M^*, Z^*)$ *obtained upon termination of the algorithm is a global optimum.*

Proof: Assume that another configuration $(\hat{\mathbf{N}}, \hat{M}, \hat{Z})$ exists, such that all constraints of (I) are satisfied and $\hat{Z} < Z^*$. First, consider the case in which $\hat{M} > M^*$. By construction, the branch corresponding to $\hat{M}$ must have been searched and, as indicated in Theorem 1, the branch would have been terminated in a finite number of steps. If this branch had produced a feasible solution with a lower cost, it would have been retained. Hence, this case is not possible.

Second, consider the case in which $\hat{M} < M^*$. From the algorithm, we know that the termination of the procedure implies that the delay constraint has been violated. We therefore know that either the branch with $\hat{M}$ produced a feasible solution with a cost larger than Z^* (or else it would have been retained), or $\hat{M}$ is smaller than the value of M when the procedure terminates. If the latter is true, then $\hat{M}$ cannot produce a feasible solution since the delay and blocking probability for $(\mathbf{N}^*, \hat{M})$ must be greater than those for $(\mathbf{N}^*, M^*)$ by Property 3; and, to reduce this delay, positions would have to be removed that would result in at least one blocking constraint being violated. Consequently, this case is a contradiction and, hence, $(\mathbf{N}^*, M^*, Z^*)$ is the optimum. Q.E.D.

One point should be noted: For the higher levels of blocking ($\gtrsim 0.05$), the solution $(\mathbf{N}^{(0)}, M_{(0)})$ is generally a feasible solution. However, as a result of the reduction in offered load to the servers because of the

blocking, $M_{(0)}$ can be reduced before any constraints are violated. This is of importance since, in practice, such systems have been engineered in such a manner that the configuration is $(\mathbf{N}^{(0)}, M_{(0)})$, which is obtained from use of the Erlang B and Erlang C formulas, as described earlier.

A more concise mathematical summary of the algorithm follows.

Optimization Algorithm:

Step 1: Initial Feasible Solution

(*i*) Determine $N_i^{(0)}$ $(i = 1, 2, \cdots, l)$ from the Erlang loss formula where $N_i^{(0)} = \min \{n | B(n, a_i) \leqq b_i\}$.

(*ii*) Determine $M_{(0)}$ from the Erlang delay formula where $M_{(0)} = \min \{m | \bar{W}(m, a) \leqq d\}$. Let $k = 1$.

kth iteration:

(*iii*) Calculate $B_i(\mathbf{N}^{(0)}, M_{(k-1)}, \mathbf{a})$ and $\bar{D}(\mathbf{N}^{(0)}, M_{(k-1)}, \mathbf{a})$. If the set of constraints (21) are satisfied, let $M^{(0)} = M_{(k-1)}$. The initial "optimal" feasible solution is $\mathbf{N}^* = \mathbf{N}^{(0)}$, $M^* = M^{(0)}$, and $Z^* = \sum_{k=1}^{l} C_i N_I^* + CM^*$. Set $j = 1$, ${}_1N^{(0)} = N^*$, and go to Step 2.

If at least one constraint of (21) is not satisfied, set $M_{(k)} = M_{(k-1)} + 1$, increment k, and return to (*iii*).

Step 2: Solution Improvement

In this step, the superscript j refers to the jth "optimal" value of M; for a given value of j, the subscript r refers to the rth value of N_i.

jth iteration:

(*i*) $NS = 1$, $G^{(j)} = 0$, $r = 1$.

(*ii*) Reduce number of servers by one, $M^{(j)} = M^{(j-1)} - 1$. If $M^{(j)} = 0$, go to Step 3.

(*iii*) Calculate $B_i({}_r\mathbf{N}^{(j-1)}, M^{(j)}, \mathbf{a})$ and $\bar{D}({}_r\mathbf{N}^{(j-1)}, M^{(j)}, \mathbf{a})$.

(*a*) If $\bar{D}({}_r\mathbf{N}^{(j-1)}, M^{(j)}, \mathbf{a}) > d$, go to Step 3.

(*b*) If (21) and (22) are satisfied and if $G^{(j)} \leqq NS \times C$, then store $({}_r\mathbf{N}^{(j-1)}, M^{(j)})$ as the new "optimal" solution. That is, set $\mathbf{N}^* = {}_r\mathbf{N}^{(j-1)}$, $M^* = M^{(j)}$ and

$$Z^* = \sum_{i=1}^{l} C_i \, {}_rN^{(j-1)} + CM^{(j)}.$$

Set ${}_1\mathbf{N}^{(j)} = {}_r\mathbf{N}^{(j-1)}$, increment j, and go to (*i*).

(*c*) If (21) and (22) are satisfied but $G^{(j)} > NS \times C$, then reduce the number of servers by subtracting 1 from $M^{(j)}$, increment NS, and return to (*iii*).

(*d*) If at least one blocking constraint of (21) is violated, i.e., $S^{(j)} = \{i \mid B_i({}_r\mathbf{N}^{(j-1)}, M^{(j)}, \mathbf{a}) > b_i\}$ is not empty, then add $\sum_{i \in S^{(j)}} C_i$ to $G^{(j)}$. Let

$$ {}_{r+1}N_i^{(j-1)} = {}_rN_i^{(j-1)} \qquad i \notin S^{(j)} $$

and

$$ {}_{r+1}N_i^{(j-1)} = {}_rN_i^{(j-1)} + 1 \qquad i \in S^{(j)}. $$

If $G^{(j)} > NS \times C$, then decrement $M^{(j)}$ by 1, increment r, and go to (*iii*). If $G^{(j)} \leqq NS \times C$, increment r, and go to (*iii*).

Step 3: Termination

If $\bar{D}({}_r\mathbf{N}^{(j-1)}, M^{(j)}, \mathbf{a}) > d$, then stop the procedure. The global optimum is $(\mathbf{N}^*, M^*, Z^*)$.

4.3 Numerical example

To illustrate the algorithm, we examine a simple two-queue system that represents an automatic call distributor system used for credit checking. The company has subscribed to two Inward WATS bands with a cost per trunk of \$800 and \$500. The two trunk groups each receive 15 erlangs of traffic in the busy hour. Calls, on the average, are 45 seconds in duration. The subscriber has requested a 5-second speed of answer and blocking objectives of P.10 and P.05, respectively. The monthly cost of an attendant is \$750. With these parameters, we begin the algorithm by using the Erlang loss formula with 15 erlangs and the delay formula with 30 erlangs to obtain $(\mathbf{N}^{(0)}, M_{(0)})$, which are (18, 20; 34). As shown in Table I, this initial solution is feasible and hence will be stored as our tentative optimal solution, $(\mathbf{N}^{(0)}, M^{(0)})$.

We proceed to Step 2 of the algorithm in an attempt to improve the initial feasible solution. By decreasing $M^{(0)}$ by one, $M^{(1)} = M^{(0)} - 1$, we obtain the system parameters (18, 20; 33) and the system characteristics (0.0912, 0.0504; 0.404), which indicate that this is not a feasible solution. Since the blocking constraint for trunk group 2 is violated, a trunk must be added to this group at a cost of \$500. The accumulated cost of the additional trunks, \$500, is less than the accumulated cost of the removed attendants, \$750. Therefore, we proceed by obtaining the system characteristics for this set of param-

Table I — An example of the optimization procedure

		N_1	N_2	M	B_1	B_2	$\bar{D}$	$Z(\$)$
Initial Feasible Solution	1	18	20	34	0.0887	0.0481	0.201	49900.
Solution Improvement	2	18	20	33	0.0912	0.0504	0.404	49150.
	3	18	21	33	0.0926	0.0373	0.546	49650.
	4	18	21	32	0.0970	0.0409	0.945	48900.
	5	18	21	31	0.1034	0.0460	1.550	48150.
	6	19	21	30	0.0939	0.0573	2.990	48200.
	* 7	19	22	30	0.0976	0.0466	3.422	48700.
	8	19	22	29	0.1120	0.0572	4.976	47950.
Termination	9	20	23	28	0.1220	0.0676	8.864	48500.

* Optimal solution.
System Parameters:

$a_1 = 15$ erlangs, $b_1 = 0.10$, $c_1 = \$800$
$a_2 = 15$ erlangs, $b_2 = 0.05$, $c_2 = \$500$
$HT = 45$ s, $d = 5$ s, $C = \$750$.

eters, i.e., (18, 21; 33). As Table I indicates, this is a feasible solution, is a cost improvement, and hence is stored as the tentative optimum.

As shown in Table I, the procedure continues from this point until (20, 23; 28), at which point the delay constraint is violated. The optimal solution is (19, 22; 30) at a cost of \$48,700. We should note that the parameters (20, 23; 29) were not examined since the accumulated cost of added trunks, \$1300, was greater than the accumulated cost of removed attendants, \$750.

The above solution is the global optimum for this constrained problem. However, practitioners might suggest that (18, 21; 31) is a more realistic design since, in fact, the blocking constraint is "essentially" satisfied (0.1034 vs 0.1000). This can be incorporated in the design procedure by allowing any solution that is within "ϵ" of a blocking objective to be retained. The algorithm can then be applied as before.

V. SUMMARY

In this paper, an analysis of a particular multiserver, multiqueue service system has been presented. Examples of this type of system are the directory assistance systems used in the telephone companies and credit verification bureaus used by the credit-card industry, which use automatic call distributors. Expressions were derived for the equilibrium-state probabilities, and four properties of the system characteristics, overall average delay, and the blocking probabilities for each queue were given.

These results were used in developing a procedure to obtain a least-cost system configuration to satisfy a given set of single-hour grade-of-

service constraints. That is, the procedure determines the number of waiting positions for each queue and the number of servers required to satisfy constraints placed on blocking probabilities and average delay at minimum cost. The work reported here should form the basis for the development of a practical method of traffic engineering and administration for small automatic call distributor systems.

REFERENCES

1. E. Brockmeyer, H. A. Halstrom, and A. Jensen, *The Life and Works of A. K. Erlang*, Trans. 2. Copenhagen: Danish Academy of Technical Sciences, 1948.
2. K. Lundkvist, "General Theory for Telephone Traffic," Ericsson Technics, *9*, No. 2, 1953, pp. 111–140.
3. H. Störmer, "Wartezeitlenkung in Handbedieten Vermittlungsanlagen," Archiv für Electronik und Übertragungstechnik, *10*, No. 2 (February 1956), pp. 58–64.
4. J. R. W. Smith and J. L. Smith, "Loss and Delay in Telephone Call Queueing Systems," A.T.E. Journal, *18*, No. 1 (January 1962), pp. 18–30.
5. J. Riordan, *Stochastic Service Systems*, New York: John Wiley, 1962, pp. 96–101.
6. L. Takács, "On a Combined Waiting Time and Loss Problem Concerning Telephone Traffic," Ann. Univ. Sci. Budapest Eötvös, Sect. Math., *1*, 1958, pp. 73–82.
7. A. Descloux, "On Markovian Servers with Recurrent Input," Sixth International Teletraffic Congress, Munich, 1970.
8. W. C. Chan, "Combined Delay- and Loss-Queueing System," Proc. IEE, *117*, No. 11 (November 1970).
9. G. Basharin, "Servicing Two Flows in a Single Queue System with a Limited Number of Places for Waiting and Absolute Priority," Eng. Cybernetics, *15*, Oct. 1967, pp. 95–105.
10. W. Wagner, "On Combined Delay and Loss Systems with Nonpreemptive Priority Service," Fifth International Teletraffic Congress, New York, 1967.
11. J. Brandt, "A Multiserver Queueing System with Preemptive Priority," Sixth International Teletraffic Congress, Munich, 1970.
12. P. Kühn, "Parallel Waiting Queues in Real-Time Computer Systems," Nachrichtentech. Z., *23*, No. 11 (November 1970), pp. 576–582.
13. P. Kühn, "Combined Delay and Loss Systems with Several Input Queues, Full and Limited Accessibility," Sixth International Teletraffic Congress, Munich, 1970.
14. P. Kühn, "Combined Delay and Loss Systems with Several Input Queues, Full and Limited Accessibility," Archiv für Electronik und Übertragungstechnik, *25*, No. 10–11 (September-October 1971), pp. 449–454.
15. R. Syski, *Introduction to Congestion Theory in Telephone Systems*, London: Oliver & Boyd, 1960, pp. 254.
16. J. D. C. Little, "A Proof for the Queueing Formula: $L = \lambda W$," Operations Research, *9*, No. 3 (May-June 1961), pp. 383–387.
17. R. Morris and E. Wolman, "A Note on 'Statistical Equilibrium,'" Operations Research, *9*, No. 5 (September-October 1961), pp. 751–753.
18. B. A. Whitaker, "An Analysis and Optimization of Multiserver, Multiqueue Systems with Finite Waiting Space in Each Queue," Ph.D. Dissertation, New York University, 1974.
19. R. B. Cooper, *Introduction to Queueing Theory*, New York: Macmillan, 1972, p. 100.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

Nonstationary Blocking in Telephone Traffic

By D. L. JAGERMAN

(Manuscript received September 10, 1974)

Blocking is considered for an N-trunk group of exponential servers with Poisson-offered load whose rate parameter varies with time. The infinite trunk case is solved by means of a rapidly convergent series of Poisson-Charlier polynomials. This solution is used to obtain practical approximations of blocking probability, transition probabilities, and recovery function for general time-variable offered load in the finite trunk-group case. An integral equation is derived satisfied by the blocking probability in the general case. In the situation of constant offered load, two additional methods are derived for providing easily computable approximations; one based on the integral equation, the other based on an approximate inversion of the Laplace transform. To aid in the latter approximation, bounds on the roots of Poisson-Charlier polynomials are obtained; in particular, an approximation is obtained for the dominant root. The inversion of the integral equation is studied with the purpose of providing the basis for future investigations of errors of approximation. Curves are provided for a number of examples permitting comparison of exact and approximate solutions.

