Signal Delay in RC Tree Networks*

Paul Penfield, Jr.**
Jorge Rubinstein***

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) to 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 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-micron minimum feature size. The importance of this technique grows as the wiring lengths increase or the feature size decreases.

---

*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, and in part by the Air Force under Contract Number AFOSR 4-9620-80-0073.

**Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Room 36-575, Cambridge, MA 02139; telephone (617) 253-2506.

***Digital Equipment Corporation, 75 Reed Road, Hudson, MA 01749; telephone (617) 568-4835.

Consider the circuit of Figure 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. Figure 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 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.

In general, the circuit response cannot be calculated in closed form. The results of this paper 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. A more complete discussion of this theory will appear elsewhere [1].

II. Statement of the Problem

An RC tree is defined as follows. Consider any resistor tree with no node at ground. From each node in this tree a capacitor to ground may be added, and any resistor may be replaced by a distributed RC line. Although nonuniform RC lines may appear in an RC tree, for simplicity, the examples in this paper involve only lumped resistors and capacitors and uniform RC lines. An RC tree has one input and any number of outputs. Side branches may or may not end in a node that is considered as an output; in fact, outputs may be taken anywhere in the tree. Nonuniform RC lines are special cases of RC trees, without any side branches. An important property of RC trees is that there is a unique path from any point in the tree to the input.

The tree representing the signal path is driven at the input with a unit step voltage. Gradually the voltages at all other nodes, and in particular at all the outputs, rise from 0 to 1 volt. 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 e, and any lumped capacitor at node k with

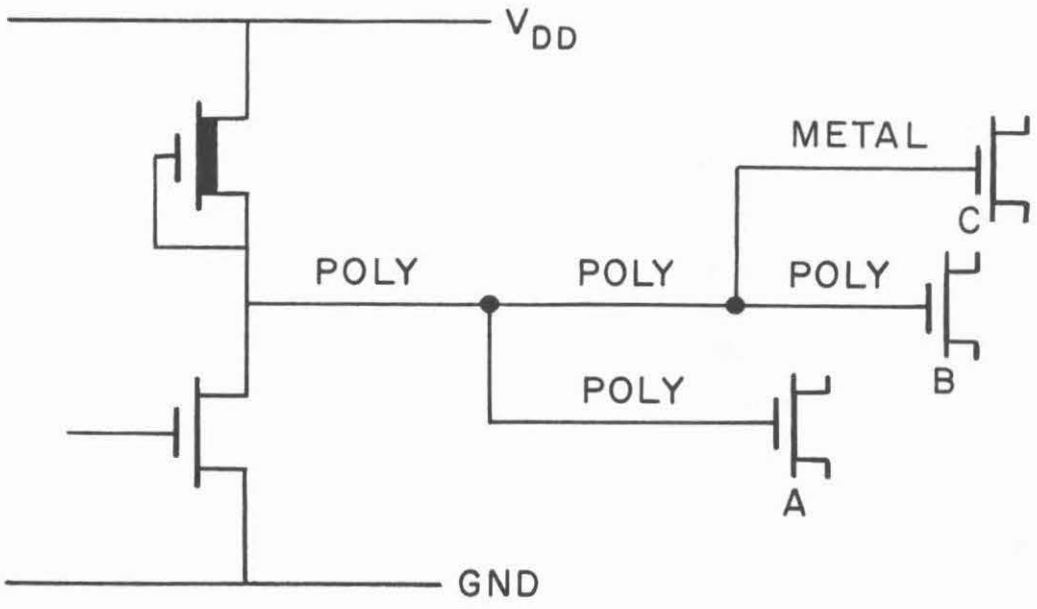

Figure 1. Typical MOS signal-distribution network. The inverter is shown driving three gates.

