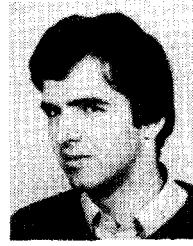




Hugo J. De Man (M'81-SM'81) was born in Boom, Belgium, on September 19, 1940. He received the electrical engineering degree and the Ph.D. degree in applied science from the Katholieke Universiteit Leuven, Heverlee, Belgium, in 1964 and 1968, respectively.

In 1968 he became a Member of the Staff of the Laboratory for Physics and Electronics of Semiconductors at the University of Leuven, working on integrated circuit technology. From 1969 to 1971 he was at the Electronic Research Laboratory, University of California, Berkeley, as an ESRO-NASA Postdoctoral Research Fellow, working on computer-aided device and circuit design. In 1971 he returned to the University of Leuven as a Research Associate of the NFWO (Belgian National Science Foundation). In 1974 he became a Professor at the University of Leuven. During the winter quarter of 1974-1975 he was a Visiting Associate Professor at the University of California, Berkeley. His actual field of research is the design of integrated circuits and computer-aided design.



Piet Stevens was born in Herk-de-Stad in 1959. He received the M.S.E.E. degree at the University of Leuven, Belgium, in 1982. His master thesis was on logic simulation.

Currently he is involved in logic and fault simulation at Silvar-Lisco.

*



Guido G. Schrooten was born in Bree, Belgium, on June 6, 1958. He received the M.S. degree from the Department of Electrical Engineering, University of Leuven, in 1982.

In August 1982 he joined the CACD group of Philips Research Laboratories at Eindhoven. He is working on switched level logic simulation and an analysis tool for MOS VLSI.

Signal Delay in RC Tree Networks

JORGE RUBINSTEIN, MEMBER, IEEE, PAUL PENFIELD, JR., FELLOW, IEEE,
AND MARK A. HOROWITZ, STUDENT MEMBER, IEEE

Abstract—In MOS integrated circuits, signals may propagate between stages with fanout. The exact calculation of signal delay through such networks is difficult. However, upper and lower bounds for delay that are computationally simple are presented in this paper. The results can be used 1) to bound the delay, given the signal threshold, or 2) to bound the signal voltage, given a delay time, or 3) certify that a circuit is "fast enough," given both the maximum delay and the voltage threshold.

I. INTRODUCTION

IN MOS INTEGRATED CIRCUITS, a given inverter or logic node may drive several gates, some of them through long wires whose distributed resistance and capacitance may not be negligible. There does not seem to be reported in the literature any simple method for estimating signal propagation delay in such circuits, nor is there any general theory of the

Manuscript received April 20, 1982; revised June 14, 1983. This work was supported in part by Digital Equipment Corporation, in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under Contract N00014-C-80-0622, in part by the Air Force under Contract AFOSR 4-9620-80-0073, and in part by the Army Research Office under Contract DAAG-29-80-K-0046.

J. Rubinstein is with the Digital Equipment Corporation, 75 Reed Road, Hudson, MA 01749.

P. Penfield, Jr. is with the Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.

M. A. Horowitz is with the Department of Electrical Engineering, Stanford University, Palo Alto, CA 94305.

properties of RC trees, as distinct from RC lines. This paper presents a computationally simple technique for finding upper and lower bounds for the delay. The technique is of importance for VLSI designs in which the delay introduced by the interconnections may be comparable to or longer than active-device delay. This can be the case for wiring lengths as short as 1 mm, with 4- μ m minimum feature size. The importance of this technique grows as the wiring lengths increase or the feature size decreases.

Consider the circuit of Fig. 1. The slowest transition (and therefore presumably the one of most interest) occurs when the driving inverter shuts off and its output voltage rises from a small value to V_{DD} . During this process, the various parasitic capacitances on the output are charged through the pullup transistor. Fig. 2 shows a simple model of this circuit for timing analysis. The pullup, which is nonlinear, is approximated by a linear resistor, and the transition is represented by a voltage source going from 0 (or a low value) to V_{DD} at time $t = 0$. (Later, for simplicity, a unit step will be considered instead.) The polysilicon lines are represented by uniform RC lines. The resistance of the metal line is neglected, but its parasitic capacitance remains. Capacitances associated with the pullup source diffusion, contact cuts, and the gates being driven are included. Any nonlinear capacitances are approximated by linear ones.

If all the resistances except the pullup can be neglected, then all the capacitors can be lumped together, and the circuit re-

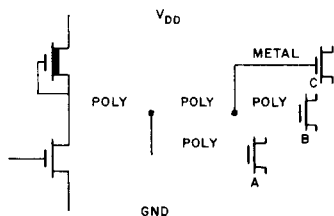


Fig. 1. Typical MOS signal-distribution network. The inverter is shown driving three gates, through a fanout network implemented in polysilicon and metal.

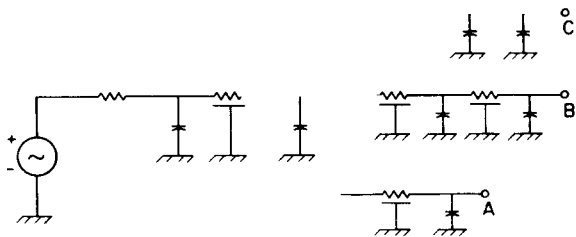


Fig. 2. Linear-circuit model for the network of Fig. 1. The voltage source is a step at time $t = 0$.

sponse may be found in closed form. The voltages at all the outputs are the same:

$$v_{\text{out}}(t) = V_{DD}(1 - e^{-t/RC_T}) \quad (1)$$

where R is the pullup resistance and C_T the total capacitance. Thus at a given time T , the output voltage $v_{\text{out}}(T)$ is given by (1), and the time at which $v_{\text{out}}(t)$ reaches some specified critical voltage V_{CR} is given by

$$T = RC_T \ln \frac{V_{DD}}{V_{DD} - V_{CR}} \quad (2)$$

However, if the resistances of the lines are comparable to that of the pullup, this solution is not correct. The circuit response cannot generally be calculated in closed form. The results below can be used to calculate upper and lower bounds to the delay that are very tight in the case where most of the resistance is in the pullup. The theory as presented here does not explicitly deal with nonlinearities and therefore does not apply to signal propagation through pass transistors.

Previous work on distributed RC circuits is summarized in the extensive bibliographies of Ghausi and Kelly [1] and Kumar [2]. There does not appear to be any treatment of RC trees, as distinct from RC lines, in these bibliographies. Perhaps the most complete treatment of the properties of RC lines is that of Protonotarios and Wing [3], [4]; some (but not all) of the theorems proved there also apply to RC trees. Most of the work cited deals with techniques to approximate the response of such networks, rather than to find bounds; an exception is that of Singhal and Vlach [5], [6]. An important early analytical approximation to delay is that by Elmore [7], who called the first moment of the impulse response the delay. This definition is inadequate because it does not define delay in terms of signal threshold.