I. INTRODUCTION

The main purpose of this paper is to present a discussion of the time behavior of blocking in a fully available N-trunk group for any initial state with exponential servers and with Poisson-offered load whose rate parameter $a(t)$ itself may be considered to vary with time; that is, the probability of j calls arriving in a time interval $(0, t)$ is assumed given by

$$\exp\left(-\int_0^t a(u)du\right)\frac{\left[\int_0^t a(u)du\right]^j}{j!}, \qquad j = 0, 1, 2, \cdots. \tag{1}$$

The service rate is taken equal to 1, so that $a(t)$ is measured in erlangs.

The problem of blocking with time-variable offered load was considered by Palm[1] for finite trunk groups and by Khintchine[2] for infinite

trunk groups. The impetus for this is the need felt for more accurate computation of blocking probabilities and correlation information[3] than can be obtained by quasi-stationary analyses, that is, by the use of equilibrium formulas in which the offered load parameter is replaced by its instantaneous value. Lack of statistical equilibrium renders this approach inaccurate. The time-variable aspect of the input stream should be carefully distinguished from other statistical descriptions such as peakedness,[4] since the effects on the system are separately identifiable. It has been reported, for example, that offered load and peakedness determinations from carried usage, peg count attempts offered to a group, and overflows are misled by the time variability of the offered load.

Palm had proposed an interesting method of accounting for the time variability of the offered load, i.e., his "slow variations" model. In this model, it was assumed that the actual ordering in time of $a(t)$, that is, the functional dependence of $a(t)$ on t could be ignored if $a(t)$ varied slowly. He replaced $a(t)$ by a random variable with an incomplete gamma function distribution.[1] Thus, traffic functions such as blocking may be obtained from their equilibrium values by averaging over the appropriate gamma distribution. This model, however, requires further elaboration in view of the investigations of Iversen,[3] who showed that the Palm approach does not correctly model the empirical data collected in the extensive Holbaek measurements of Danish telephone traffic. Iversen found that the correct time variation of the traffic could not be ignored.

The trunk provisioning procedure whereby one uses the average offered load over a busy hour to achieve a required grade of service results, in some cases that were considered, in only a small underestimate of the required number of trunks as calculated by the methods of this paper. Since the standard method is convenient, this may be viewed as substantiation of the approach.

Essential for the methods of this paper is a Volterra integral equation derived in Section II satisfied by the blocking probability, $P_N(t, N)$, experienced by a load of $a(t)$ erlangs offered to an N-trunk group. Exact analytical solution of this equation is not useful, but numerical methods may be advantageously used. An important feature of the equation, nonetheless, is that it permits studying errors of approximation and, in one instance (Appendix D), was directly used in the construction of an approximate operator for studying the transient response in the case of constant $a(t)$. Appendix A presents an explicit representation of $P_N(t, N)$ for general $a(t)$ by means of an infinite Neumann expansion. Inequalities for $P_N(t, N)$ and truncation error estimates for the Neumann series are also given.

The infinite trunk group case forms the basis of the approximations developed in Section IV that are applicable to the general case of time variable $a(t)$. Although this case was solved by Khintchine,[2] Section III presents new representations in terms of rapidly convergent Poisson-Charlier expansions. Truncation error estimates are obtained, and the rate at which the state probabilities approach the Poisson form is assessed. To aid in the use of Poisson-Charlier polynomials, Appendix B provides a short discussion of their properties, especially providing convenient means of expanding a function into a Poisson-Charlier series. Since the state probabilities of the infinite trunk group system are often close to Poisson, this form of representation is very useful. The Poisson-Charlier expansion expresses the deviation of a function from the Poisson form. Further, in Section IV, the Poisson-Charlier polynomials are used to express the transition probabilities in explicit, closed form.

The approximations of Section IV are applicable to time variable $a(t)$, and are developed from the infinite trunk group solution by renormalization appropriate to the finite trunk group. To facilitate the use of the approximations, closed expressions are obtained for the relevant infinite trunk group solutions. This approximation procedure gives rise to the useful notion of a "modified offered load." One of the approximations obtained was, in fact, already obtained by Palm.[1] This approximation is particularly interesting because it uses the Erlang loss function, $B(N, a)$, for which rapid methods of computation are available.[5,6] A special case of the approximations for transition probabilities is that for the recovery function, which is important in the discussion of correlation properties[7] and, hence, in the determination of variances of traffic parameter estimators.

The constant offered load case is studied in Appendix C. Although the solution for the state probabilities is known,[8] the integral equation formulation appears to be new. Certain advantages are obtainable from this formulation. The errors of approximations to the state probabilities satisfy the same integral equation but with a different inhomogeneous term; thus, the more general integral equation is studied, leading to methods for investigating the quality of approximation. For this purpose, a natural Banach space (uniform norm over $[0, \infty]$) is introduced, in which the integral operator is bounded and has a bounded inverse. Of course, the known Laplace transform of the transition probabilities is immediately obtained as a corollary. The integral equation is also used in Appendix D as a tool for the construction of an approximate solution (the scaling method) corresponding to an arbitrary initial state. Appendix C also presents several bounds on the required roots of Poisson-Charlier polynomials by

obtaining general bounds on the largest and smallest members of sets of positive numbers subject to isoperimetric constraints; the results are then specialized to the Poisson-Charlier polynomials. One of the bounds for the dominant root is explicit and easily calculable. Its accuracy appears to be good (see Fig. 8).

Appendix D provides two approximations for the case of constant offered load. The scaling approximation, whose genesis was suggested by S. Horing, is constructed by obtaining an approximate time invariant of the transition blocking probability from the initially empty state. The subsequent generalization of this approximation to arbitrary initial states is then obtained by means of the integral equation. The Laplace transform approximation is constructed by an adaptation of the Widder formula[9] for the inversion of the Laplace transform. It requires the determination of the dominant root; but, depending on the needs, it may be made arbitrarily accurate. It has the interesting property that, under certain conditions, it provides bounds for the exact solution.

A discussion of some results and graphical illustrations is given in Section V. In testing the quality of the modified offered load approximations, high change rates of $a(t)$ were chosen, in fact, much higher than would occur in practice. The errors of approximation increase with increasing rate of change of $a(t)$, hence, the examples chosen indicate much higher errors than one would expect to encounter in the practical application of these methods.

II. INTEGRAL EQUATION FOR BLOCKING

It is the object of this section to establish Theorems 1 and 2, which provide integral representations of the binomial moments (13) of the probability distribution (2) of the number of busy trunks, and corollary 1 of Theorem 2, in which an integral equation is given for the probability that all trunks are busy at a given time.

Let $\xi = \xi(t, N)$ be the number of trunks busy at time t in an N-trunk group, and $P_j = P_j(t, N)$ the corresponding probability,

$$P[\xi(t, N) = j] = P_j(t, N). \tag{2}$$

The probability generating function $g(t, \zeta, N)$ is given by

$$g = E\zeta^{\xi(t,N)} = \sum_{j=0}^{N} P_j(t, N)\zeta^j. \tag{3}$$

At the point $t + dt$, one has

$$g(t + dt, \zeta, N) = g + \partial g, \qquad \xi(t + dt, N) = \xi + d\xi \tag{4}$$

and, hence,

$$g + \partial g = E_\xi\{\zeta^\xi E[\zeta^{d\xi} | \xi]\}, \tag{5}$$

in which $E[\zeta^{d\xi} | \xi]$ is the conditional expectation of $\zeta^{d\xi}$ given ξ, and E_ξ is the expectation over the probability distribution of ξ. The boundary at $j = N$ necessitates a further analysis of (5). One has

$$\begin{aligned} g + \partial g = (1 - P_N)E_\xi\{\zeta^\xi E[\zeta^{d\xi} | \xi < N]\} \\ + P_N E_\xi\{\zeta^\xi E[\zeta^{d\xi} | \xi = N]\}. \end{aligned} \tag{6}$$

The probability distribution of $d\xi$ is

$$\begin{aligned} &P[d\xi = -1 | \xi = j, 0 \leqq j < N] = jdt, \\ &P[d\xi = 0 | \xi = j, 0 \leqq j < N] = 1 - (a + j)dt, \\ &P[d\xi = 1 | \xi = j, 0 \leqq j < N] = adt, \\ &P[d\xi = 0 | \xi = N] = 1 - Ndt, \\ &P[d\xi = -1 | \xi = N] = Ndt; \end{aligned} \tag{7}$$

hence,

$$\begin{aligned} &E[\zeta^{d\xi} | \xi = j, 0 \leqq j < N] = 1 + (a\zeta - a - j + j\zeta^{-1})dt, \\ &E[\zeta^{d\xi} | \xi = N] = 1 - N(1 - \zeta^{-1})dt. \end{aligned} \tag{8}$$

From (6) and (8), one has

$$\begin{aligned} \frac{\partial g}{\partial t} &= (1 - P_N)E_\xi[\zeta^\xi(a\zeta - a - \xi + \xi\zeta^{-1}) | \xi < N] \\ &\quad - P_N E_\xi[\xi\zeta^{\xi-1}(\zeta - 1) | \xi = N], \end{aligned} \tag{9}$$

$$\frac{\partial g}{\partial t} = (1 - P_N)a(\zeta - 1)E_\xi[\zeta^\xi | \xi < N] - (\zeta - 1)E_\xi[\xi\zeta^{\xi-1}];$$

hence,

$$\frac{\partial g}{\partial t} + (\zeta - 1)\frac{\partial g}{\partial \zeta} = a(\zeta - 1)g - a(\zeta - 1)\zeta^N P_N. \tag{10}$$

The infinite trunk group does not require the analysis of (6) nor the boundary conditions ($\xi = N$) of (7), hence (9) becomes

$$\frac{\partial g}{\partial t} = a(\zeta - 1)g - (\zeta - 1)E_\xi[\xi\zeta^{\xi-1}]. \tag{11}$$

Thus, the corresponding equation for the infinite trunk group is

$$\frac{\partial g}{\partial t} + (\zeta - 1)\frac{\partial g}{\partial \zeta} = a(\zeta - 1)g. \tag{12}$$

The binomial moments $\beta_s(t, N)$ are defined by

$$\beta_s(t, N) = \sum_{j=s}^{\infty} P_j(t, N)\binom{j}{s}, \tag{13}$$

in which $\binom{j}{s}$ designates the binomial function defined by

$$\begin{aligned}\binom{j}{s} &= \frac{j(j-1)\cdots(j-s+1)}{s!}, \qquad s \geqq 1,\\ \binom{j}{0} &= 1.\end{aligned} \tag{14}$$

The binomial moment generating function $l(t, w, N) = \sum_{s=0}^{\infty} \beta_s(t, N)w^s$ is given by

$$l(t, w, N) = g(t, 1 + w, N). \tag{15}$$

Thus, the differential equation satisfied by l is

$$\frac{\partial l}{\partial t} + w\frac{\partial l}{\partial w} = awl - aw(1 + w)^N P_N. \tag{16}$$

The corresponding equation for the infinite trunk group is

$$\frac{\partial l}{\partial t} + w\frac{\partial l}{\partial w} = awl. \tag{17}$$

Equation (16) is a linear partial differential equation that can be solved by the following device (method of characteristics). Let θ be a new, independent variable and set

$$\begin{aligned} l &= l(\theta),\\ w &= w(\theta),\\ t &= t(\theta).\end{aligned} \tag{18}$$

Then, comparison of

$$\frac{dl}{d\theta} = \frac{\partial l}{\partial t}\frac{dt}{d\theta} + \frac{\partial l}{\partial w}\frac{dw}{d\theta} \tag{19}$$

with (16) yields the set of equations

$$\begin{aligned}\frac{dl}{d\theta} &= aw - aw(1 + w)^N P_N,\\ \frac{dw}{d\theta} &= w,\\ \frac{dt}{d\theta} &= 1,\end{aligned} \tag{20}$$

whose solution for l is then obtained. To exhibit the solution conveniently, let

$$\alpha(t, \tau) = e^{-t}\int_{\tau}^{t} e^s a(s)ds, \tag{21}$$

and

$$\Lambda = \Lambda(t) = \alpha(t, 0) = e^{-t} * a(t), \tag{22}$$

where $*$ designates convolution product over $(0, t)$. The solution of (16) takes the form

$$l(t, w, N) = l(0, we^{-t}, N)e^{\Lambda w} - w\int_0^t e^{\alpha w + \tau - t}[1 + we^{\tau - t}]^N a(\tau)P_N(\tau, N)d\tau. \tag{23}$$

Similarly, the solution of (17) is

$$l(t, w, \infty) = l(0, we^{-t}, \infty)e^{\Lambda w}. \tag{24}$$

To obtain the binomial moments themselves, it is convenient to introduce the Volterra operator K_s,

$$K_s f = \int_0^t K_s(t, \tau)f(\tau)d\tau, \tag{25}$$

defined by the kernel

$$K_s(t, \tau) = \sum_{j=0}^{s-1} \frac{\alpha^j}{j!}\binom{N}{s-1-j} e^{-(s-j)(t-\tau)}a(\tau). \tag{26}$$

Since the Laguerre polynomial $L_n^{(\alpha)}$ $(-x)$[10] is given by

$$L_n^{(\alpha)}(-x) = \sum_{j=0}^{n} \frac{x^j}{j!}\binom{n+\alpha}{n-j}, \tag{27}$$

the kernel $K_s(t, \tau)$ may be written more compactly; thus,

$$K_s(t, \tau) = a(\tau)e^{-s(t-\tau)}L_{s-1}^{(N-s+1)}(-\alpha e^{t-\tau}). \tag{28}$$

The following theorems may now be stated.

Theorem 1: *The binomial moments, $\beta_s(t, \infty)$, for the infinite trunk group are given by*

$$\beta_s(t, \infty) = \sum_{j=0}^{s} \beta_{s-j}(0, \infty)e^{-(s-j)t}\frac{\Lambda^j}{j!}.$$

Proof: The coefficient of w^s in the expansion of the right-hand side of (24) yields the result.