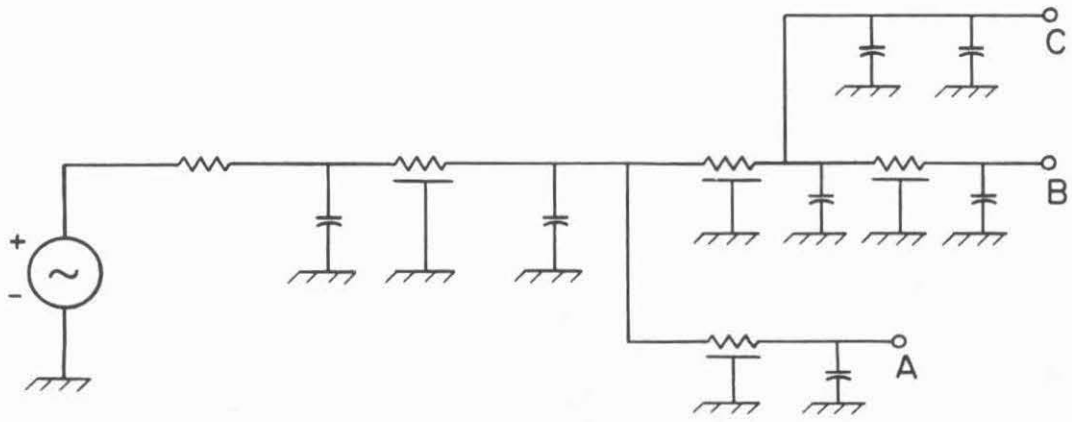

Figure 2. Linear-circuit model for the network of Figure 1. The voltage source is a step at time $t = 0$.

capacitance C_k. For the moment consider only lumped capacitors; the theory is similar if the distributed lines are considered also. One may think of many-stage approximations for the distributed lines, or one may convert some summations in the formulas below to a form including both summations over lumped capacitors and integrals over distributed ones.

The resistance R_{ke} is defined as the resistance of the portion of the (unique) path between the input and e, that is common with the (unique) path between the input and node k. In particular, R_{ee} is the resistance between input and output e, and R_{kk} is the resistance between the input and node k. Thus $R_{ke} \leq R_{kk}$ and $R_{ke} \leq R_{ee}$. For an example, see Figure 3.

The sum (over all the capacitors in the network)

$$T_{De} = \sum_k R_{ke} C_k \tag{1}$$

has the dimensions of time and is in general different for each output. It is equal to the first-order moment of the impulse response, which has been called "delay" by Elmore [2]. This constant alone can be used to generate a lower bound to the step response. Two other time constants and a set of tighter upper and lower bounds using them will be given at the end of this section.

Let $v_k(t)$ and $v_e(t)$ be the voltages at the node k and output e respectively, in response to a unit step excitation. The current $C_k\, dv_k/dt$ that feeds the capacitor C_k contributes to the voltage drop between the input and the output e by the amount $R_{ke} C_k\, dv_k/dt$ as it flows through the resistance R_{ke}. The net voltage drop $1 - v_e(t)$ is obtained by adding the

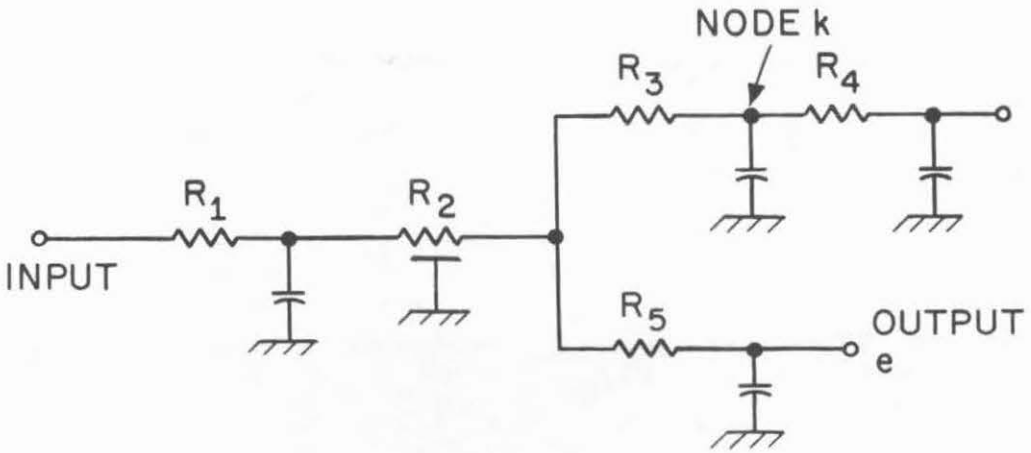

Figure 3. Illustration of resistance terms. For this network, $R_{ke} = R_1 + R_2$, $R_{kk} = R_1 + R_2 + R_3$, and $R_{ee} = R_1 + R_2 + R_5$.