Preliminary, restricted versions of some of the results given below have been presented before by the two senior authors

[8], [9], and utilized in at least two working timing analyzers [10]–[12]. The junior author simplified the derivation and tightened some of the bounds.

II. STATEMENT OF THE PROBLEM

An RC tree, as considered in this paper, is a generalization of the well-known RC lines [3], [4]. It may be defined recursively as follows. There are three primitive elements. First, a lumped capacitor between ground and another node is an RC tree. Second, a lumped resistor between two nonground nodes is an RC tree. Third, a (distributed) RC line, uniform or nonuniform, in the configuration with no dc path to ground, is an RC tree. Finally, any two RC trees with common ground, and one nonground node from each connected together, form a new RC tree. This definition does not permit resistor loops, so that the resistors (including those in the distributed RC lines) form a topological tree that does not include the ground node. All of the capacitors (including the distributed capacitances in the RC lines) are connected to ground. One of the nonground nodes of the final tree is assumed to be the input, and one or more nodes the outputs.

In many cases, each branch of the tree except the input terminates in an output; however, this is not required, and in this paper the outputs may be defined at any of the nonground nodes.

As a consequence of this definition, there is a unique path through the resistive part of the network from any nonground node to the input.

For simplicity, most of the theory below will be presented for the special case with only lumped resistors and capacitors. However, the generalization to include distributed RC lines (uniform or nonuniform) is straightforward. All the results apply in the form given, except that the summations in the formulas for T_P , T_{Di} , and T_{Ri} are replaced by a combination of summations and integrals. The easiest way to picture the result is to think of each RC line as represented by a finite number of lumped RC sections, so that the derivations apply, and then consider the limit as the number of sections used to represent each line goes to infinity. All the summations are well behaved in the limit. The required integrals are given explicitly in Appendix A for both uniform and nonuniform distributed lines.

The RC tree representing the signal path is, without loss of generality, assumed to be driven at the input with a unit step voltage (henceforth all voltages may be thought of as normalized to the magnitude of the step excitation). Gradually the voltages at all other nodes, and in particular at all the outputs, rise from 0 to 1 V. It is assumed that the output voltages cannot be calculated easily. The problem is to find simple upper and lower bounds for the output voltages, or, equivalently, to find upper and lower bounds for the delay associated with each output.

III. ANALYTICAL THEORY

Consider any output node i (in this paper, i will be used as an index selecting an output node) and any lumped capacitor at node k with capacitance C_k . The resistance R_{ki} is defined as the resistance of the portion of the (unique) path between

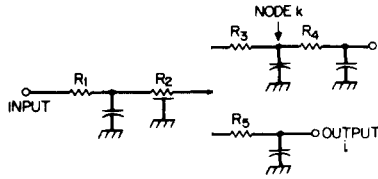


Fig. 3. Illustration of resistance terms. For this network, $R_{ki} = R_1 + R_2$, $R_{kk} = R_1 + R_2 + R_3$, and $R_{ii} = R_1 + R_2 + R_5$.

the input and i , that is common with the (unique) path between the input and node k . In particular, R_{ii} is the resistance between input and output i , and R_{kk} is the resistance between the input and node k . Thus $R_{ki} \leq R_{kk}$ and $R_{ki} \leq R_{ii}$. For an example, see Fig. 3.

The sum (over all the capacitors in the network)

$$T_P = \sum_k R_{kk} C_k \quad (3)$$

has the dimensions of time. Next, define for each output i two quantities that also have the dimensions of time

$$T_{Di} = \sum_k R_{ki} C_k \quad (4)$$

$$T_{Ri} = \left(\sum_k R_{ki}^2 C_k \right) / R_{ii}. \quad (5)$$

These quantities play a role in the final delay formulas but none of them is equal to the delay, although T_{Di} is equal to the first-order moment of the impulse response (see Appendix B), which has been called "delay" by Elmore [7]. Note that the network has one value of T_P , but each output of the network has a separate T_{Di} and T_{Ri} . It is easily shown from the definitions that

$$T_{Ri} \leq T_{Di} \leq T_P. \quad (6)$$

For RC trees without side branches, $T_{Di} = T_P$. An interpretation of T_P and T_{Di} in terms of the system function of the network appears in Appendix B.

The voltage at each output i (and in fact at each node) is a monotonic function of time during the transient, as proved in Appendix C. Also, the analog of the well-known fact that voltage along an RC line is a concave function of distance (suitably defined) is the following general result (proved in Appendix D):

$$R_{ii} [1 - v_k(t)] \geq R_{ki} [1 - v_i(t)]. \quad (7)$$

A similar result is found by interchanging i and k subscripts

$$R_{ki} [1 - v_k(t)] \leq R_{kk} [1 - v_i(t)]. \quad (8)$$

These results apply to any output i and any node k , whether the output is "upstream" or "downstream" from the node k .

At any instant of time, the voltage difference between the input and any output i may be calculated by summing the voltage drops along the (unique) path between input and output. Each such drop may be expressed as the resistance times the current feeding all "downstream" capacitors. Alternatively, this double sum may be expressed as a sum over all capacitors in the network, of the current through each capacitor

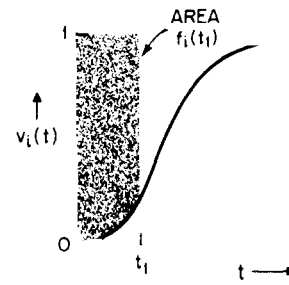


Fig. 4. Interpretation of $f_i(t)$ as the integral above the response $v_i(t)$. Note that $f_i(\infty) = T_{Di}$.

times that portion of the "upstream" resistance that also happens to lie along the path to the output i . This resistance is what has been defined as R_{ki} , so

$$1 - v_i(t) = \sum_k R_{ki} C_k \frac{dv_k}{dt}. \quad (9)$$

Equation (9) is integrated between 0 and t , and the result denoted $f_i(t)$:

$$\begin{aligned} f_i(t) &= \int_0^t [1 - v_i(t')] dt' \\ &= \sum_k R_{ki} C_k v_k(t) \\ &= T_{Di} - \sum_k R_{ki} C_k [1 - v_k(t)]. \end{aligned} \quad (10)$$

This integral plays a central role in the derivation of the bounds. A graphical interpretation appears in Fig. 4, which shows a typical step response. The area above the response but below the unit input is $f_i(t)$. As t approaches infinity, this approaches T_{Di} .