Theorem 2: *The binomial moments, $\beta_s(t, N)$, for the finite trunk group satisfy*

$$\beta_s(t, N) = \beta_s(t, \infty) - K_s P_N.$$

Proof: The coefficient of w^s in the expansion of the right-hand side of (23) provides the required result.

Corollary 1: *The probability, $P_N(t, N)$, that all trunks are busy satisfies the integral equation*

$$P_N(t, N) = \beta_N(t, \infty) - K_N P_N,$$

in which

$$\beta_N(t, \infty) = \sum_{j=0}^{N} \beta_{N-j}(0, N) e^{-(N-j)t} \frac{\Lambda^j}{j!}$$

$$K_N(t, \tau) = a(\tau) e^{-N(t-\tau)} L_{n-1}^{(1)}(-\alpha e^{t-\tau}).$$

Proof: For the finite trunk group (13) shows that

$$\beta_N(t, N) = P_N(t, N).$$

Hence, the integral equation follows from Theorem 2. The explicit expressions for $\beta_N(t, \infty)$ and $K_N(t, \tau)$ are obtained from Theorem 1 and (28), respectively.

The special case of all trunks free initially leads to a somewhat simpler integral equation for $P_N(t, N)$. This is given in the following corollary.

Corollary 2: *When all trunks are initially free, $P_N(t, N)$ satisfies*

$$P_N(t, N) = \frac{\Lambda^N}{N!} - K_N P_N.$$

Proof: The initial probability distribution $P_j(0, N)$ has the form

$$\begin{aligned} P_j(0, N) &= 1 \qquad j = 0, \\ &= 0 \qquad j > 0. \end{aligned} \tag{29}$$

Hence, the binomial moments satisfy

$$\begin{aligned} \beta_s(0, N) &= 1 \qquad s = 0, \\ &= 0 \qquad s > 0. \end{aligned} \tag{30}$$

The equation for $\beta_N(t, \infty)$ given in the first corollary now yields

$$\beta_N(t, \infty) = \frac{\Lambda^N}{N!}. \tag{31}$$

The result of the corollary follows.

The probability $P_N(t, N)$ corresponding to all trunks initially busy is called the recovery function; it satisfies the following integral equation.

Corollary 3: *When all trunks are initially busy, one has*

$$P_N(t, N) = e^{-Nt} L_N(-e^t \Lambda) - K_N P_N.$$

Proof: We have

$$\begin{aligned} P_j(0, N) &= 0 \qquad 0 \leqq j < N, \\ &= 1 \qquad j = N. \end{aligned} \tag{32}$$

Hence, the required binomial moments are

$$\beta_s(0, N) = \binom{N}{s}. \tag{33}$$

The result follows from the equation for $\beta_N(t, \infty)$ and from (27).

A noteworthy case occurs when a is constant, then the integral equation for $P_N(t, N)$ becomes of convolution type. Thus, let

$$K_N(t) = ae^{-Nt}L_{N-1}^{(1)}[-a(e^t - 1)]. \tag{34}$$

Then one has Corollary 4.

Corollary 4: *When the offered load a is constant, $P_N(t, N)$ satisfies*

$$P_N(t, N) = \beta_N(t, \infty) - K_N * P_N.$$

Proof: Use of (21) and $K_N(t, \tau)$ as given in Corollary 1 yields the result.

For constant a, an equilibrium distribution $P_j(\infty, N)$ exists.[2] Let

$$S_N(a) = \sum_{k=0}^{N} \frac{a^k}{k!}. \tag{35}$$

Then

$$P_j(\infty, N) = \frac{a^j}{j!\, S_N(a)}, \qquad 0 \leqq j \leqq N. \tag{36}$$

The notation $B(N, a)$ is used for the blocking probability $P_N(\infty, N)$ and is referred to as Erlang's loss formula.[6] Corresponding to the equilibrium distribution, one has the binomial moments $\beta_s(0, N)$ and, hence, the moments for the infinite trunk group given in Theorem 1. These moments will be denoted by $\beta_s^e(t, a)$. It may be noted that

$$\lim_{t \to \infty} \beta_s^e(t, a) = \frac{a^s}{s!}. \tag{37}$$

The integral of $K_N(t)$ that is useful in error analyses may now be easily obtained.

Theorem 3: *When the offered load is constant, we have*

$$\int_0^t K_N(\tau)d\tau = \frac{\beta_N^e(t, a)}{B(N, a)} - 1,$$

$$\int_0^\infty K_N(\tau)d\tau = S_N(a) - 1,$$

in which $K_N(\tau)$ is given in (34).

Proof: Since $B(N, a)$ is the equilibrium solution of the integral equation of Corollary 4, and since $\beta_N^e(t, a)$ corresponds to the equilibrium state, the solution of the integral equation is constant and equal to $B(N, a)$; thus,

$$B(N, a) + K_N * B(N, a) = \beta_N^e(t, a).$$

This immediately implies the first equation of the theorem. The second follows on considering $t \to \infty$, and using (36) and (37).

It may be useful to observe that the positive character of the general operator K_N in Corollary 1 immediately supplies the inequalities

$$\beta_N(t, \infty) - K_N\beta_N(t, \infty) < P_N(t, N) < \beta_N(t, \infty). \tag{38}$$

A Neumann-series solution of the integral equation of Corollary 1 is discussed in Appendix A. Higher-order inequalities of type (38) are also given.

III. THE INFINITE TRUNK SOLUTION

It will be convenient to express the solution for $P_x(t, \infty)$ in terms of Poisson-Charlier polynomials[11] whose relevant properties are discussed in Appendix B. The probability distribution of the number of busy trunks for the infinite trunk case was considered by Khintchine[2] and, for constant offered load, by Karlin and McGregor.[12] Theorem 4 presents a rapidly convergent form of the solution in terms of Poisson-Charlier polynomials valid for any initial state. This solution will be the main tool for the construction of approximations to distributions in the finite trunk case.

From (15), let $l(0, w, \infty)$ be given by

$$l(0, w, \infty) = \sum_{j=0}^{\infty} \beta_j(0, \infty) w^j, \tag{39}$$

then the binomial moment generating function for the infinite trunk case is, from (24),

$$l(t, w, \infty) = e^{\Lambda w} \sum_{j=0}^{\infty} \beta_j(0, \infty) e^{-jt} w^j. \tag{40}$$

The mean, μ, of this distribution is the coefficient of w; hence,

$$\mu = \Lambda + \mu_0 e^{-t}, \tag{41}$$

in which μ_0 is the mean of the initial distribution. One may now state Theorem 4.

Theorem 4: *The probability distribution,* $P_x(t, \infty)$, *of the number of busy trunks in an infinite trunk group has the convergent representation*

$$P_x(t, \infty) = \psi(x, \mu) \sum_{j=0}^{\infty} e^{-jt} d_j G_j(x, \mu),$$

$$d_j = \sum_{\nu=0}^{j} \frac{(-1)^\nu}{\nu!} \mu^\nu \beta_{j-\nu}(0, \infty),$$

provided that

$$l(0, z-1, \infty) = \sum_{j=0}^{\infty} \beta_j(0, \infty)(z-1)^j = \sum_{j=0}^{\infty} P_j(0, \infty) z^j$$

converges for $|z| < r (r > 2)$.

Proof: Use of (40) in (101) which the choice $\lambda = \mu$ as given in (41) yields

$$C(w) = e^{-\mu_0 w e^{-t}} \sum_{j=0}^{\infty} \beta_j(0, \infty) e^{-jt} w^j. \tag{42}$$

Hence,

$$c_j = e^{-jt} \sum_{\nu=0}^{j} \frac{(-1)^\nu}{\nu!} \mu_0^\nu \beta_{j-\nu}(0, \infty) \tag{43}$$

and, from (95),

$$P_x(t, \infty) = \psi(x, \mu) \sum_{j=0}^{\infty} c_j G_j(x, \mu). \tag{44}$$

Let

$$d_j = \sum_{\nu=0}^{j} \frac{(-1)^\nu}{\nu!} \mu_0^\nu \beta_{j-\nu}(0, \infty), \tag{45}$$

then

$$c_j = e^{-jt} d_j, \tag{46}$$

and the formula of the theorem follows. A theorem of Uspensky[13] states that the general representation of (95) is valid in the sense that the series converges to $F(x)$ if the series $\sum_{x=0}^{\infty} F(x) z^x$ has radius of convergence greater than 2. Since

$$l(t, z-1, \infty) = e^{-\Lambda + \Lambda z} l[0, (z-1)e^{-t}, \infty], \tag{47}$$

the radius of convergence of $l(t, z-1, \infty)$ is greater than 2 by the condition on $l(0, z-1, \infty)$ stated in the theorem. Hence, by Uspensky's theorem, the representation of (44) is valid for all $t \geqq 0$.

A truncation error estimate for the series representation of Theorem 4 is given in Theorem 5.

Theorem 5:

$$|P_x(t, \infty) - \psi(x, \mu) - \psi(x, \mu) \sum_{j=2}^{k} e^{-jt} d_j G_j(x, \mu)|$$

$$\leqq \left(\frac{2}{w}\right)^{k+1} \frac{e^{-(k+1)t}}{1 - (2/w)e^{-t}} e^{\mu_0 w} l(0, w, \infty),$$

$$k \geqq 1, \qquad R > w > 0, \qquad t > \ln \frac{2}{w},$$

in which R is the radius of convergence of $l(0, w, \infty)$.

Proof: Since

$$|\psi(x, \mu) G_j(x, \mu)| = |\psi^{(j)}(x, \mu)| \leqq 2^j, \qquad j \geqq 0, \tag{48}$$

we have

$$\sum_{j>k} e^{-jt} |d_j| \, |\psi(x, \mu) G_j(x, \mu)| \leqq \sum_{j>k} e^{-jt} 2^j |d_j|. \tag{49}$$

Also, from (45),

$$|d_j| \leqq \sum_{j=0}^{\nu} \frac{\mu_0^{\nu}}{\nu!} \beta_{j-\nu}(0, \infty) \leqq e^{\mu_0 w} l(0, w, \infty) w^{-j}. \tag{50}$$

Thus,

$$\sum_{j>k} e^{-jt} |d_j| \, |\psi(x, \mu) G_j(x, \mu)| \leqq e^{\mu_0 w} l(0, w, \infty) \sum_{j>k} e^{-jt} 2^j w^{-j} \tag{51}$$

$$\leqq e^{\mu_0 w} l(0, w, \infty) \left(\frac{2}{w}\right)^{k+1} \frac{e^{-(k+1)t}}{1 - (2/w)e^{-t}}. \tag{52}$$

The conditions of the theorem ensure the convergence of $l(0, w, \infty)$ and of the series of (51).

The corollaries below follow directly under the conditions of Theorem 5.

Corollary 1:

$$P_x(t, \infty) = \psi(x, \mu) \sum_{j=0}^{k} e^{-jt} d_j G_j(x, \mu) + 0(e^{-(k+1)t}).$$

Corollary 2:

$$|P_x(t, \infty) - \psi(x, \mu)| \leqq \frac{4}{w^2} \frac{e^{-2t}}{1 - (2/w)e^{-t}} e^{\mu_0 w} l(0, w, \infty).$$

Corollary 3:

$$P_x(t, \infty) = \psi(x, \mu) + 0(e^{-2t}).$$

Thus, the distribution quickly becomes nearly Poisson with the time

variable parameter μ, regardless of the initial distribution. In fact, if the initial distribution is Poisson, one might anticipate $P_x(t, \infty)$ to remain Poisson for all $t \geqq 0$. This is asserted by Theorem 6.

Theorem 6: *If* $P_x(0, \infty)$ *is Poisson with parameter* μ_0, *then*

$$P_x(t, \infty) = \psi(x, \mu).$$

Proof: The binomial moments, $\beta_s(0, \infty)$, are

$$\beta_s(0, \infty) = \frac{\mu_0^s}{s!}, \qquad s \geqq 0. \tag{53}$$

It follows, from (45), that

$$d_j = \frac{\mu_0^j}{j!} \sum_{\nu=0}^{j} \binom{j}{\nu} (-1)^\nu = 0, \qquad j > 0. \tag{54}$$

The result is now obtained from Theorem 3.

IV. APPROXIMATIONS

The Neumann series (75) is an explicit solution of the integral equation of blocking given in Corollary 1, Theorem 2. For constant offered load a, the resolvent kernel solution (108) and Theorem 11 are also available; however, especially when N is large, these solutions are not convenient. It is therefore important to have approximations that lend themselves to computation for large N in a sufficiently convenient manner. Three such approximations have been developed, namely: the "modified offered load" approximation that is useful and fairly accurate in the general case, that is, for time variable offered load, the "scaling" approximation, and the "Laplace transform" approximation, which are applicable only to transient phenomena under constant offered load. The scaling approximation is also fairly accurate and does not require factorization of polynomials. The Laplace transform approximation consists in fact of an infinite set of approximations of arbitrarily high accuracy. It usually requires finding a single root—the so-called dominant root—of an appropriate polynomial. The scaling and Laplace transform approximations are discussed in Appendix D. Appendix C provides approximations for the required dominant root. The modified offered load approximation is presented below.