contributions from all the currents in the tree:

$$1 - v_e(t) = \sum_k R_{ke} C_k \frac{dv_k}{dt} . \tag{2}$$

Integration of the right-hand side of (2) from 0 to ∞ yields T_{De}, since the voltages are assumed to rise from 0 at $t = 0$ to 1 when t approaches ∞, everywhere in the tree. Thus T_{De} is equal to the area above the unit step response $v_e(t)$ but below 1, as indicated in Figure 4. Since $v_e(t)$ increases monotonically (a fact proven elsewhere [1]), no rectangle with one corner on $v_e(t)$ and bounded by the lines $t = 0$ and $v_e(t) = 1$ can have an area greater than T_{De} (see Figure 4), i.e.,

$$t[1 - v_e(t)] \leq T_{De}. \tag{3}$$

This expression yields a lower bound for $v_e(t)$,

$$v_e(t) \geq 1 - \frac{T_{De}}{t} . \tag{4}$$

This result illustrates how a suitably defined time constant T_{De} can be used in a bound for the step response. The computation of T_{De} and the bound are much simpler than the exact calculation of the response, especially for RC trees with distributed lines.

More complete and tighter bounds require two additional time constants T_p and T_{Re} to be defined:

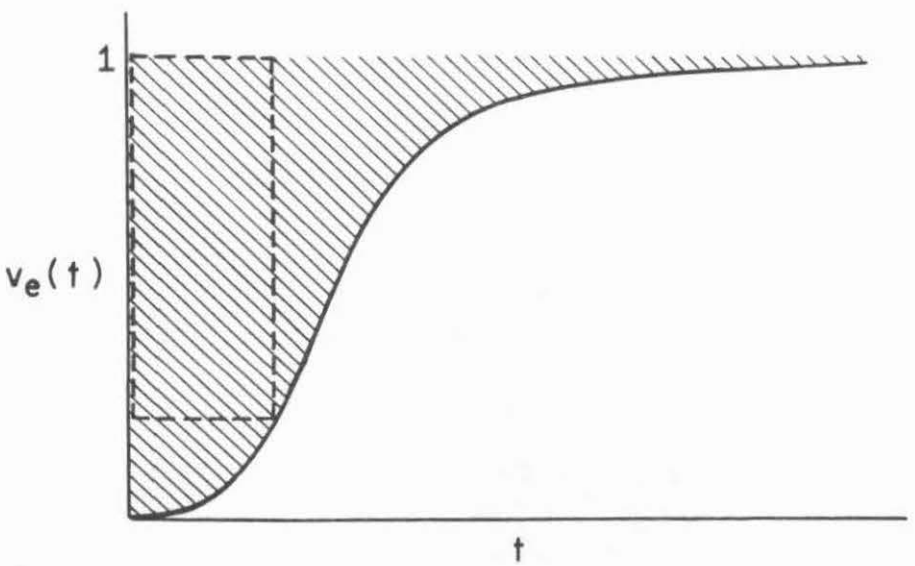

Figure 4. The shaded area is equal to T_{De} and the rectangle has smaller area.

$$T_P = \sum_k R_{kk}C_k \tag{5}$$

$$T_{Re} = \left(\sum_k R_{ke}^2 C_k\right)/R_{ee} \ . \tag{6}$$

Both summations extend over all the capacitors of the network. As with T_{De}, these additional time constants can be computed easily, even in the presence of distributed lines, and while T_{Re} is in general different for different output nodes, T_P is the same for all outputs. It is easily seen that

$$T_{Re} \leq T_{De} \leq T_P. \tag{7}$$

For nonuniform RC lines (i.e., RC trees without side branches) $T_{De} = T_P$. For a single uniform RC line, $T_P = T_{De} = RC/2$, and $T_{Re} = RC/3$. Lower bounds, tighter than (4), and upper bounds can both be derived in terms of these three characteristic times. A detailed derivation [1] leads to the upper bounds

$$v_e(t) \leq 1 - \frac{T_{De} - t}{T_P} \tag{8}$$

$$v_e(t) \leq 1 - \frac{T_{De}}{T_P}\, e^{-t/T_{Re}} \tag{9}$$

and lower bounds

$$v_e(t) \geq 0 \tag{10}$$

$$v_e(t) \geq 1 - \frac{T_{De}}{t + T_{Re}} \tag{11}$$

$$v_e(t) \geq 1 - \frac{T_{De}}{T_P}\, e^{(T_P - T_{Re})/T_P}\, e^{-t/T_P} \tag{12}$$