If (7) and (8) are used in (10), the result is

$$T_{Ri} [1 - v_i(t)] \leq T_{Di} - f_i(t) \leq T_P [1 - v_i(t)] \quad (11)$$

which is equivalent to

$$\frac{T_{Di} - f_i(t)}{T_P} \leq \frac{df_i(t)}{dt} \leq \frac{T_{Di} - f_i(t)}{T_{Ri}} \quad (12)$$

which, when integrated between times t_1 and $t_2 \geq t_1$, yields

$$\begin{aligned} [T_{Di} - f_i(t_1)] e^{-(t_2-t_1)/T_{Ri}} &\leq [T_{Di} - f_i(t_2)] \\ &\leq [T_{Di} - f_i(t_1)] e^{-(t_2-t_1)/T_P}. \end{aligned} \quad (13)$$

Since $v_i(t)$ is monotonic nondecreasing

$$(t_4 - t_3) [1 - v_i(t_4)] \leq f_i(t_4) - f_i(t_3) \quad (14)$$

for any nonnegative t_3 and t_4 .

The voltage bounds are now easily derived. Of course

$$v_i(t) \geq 0 \quad (15)$$

but, in addition, from (11) and (14) with $t_3 = 0$ and $t_4 = t$

$$v_i(t) \geq 1 - \frac{T_{Di}}{t + T_{Ri}} \quad (16)$$

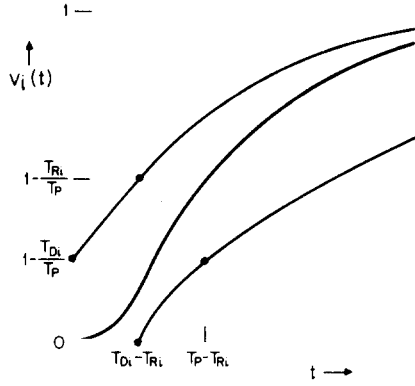


Fig. 5. Form of the bounds, with the distances from the exact solution exaggerated for clarity.

and, from the first inequality in (11), (14) with $t_3 = t - T_P + T_{Ri}$ and $t_4 = t$, and the second inequality in (13) with $t_1 = 0$ and $t_2 = t_3$

$$v_i(t) \geq 1 - \frac{T_{Di}}{T_P} e^{(T_P - T_{Ri})/T_P} e^{-t/T_P} \quad (17)$$

which holds only for $t \geq T_P - T_{Ri}$. The best lower bound is (15) for $t \leq T_{Di} - T_{Ri}$, (16) for $T_{Di} - T_{Ri} \leq t \leq T_P - T_{Ri}$, and (17) for $T_P - T_{Ri} \leq t$. The upper bounds on voltage are, from (11) and (14) with $t_3 = t$ and $t_4 = 0$

$$v_i(t) \leq 1 - \frac{T_{Di} - t}{T_P} \quad (18)$$

and, from the second inequality in (11), the first inequality in (13) with $t_1 = T_{Di} - T_{Ri}$ and $t_2 = t$, and (14) and $t_3 = T_{Di} - T_{Ri}$ and $t_4 = 0$

$$v_i(t) \leq 1 - \frac{T_{Ri}}{T_P} e^{(T_{Di} - T_{Ri})/T_{Ri}} e^{-t/T_{Ri}} \quad (19)$$

which holds only for $t \geq T_{Di} - T_{Ri}$. The best upper bound for voltage is (18) for $t \leq T_{Di} - T_{Ri}$ and (19) for $T_{Di} - T_{Ri} \leq t$.

Bounds for the time, given the voltage, are possible because the voltage is a monotonic function of time. Of course

$$t \geq 0 \quad (20)$$

and in addition, (18) and (19) can be inverted to yield

$$t \geq T_{Di} - T_P [1 - v_i(t)] \quad (21)$$

$$t \geq T_{Di} - T_{Ri} + T_{Ri} \ln \frac{T_{Ri}}{T_P [1 - v_i(t)]} \quad (22)$$

and (16) and (17) yield

$$t \leq \frac{T_{Di}}{1 - v_i(t)} - T_{Ri} \quad (23)$$

$$t \leq T_P - T_{Ri} + T_P \ln \frac{T_{Di}}{T_P [1 - v_i(t)]} \quad (24)$$

where (22) applies only if $v_i(t) \geq 1 - T_{Ri}/T_P$, and (24) only if $v_i(t) \geq 1 - T_{Di}/T_P$. The general form of all these bounds is illustrated in Fig. 5.

These bounds, (15)-(19) for voltage, and (20)-(24) for time, constitute the major result of this paper.

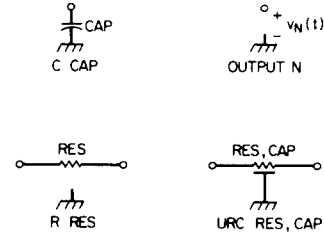


Fig. 6. Elements sufficient for describing RC trees. For simplicity, only uniform distributed RC lines are included. The parameters are CAP, RES, and N, and the functions which return networks are C, OUTPUT, R, and URC. The capacitor and output designation are one-port elements, and the resistor and uniform line are two-port elements.

IV. PRACTICAL HIERARCHICAL ALGORITHMS

Use of hierarchy is a powerful way to deal with complexity in design of large systems. Computation cost is usually less with hierarchical algorithms, and analysis of part of a design can be done before the rest is known. In this section, programs are given for calculating the voltage and time bounds of this paper hierarchically. Although intended for exposition, these programs are complete and do work. They may be used interactively, without any changes whatever, for small or moderate size networks, or, for large networks, they may be incorporated into systems that deal with machine-readable network descriptions.

One way to use the inequalities of this paper is to consider the overall RC tree, and compute for each capacitor the appropriate R_{ki} and R_{kk} so that T_P , T_{Di} , and T_{Ri} for each output can be found. Of course, for networks with distributed lines, the sums are augmented with integrals as discussed in Appendix A. In this approach, the calculations for each output require time proportional to the square of the number of elements.

An alternate scheme is to build up the network by construction, and calculate independently for each of the partially constructed networks enough information to permit the final calculation of T_P , T_{Di} , and T_{Ri} . A recursive definition of RC trees is given below, and if the network is expressed in these terms rather than in the form of a schematic diagram, the resulting expression can be used as a guide for the calculations. The computation time for each output is proportional to the number of elements, rather than the square of the number. Programs that implement this approach appear below.