Let $P_{i,x}(t, N)$ be the probability that the N-trunk group started from state i at time 0 and proceeded to state x at time t, then $P_{i,x}(t, \infty)$ may be computed from Theorem 4 using

$$\beta_j(0, \infty) = \binom{i}{j}. \tag{55}$$

An approximation, $\bar{P}_{i,x}(t, N)$, for $P_{i,x}(t, N)$ is given by

$$\bar{P}_{i,x}(t, N) = \frac{P_{i,x}(t, \infty)}{\sum_{\nu=0}^{N} P_{i,\nu}(t, \infty)}. \tag{56}$$

This approximation is suggested by the following considerations. One has

$$\lim_{N\to\infty} \bar{P}_{i,x}(t, N) = P_{i,x}(t, \infty). \tag{57}$$

Hence, the approximation should be accurate even for time-varying offered load when N is large. Furthermore, when a is constant, since $\lim_{t\to\infty} \mu = a$, one has

$$\lim_{t\to\infty} \bar{P}_{i,x}(t, N) = \frac{a^x}{x!\, S_N(a)}, \tag{58}$$

which, as previously indicated, is the exact equilibrium distribution; hence, the approximation should be accurate when t is large even for time-varying $a(t)$, provided $\dot{a}(t)$ is small. Since, by the law of total probability,

$$P_x(t, N) = \sum_{i=0}^{N} P_i(0, N) P_{i,x}(t, N), \tag{59}$$

one can construct an approximation, $\bar{P}_x(t, N)$, to $P_x(t, N)$ by use of $\bar{P}_{i,x}(t, N)$; thus,

$$\bar{P}_x(t, N) = \sum_{i=0}^{N} P_i(0, N) \bar{P}_{i,x}(t, N). \tag{60}$$

To facilitate the use of (60), Theorem 7 expresses $P_{i,x}(t, \infty)$ in finite form.

Theorem 7:

$$P_{i,x}(t, \infty) = (1 - e^{-t})^i e^{-\Lambda} \frac{(-\Lambda)^x}{x!} G_x[i, -(e^t - 1)\Lambda].$$

Proof: From Theorem 4 and (55), we have

$$P_{i,x}(t, \infty) = \psi(x, \mu) \sum_{j=0}^{\infty} e^{-jt} d_j G_j(x, \mu), \tag{61}$$

$$d_j = \sum_{\nu=0}^{j} \frac{(-1)^\nu}{\nu!} i^\nu \binom{i}{j-\nu}. \tag{62}$$

Comparison of (62) with (90) shows that

$$d_j = \frac{i^j}{j!} G_j(i, i). \tag{63}$$

Let

$$C_j = \frac{(ie^{-t})^j}{j!} G_j(i, i), \tag{64}$$

then

$$C(w) = \sum_{j=0}^{\infty} \frac{(ie^{-t})^j}{j!} G_j(i, i)w^j, \tag{65}$$

which, by comparison with (89), may be rewritten

$$C(w) = e^{-ie^{-t}w}(1 + e^{-t}w)^i. \tag{66}$$

Let

$$B(w) = e^{\mu w}C(w) = e^{\Lambda w}(1 + e^{-t}w)^i \tag{67}$$

and

$$g(z) = B(z - 1) = e^{-\Lambda}e^{\Lambda z}(1 - e^{-t} + e^{-t}z)^i. \tag{68}$$

Then

$$P_{i,x}(t, \infty) = (1 - e^{-t})^i(e^t - 1)^{-x}e^{-\Lambda} \sum_{\nu=0}^{x} \binom{i}{x - \nu} \frac{[(e^t - 1)\Lambda]^\nu}{\nu!}. \tag{69}$$

Hence, comparison of (69) with (90) finally yields

$$P_{i,x}(t, \infty) = (1 - e^{-t})^i e^{-\Lambda} \frac{(-\Lambda)^x}{x!} G_x[i, -(e^t - 1)\Lambda]. \tag{70}$$

Immediate corollaries are the following:

Corollary 1:

$$P_x(t, \infty) = e^{-\Lambda} \frac{(-\Lambda)^x}{x!} \sum_{i=0}^{\infty} P_i(0, \infty)(1 - e^{-t})^i G_x[i, -(e^t - 1)\Lambda].$$

Corollary 2:

$$\bar{P}_{i,x}(t, N) = \frac{G_x[i, -(e^t - 1)\Lambda]}{\sum_{\nu=0}^{N} (x!/\nu!)(-\Lambda)^{\nu-x} G_\nu[i, -(e^t - 1)\Lambda]}.$$

Of particular interest are the functions $P_{0,N}(t, N)$ and $P_{N,N}(t, N)$; the first describes the progression of the system from initially empty to blocked, and the second describes the recovery of the system from an initially blocked condition to the blocked condition again. The latter function is called the "recovery function." [7] The following formulas are obtained from Corollary 2.

$$\bar{P}_{0,N}(t, N) = B(N, \mu). \tag{71}$$

The general principle of approximation employed, namely, the renormalization of an appropriate solution for the infinite trunk group

case, allows one to state, by use of Theorem 6, that, whenever one starts from an approximately Poisson initial state, an approximation to $P_N(t, N)$ is

$$P_N(t, N) = B(N, \mu). \tag{72}$$

This approximation was known to C. Palm.[1] The parameter μ is regarded as a modified offered load.

The recovery function approximation obtained from Corollary 2 is

$$\bar{P}_{N,N}(t, N) = \frac{G_N[N, -(e^t - 1)\Lambda]}{\sum_{\nu=0}^{N} (N!/\nu!)(-\Lambda)^{\nu-N} G_\nu[N, -(e^t - 1)\Lambda]}. \tag{73}$$

V. NUMERICAL EXAMPLES

For the purpose of providing some idea of the accuracy of the approximations developed in Section IV and Appendix D, curves were drawn up comparing exact and approximate solutions for a group of ten trunks. These curves illustrate nonstationary behavior. Figure 1 shows the scaling and modified offered load approximations for a step input problem in which $a = 7$ erlangs is offered to an initially empty group. The scaling approximation of (166) was used, and (71) was used for the modified offered load approximation. Apparently, for this situation, the scaling approximation is somewhat more accurate.

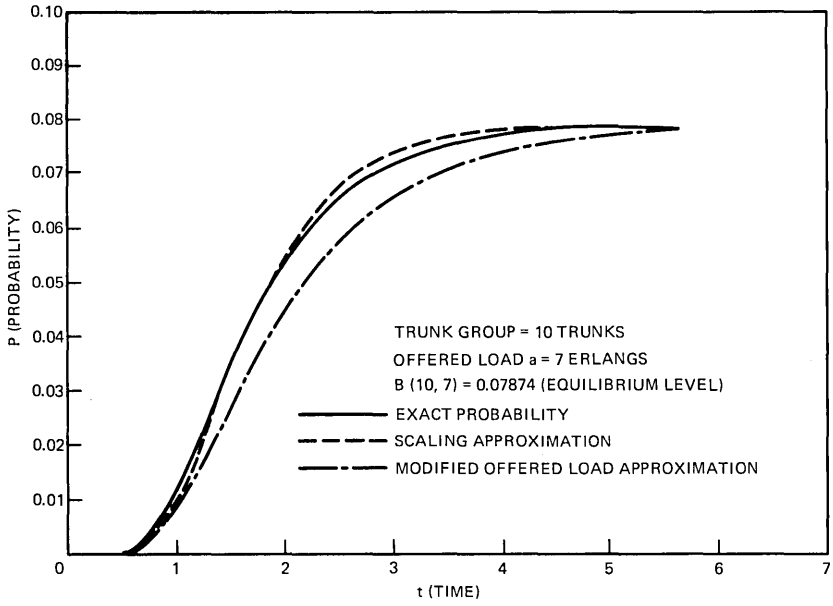

Fig. 1—All trunks empty initially—scaling and modified offered load approximations.

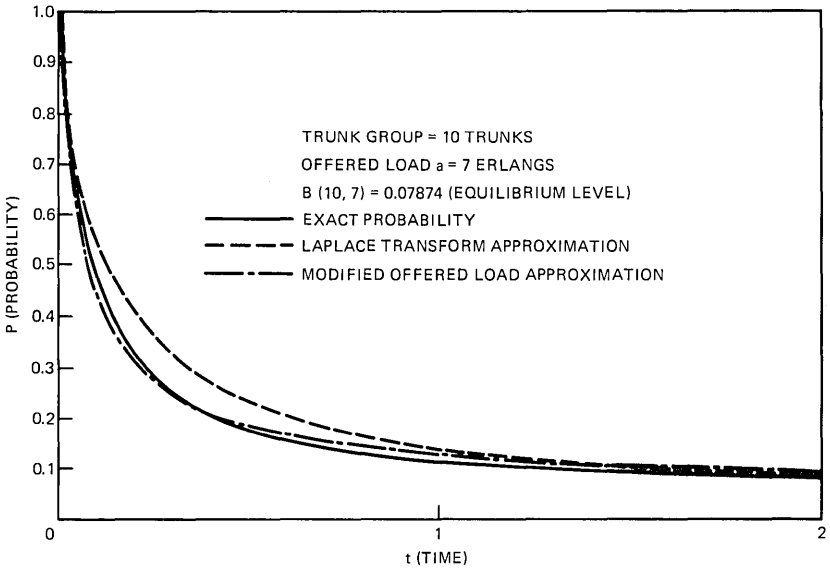

Fig. 2—Recovery function—Laplace transform and modified offered load approximations.

The recovery function approximations of (73) and (187) are compared to the exact solution of (185) in Fig. 2. The approximations are correct at the extremes $t = 0, t = \infty$, and track the exact curve reasonably well. The approximation of (73) is more accurate initially and is, of course, also applicable when the offered load is time-variable; how-

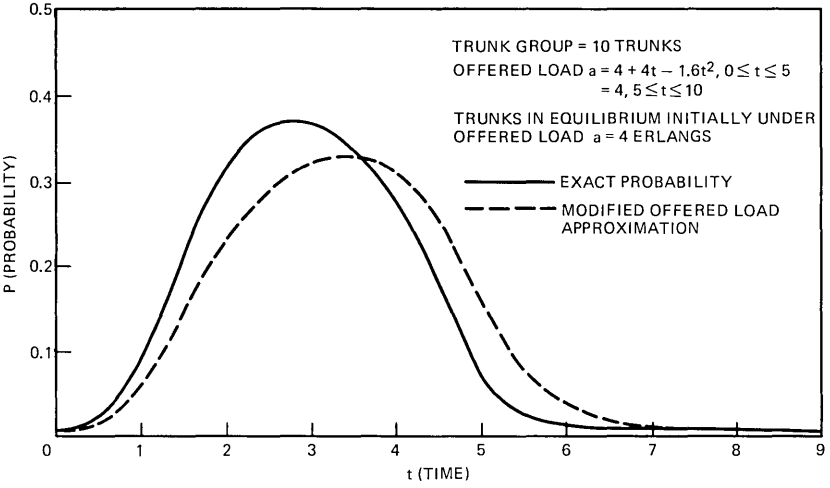

Fig. 3—Pulse response—modified offered load approximation.

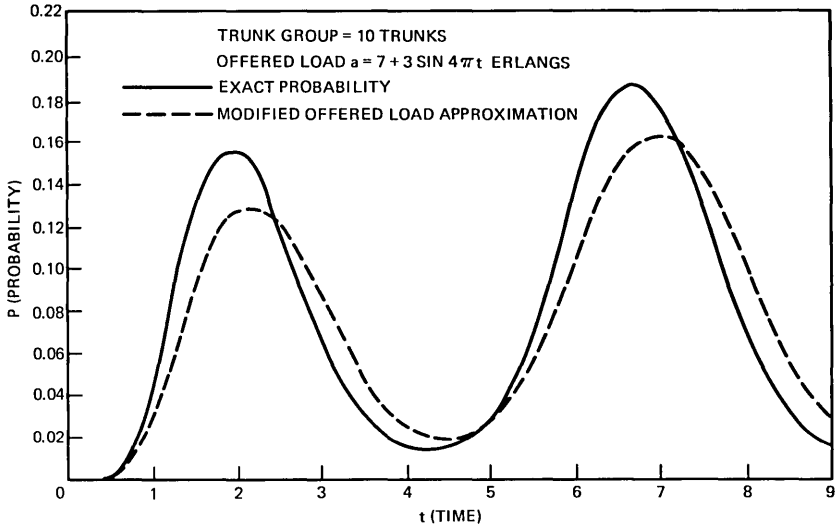

Fig. 4—All trunks empty initially—modified offered load approximation.

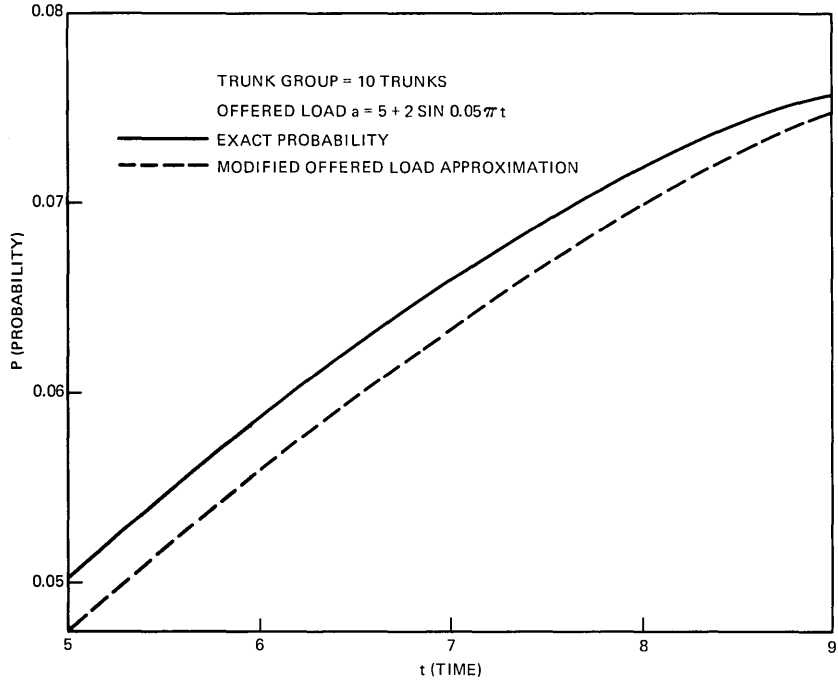

Fig. 5—All trunks empty initially—modified offered load approximation.