where (12) applies if $t \geq T_P - T_{Re}$. The tightest upper bounds are (8) for small t and (9) for large t. The tightest lower bounds are (10) for $t \leq T_{De} - T_{Re}$, (11) for $T_{De} - T_{Re} \leq t \leq T_P - T_{Re}$, and (12) for $T_P - T_{Re} \leq t$.

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

$$t \geq 0 \tag{13}$$

and in addition, (8) and (9) can be inverted to yield

$$t \geq T_{De} - T_P[1 - v_e(t)] \tag{14}$$

$$t \geq T_{Re} \ln \frac{T_{De}}{T_P[1 - v_e(t)]} \tag{15}$$

and (11) and (12) yield

$$t \leq \frac{T_{De}}{1 - v_e(t)} - T_{Re} \tag{16}$$

$$t \leq T_P - T_{Re} + T_P \ln \frac{T_{De}}{T_P[1 - v_e(t)]} \tag{17}$$

where (17) only applies if $v_e(t) \geq 1 - T_{De}/T_P$. The general form of all these bounds is illustrated in Figure 5.

These bounds, (8) to (12) for voltage, and (13) to (17) for time, constitute the major result of this paper.

IV. Practical Algorithms

One way to use the inequalities of the previous section is to consider the overall RC tree, and compute for each capacitor the appropriate R_{ke} and

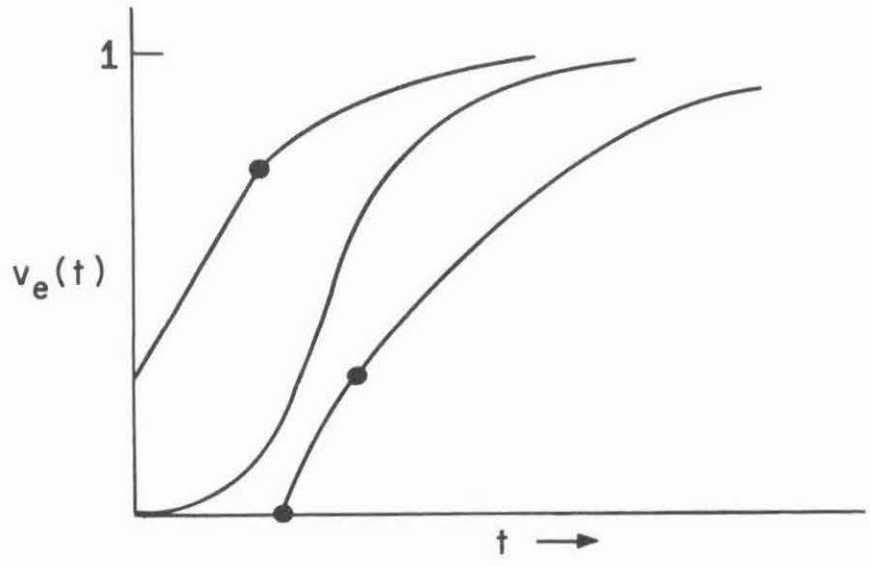

Figure 5. Form of the bounds, with the distances from the exact solution exaggerated for clarity.

R_{kk} so that T_P, T_{De}, and T_{Re} for each output can be found. Of course for distributed lines the sums are replaced by appropriate integrals. In this approach, the calculations necessary for each output require time proportional to the square of the number of elements.

An alternate approach 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_{De}, and T_{Re}. 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.

For simplicity, the tree is assumed to consist of lumped capacitors, lumped resistors, and uniform (not nonuniform) RC lines. Only one output is considered; a more general set of programs is described elsewhere [1]. Only one primitive element, a uniform line (URC) is necessary. If either the resistance of the line or the capacitance is zero, the line reduces to a lumped capacitor or resistor. The line is denoted URC R,C where R,C is a vector of length 2 consisting of the resistance and capacitance of the line, in that order. A capacitor is written URC 0,C and a resistor URC R,0. Figure 6 shows a way of converting a subtree into a side branch and a way of cascading two subtrees. The topology of any RC tree can be denoted by an expression using only these two functions, WB and WC.