Fig. 6 shows the four building blocks: lumped capacitor, lumped resistor, uniform RC line, and declaration of output. The capacitor and the output label are considered as two-terminal, or one-port networks. The RC line and the resistor are considered as two-port networks. If desired, particular nonuniform RC lines, such as exponentially or linearly tapered lines, can be included also. Fig. 7 shows the five permissible ways of wiring these building blocks, or previously wired sub-networks, together. Any RC tree can be denoted by an expression using only these wiring functions. The syntax shown is identical to APL syntax, and the programs below are written in APL. Note that Figs. 6 and 7 do not give a minimal set of elements or wiring functions, since some can be expressed in

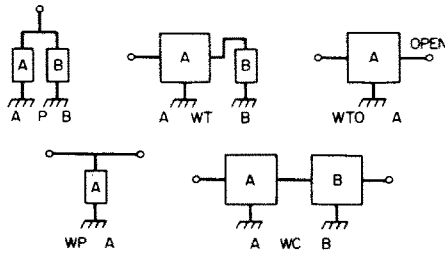


Fig. 7. Wiring functions for interconnecting elements or subtrees. The functions which return one-port networks are *P*, *WT*, and *WTO*, and those that return two-port networks are *WP* and *WC*. Here *A* and *B* are previously defined *RC* trees.

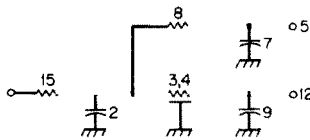


Fig. 8. Example network. Parameter values are in ohms and farads. The characteristic times (in sec.) are $T_P = 419$, $T_{D5} = 386$, $T_{R5} = 307.7$, $T_{D12} = 363$, and $T_{R12} = 335.2$.

terms of others. The names for the wiring functions are taken from the notation of the program MARTHA [13] - [16].

Example: The network shown in Fig. 8 may be denoted

$$(R\ 15)\ WT(C\ 2)\ P((R\ 8)\ WT(C\ 7)\ P\ OUTPUT\ 5) \\ P((URC\ 3\ 4)\ WC\ WPC\ 9)\ WT\ OUTPUT\ 12 \quad (25)$$

and is a one-port network, with two declared outputs.

For convenience, this notation allows a network with only one output to be expressed as a two-port with the second port an implicit output, without any explicit output declaration. The explicit declaration of outputs is handy because often side branches do not represent outputs of interest.

If an expression such as (25) is to be used as a guide for the calculations, then each function shown must correspond to the calculation of partial results which are sufficient to allow further calculations. The following information is adequate at each stage in the construction of the network:

- (1) Total capacitance C_T .
- (2) T_P of the network as constructed so far.
- (3) For a two-port, considering port 2 as an implicit output, R_{22} , T_{D2} , and T_{R2} . (For convenience, the product $R_{22}T_{R2}$ is used in the programs below instead of T_{R2} .)
- (4) For each declared output in a one-port or two-port network, R_{ii} , T_{Di} , and T_{Ri} . (For convenience, $R_{ii}T_{Ri}$ is used rather than T_{Ri} .)
- (5) For each declared output in a two-port network, R_{2i} .

Each of the quantities identified above pertains to the particular subnetwork and can be calculated from a knowledge of that subnetwork alone, independent of how the subnetwork may later be wired together with other subnetworks. As an example of the use of these quantities during construction of the network, consider the cascade operation *WC*. The objective is to find C_T , T_P , R_{22} , T_{D2} , T_{R2} , and all R_{ii} , T_{Di} , T_{Ri} , and R_{2i} of the cascade *A WC B* from the corresponding quantities for its two arguments *A* and *B*. The formulas for calculating these are

```

∇ Z+C CAP
[1] +'CAPACITANCE' ERRORIF 1≠pCAP+,CAP
[2] Z+1,CAP,0
∇
∇ Z+OUTPUT N
[1] +'OUTPUT LABEL' ERRORIF 1≠pN+,N
[2] Z+ 1 0 0 ,N, 0 0 0
∇
∇ Z+R RES
[1] +'RESISTANCE' ERRORIF 1≠pRES+,RES
[2] Z+ 2 0 0 ,RES, 0 0
∇
∇ Z+URC RC;RES;CAP
[1] +'RES, CAP' ERRORIF 2≠pRC+,RC
[2] RES+1+RC
[3] CAP+1+RC
[4] Z+2,CAP,(CAP*RES+2),RES,(CAP*RES+2),CAP*RES*RES+3
∇
    
```

Fig. 9. APL functions for the elements.

$$C_T = C_{TA} + C_{TB} \quad (26)$$

$$T_P = T_{PA} + T_{PB} + R_{22A}C_{TB} \quad (27)$$

$$R_{22} = R_{22A} + R_{22B} \quad (28)$$

$$T_{D2} = T_{D2A} + T_{D2B} + R_{22A}C_{TB} \quad (29)$$

$$T_{R2}R_{22} = T_{R2A}R_{22A} + T_{R2B}R_{22B} \\ + 2R_{22A}T_{D2B} + R_{22A}^2C_{TB} \quad (30)$$

$$R_{ii} = R_{iiA}, R_{iiB} + R_{22A} \quad (31)$$

$$T_{Di} = (T_{DiA} + R_{2iA}C_{TB}), T_{DiB} \\ + R_{22A}C_{TB} + T_{D2A} \quad (32)$$

$$T_{Ri}R_{ii} = (T_{RiA}R_{iiA} + R_{2iA}^2C_{TB}), T_{RiB}R_{iiB} \\ + T_{R2A}R_{22A} + 2R_{22A}T_{DiB} + R_{22A}^2C_{TB} \quad (33)$$

$$R_{2i} = R_{2iA}, R_{2iB} + R_{22A} \quad (34)$$

The corresponding formulas for the other wiring functions are similar, but not as complicated.

A set of APL functions which implement this scheme appear in Figs. 9-12. The necessary data is passed around in the form of vectors. A one-port network is represented by the vector 1, C_T , T_P , followed by zero or more sets of four numbers, N_i , R_{ii} , T_{Di} , $T_{Ri}R_{ii}$. The number 1 starting off the vector is the number of ports, and N_i is the (numerical) label for each output. It is not necessary to pass along the number of declared outputs since that can be calculated from the length of the vector. In a similar way, a two-port network is represented by the vector 2, C_T , T_P , R_{22} , T_{D2} , $T_{R2}R_{22}$, followed by zero or more sets of five numbers, N_i , R_{ii} , T_{Di} , $T_{Ri}R_{ii}$, and R_{2i} , one set for each declared output.

The background functions in Fig. 12 provide 1) some error control (with automatic abort in case of error), 2) calculation of the number of ports of a network, 0 being returned for ill-formed arguments, 3) a matrix with the data for the declared outputs, and 4) extraction of T_P , T_{Di} , and T_{Ri} . The four elements appear in Fig. 9, and the wiring functions in Fig. 10. The listing of *WC*, for example, shows after error checking, the calculation of the required output, term by term, from the arguments. This function can be compared with (26)-(34).