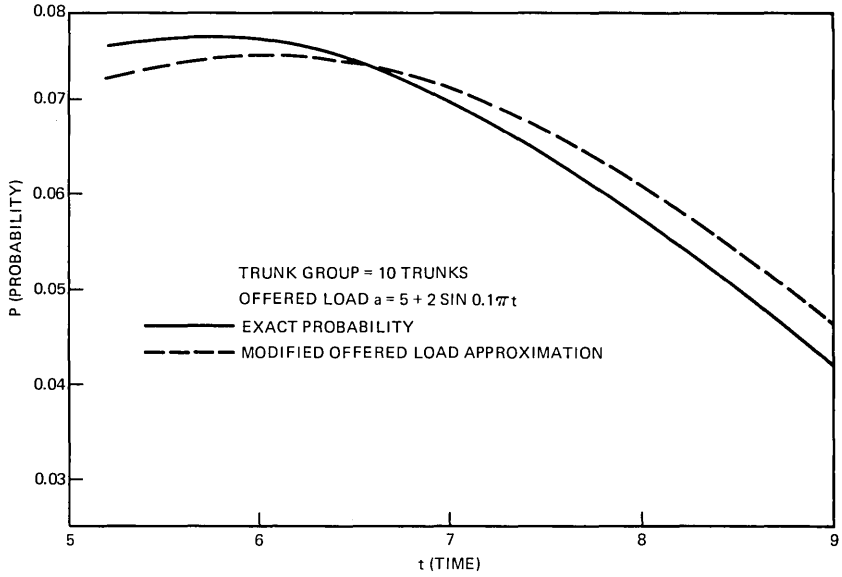

Fig. 6—All trunks empty initially—modified offered load approximation.

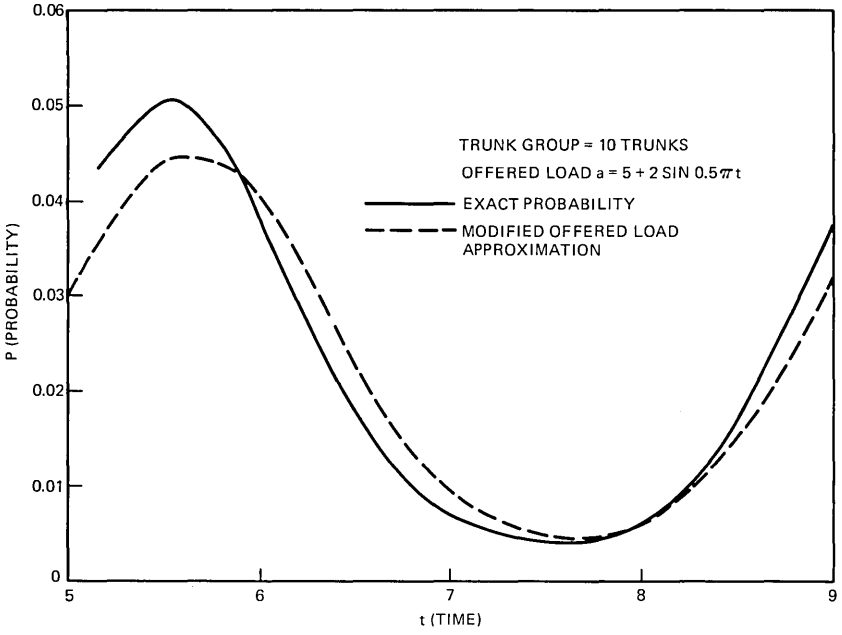

Fig. 7—All trunks empty initially—modified offered load approximation.

ever, the approximation of (187) is simply the case $n = 0$ of Theorem 16. Considerable enhancement of accuracy may, for example, be obtained by using $n = 1$ or even higher values of n.

The modified offered load approximation, (72), is compared to a pulse response in Fig. 3. It is seen that, despite the rapid variation of $a(t)$ (as high as 14 erlangs/call duration) and the large range of probability values, the approximation well imitates the course of the response.

Figures 4, 5, 6, and 7 illustrate the modified offered load approximation applied to sinusoidal inputs. In all cases, the trunk group was initially empty. Figure 4 shows the response, starting from $t = 0$, to $a = 7 + 3 \sin 4\pi t$. This may be considered to have wide excursions compared to the constant term 7, and rapid oscillations, i.e., the period, T, is 5. The exact curve is seen to be well imitated by the approximation.

One may consider the total error to consist of two components, an evanescent part arising from the specific initial state and a component

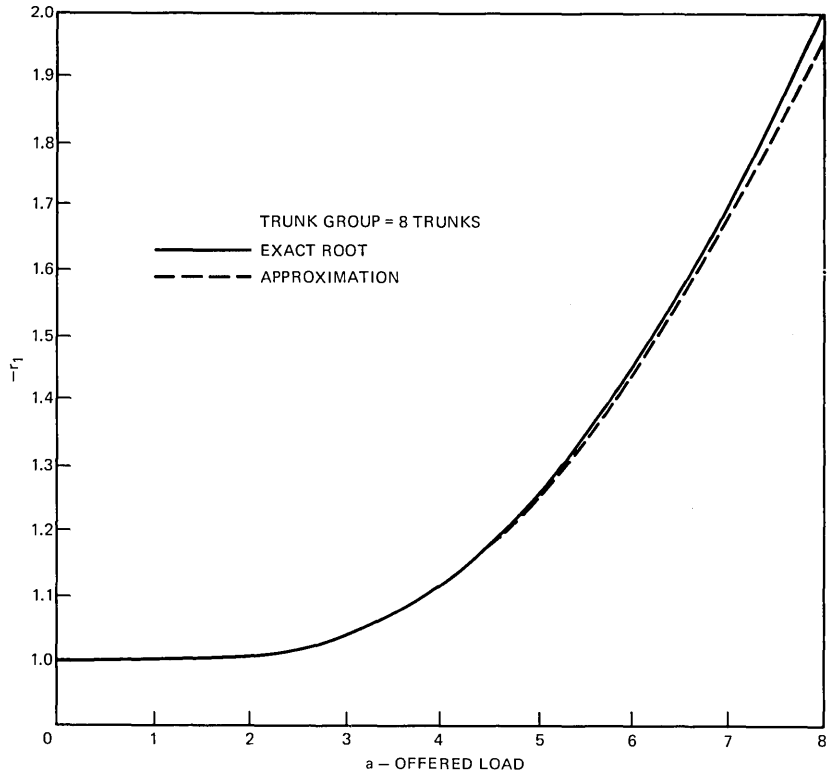

Fig. 8—Upper bound approximation to dominant root.

resulting from the rate of change of the offered load itself, that is, a function of $\dot{a}(t)$ which vanishes when $\dot{a}(t) \equiv 0$. Figures 5, 6, and 7 are intended to illustrate the latter component above; hence, the time scale starts at five. The periods of 40, 20, and 4, respectively, were chosen to reflect the effect of $\dot{a}(t)$ on the approximations. To provide clearer comparison, the probability scales of the graphs have been expanded.

The lower bound of Theorem 13 provides an upper bound on the dominant root. A comparison with the exact values for an eight-trunk group taken from Beneš[7] is given in Fig. 8.

VI. NEEDED INVESTIGATIONS

Much work remains to be done to provide a satisfying and viable tool for fully available trunk group analyses. One may mention error estimation of the approximations suggested in Section IV and Appendix D, the investigation of new approximations, such as studying the consequences of using a refined scaling approximation or an improved modified offered load, and the study of $(I + K_N)^{-1}$ in the Banach space X, for variable $a(t)$. This, in turn, would permit new approximations to be constructed and would provide improved means of error investigation. Relatively little is known about the behavior of the zeros of Poisson-Charlier polynomials, especially in the present context, as functions of a, N. It is hoped this paper will provide an impetus for further investigation.

VII. ACKNOWLEDGMENTS

It is my pleasure to acknowledge the helpful discussions of this material with S. Horing and the critical examination of the paper by L. J. Forys and R. P. Marzec. I wish to thank S. E. Miller and M. D. Vizgirda for the numerical calculations on which the comparative curves are based.

APPENDIX A

Neumann-Series Solution

The Neumann-series solution of the integral equation for blocking,

$$P_N(t, N) = \beta_N(t, \infty) - K_N P_N, \tag{74}$$

is

$$P_N(t, N) = \beta_N(t, \infty) - K_N\beta_N(t, \infty) + K_N^2\beta_N(t, \infty) - \cdots, \tag{75}$$

which, of course, is convergent for all t. The positivity of K_N implies

the system of inequalities

$$\sum_{j=0}^{2k+1} (-1)^j K_N^j \beta_N(t, \infty) < P_N(t, N) < \sum_{j=0}^{2k} (-1)^j K_N^j \beta_N(t, \infty), \; k \geqq 0, \tag{76}$$

which generalize (38).

Truncation error estimates for (75) will now be obtained. The inequalities (76) yield

$$-K_N^{2k+1}\beta_N(t, \infty) < P_N(t, N) - \sum_{j=0}^{2k} (-1)^j K_N^j \beta_N(t, \infty) < 0, \tag{77}$$

$$0 < P_N(t, N) - \sum_{j=0}^{2k+1} (-1)^j K_N^j \beta_N(t, \infty) < K_N^{2k+1}\beta_N(t, \infty). \tag{78}$$

Hence, it is only necessary to bound $K_N^{2k+1}\beta_N(t, \infty)$. Let

$$a(t) \leqq \bar{a}, \tag{79}$$

then, from (27) and (34),

$$K_N(t, \tau; a) \leqq K_N(t - \tau; \bar{a}) \leqq Ae^{-(t-\tau)}, \tag{80}$$

$$A = \bar{a} L_{N-1}^{(1)}(-\bar{a}). \tag{81}$$

The dependence on a is explicitly shown in (80). One similarly obtains

$$\beta_N(t, \infty ; a) \leqq \bar{\beta} = \sum_{j=0}^{N} \beta_{N-j}(0, N) \frac{\bar{a}^j}{j!}, \qquad t \geqq 0. \tag{82}$$

One may now state Theorem 8.

Theorem 8:

$$K_N^r \beta_N(t, \infty) \leqq \bar{\beta} \frac{(At)^r}{r!} e^{-t}.$$

Proof: One has

$$K_N(t; \bar{a}) \leqq Ae^{-t}. \tag{83}$$

Hence,

$$K_{N,r}(t; \bar{a}) \leqq A^r \frac{t^{r-1}}{(r-1)!} e^{-t}, \tag{84}$$

in which $K_{N,r}(t; \bar{a})$ is the r-fold convolution of $K_N(t; \bar{a})$ with itself. Convolving this with $\beta_N(t, \infty)$ finally yields

$$K_{N,r}(t; \bar{a}) * \beta_N(t, \infty) \leqq \bar{\beta} \frac{(At)^r}{r!} e^{-t}. \tag{85}$$

APPENDIX B

Poisson-Charlier Polynomials

Some properties of Poisson-Charlier polynomials[11] are developed in this appendix, especially with a view to convenient representation of functions by series expansion. Let

$$\psi(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \cdots. \tag{86}$$

Then the polynomials $G_j(x, \lambda)$ are defined by

$$\frac{d^j}{d\lambda^j} \psi(x, \lambda) = \psi^{(j)}(x, \lambda) = G_j(x, \lambda)\psi(x, \lambda). \tag{87}$$

The Taylor expansion

$$\psi(x, \lambda + \zeta) = \sum_{j=0}^{\infty} \frac{\zeta^j}{j!} \psi^{(j)}(x, \lambda) \tag{88}$$

yields the generating function

$$e^{-\zeta} \left(1 + \frac{\zeta}{\lambda}\right)^x = \sum_{j=0}^{\infty} G_j(x, \lambda) \frac{\zeta^j}{j!}. \tag{89}$$

Thus, explicit formulas for $G_j(x, \lambda)$ are

$$\begin{aligned} G_j(x, \lambda) &= \frac{j!}{\lambda^j} \sum_{\nu=0}^{j} (-1)^\nu \binom{x}{j-\nu} \frac{\lambda^\nu}{\nu!} \\ &= \sum_{\nu=0}^{j} (-1)^{j-\nu} \binom{j}{\nu} \nu! \lambda^{-\nu} \binom{x}{\nu}. \end{aligned} \tag{90}$$

The first few polynomials are

$$\begin{aligned} G_0(x, \lambda) &= 1, \\ G_1(x, \lambda) &= \frac{1}{\lambda} (x - \lambda), \\ G_2(x, \lambda) &= \frac{1}{\lambda^2} [x^2 - (2\lambda + 1)x + \lambda^2], \\ G_3(x, \lambda) &= \frac{1}{\lambda^3} [x^3 - 3(\lambda + 1)x^2 + (3\lambda^2 + 3\lambda + 2)x - \lambda^3]. \end{aligned} \tag{91}$$

A recurrence relation derived from (89) is

$$G_{j+1}(x, \lambda) = \frac{x - j - \lambda}{\lambda} G_j(x, \lambda) - \frac{j}{\lambda} G_{j-1}(x, \lambda). \tag{92}$$

The following inner product is defined for functions $f(x)$, $g(x)$ of the discrete variable x:

$$(f, g) = \sum_{x=0}^{\infty} \psi(x, \lambda) f(x) g(x). \tag{93}$$

The Poisson-Charlier polynomials are orthogonal with respect to this inner product[11]

$$\begin{aligned} (G_j, G_k) &= 0, \qquad j \neq k \\ (G_j, G_j) &= \frac{j!}{\lambda^j}. \end{aligned} \tag{94}$$

Accordingly, the coefficients c_j in the expansion of a function $f(x)$ in the form

$$f(x) = \psi(x, \lambda) \sum_{j=0}^{\infty} c_j G_j(x, \lambda) \tag{95}$$

are given by

$$c_j = \frac{\lambda^j}{j!} \sum_{x=0}^{\infty} G_j(x, \lambda) f(x). \tag{96}$$