Example: The network shown in Figure 7 is a tree with one side branch and may be denoted

```
(URC 15 0) WC (URC 0 2) WC (WB (URC 8 0) WC URC 0 7) WC (URC 3 4)
    WC URC 0 9.                                                    (18)
```

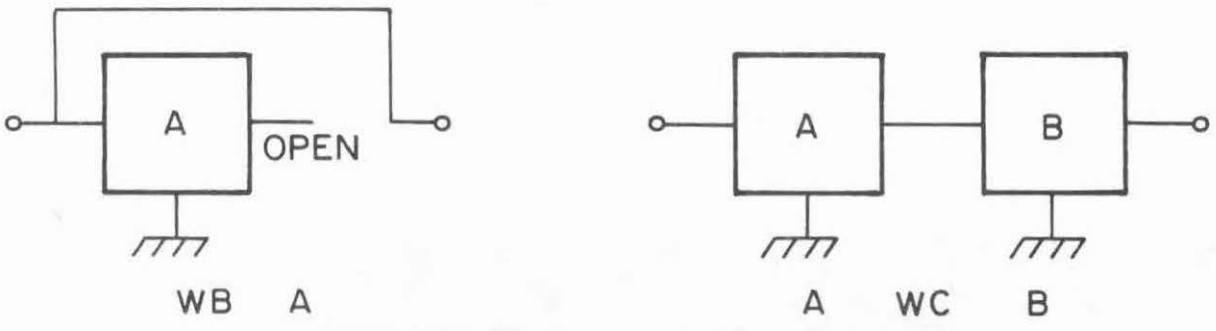

Figure 6. Wiring functions for interconnecting elements or subtrees. Here A and B are previously defined RC trees.

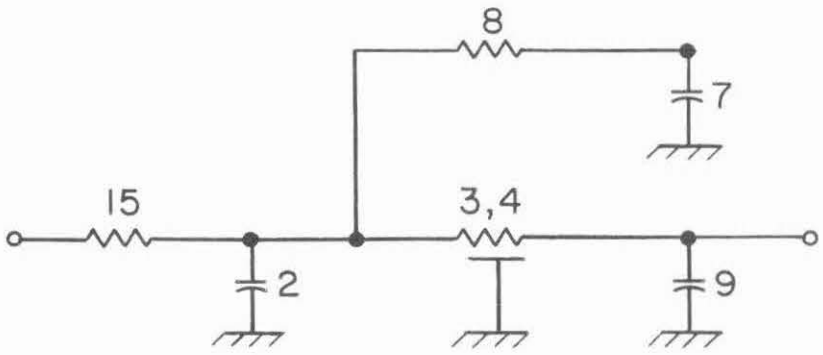

Figure 7. Example network. Parameter values are in ohms and farads.

An expression such as (18) can be used as a guide for the calculations if each function shown corresponds 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) Considering port 2 as the 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}.)

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}, and T_{R2}, of the cascade A WC B from the corresponding quantities for its two arguments, A and B. The formulas for calculating these are

$$C_T = C_{TA} + C_{TB} \tag{19}$$

$$T_P = T_{PA} + T_{PB} + R_{22A}C_{TB} \tag{20}$$

$$R_{22} = R_{22A} + R_{22B} \tag{21}$$

$$T_{D2} = T_{D2A} + T_{D2B} + R_{22A}C_{TB} \tag{22}$$

$$T_{R2}R_{22} = T_{R2A}R_{22A} + T_{R2B}R_{22B} + 2R_{22A}T_{D2B} + R_{22A}^2C_{TB}. \tag{23}$$

The corresponding formulas for WB are even simpler:

$$C_T = C_{TA} \tag{24}$$

$$T_P = T_{PA} \tag{25}$$

$$R_{22} = 0 \tag{26}$$

$$T_{D2} = 0 \tag{27}$$

$$T_{R2}R_{22} = 0. \tag{28}$$

A set of APL functions which implement this approach appear in Figures 8 and 9. The necessary data is passed around in the form of vectors. A two-port network is represented by the vector C_T, T_P, R_{22}, T_{D2}, $T_{R2}R_{22}$. The listing of WC, for example, shows the calculation of the required output, term by term, from the arguments. This function can be compared with (19) to (23).