Fig. 11 shows five functions intended to calculate the bounds for any network. The convention followed is that if the argument for any of these is a two-port network, the second port

```

▽ Z+A P B
[1] + '1-PORT' ERRORIF(1≠NPORTS A)∨1≠NPORTS B
[2] Z+1,(A[2 3]+B[2 3]),(3+A),3+B
▽
▽ Z+A WT B;E
[1] + '2-PORT' ERRORIF 2≠NPORTS A
[2] + '1-PORT' ERRORIF 1≠NPORTS B
[3] Z+1,(A[2]+B[2]),A[3]+B[3]+B[2]×A[4]
[4] E+OUTPUTS A
[5] E[;3]+E[;3]+B[2]×E[;5]
[6] E[;4]+E[;4]+B[2]×E[;5]×2
[7] Z+Z,, 0 1 +E
[8] E+OUTPUTS B
[9] E[;2]+E[;2]+A[4]
[10] E[;4]+E[;4]+A[6]+(2×A[4]×E[;3])+B[2]×A[4]×2
[11] E[;3]+E[;3]+A[5]+B[2]×A[4]
[12] Z+Z,,E
▽
▽ Z+WTO A
[1] + '2-PORT' ERRORIF 2≠NPORTS A
[2] Z+1,A[2 3],, 0 1 +OUTPUTS A
▽
▽ Z+A WC B;E
[1] + '2-PORT' ERRORIF(2≠NPORTS A)∨2≠NPORTS B
[2] Z+2,(A[2]+B[2]),(A[3]+B[3]+B[2]×A[4]),(A[4]+B[4]),A[5]+B[5]+B[2]×A[4]
[3] Z+Z,A[6]+B[6]+(2×B[5]×A[4])+B[2]×A[4]×2
[4] E+OUTPUTS A
[5] E[;3]+E[;3]+B[2]×E[;5]
[6] E[;4]+E[;4]+B[2]×E[;5]×2
[7] Z+Z,,E
[8] E+OUTPUTS B
[9] E[; 2 5]+E[; 2 5]+A[4]
[10] E[;3]+E[;3]+A[5]+B[2]×A[4]
[11] E[;4]+E[;4]+A[6]+(2×B[5]×A[4])+B[2]×A[4]×2
[12] Z+Z,,E
▽
▽ Z+WP A
[1] + '1-PORT' ERRORIF 1≠NPORTS A
[2] Z+2,A[2 3], 0 0 0 ,, (OUTPUTS A),0
▽

```

Fig. 10. APL functions for the wiring functions.

```

▽ Z+VMIN A;II;TD;TP;TR
[1] +T SETUP A
[2] Z+(II≧TP-TR)×1-(TD+TP)×*1-(II+TR)+1E-30[TP
[3] Z+Z[1-TD+1E-30[II+TR
▽
▽ Z+VMAX A;II;TD;TP;TR
[1] +T SETUP A
[2] Z+TR×*0[1+(TD-II)+1E-30[TR
[3] Z+1-(Z[TD-II)+1E-30[TP
▽
▽ Z+TMIN A;II;TD;TP;TR
[1] +V SETUP A
[2] Z+TD-TP×1-II
[3] Z+Z[(TR≧TP×1-II)×TD-TR×1-TR+1E-20[TP×1-II
▽
▽ Z+TMAX A;II;TD;TP;TR
[1] +V SETUP A
[2] Z+(TD+1E-20[1-II)-TR
[3] Z+Z[(TP-TR)+0[TP×TD+1E-20[TP×1-II
▽
▽ Z+OK A
[1] + 'CIRCUIT' ERRORIF~(NPORTS A)ε 1 2
[2] Z+(T°.≧TMAX A)-T°.<TMIN A
▽
▽ Z+M ERRORIF B
[1] Z+10
[2] +(-1εB)/0
[3] M, 'ERROR.'
[4] Z+0
▽
▽ Z+NPORTS A
[1] Z+0
[2] +(1≠ρρA)/0
[3] Z+1
[4] +((1=1+A)∧3=4|ρA)/0
[5] Z+2
[6] +((2=1+A)∧(6≤ρA)∧1=5|ρA)/0
[7] Z+0
▽
▽ E+OUTPUTS A
[1] E+ 0 0 ρ0
[2] +(1 2 =NPORTS A)/L1,L2
[3] +0
[4] L1:E+(((ρA)-3)±4),4)ρ3+A
[5] +0
[6] L2:E+(((ρA)-6)±5),5)ρ6+A
▽
▽ Z+I SETUP A;E;N;S
[1] N≠NPORTS A
[2] +Z+'CIRCUIT' ERRORIF~Nε 1 2
[3] +(N=2)/L2
[4] E+OUTPUTS A
[5] S+(ρI),1+ρE
[6] TP+A[3]
[7] TD+SpE[;3]
[8] TR+Sp(E[;2]=0)×E[;4]+E[;2]
[9] II+ϕ(ϕS)ρϕI
[10] +0
[11] L2:TP+A[3]
[12] TD+A[5]
[13] TR+(A[4]=0)×A[6]+A[4]
[14] II+I
▽

```

Fig. 11. Response functions. The very small numbers in the functions guard against errors for pathological networks and certain limiting values for voltage and time.

is taken as the desired output, and the declared outputs are ignored. If the argument is a one-port network, then the declared outputs are used. The two functions *TMIN* and *TMAX* calculate the lower and upper bounds for delay, and refer to a global variable named *V* which contains the threshold, a number (or array of numbers) between 0 and 1. The functions *VMIN* and *VMAX* calculate the lower and upper bounds for signal voltage and refer to a global variable *T* containing an ar-

Fig. 12. APL background functions to support the functions in Figs. 9, 10, and 11.

EXAMPLE OF THE USE OF RC-TREE DELAY CALCULATIONS:

```
BRANCH1 + (R 8) WT (C 7) P OUTPUT 5
BRANCH2 + (URC 3 4) WT (C 9) P OUTPUT 12
NET + (R 15) WT (C 2) P BRANCH1 P BRANCH2
```

NOW THE NETWORK IS DEFINED.

```
V + 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
```

NOW THE VECTOR OF THRESHOLD VOLTAGES IS DEFINED.