For the purpose of the present investigation, a more convenient mode of determining c_j is achieved by obtaining their generating function; that is,

$$C(w) = \sum_{j=0}^{\infty} c_j w^j. \tag{97}$$

From (96), one has

$$C(w) = \sum_{x=0}^{\infty} f(x) \sum_{j=0}^{\infty} \frac{(\lambda w)^j}{j!} G_j(x, \lambda). \tag{98}$$

Hence, from (89),

$$C(w) = e^{-\lambda w} \sum_{x=0}^{\infty} f(x)(1 + w)^x. \tag{99}$$

Let

$$B(w) = \sum_{x=0}^{\infty} f(x)(1 + w)^x. \tag{100}$$

Then $B(w)$ is the binomial moment generating function of $f(x)$ and

$$C(w) = e^{-\lambda w} B(w). \tag{101}$$

From (101), the first few coefficients c_j are obtained in terms of the

corresponding binomial moments β_j of $f(x)$; thus,

$$\begin{aligned} c_0 &= \beta_0, \\ c_1 &= \beta_1 - \lambda\beta_0, \\ c_2 &= \beta_2 - \lambda\beta_1 + \tfrac{1}{2}\lambda^2\beta_0, \\ c_3 &= \beta_3 - \lambda\beta_2 + \tfrac{1}{2}\lambda^2\beta_1 - \tfrac{1}{6}\lambda^3\beta_0. \end{aligned} \tag{102}$$

A useful choice of the parameter λ is suggested by (102), namely,

$$\lambda = \beta_1/\beta_0, \tag{103}$$

which yields $c_1 = 0$ and

$$\begin{aligned} c_2 &= \beta_2 - \frac{\beta_1^2}{2\beta_0}, \\ c_3 &= \beta_3 - \frac{\beta_1\beta_2}{\beta_0} + \frac{\beta_1^3}{2\beta_0^2}. \end{aligned} \tag{104}$$

For a probability distribution, one has $\beta_0 = 1$, and the choice (103) for λ implies that λ is equated to the mean of the distribution. In this case, one has

$$\begin{aligned} c_2 &= \tfrac{1}{2}(\sigma^2 - \mu), \\ c_3 &= \tfrac{1}{6}(\alpha - 3\sigma^2 + 2\mu), \end{aligned} \tag{105}$$

in which μ is the mean, σ^2 the variance, and α the third moment about the mean of the distribution.

APPENDIX C

Constant Offered Load—Dominant Roots

In this appendix, the integral equation for the constant a case, namely,

$$\begin{aligned} P_N(t, N) &= \beta_N(t, \infty) - K_N * P_N(t, N), \\ K_N(t) &= ae^{-Nt}L_{N-1}^{(1)}[-a(e^t - 1)], \\ \beta_N(t, \infty) &= e^{-Nt}\sum_{j=0}^{N}\beta_{N-j}(0, N)\frac{[a(e^t - 1)]^j}{j!}, \end{aligned} \tag{106}$$

is studied. In fact, the somewhat more general equation,

$$f(t) + K_N * f(t) = g(t), \tag{107}$$

is resolved. This presents a considerable advantage over the solution of (106), since the errors of approximations to $P_N(t, N)$ satisfy (107), and hence may be studied by means of Theorem 10 and its corollary. Solutions for blocking have been obtained in the literature,[7,8] but do not provide a means for error analysis. For the practical utilization of the solutions, bounds for the exponents occurring in the explicit representation of the resolvent kernel will also be obtained.

Integral equation theory[14] asserts that a resolvent kernel, $Q_N(t)$, exists with the property

$$f(t) = g(t) + Q_N * g(t). \tag{108}$$

Laplace transformation will be used to study (107) and (108).

Theorem 9: *The Laplace transforms, $\tilde{K}_N(s)$, $\tilde{Q}_N(s)$, of $K_N(t)$, $Q_N(t)$, respectively, are*

$$\tilde{K}_N(s) = \frac{(-1)^N a^N G_N(-s-1, a)}{(s+1)\cdots(s+N} - 1,$$

$$\tilde{Q}_N(s) = \frac{(s+1)\cdots(s+N) - (-1)^N a^N G_N(-s-1, a)}{(-1)^N a^N G_N(-s-1, a)}.$$

Proof: One has, from (27),

$$L_{N-1}^{(1)}(-X) = \sum_{j=0}^{N-1} \binom{N}{j+1} \frac{X^j}{j!}, \tag{109}$$

and hence,

$$K_N(t) = a e^{-Nt} \sum_{j=0}^{N-1} \binom{N}{j+1} \frac{a^j}{j!} (e^t - 1)^j. \tag{110}$$

Thus,

$$\tilde{K}_N(s) = a \sum_{j=0}^{N-1} \binom{N}{j+1} \frac{a^j}{j!} \int_0^\infty e^{-(N+s-j)t}(1 - e^{-t})^j dt. \tag{111}$$

Letting $X = e^{-t}$, one has

$$\tilde{K}_N(s) = a \sum_{j=0}^{N-1} \binom{N}{j+1} \frac{a^j}{j!} \int_0^1 X^{N+s-j-1}(1 - X)^j dx. \tag{112}$$

The integral in (112) is the beta function, $B(N + s - j, j + 1)$. Hence,

$$\tilde{K}_N(s) = a \sum_{j=0}^{N-1} \binom{N}{j+1} a^j \frac{\Gamma(N + s - j)}{\Gamma(N + s + 1)}. \tag{113}$$

One has the following transformations:

$$\tilde{K}_N(s) = \sum_{j=1}^{N} \binom{N}{j} \frac{a^j}{(N+s)\cdots(N+s-j+1)}, \tag{114}$$

$$\tilde{K}_N(s) = \frac{1}{(s+1)\cdots(s+N)} \sum_{j=1}^{N} \binom{N}{j} a^j (s+1)\cdots(s+N-j), \tag{115}$$

$$\tilde{K}_N(s) = \frac{a^N}{(s+1)\cdots(s+N)} \sum_{j=1}^{N} \binom{N}{j} a^{-j}(s+1)\cdots(s+j), \tag{116}$$

$$\tilde{K}_N(s) = \frac{a^N}{(s+1)\cdots(s+N)} \sum_{j=1}^{N} (-1)^j \binom{N}{j} j!\, a^{-j} \binom{-s-1}{j}. \tag{117}$$

Thus, the required formula for $\tilde{K}_N(s)$ is obtained from (117) by comparison with (90). From (107), one has

$$\tilde{f} + \tilde{K}_N * \tilde{f} = \tilde{g}. \tag{118}$$

Hence,

$$\tilde{f} = \frac{\tilde{g}}{1 + \tilde{K}_N} = \tilde{g} - \frac{\tilde{K}_N}{1 + \tilde{K}_N}\tilde{g}; \tag{119}$$

thus,

$$\tilde{Q}_N(s) = -\frac{\tilde{K}_N(s)}{1 + \tilde{K}_N(s)}. \tag{120}$$

This, together with the formula for $\tilde{K}_N(s)$, yields the required expression for $\tilde{Q}_N(s)$.

Corollary:

$$\tilde{P}_{i,N}(s, N) = (-1)^{N-i}\frac{G_i(-s, a)}{sG_N(-s-1, a)}.$$

Proof: Use of Theorem 2, Corollary 4, Theorem 9, and eq. (178) with $\beta_j(0, N) = \binom{i}{j}$.

An expression equivalent to this corollary was given by Takács.[15]

It will be useful now to introduce the Banach space, X, of functions $f(t)$ that are bounded and measurable over $(0, \infty)$ and normed by

$$\|f\| = \sup_{t \geq 0} |f(t)|. \tag{121}$$

One may now state Theorem 10.

Theorem 10: *The operators* $I + K_N$ *and* $(I + K_N)^{-1}$ *are bounded; further,*

$$\|I + K_N\| = S_N(a),$$

$$\|(I + K_N)^{-1}\| = 1 + \int_0^\infty |Q_N(t)|\,dt,$$

in which the operator norms are those induced by (*121*).

Proof: The quantity $\|I + K_N\|$ is obtained directly from Theorem 3 and the formula

$$\|I + K_N\| = 1 + \int_0^\infty K_N(t)dt. \tag{122}$$

Since the polynomials $G_N(x, a)$ are orthogonal over $(0, \infty)$, it follows that the zeros, $P_1, \cdots, P_N$ are distinct, real, and positive; hence, the zeros, $r_1, \cdots, r_N$, of $G_N(-s-1, a)$ are distinct, real, and less than minus one. The Paley-Wiener theorem[9] applied to $\tilde{Q}_N(s)$ now asserts

that $I + K_N$ has a bounded inverse with norm C given by

$$C = \|(I + K_N)^{-1}\| = 1 + \int_0^\infty |Q_N(t)|\,dt. \tag{123}$$

An immediate corollary is the following.

Corollary: *The integral equation*

$$f + K_N * f = g, \qquad g \in X,$$

possesses a solution $f \in X$ *satisfying*

$$\|f\| \leqq C\|g\|,$$
$$\overline{\lim_{t\to\infty}} |f(t)| \leqq C \overline{\lim_{t\to\infty}} |g(t)|.$$

In particular, if $\lim_{t\to\infty} g(t) = 0$, *then* $\lim_{t\to\infty} f(t) = 0$.

Proof: The result follows from

$$f = g + Q_N * g \tag{124}$$

and the definition of C.

The following theorem provides a representation of $Q_N(t)$ and an estimate of C.

Theorem 11:

$$Q_N(t) = \sum_{j=1}^{N} \frac{\prod_{\nu=1}^{N} (r_j + \nu)}{\prod_{\substack{\nu=1 \\ \nu\neq j}}^{N} (r_j - r_\nu)} e^{r_j t}, \qquad C \leqq 1 + \sum_{j=1}^{N} \frac{\prod_{\nu=1}^{N} |r_j + \nu|}{|r_j| \prod_{\substack{\nu=1 \\ \nu\neq j}}^{N} |r_j - r_\nu|}.$$

Proof: One has

$$(-1)^N a^N G_N(-s - 1, a) = (s - r_1)\cdots(s - r_N) \tag{125}$$

and, hence, from Theorem (9),

$$\tilde{Q}_N(s) = \frac{(s + 1)\cdots(s + N) - (s - r_1)\cdots(s - r_N)}{(s - r_1)\cdots(s - r_N)}. \tag{126}$$

The partial fraction expression for $\tilde{Q}_N(s)$ is

$$\tilde{Q}_N(s) = \sum_{j=1}^{N} \frac{\prod_{\nu=1}^{N} (r_j + \nu)}{\prod_{\substack{\nu=1 \\ \nu\neq j}}^{N} (r_j - r_\nu)} \frac{1}{s - r_j}. \tag{127}$$

Thus,

$$Q_N(t) = \sum_{j=1}^{N} \frac{\prod_{\nu=1}^{N} (r_j + \nu)}{\prod_{\substack{\nu=1 \\ \nu \neq j}}^{N} (r_j - r_\nu)} e^{r_j t}. \tag{128}$$

Also, one has

$$|Q_N(t)| \leqq \sum_{\nu=1}^{N} \frac{\prod_{\nu=1}^{N} |r_j + \nu|}{\prod_{\substack{\nu=1 \\ \nu \neq j}}^{N} |r_j - r_\nu|} e^{r_j t}, \tag{129}$$

and, accordingly,

$$C = 1 + \int_0^\infty |Q_N(t)|\, dt \leqq 1 + \sum_{j=1}^{N} \frac{\prod_{\nu=1}^{N} |r_j + \nu|}{|r_j| \prod_{\substack{\nu=1 \\ \nu \neq j}}^{N} |r_j - r_\nu|}. \tag{130}$$

To make the results of Theorem 11 accessible for estimation, particularly in numerical applications, and for the Laplace transform approximation of Appendix D, bounds for the roots, r_j, will be obtained. The generating function, $g(z)$, for the equilibrium probability distribution, $P_j(\infty, N)$, of (36) may be written as

$$g(z) = \frac{S_N(az)}{S_N(a)}. \tag{131}$$

Thus, the mean, m, and variance, σ^2, are

$$m = a[1 - B(N, a)], \tag{132}$$

$$\sigma^2 = m - (N - m)aB(N, a). \tag{133}$$

We now have Theorem 12.

Theorem 12: (*Beneš*) $r_1 > -(m/\sigma^2)$.

To obtain further bounds, the following lemmas are needed.

Lemma 1: $\rho_N \geqq \cdots \geqq \rho_1 > 0$,

$$\rho_1 + \cdots + \rho_N = s_1, \qquad \rho_1^2 + \cdots + \rho_N^2 = s_2, \qquad \rho_1 \cdots \rho_N = D,$$

then

$$\rho \leqq \rho_1 \leqq \cdots \leqq \rho_N \leqq \frac{s_1 + \sqrt{(N-1)(Ns_2 - s_1^2)}}{N},$$

in which ρ *is the small positive root of*

$$\rho(s_1 - \rho)^{N-1} = D(N-1)^{N-1}.$$

These bounds are sharp.

Proof: The equations

$$\rho_1 + \cdots + \rho_N = s_1, \tag{134}$$

$$\rho_1^2 + \cdots + \rho_N^2 = s_2 \tag{135}$$

may be written

$$\rho_1 + \cdots + \rho_{N-1} = s_1 - \rho_N, \tag{136}$$

$$\rho_1^2 + \cdots + \rho_{N-1}^2 = s_2 - \rho_N^2. \tag{137}$$

Let

$$\rho_N^* = \max \rho_N \tag{138}$$

over all allowable sequences. Then the sum (137) is minimum when ρ_N is replaced by ρ_N^*. This occurs, however, only when

$$\rho_1 = \cdots = \rho_{N-1} = \frac{s_1 - \rho_N^*}{N-1}. \tag{139}$$

Thus, from (137),

$$(N-1)\left(\frac{s_1 - \rho_N^*}{N-1}\right)^2 = s_2 - \rho_N^{*2}. \tag{140}$$

The solution of (140) for ρ_N^* is

$$\rho_N^* = \frac{s_1 + \sqrt{(N-1)(Ns_2 - s_1^2)}}{N}. \tag{141}$$

This proves the upper bound of the lemma. The inequality is attained for the vector $(\rho_1, \cdots, \rho_N)$ defined by (139) and (141).