Figure 9 shows five functions intended to calculate the bounds for any network. 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 array of 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 Figure 7 is shown in Figures 10 and 11.

Because these functions were written for exposition, no protection is included against meaningless values of V or T. In addition, these fail for networks without any resistances or capacitances, and for V = 0 or T = 0.

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

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

```
    ∇ Z←URC X
[1]   Z←X[2],(X[1]×X[2]÷2),X[1],(X[1]×X[2]÷2),X[1]×X[1]×X[2]÷3
    ∇


    ∇ Z←WB A
[1]   Z←A[1 2], 0 0 0
    ∇


    ∇ Z←A WC B
[1]   Z←(A[1]+B[1]),(A[2]+B[2]+A[3]×B[1]),(A[3]+B[3]),A[4]+B[4]+A[3]×B[1]
[2]   Z←Z,A[5]+B[5]+(2×A[3]×B[4])+A[3]×A[3]×B[1]
    ∇
```

Figure 8. APL functions for the element and wiring functions.

```
    ∇ Z←VMIN A
[1]   Z←(T≥A[2]-A[5]÷A[3])×1-(*-(T-A[2]-A[5]÷A[3])÷A[2])×A[4]÷A[2]
[2]   Z←0⌈Z⌈1-A[4]÷T+A[5]÷A[3]
    ∇


    ∇ Z←VMAX A
[1]   Z←(1-(A[4]÷A[2])×*-T÷A[5]÷A[3])⌊(T+A[2]-A[4])÷A[2]
    ∇


    ∇ Z←TMIN A
[1]   Z←-⊛A[2]×(1-V)÷A[4]
[2]   Z←0⌈(Z×A[5]÷A[3])⌈A[4]-A[2]×1-V
    ∇


    ∇ Z←TMAX A
[1]   Z←(A[4]÷1-V)-A[5]÷A[3]
[2]   Z←Z⌊(A[2]-A[5]÷A[3])-0⌊A[2]×⊛A[2]×(1-V)÷A[4]
    ∇


    ∇ Z←OK A
[1]   Z←(T≥TMAX A)-T<TMIN A
    ∇
```

Figure 9. Response functions.

```
⍝ AN EXAMPLE OF THE USE OF THE RC TREE DELAY CALCULATIONS.

BRANCH ← WB (URC 8 0) WC URC 0 7
NET ← (URC 15 0) WC (URC 0 2) WC BRANCH WC (URC 3 4) WC URC 0 9

⍝ NOW THE NETWORK IS DEFINED.

V← 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),[1.5] TMAX NET
 0.1        0         68.167
 0.2       27.8      117.22
 0.3       71.46     173.17
 0.4      123.13     237.76
 0.5      184.23     314.15
 0.6      259.02     407.65
 0.7      355.45     528.18
 0.8      491.34     698.07
 0.9      723.66     988.5

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

T← 20 40 60 80 100 200 300 400 500 1000 2000

T, (VMIN NET),[1.5] VMAX NET
  20          0          0.18138
  40          0.03243    0.22912
  60          0.0814     0.27565
  80          0.12565    0.31761
 100          0.16644    0.35714
 200          0.34342    0.52297
 300          0.48283    0.64603
 400          0.59263    0.73734
 500          0.67913    0.8051
1000          0.90271    0.95615
2000          0.99105    0.99778
```

Figure 10. Example of the use of the fast calculation scheme to find upper and lower bounds on delay and response voltage.

A function which returns a network with N minterms is shown in Figure 12. The results of calculating the delay as a function of the number of minterms are shown in Figure 13. The voltage threshold was taken to be 0.7 times V_{DD}. On this log-log plot the quadratic dependence of delay on number