NEXT TO FIND THE MINIMUM AND MAXIMUM BOUNDS FOR DELAY:

V, (TMIN NET), TMAX NET				
0	0	0	78.261	27.833
0.1	8.9	0	121.03	68.167
0.2	50.8	27.8	170.39	117.22
0.3	93.05	72.555	226.34	173.17
0.4	140.49	124.22	290.92	237.76
0.5	196.6	185.33	367.32	314.15
0.6	265.27	260.12	460.81	407.65
0.7	353.8	356.54	581.35	528.18
0.8	478.57	492.44	751.24	698.07
0.9	691.88	724.76	1041.7	988.5
V	TMIN	TMAX	TMIN	TMAX
	OUTPUT 5	OUTPUT 12	OUTPUT 5	OUTPUT 12

NOW TO DEFINE A DELAY VECTOR AND GET THE BOUNDS ON VOLTAGE:

```
T + 0 20 40 60 80 100 200 300 400 500 1000 2000
```

T, (VMIN NET), VMAX NET				
0	0	0	0.078759	0.13365
20	0	0	0.12649	0.18138
40	0	0.03243	0.17422	0.2286
60	0	0.0814	0.22196	0.27328
80	0.0044853	0.12565	0.26968	0.31538
100	0.053316	0.16644	0.31563	0.35503
200	0.25459	0.34342	0.5055	0.52141
300	0.41286	0.48283	0.64269	0.64487
400	0.53752	0.59263	0.74182	0.73648
500	0.63571	0.67913	0.81345	0.80446
1000	0.88954	0.90271	0.96326	0.95601
2000	0.98984	0.99105	0.99857	0.99777
T	VMIN	VMIN	VMAX	VMAX
	OUTPUT 5	OUTPUT 12	OUTPUT 5	OUTPUT 12

Fig. 13. Example of the use of the fast calculation scheme to find upper and lower bounds on delay and response voltage.

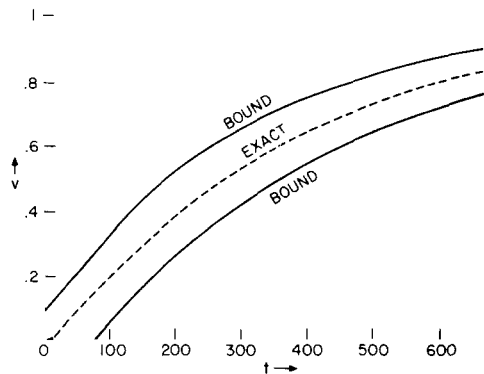


Fig. 14. Upper and lower bounds for output 5, as calculated in Fig. 13. The exact solution, found from circuit simulation, is shown also.

ray of (positive) delay times. The final function, *OK*, refers to both *V* and *T* and returns 1 if all is well, that is, if $TMAX \leq T$, or -1 if the network definitely will fail, that is if $T < TMIN$, or 0 if the bounds are not tight enough to tell for sure, that is if $TMIN \leq T < TMAX$. An example of the use of these functions to test the network in Fig. 8 is shown in Figs. 13 and 14.

V. APPLICATION TO PLA SPEED ESTIMATES

These bounds are applied, as an example, to polysilicon lines driving the AND plane of a PLA, to determine whether or not the dominant delay occurs here. It is assumed that a strong

```
V Z+PLALINE N:A
[1] A+(URC 180 0.01)WC URC 30 0.013
[2] A IS A SINGLE SECTION ACCOUNTING FOR TWO MINTERMS
[3] Z+(R 380)WC WP C 0.04
[4] Z IS THE PULLUP R AND C FOR SUPERBUFFER DRIVER
[5] LOOP:+(N<=0)/0
[6] Z+Z WC A
[7] N+N-2
[8] +LOOP
```

Fig. 15. APL function which returns a model of a PLA line with *N* minterms.

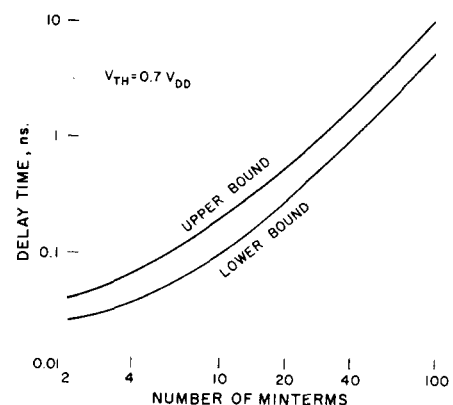


Fig. 16. Upper and lower bounds on response time of the network of Fig. 15, shown as a function of the number of minterms in the PLA.

superbuffer driver drives the line, and that every other minterm has a transistor present. The gates are assumed to be 4-microns square, separated by 24 μ m of RC line. The poly resistance is assumed to be 30- Ω per square, the gate oxide thickness 400 \AA , and the field-oxide thickness 3000 \AA .

These numbers lead to a capacitance of 0.01 pF and resistance 180 Ω between gates, and a resistance of 30 Ω and capacitance of 0.013 pF for each gate. The network is driven by a source resistance of 380 Ω and the effective capacitance of the output of the driver is estimated as 0.04 pF.

A function which returns a network with *N* minterms is shown in Fig. 15. The results of calculating the delay as a function of the number of minterms are shown in Fig. 16. The voltage threshold was taken to be 0.7 times V_{DD} . On this log-log plot the quadratic dependence of delay on number of minterms (as a measure of the length of the line) is evident. Also evident is the fact that even with as many as a hundred minterms, the delay is guaranteed to be no worse than 10 ns. This suggests that the dominant delay in a PLA occurs elsewhere.

VI. CONCLUSIONS

A computationally efficient method for calculating the signal delay through MOS interconnect lines with fanout has been described. Tight upper and lower bounds for the step response of RC trees have been presented, together with linear-time algorithms for these bounds from an algebraic description of the tree. Substantial computational simplicity is achieved even in the presence of RC distributed lines by representing the RC tree by a small set of suitably defined characteristic times, which can be calculated easily and used to generate the bounds.

Although only the step response is considered here, the results can be extended to upper and lower bounds for arbitrary

excitation by use of the superposition integral. This extension is discussed in Appendix E. An example of this calculation appears in [9].

Extensions of the theory to RC trees with nonlinear elements (similar to the work of Glasser [17] for nonlinear MOS inverters) would be desirable for better modeling of MOS circuits. Investigations of RC trees with nonlinear capacitors and resistors are now under way, along with attempts to unify the modeling of gates and interconnects, and in particular to include pass transistors in the interconnects. Tighter bounds are also being looked for.

APPENDIX A

The results of this paper are valid for RC trees that contain distributed RC lines. All results apply without change, except that the definitions of T_P , T_{Di} , and T_{Ri} in (3)–(5) are replaced by (36)–(38) below.

The summations in (3)–(5) are for the case of lumped capacitors only, and the index k runs over all lumped capacitors in the network. The form for networks with distributed RC lines is similar; the index k runs over both lumped capacitors and RC lines. The terms for lumped capacitors are unchanged from (3)–(5), but for distributed lines additional terms appear in (36)–(38).