Similarly, one may write

$$\rho_2 + \cdots + \rho_N = s_1 - \rho_1, \tag{142}$$

$$\rho_2 \cdots \rho_N = D/\rho_1. \tag{143}$$

Let

$$\rho_1^* = \min \rho_1 \tag{144}$$

over all allowable sequences. Then the product (143) is maximum when ρ_1 is replaced by ρ_1^*. This occurs only when

$$\rho_2 = \cdots = \rho_N = \frac{s_1 - \rho_1^*}{N-1}. \tag{145}$$

Thus,

$$\left[\frac{s_1 - \rho_1^*}{N-1}\right]^{N-1} = D/\rho_1^* \tag{146}$$

and the equation of the lemma defining ρ has been established. The

inequality is attained for the vector $(\rho_1, \cdots, \rho_N)$ defined by (145) and (146).

Lemma 2: $\rho_N \geqq \cdots \geqq \rho_1 > 0$,

$$s_{-1} = \frac{1}{\rho_1} + \cdots + \frac{1}{\rho_N}, \qquad D = \rho_1 \cdots \rho_N.$$

Then

$$\rho_N \geqq \cdots \geqq \rho_1 \geqq \rho,$$

in which ρ *is the small positive root of*

$$D\left(s_{-1} - \frac{1}{\rho}\right)^{N-1} = \rho(N-1)^{N-1}.$$

The bound is sharp.

Proof: One may write

$$\frac{1}{\rho_2} + \cdots + \frac{1}{\rho_N} = s_{-1} - \frac{1}{\rho_1}, \tag{147}$$

$$\rho_2 \cdots \rho_N = D/\rho_1. \tag{148}$$

Let

$$\rho_1^* = \min \rho_1 \tag{149}$$

over all allowable sequences. Then the product (148) is maximum when ρ_1 is replaced by ρ_1^*. This occurs when

$$\rho_2 = \cdots = \rho_N = \frac{N-1}{s_{-1} - \frac{1}{\rho_1^*}}. \tag{150}$$

Thus,

$$\left[\frac{N-1}{s_{-1} - \frac{1}{\rho_1^*}}\right]^{N-1} = D/\rho_1^* \tag{151}$$

and the equation of the lemma is established. The inequality is attained for the vector $(\rho_1, \cdots, \rho_N)$ defined by (150) and (151).

The application of the lemmas to the polynomials $G_N(x, a)$ is accomplished by identifying $\rho_1, \cdots, \rho_N$ with its zeros. For this purpose, the form of $G_N(x, a)$ given in (90) will be recast, by the help of the Stirling numbers of first kind,[11] into standard form; that is,

$$G_N(x, a) = \sum_{m=0}^{N} a_{N-m} x^m. \tag{152}$$

The Stirling numbers, S_j^m, are defined by

$$\prod_{\nu=0}^{j-1} (x - \nu) = \sum_{m=1}^{j} S_j^m x^m. \tag{153}$$

Thus, one has

$$G_N(x, a) = (-1)^N + \sum_{m=1}^{N} x^m \sum_{\nu=0}^{N-m} (-1)^\nu a^{\nu-N} \binom{N}{\nu} S_{N-\nu}^m. \tag{154}$$

The sums s_1, s_2 are accordingly given by

$$s_1 = \binom{N}{2} + aN, \tag{155}$$

$$s_2 = s_1^2 - 6\binom{N}{4} - (6a + 4)\binom{N}{3} - 2a^2\binom{N}{2}. \tag{156}$$

The reciprocal polynomial, that is, the polynomial whose zeros are $\rho_1^{-1}, \cdots, \rho_N^{-1}$ is given by

$$X^N G_N(x^{-1}, a) G_N(x, a) = \sum_{m=0}^{N} a_m x^m. \tag{157}$$

Hence, the analogous quantities s_{-1} and $s_{-2} = \rho_1^{-2} + \cdots + \rho_N^{-2}$ are given by

$$s_{-1} = \sum_{\nu=1}^{N} \frac{1}{\nu} N^{(\nu)} a^{-\nu}, \tag{158}$$

$$s_{-2} = s_{-1}^2 - 2 \sum_{\nu=2}^{N} \frac{1}{\nu} N^{(\nu)} a^{-\nu} \sum_{j=1}^{\nu-1} \frac{1}{j}, \tag{159}$$

in which

$$N^{(0)} = 1, \qquad N^{(\nu)} = N(N-1)\cdots(N-\nu+1), \qquad \nu > 0. \tag{160}$$

The upper bound of Lemma 1 now establishes Theorem 13.

Theorem 13:

$$\frac{N}{s_{-1} + \sqrt{(N-1)(Ns_{-2} - s_{-1}^2)}} \leqq \rho_1 < \cdots < \rho_N \leqq \frac{s_1 + \sqrt{(N-1)(Ns_2 - s_1^2)}}{N}.$$

Proof: The upper bound is immediate. The lower bound is obtained by applying the upper bound of Lemma 1 to the reciprocal equation.

A numerical illustration of Theorem 13 is provided by the zeros of $G_{10}(x, 7)$ used in obtaining the recovery function plotted in Fig. 2. They are 0.332811, 2.05847, 4.06653, 6.31227, 8.81308, 11.5197,

14.6407, 18.0255, 22.0872, and 27.1438. The lower and upper bounds given by the theorem are 0.32964 and 36.82292, respectively. It appears that the lower bound may well be usable as an approximation to ρ_1. The accuracy of this when used to approximate $-r_1 = 1 + \rho_1$ is illustrated in Fig. 8. The exact values of $-r_1$ are taken from Beneš.[7] This provides an upper bound for r_1 which, together with the available lower bounds, is useful for error investigations.

APPENDIX D

Approximations—Constant Offered Load

It was suggested by S. Horing that an appropriate scaling between $P_N(t, N)$ and $P_k(t, k)$ may exist; that is, a function $F[P_N(t, N), P_k(t, k)]$ may exist, which would be approximately independent of t and which may, therefore, permit the approximate determination of $P_N(t, N)$ in terms of $P_1(t, 1)$, for example, thus permitting large trunk groups to be studied in terms of the behavior of small ones. Since it is feasible to use (108) for small trunk groups, this would constitute an approximation of the solution of (106) for large trunk groups.

Consider the following

$$\frac{B(kl, a)}{B(l, a/k)^k} = \frac{l!k^{kl}}{(kl)!} \frac{S_l(a/k)^k}{S_{kl}(a)}. \tag{161}$$

The ratio $S_l(a/k)^k/S_{kl}(a)$ is approximately independent of a since $S_l(a/k)^k \backsimeq S_{kl}(a) \backsimeq e^a$ for k large. Thus,

$$\frac{B(kl, a)}{B(l, a/k)^k} \tag{162}$$

is approximately independent of a. It would seem, therefore, that the ratio (162) is approximately a time invariant of (106), especially for large t; that is, the function

$$\frac{P_{kl}(t, kl; a)}{P_l(t, l; a/k)^k} \tag{163}$$

is approximately equal to the ratio (162) when t is large; thus,

$$P_{kl}(t, l; a) \backsimeq B(kl, a) \left[\frac{P_l(t, l; a/k)}{B(l, a/k)} \right]^k. \tag{164}$$

We have, from (107) and Theorem 11,

$$P_1(t, 1; a) = \frac{a}{a+1} \{1 - e^{-(a+1)t}\}, \tag{165}$$

when the trunk is initially empty. Hence, from (164) with $l = 1$,

$k = N$,

$$P_N(t, N) \leqq B(N, a)\{1 - e^{-[(a/N)+1]t}\}^N \tag{166}$$

for the case when all N trunks are initially empty.

It is desired now to extend the approximate solution (166) for application to any initial condition. Define $q_N(t)$ by

$$q_N(t) = B(N, a)\{1 - e^{-[(a/N)+1]t}\}^N. \tag{167}$$

Then a convolution operator with kernel $L_N(t)$ will be constructed so that one has

$$q_N + L_N * Q_N = \frac{\Lambda^N}{N!}, \tag{168}$$

exactly. The operator L_N will then be taken to approximate the operator K_N of (106). Accordingly, an approximation $\bar{P}_N$ to P_N corresponding to initial states other than all trunks empty is defined by the equation

$$\bar{P}_N + L_N * \bar{P}_N = \beta_N. \tag{169}$$

Theorem 14: *The Laplace transform, $\tilde{L}_N(s)$, of the approximating $L_N(t)$ is*

$$1 + \tilde{L}_N = \left(\frac{a}{N} + 1\right) S_N(a) \frac{\Gamma(s)\Gamma\left(N + \frac{s}{(a/N)+1} + 1\right)}{\Gamma(s + N + 1)\Gamma\left(\frac{s}{(a/N)+1}\right)}$$

$$= \left(\frac{a}{N} + 1\right)^{-N} S_N(a) \prod_{\nu=1}^{N} \frac{s + \nu[(a/N) + 1]}{s + \nu}.$$

Proof: One has

$$\tilde{q}_N = B(N, a) \int_0^\infty e^{-st}\{1 - e^{-[(a/N)+1]t}\}^N dt. \tag{170}$$

Let $u = [(a/N) + 1]t$, then

$$\tilde{q}_N = \frac{B(N, a)}{(a/N) + 1} \int_0^\infty e^{-[s/(a/N+1)]u}(1 - e^{-u})^N du. \tag{171}$$

The substitution $x = e^{-u}$ yields

$$\tilde{q}_N = \frac{B(N, a)}{(a/N) + 1} \int_0^1 x^{[s/(a/N+1)]-1}(1 - x)^N dx; \tag{172}$$

thus,

$$\tilde{q}_N = \frac{B(N, a)}{(a/N) + 1} \frac{N!\,\Gamma\left[\frac{s}{(a/N)+1}\right]}{\Gamma\left[N + \frac{s}{(a/N)+1} + 1\right]}. \tag{173}$$

One also has

$$\int_0^\infty e^{-st}(1 - e^{-t})^N dt = \frac{N!\,\Gamma(s)}{\Gamma(N + s + 1)}. \tag{174}$$

The Laplace transform of each term of (168) now yields

$$\tilde{q}_N(1 + \tilde{L}) = \frac{a^N \Gamma(s)}{\Gamma(N + s + 1)}. \tag{175}$$

Use of (173) yields the result of the theorem.

It may be noted that L_N has an impulsive component at $t = 0$ whose value is

$$\left(\frac{a}{N} + 1\right)^{-N} S_N(a) - 1. \tag{176}$$

For large N, this is nearly zero. The effect of this is to create an error at $t = 0$; that is, $\tilde{P}_N(0, N) \neq P_N(0, N)$ if $\beta_N(0, N) \neq 0$. The larger N is (for fixed a), the smaller the discrepancy.

Equation (169) is studied in Theorem 15.

Theorem 15: *The solution of (169) is*

$$\tilde{\tilde{P}}_N = \frac{N!}{(a/N) + 1} B(N, a) \frac{\Gamma\left[\dfrac{s}{(a/N) + 1}\right]}{\Gamma(s)\Gamma\left[N + \dfrac{s}{(a/N) + 1} + 1\right]} \cdot \sum_{j=0}^{N} \beta_j(0, N) a^{-j} \Gamma(s + j),$$

$$\tilde{\tilde{P}}_N = \sum_{j=0}^{N} \beta_j(0, N) a^{-j} \frac{\Gamma(D + j)}{\Gamma(D)} q_N(t), \qquad D \equiv \frac{d}{dt}.$$

Proof: From Corollary 1, Theorem 2, one has

$$\beta_N = \sum_{j=0}^{N} \beta_{N-j}(0, N) \frac{a^j}{j!} e^{-(N-j)t}(1 - e^{-t})^j; \tag{177}$$

hence,

$$\tilde{\beta}_N = \sum_{j=0}^{N} \beta_j(0, N) a^{N-j} \frac{\Gamma(s + j)}{\Gamma(s + N + 1)}. \tag{178}$$

Transforming the terms of (169) yields

$$\tilde{\tilde{P}}_N(1 + \tilde{L}_N) = \tilde{\beta}_N, \tag{179}$$

and hence the result of the theorem is obtained from Theorem 14 and eq. (178). The inversion of the transform, $\tilde{\tilde{P}}_N$, by use of differentiation follows on use of (173) in $\tilde{\tilde{P}}_N$.

Another method of approximation useful for constant offered load is based on an approximate inversion of the Laplace transform. Let

$$\tilde{f}(s) = \int_0^\infty e^{-su} f(u) du, \qquad s > 0; \tag{180}$$

then

$$\frac{(-1)^n}{n!} s^{n+1} \tilde{f}^{(n)}(s) = \frac{s^{n+1}}{n!} \int_0^\infty e^{-su} u^n f(u) du. \tag{181}$$

The function $(s^{n+1}/n!)e^{-su}u^n$ is a probability density function for any $s > 0$, $n \geqq 0$ whose mean is $(n + 1)/s$ and variance $(n + 1)/s^2$. Letting $s = (n + 1)/t$, the mean and variance are t and $t^2/(n + 1)$, respectively; hence, Korovkin's theorem on sequences of positive functionals[16] establishes the Widder inversion formula[9]:

$$\lim_{n \to \infty} \frac{(-1)^n}{n!} s^{n+1} \tilde{f}^{(n)}(s) \Big|_{s=(n+1)/t} = f(t). \tag{182}$$

The above discussion forms the basis for Theorem 16.