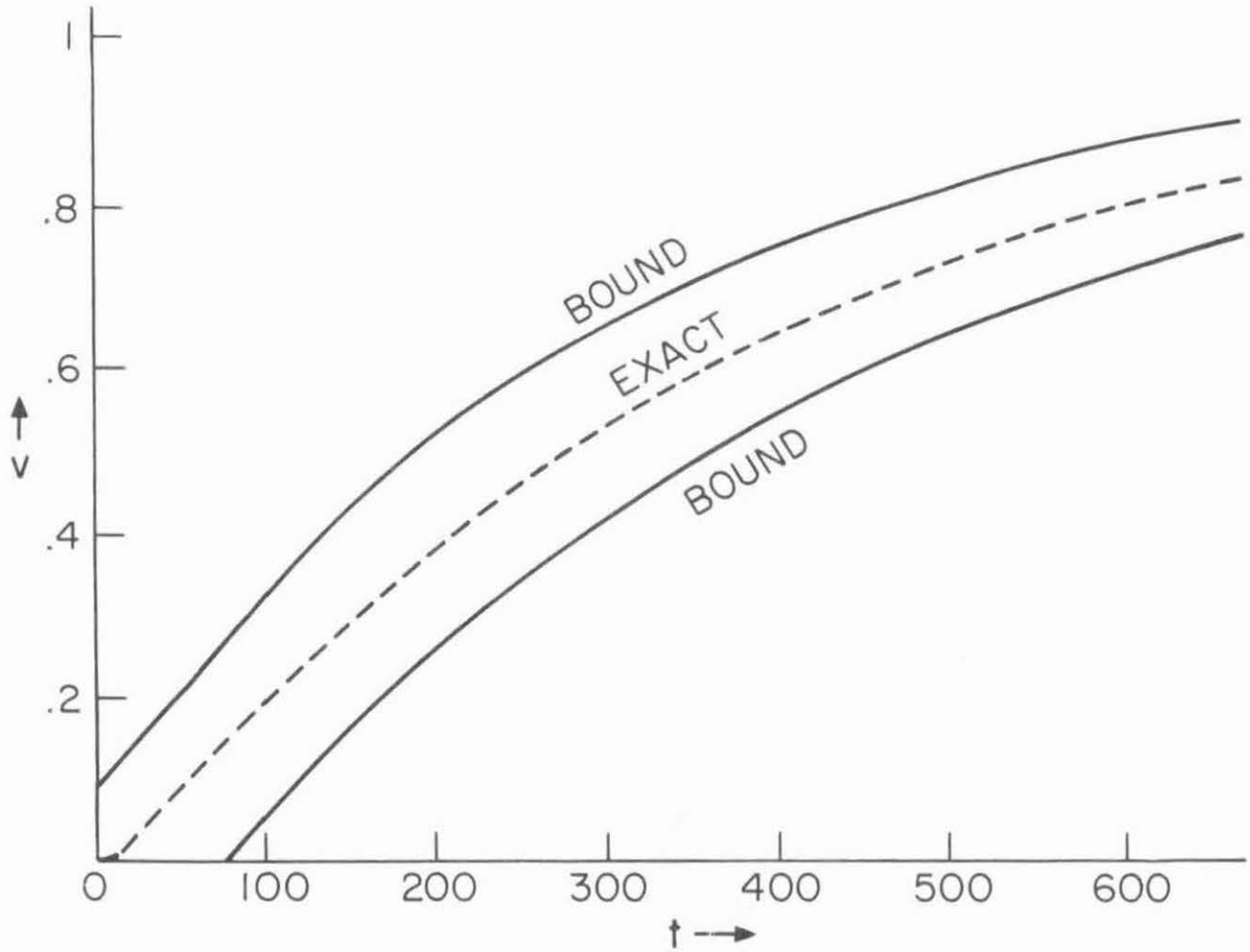

Figure 11. Upper and lower bounds as calculated in Figure 10. The exact solution, found from circuit simulation, is shown also.

```
    ∇ Z←PLALINE N;A
[1]   A←(URC 180 0.0107)WC URC 30 0.0134
[2]   ⍝ A IS A SINGLE SECTION ACCOUNTING FOR TWO MINTERMS
[3]   Z←(URC 378 0)WC URC 0 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
    ∇
```

Figure 12. APL function which returns a model of a PLA line with N minterms.