Each line k has a total capacitance C_k and appears in the network with one end (say the left-hand end) nearer the input of the network. Along the line, the capacitance is distributed, but the cumulative capacitance c is a function of position, and has a value between 0 (at the left end) and C_k (at the right end). For each value of c , there is a value of cumulative resistance $r(c)$ monotonically increasing with c ; $r(0) = 0$ and $r(C_k)$ is the total resistance of the line. For uniform lines, $r(c)$ is a linear function, and for nonuniform lines $r(c)$ has other shapes. Define the series of integrals

$$I_k^{(n)} = \int_0^{C_k} [r(c)]^n dc. \quad (35)$$

Note that if the line k is interpreted as a simple RC tree without any additional elements, then its T_P and T_D are $I_k^{(1)}$ and its T_{Ri} is $I_k^{(2)}/r(C_k)$. For a uniform line, $I_k^{(1)} = r(C_k) C_k/2$ and $I_k^{(2)} = [r(C_k)]^2 C_k/3$.

For each distributed line k let R_{kk} be the resistance between the left-hand end of the line and the input of the network, and R_{ki} be the portion of that resistance that also lies on the (unique) path between the input and any output i . Then the expressions for T_P , T_{Di} , and T_{Ri} are

$$T_P = \sum_k R_{kk} C_k + \sum_k I_k^{(1)} \quad (36)$$

$$T_{Di} = \sum_k R_{ki} C_k + \sum_k I_k^{(1)} \quad (37)$$

$$T_{Ri} = \left(\sum_k R_{ki}^2 C_k + 2 \sum_k R_{ki} I_k^{(1)} + \sum_k I_k^{(2)} \right) / R_{ii} \quad (38)$$

where the first sum in all three expressions is over both lumped capacitors and distributed lines; the second sum in (36) is over all distributed lines; and the second sum in (37) and the

second and third sums in (38) are over only those distributed lines which lie along the (unique) path between the input and output i .

APPENDIX B

From (3) it is evident that T_P is equal to the sum of all the open-circuit time constants of the network, a quantity that is well known in the analysis of multistage amplifiers, and that has been shown to be equal to the negative of the sum of the inverse of all the transmission poles [18]. That is, if the normalized system function $H_i(s)$ for the output i is

$$H_i(s) = \frac{1 + b_1 s + b_2 s^2 + \dots}{1 + a_1 s + a_2 s^2 + \dots} \quad (39)$$

then $T_P = a_1$. Also, it can be shown that $T_{Di} = a_1 - b_1$. To prove this, one starts from the equality [19] between $a_1 - b_1$ and the first-order time moment $\int_0^\infty h_i(t) t dt$ of the impulse response $h_i(t)$. Integrating by parts, one can show that

$$\int_0^\infty h_i(t) t dt = \int_0^\infty [1 - v_i(t)] dt,$$

and therefore is equal to $f_i(\infty) = T_{Di}$, as given by (10). This completes the proof, and shows that T_{Di} is equal to the first-order time-moment of the impulse response.

APPENDIX C

It is proved here that when a linear RC tree is excited with a step input from an initial rest condition, the voltage at each node is a monotonic function of time. This condition is identical to the condition that the impulse response of the same network is nonnegative. The proof is to assume that, with an impulse applied at the input, the voltage on one or more nodes is, at some instant of time, negative, and then show that this assumption leads to a contradiction.

It is assumed that distributed RC lines in the tree can be replaced by finite lumped ladder approximations with arbitrarily close impulse responses, so that if one or more nodes of the original network has a negative voltage, then so does one or more nodes of an approximate lumped network. For the remainder of this appendix, this lumped network is dealt with.

At time $t = 0+$, the voltages on all nodes are nonnegative. Let $v_{\min}(t)$ denote the lowest voltage of any node in the network at time t . Assume that at some time $t_0 > 0$, $v_{\min}(t_0) < 0$. Then there must be some prior time t_1 when v_{\min} and its derivative with respect to time are both negative.

At time t_1 , $v_{\min}(t_1)$ is achieved by at least one node. If it is achieved by more than one, at least one must have a negative derivative. This node is characterized by having the lowest voltage in the network, and also a negative derivative. The net current flowing into this node from other nodes in the RC tree is nonnegative, because the adjacent nodes, to which this node is directly connected through resistors, are not at a lower potential. This net current must flow into the capacitor at that node, and therefore the rate of change of the node voltages is nonnegative. This contradiction proves the impossibility of the assumption above that a node voltage is negative.

APPENDIX D

Equation (7) is to be proved. Note first, for any three nodes i , j , and k in the network, that R_{ii} is at least as large as both R_{ki} and R_{ji} . Also, R_{jk} is at least as large as the lesser of R_{ki} and R_{ji} . Thus

$$R_{ii}R_{jk} \geq R_{ki}R_{ji}. \quad (40)$$

Now note that, similar to (9)

$$1 - v_k(t) = \sum_j R_{jk}C_j \frac{dv_j}{dt} \quad (41)$$

$$1 - v_i(t) = \sum_j R_{ji}C_j \frac{dv_j}{dt} \quad (42)$$

so that, because of (40) and the fact that $v_j(t)$ is monotonic

$$\begin{aligned} & R_{ii}[1 - v_k(t)] - R_{ki}[1 - v_i(t)] \\ &= \sum_j (R_{ii}R_{jk} - R_{ki}R_{ji})C_j \frac{dv_j}{dt} \\ &\geq 0 \end{aligned} \quad (43)$$

which immediately implies (7).

APPENDIX E

Bounds for the response $y_i(t)$ of an RC tree to an arbitrary excitation $x(t)$ can be obtained from the upper and lower bounds $v_{ui}(t)$ and $v_{li}(t)$ derived for the unit step response $v_i(t)$.

First, the superposition integral can be used to obtain $y_i(t)$ as

$$\begin{aligned} y_i(t) &= \int_0^t v_i(t-t') \frac{dx(t')}{dt'} dt' \\ &= v_i(t) * dx/dt \end{aligned} \quad (44)$$

where * denotes time convolution. From

$$v_{li}(t) \leq v_i(t) \leq v_{ui}(t) \quad (45)$$

one obtains

$$v_{li}(t) * dx/dt \leq y_i(t) \leq v_{ui}(t) * dx/dt, \quad dx/dt \geq 0 \quad (46)$$

$$v_{ui}(t) * dx/dt \leq y_i(t) \leq v_{li}(t) * dx/dt, \quad dx/dt \leq 0 \quad (47)$$

where $v_{ui}(t)$ and $v_{li}(t)$ are known analytically. From (46) it can be seen that bounds for the ramp response can be obtained simply by integrating the unit step bounds. Equations (46) and (47) apply for monotonic inputs.