Theorem 16: *For $\epsilon > 0$, let the transform of $e^{\epsilon t} f(t)$, namely, $\tilde{f}(s - \epsilon)$, exist for $s > 0$, and let $e^{\epsilon t} f(t)$ be convex in $t > 0$, then*

$$f(t) \leqq e^{-\epsilon t} \frac{(-1)^n}{n!} s^{n+1} \tilde{f}^{(n)}(s - \epsilon) \Big|_{s=(n+1)/t}, \qquad n \geqq 0, \qquad t > 0.$$

Proof: Jensen's inequality applied to (181) in the form

$$\frac{(-1)^n}{n!} s^{n+1} \tilde{f}^{(n)}(s - \epsilon) = \frac{s^{n+1}}{n!} \int_0^\infty e^{-su} u^n e^{\epsilon u} f(u) du \tag{183}$$

establishes the theorem. By virtue of (182), when similarly modified for the function $e^{\epsilon t} f(t)$, the dexter of the inequality always provides an approximation to $f(t)$ even when $e^{\epsilon t} f(t)$ is not convex.

Corollary: *$\tilde{f}(s - \epsilon)$ exists for $s > 0$, and $e^{\epsilon t} f(t)$ is convex in $t > 0$ implies*

$$f(t) \leqq \frac{1}{t} e^{-\epsilon t} \tilde{f}\left(\frac{1}{t} - \epsilon\right), \qquad t > 0.$$

Proof: The case $n = 0$ of Theorem 16.

If $\tilde{f}(s)$ should have a dominant pole, it is usually advantageous to choose ϵ equal to the negative of that pole.

The above corollary will now be applied to obtaining an inequality for the recovery function. The corollary to Theorem 9 shows that

$$\tilde{P}_{N,N}(s, N) = \frac{G_N(-s, a)}{sG_N(-s - 1, a)}. \tag{184}$$

The inversion of $\tilde{P}_{N,N}(s, N)$ is readily accomplished; the result is[7]

$$P_{N,N}(t, N) = B(N, a) - \sum_{j=1}^{N} \frac{e^{r_j t}}{r_j} \prod_{i \neq j} \left(1 - \frac{1}{r_j - r_i}\right). \quad (185)$$

It follows from (185) that

$$e^{-r_1 t}[P_{N,N}(t, N) - B(N, a)] \quad (186)$$

is convex for $t > 0$ and that the corresponding Laplace transform exists for $s > 0$, hence, by the above corollary, one obtains

$$P_{N,N}(t, N) \leqq B(N, a) + e^{r_1 t}\left[\frac{G_N[-(1/t) - r_1, a]}{(1 + r_1 t)G_N[-(1/t) - r_1 - 1, a]} - \frac{B(N, a)}{1 + r_1 t}\right]. \quad (187)$$

REFERENCES

1. B. Wallström, "Congestion Studies in Telephone Systems with Overflow Facilities," Ericsson Technics, No. 3.
2. A. Y. Khintchine, *Mathematical Methods in the Theory of Queuing*, New York: Hafner, 1969.
3. V. B. Iversen, "Analysis of Traffic Processes Based on Data Obtained by the Scanning Method," Lyngby, Denmark: Technical University of Denmark.
4. R. I. Wilkinson, "Theories for Toll Traffic Engineering in the U.S.A.," B.S.T.J., *35*, No. 2 (March 1956), pp. 421–514.
5. D. L. Jagerman, "An Approximation Theorem of Central Limit Type," unpublished work.
6. D. L. Jagerman, "Some Properties of the Erlang Loss Function," B.S.T.J., *53*, No. 3 (March 1974), pp. 525–551.
7. V. E. Beneš, *Mathematical Theory of Connecting Networks and Telephone Traffic*, New York: Academic Press, 1965, Chapter 6.
8. J. Riordan, *Stochastic Service Systems*, New York: Wiley, 1962.
9. R. E. A. C. Paley and N. Wiener, "Fourier Transforms in the Complex Domain," Amer. Math. Soc., 1934.
10. G. Sansone, *Orthogonal Functions*, New York: Interscience, 1959, Chapter IV.
11. C. Jordon, *Calculus of Finite Differences*, New York: Chelsea, 1947, Chapter VIII.
12. S. Karlin and J. L. McGregor, "The Differential Equations of Birth-and-Death Processes and the Stieltjes Moment Problem," Trans. Amer. Math. Soc., *85*, 1957, pp. 489–546.
13. E. Schmidt, "Über die Charlier-Jordansche Entwicklung einer Willkürlichen Funktion nach der Poissonschen Funktion und ihren Ableitungen," Zeitschift für Ange. Mathematik und Mechanik, V. 13 (1933).
14. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, New York: Oxford, 1948.
15. L. Takács, *Introduction to the Theory of Queues*, New York: Oxford, 1962, Chapter 4.
16. P. P. Korovkin, "Linear Operators and Approximation Theory," New York: Gordon and Breach, 1960, p. 14.

Contributors to This Issue

Dan L. Bisbee, B.S., 1965, Monmouth College; Bell Laboratories, 1955—. Mr. Bisbee has been involved in the measurement of optical transmission losses in bulk glass and optical fibers. He is presently engaged in developing techniques for splicing cables made up of optical fiber waveguides.

Edwin L. Chinnock, Stevens Institute of Technology; Bell Laboratories, 1939—. Mr Chinnock has worked on microwave components, microwave radio relay, and helix waveguide fabrication. He is presently working on optical waveguide components.

Richard W. Daniels, B.S., 1964, Brown University; M.S., 1965, Massachusetts Institute of Technology; Ph.D., 1969, Northeastern University; Bell Laboratories, 1964—. Mr. Daniels has worked on various facets of filter design, including exploratory work on active filters and design of crystal filters. Currently, he is assigned to Tennessee State University on the Aid to Black Colleges Program. Author, *Approximation Methods for Electronic Filter Design.* Member, IEEE, Tau Beta Pi, Sigma Xi.

Robert J. Dow, B.S. (Physics), 1959, University of Massachusetts; M.S. (Communication Theory), 1961, Northeastern University; Bell Laboratories, 1959—. Mr. Dow has been involved in the design and development of thin film components. His current work includes the evaluation of laser machining for thin and thick film applications.

James L. Flanagan, B.S. (Electrical Engineering), 1948, Mississippi State University; S.M., 1950, and Sc.D., 1955, Massachusetts Institute of Technology. Faculty of Electrical Engineering, Mississippi State University, 1950–1952; Air Force Cambridge Research Center, 1954–1957; Bell Laboratories, 1957—. Mr. Flanagan has worked in digital communications, computer techniques, and acoustics research. He is presently Head, Acoustics Research Department. Fellow, IEEE, Acoustical Society of America; Board of Governors, American Institute of Physics; member, Tau Beta Pi, Sigma Xi.

Robert A. Friedenson, B.E.E., 1965, M.S., 1966, and Ph.D., 1969, Cornell University; Bell Laboratories, 1969—. Mr. Friedenson was engaged in the design and development of RC active filters for PCM channel banks. In 1971 he was appointed supervisor of a group responsible for computer aids to circuit design and testing. His current interests include computer aids to digital system simulation, PCM repeater design, and analog and digital circuit simulation and testing. Member, IEEE, Eta Kappa Nu, Tau Beta Pi.

Richard D. Gitlin, B.E.E., 1964, City College of New York; M.S., 1965, and D.Eng.Sc., 1969; Columbia University; Bell Laboratories, 1969—. Mr. Gitlin is presently concerned with problems in data transmission. Member, IEEE, Sigma Xi, Eta Kappa Nu, Tau Beta Pi.

D. Gloge, Dipl. Ing., 1961, Dr. Ing., 1964, Technical University of Braunschweig, Germany; Bell Laboratories, 1965—. Mr Gloge's work has included the design and field testing of various optical transmission media and the application of ultra-fast measuring techniques to optical component studies. He is presently engaged in transmission research related to optical fiber communication systems.

Jeremiah F. Hayes, B.E.E., 1956, Manhattan College; M.S., 1961, New York University; Ph.D., 1966, University of California, Berkeley; Faculty Member, Purdue University, 1966–1969; Bell Laboratories, 1969—. Mr. Hayes is currently working in the area of computer communications networks. Member, IEEE, Sigma Xi, Eta Kappa Nu.

Otto Herrmann, Dipl.-Ing (Electrical Engineering), 1956, and Dr.-Ing. (Electrical Engineering), 1965, University of Aachen, Germany; venia legendi, 1971, University of Erlangen, Nuremberg, Germany. Mr. Herrmann has worked on problems concerning approximation theory as applied to analog and digital filter design. From 1959 to 1971 he was a Teaching and Research Assistant at the University of Aachen, University of Karlsruhe, and University of Erlangen. He was at Bell Laboratories during the summer of 1972 on leave from the Technical Faculty at the University of Erlangen. Presently, he teaches courses in communications, analog computation, and digital signal processing at the University of Erlangen. Member, Nachrichtentechnische Gesellschaft.

Kenzo Ishizaka, B.S., 1953, M.S., 1955, and Ph.D. (Engineering), 1972, Tohoku University, Sendai, Japan; Toyo Communication Equipment Co., 1955–1962; University of Electro-Communications, Tokyo, Japan, 1962—; on leave to Bell Laboratories, 1970–1971 and 1972—. Mr. Ishizaka has worked on computer simulation of speech production in the Acoustics Research Department. Member, Acoustical Society of America, Acoustical Society of Japan, Institute of Electronics and Communication Engineers of Japan.

D. L. Jagerman, B.E.E., 1949, Cooper Union; M.S., 1954, and Ph.D. (Mathematics), 1962, New York University; Bell Laboratories, 1964—. Mr. Jagerman has been engaged in mathematical research on numerical quadrature theory, interpolation, mathematical properties of pseudorandom number generators, dynamic programming, approximation theory, and widths and entropy with application to the storage and transmission of information. His recent work concerns the theory of queuing systems and its applications to telephone traffic problems. Member, Pi Mu Epsilon.

Patricia H. McDonald, B.A. (Mathematics), 1963, Trinity College, Washington, D. C.; M.A.T., 1964, University of Massachusetts; Bell Laboratories, 1964—. Mrs. McDonald has developed computer programs for general purpose optimization, tolerance analysis applications, switched network analysis, and simulation of digital systems. She has recently been involved with the design of an interactive filter synthesis program.

Lawrence R. Rabiner, S.B., S.M., 1964, Ph.D., 1967, Massachusetts Institute of Technology; Bell Laboratories, 1962—. Mr. Rabiner has 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. Coauthor, *Theory and Application of Digital Signal Processing.* Member, Eta Kappa Nu, Sigma Xi, Tau Beta Pi; Fellow, Acoustical Society of America; President, IEEE G-ASSP Ad Com; member, G-ASSP Technical Committee on Digital Signal Processing, G-ASSP Technical Committee on Speech Communication, IEEE Proceedings Editorial Board, Technical Committee on Speech Communication of the Acoustical Society; former Associate Editor of the G-ASSP Transactions.

Ronald W. Schafer, B.S. (E.E.), 1961, and M.S. (E.E.), 1962, University of Nebraska; Ph.D., 1968, Massachusetts Institute of

Technology; Bell Laboratories, 1968—. At Bell Laboratories, Mr. Schafer has been engaged in research on speech analysis and synthesis, digital signal processing techniques, and digital waveform coding. Presently, he is on leave from Bell Laboratories as John O. McCarty/Audichron Professor of Electrical Engineering at Georgia Institute of Technology. Coauthor, *Digital Signal Processing*. Senior member, IEEE; member, Acoustical Society of America, ADCOM of the IEEE Acoustics, Speech, and Signal Processing Group; Associate Editor, ASSP Transactions.

Kathleen L. Shipley, B.A. (Mathematics) 1970, Douglass College; Bell Laboratories, 1970—. Mrs. Shipley is a member of the Acoustics Research Department and has worked on scientific programming for laboratory computer systems dedicated to research in communications acoustics. Member, Pi Mu Epsilon.

Peter W. Smith, B.Sc., Mathematics and Physics, 1958, and M.Sc. and Ph.D., Physics, 1961 and 1964, McGill University; Visiting Mackay Lecturer in Electrical Engineering, 1970, University of California, Berkeley; Bell Laboratories, 1964—. Mr. Smith has investigated a number of systems for obtaining single-frequency laser operation and is currently investigating the use of waveguide techniques for producing miniature gas lasers. Member, American Physical Society, Optical Society of America, IEEE.

James Tow, B.S. (E.E.), 1960, M.S. (E.E.), 1962, and Ph.D. (E.E.), 1966, University of California, Berkeley; Bell Laboratories, 1966—. Mr. Tow is a member of the Network Analysis and Synthesis Department. His present interests include computer-aided network analysis and design, active filter realization, and microprocessor applications. Member, IEEE, Eta Kappa Nu, Phi Beta Kappa.

B. A. Whitaker, B.S. (Applied Mathematics), 1967, North Carolina State University; M.S. (Operations Research), 1968, Johns Hopkins University; Ph.D. (Operations Research), 1974, New York University; Bell Laboratories, 1967—. Mr. Whitaker has worked on various aspects of traffic network planning. Member, ORSA, Pi Mu Epsilon, Phi Kappa Phi, Phi Eta Sigma.

THE BELL SYSTEM TECHNICAL JOURNAL is abstracted or indexed by *Abstract Journal in Earthquake Engineering. Applied Mechanics Review, Applied Science & Technology Index, Chemical Abstracts, Computer Abstracts, Computer & Control Abstracts, Current Papers in Electrical & Electronic Engineering, Current Papers on Computers & Control, Electrical & Electronic Abstracts, Electronics & Communications Abstracts Journal, The Engineering Index, International Aerospace Abstracts, Journal of Current Laser Abstracts, Language and Language Behavior Abstracts, Mathematical Reviews, Metals Abstracts, Science Abstracts,* and *Solid State Abstracts Journal.* Reproductions of the Journal by years are available in microform from University Microfilms, 300 N. Zeeb Road, Ann Arbor, Michigan 48106.


Bell System