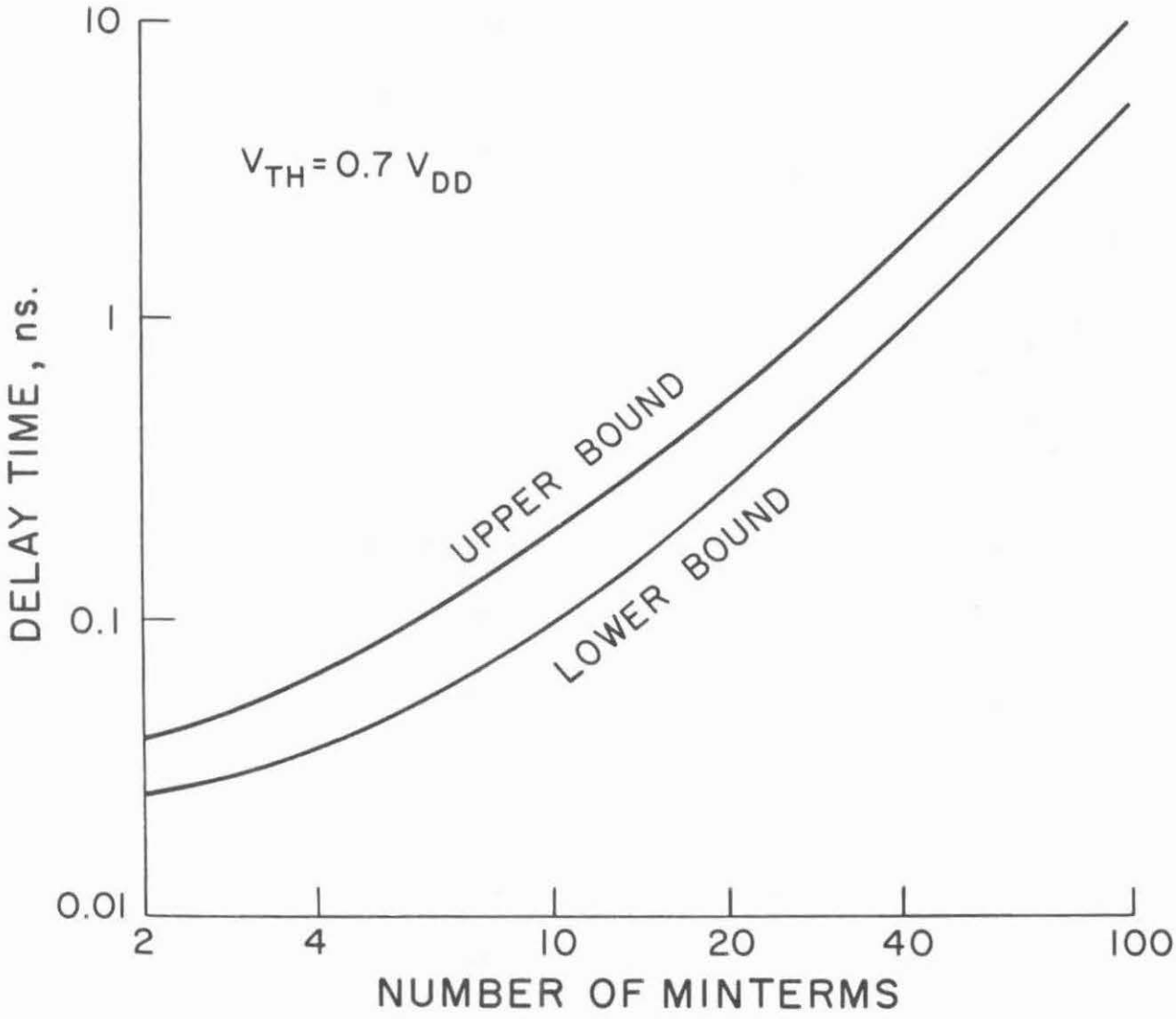

Figure 13. Upper and lower bounds on response time of the network of Figure 12, shown as a function of the number of minterms in the PLA.

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 nsec. 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 by inspection 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 [1].

Extensions of the theory to RC trees with nonlinear elements (similar to the work of Glasser [3] 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.

VII. Acknowledgements

The authors are pleased to acknowledge discussions with Steven Greenberg, Llanda Richardson, and Lance Glasser, and help from Barbara Lory in manuscript preparation.

References

[1] J. Rubinstein and P. Penfield, Jr.; to be published.

[2] W. C. Elmore, "The Transient Response of Damped Linear Networks with Particular Regard to Wide-Band Amplifiers," Journal of Applied Physics, vol. 19, no. 1, pp. 55-63; January, 1948.

[3] L. A. Glasser, "The Analog Behavior of Digital Integrated Circuits," private communication; December, 1980.