For the general case where the excitation $x(t)$ has both positive and negative slopes, one can define the following functions:

$$v_{MAXi}(t, t') = \begin{cases} v_{ui}(t-t'), & dx/dt' > 0 \\ v_{li}(t-t'), & dx/dt' < 0 \\ 0, & dx/dt' = 0 \end{cases} \quad (48)$$

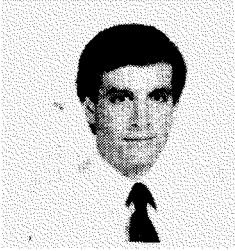
$$v_{MINi}(t, t') = \begin{cases} v_{li}(t-t'), & dx/dt' > 0 \\ v_{ui}(t-t'), & dx/dt' < 0 \\ 0, & dx/dt' = 0. \end{cases} \quad (49)$$

The response $y_i(t)$ is then bounded by

$$\begin{aligned} \int_0^t v_{MINi}(t, t') \frac{dx(t')}{dt'} dt' &\leq y_i(t) \\ &\leq \int_0^t v_{MAXi}(t, t') \frac{dx(t')}{dt'} dt'. \end{aligned} \quad (50)$$

REFERENCES

- [1] M. S. Ghauri and J. J. Kelly, *Introduction to Distributed-Parameter Networks*. New York: Holt, Rinehart and Winston, 1968.
- [2] U. Kumar, "A bibliography of distributed-RC networks," *IEEE Circuits Syst. Mag.*, vol. 2, no. 2, pp. 9-19, June, 1980.
- [3] E. N. Protonotarios and O. Wing, "Theory of nonuniform RC lines, Part I: Analytic properties and realizability conditions in the frequency domain," *IEEE Trans. Circuit Theory*, vol. CT-14, no. 1, pp. 2-12, Mar. 1967.
- [4] E. N. Protonotarios and O. Wing, "Theory of nonuniform RC lines, Part II: Analytic properties in the time domain," *IEEE Trans. Circuit Theory*, vol. CT-14, no. 1, pp. 13-20, Mar. 1967.
- [5] K. Singhal and J. Vlach, "Calculation of time responses of nonuniform RC-distributed active networks," in *Dig. Tech. Pap., London 1971 IEEE Int. Symp. Electrical Network Theory*, (London, England), pp. 75-76, Sept. 6-10, 1971.
- [6] K. Singhal and J. Vlach, "Approximation of nonuniform RC-distributed networks for frequency- and time-domain computations," *IEEE Trans. Circuit Theory*, vol. CT-19, no. 4, pp. 347-354, July 1972.
- [7] W. C. Elmore, "The transient response of damped linear networks with particular regard to wide-band amplifiers," *J. Appl. Phys.*, vol. 19, no. 1, pp. 55-63, Jan. 1948.
- [8] Paul Penfield, Jr. and Jorge Rubinstein, "Signal delay in RC tree networks," in *Proc. Second Caltech Conf. on Very Large Scale Integration*, (Pasadena, CA), pp. 269-283, Jan. 19-21, 1981.
- [9] Paul Penfield, Jr. and Jorge Rubinstein, "Signal delay in RC tree networks," in *Proc. Eighteenth Design Automation Conf.*, (Nashville, TN), pp. 613-617, June 29-July 1, 1981.
- [10] Rathin Putatunda, "AUTO-DELAY: A program for automatic calculation of delay in LSI/VLSI chips," in *Proc. Nineteenth Design Automation Conf.*, (Las Vegas, NV), pp. 616-621, June 14-16, 1982.
- [11] Rathin Putatunda, "Automatic calculation of delay in custom generated LSI/VLSI chips," in *Proc. IEEE Int. Conf. Circuits and Computers*, (New York, NY), pp. 193-196, Sept. 28-Oct. 1, 1982.
- [12] N. P. Jouppi, "TV: An nMOS timing analyzer," in *Proc. Third Caltech Conf. on Very Large Scale Integration*, R. Bryant, ed. Rockville, MD: Computer Science Press, 1983, pp. 71-85.
- [13] P. Penfield, Jr., *MARTHA User's Manual*. Cambridge, MA: MIT Press, 1971.
- [14] P. Penfield, Jr., "Description of electrical networks using wiring operators," *Proc. IEEE*, vol. 60, pp. 49-53, Jan. 1972.
- [15] P. Penfield, Jr., *MARTHA User's Manual, 1973 Addendum*. Cambridge, MA: MIT Dep. of Elect. Eng., 1974.
- [16] P. Penfield, Jr., "MARTHA," in *Handbook of Circuit Analysis Languages and Techniques*, R. W. Jensen and L. P. McNamee, eds. Englewood Cliffs, NJ: Prentice-Hall, 1976, pp. 496-535.
- [17] L. A. Glasser, "The analog behavior of digital integrated circuits," in *Proc. Eighteenth Design Automation Conf.*, (Nashville, TN), pp. 603-612, June 29-July 1, 1981.
- [18] R. D. Thornton, C. L. Searle, D. O. Pederson, R. B. Adler, and E. J. Angelo, Jr., *Multistage Transistor Circuits*. New York: Wiley, 1965, sec. 1.2.
- [19] R. D. Thornton et al., *Multistage Transistor Circuits*. New York: Wiley, 1965, sec. 8.2.6.



Jorge Rubinstein (M'81) was born in Mexico City. He received the degree Ingeniero Mecánico Electricista in 1977 from the Universidad Anáhuac in Mexico City. He received the S.M. degree from the Massachusetts Institute of Technology, Cambridge, MA in 1979.

Since 1980 he has been employed by Digital Equipment Corporation in Massachusetts.

*



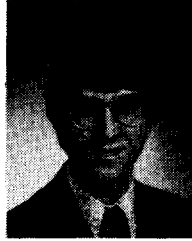
Paul Penfield, Jr. (S'57-M'62-F'72) was born in Detroit, MI, on May 28, 1933. He received the B.A. degree in physics from Amherst College, Amherst, MA, in 1955, and the Sc.D. in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1960.

At present he is a Professor of Electrical Engineering at MIT, overseeing its interdisciplinary VLSI research program. Between 1974 and 1978 he served as Associate Head of the MIT Department of Electrical Engineering and Com-

puter Science. His professional interests have included noise and thermodynamics, solid-state microwave circuits, electrodynamics of moving media, circuit theory, computer-aided design, APL language extensions, and integrated-circuit design automation.

He is a member of Sigma Xi, American Physical Society, Audio Engineering Society, and Association for Computing Machinery.

*



Mark A. Horowitz (S'78) received the B.S. and S.M. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1978. He is currently with the Department of Electrical Engineering at Stanford University, Palo Alto, CA, where his Ph.D. research is on computer aids for integrated circuit design.

Since 1979 he has been employed part time by Philips Research Labs, Sunnyvale, CA, working on device modeling and digital circuit design.

Mr. Horowitz is a member of Eta Kappa Nu and Sigma Xi.