AN EXTENSION OF ELMORE'S DELAY AND ITS APPLICATION FOR TIMING ANALYSIS OF MOS PASS TRANSISTOR NETWORKS

Pak K. Chan

February 1986
CSD-860056

An Extension of Elmore's Delay and its Application for Timing Analysis of MOS Pass Transistor Networks

Pak K. Chan

Computer Science Department
Boelter Hall, University of California, Los Angeles
Los Angeles, California 90024
(213)825-2266

Abstract - This work extends the notion of delay defined by Elmore [2] to accommodate the effect of non-unit-step (slow) excitations for MOS pass networks. Following Horowitz [3], each pass transistor in a pass network is modelled as a nonlinear device. A single value of delay for each node in a pass network is then derived by performing a state-space analysis on the network model. This approach provides a compact delay expression in closed-form.

I. A Model for Pass Transistors

We follow the nonlinear pass transistor model proposed in [3]. The quadratic equation which describes the behavior of an MOS transistor in non-saturation is given by

$$i_{ds} = \beta \,(2(V_{gs} - V_T)V_{ds} - V_{ds}^2), \tag{1}$$

where i_{ds} is the current flowing from the drain to source, β is a constant, V_{gs} is the voltage across the gate and source, V_T is the threshold voltage, and V_{ds} is the voltage across the drain and source. If we supply a unit voltage to the gate to turn on the pass transistor, i.e., $V_{gs} = 1 - V_s$, then equation (1) can be written as:

$$i = \frac{(1 - V_T)}{R_{eff}} \,[\, f(V_d) - f(V_s) \,] \tag{2}$$

$$f(V) \equiv 1 - (1 - V)^2 \;\; ; \;\; R_{eff} \equiv \frac{2}{\beta\,(1 - V_T)} ,$$

where R_{eff} is the effective resistance of the transistor. The effective conductance $\frac{1}{R_{eff}}$ is denoted as G_{eff}.

For a network of pass transistors, we apply Kirchhoff's current law to relate the current flow in the pass network, giving us

$$\mathbf{C}\,\dot{\mathbf{x}}(t) = \mathbf{G}\,f\,(\mathbf{v}(t)) + \mathbf{D}\,f\,(\mathbf{e}(t)), \tag{3.a}$$

where $\mathbf{C}$ is the capacitance (diagonal) matrix, $\mathbf{x}(t) \equiv \frac{1}{(1 - V_T)}\mathbf{v}(t)$ is the state vector, $\dot{\mathbf{x}}(t)$ is the time derivative of $\mathbf{x}(t)$, $\mathbf{v}(t)$ is the voltage at each capacitor, and $\mathbf{D}$ is the effective conductance (diagonal) matrix with $D_{ii}(t)$ representing the effective conductance connected at node n_i to excitation $e_i(t)$. $\mathbf{G}$ is the effective node-conductance matrix with components G_{ij}. For $j \neq i$, G_{ij} is the branch effective conductance between nodes n_i and n_j, whereas G_{ii} is the negative sum of all branch effective conductances at node n_i. For $j \neq i$, G_{ij} is equal to G_{ji} by reciprocity. Therefore $-\mathbf{G}$ is a *Stieltjes matrix.*

For a network which has only single *type* of input (this doesn't preclude multiple inputs which have the same excitation function $e(t)$) we can write (3.a) as

$$\mathbf{C}\,\dot{\mathbf{x}}(t) = \mathbf{G}\,f\,(\mathbf{v}(t)) + \mathbf{d}\,f\,(e(t)), \tag{3.b}$$

where d_i is the effective conductance connected at node n_i to the excitation $e(t)$. For the sake of simplicity for presentation, we'll focus on pass networks with a single *type* of excitation source.

If $\mathbf{G}^{-1}\mathbf{d} = -\mathbf{1}$, equation (3.b) can be rewritten as,

$$\mathbf{RC}\,\dot{\mathbf{x}}(t) = \mathbf{1}f\,(e(t)) - f\,(\mathbf{v}(t)), \tag{4}$$

where $\mathbf{R} = -\mathbf{G}^{-1}$. The condition $\mathbf{G}^{-1}\mathbf{d} = -\mathbf{1}$ states a topological property for the class of pass network we are analyzing. This property holds for all tree networks and a restricted class of non-tree networks which have no d.c. path to ground (except the excitation source). This condition states as $\mathbf{G}^{-1}\mathbf{D}\mathbf{1} = -\mathbf{1}$ with respect to equation (3.a). $\mathbf{G}$ is found by inspection. Except for tree networks where $\mathbf{R}$ can be found by inspection, finding $\mathbf{R}$ involves numerical methods to evaluate $\mathbf{G}^{-1}$ [1]. Equation (4) represents the state equation of a nonlinear system. In general, finding an exact solution for this type of nonlinear system is difficult. However, we are only interested in finding a quantity which we call *delay* implicitly described by this equation, as we'll see in the next section.

II. An Extension of Elmore's Delay

Delay is the manifestation of the inertia of a system. One way to quantify delay as suggested by Elmore [2] is to take the first-order moment of the impulse response $h(t)$, commonly known as the inertia, as the delay, i.e.,

$$T_D \equiv \int_0^\infty h(t)t \, dt.$$

In [1], Elmore's definition of delay is extended to be

$$T_D \equiv \int_0^\infty [u(t) - v(t)] \, dt \;, \tag{5}$$

where $v(t)$ is the voltage response due to excitation $e(t)$ which is not necessary a unit-step function. The above quantity has been demonstrated to be a consistent definition of delay for linear *RC* networks.

Nevertheless, the above definition doesn't hold for nonlinear systems such as the one presented in equation (4). For instance, the quantity $\int_0^\infty [u(t) - v(t)] \, dt$ doesn't converge even for the simplest pass network when excited by $u(t)$.

For the class of pass network we are investigating, we define *delay* as,

$$T_D \equiv \int_0^\infty [\, u(t) - f(v(t)) \,] \, dt \;.$$

To extend the above definition of delay for each node n_i of a pass network we attach a subscript i to the above definition and obtain

$$T_{Di} \equiv \int_0^\infty [\, u(t) - f(v_i(t)) \,] \, dt \;.$$

Namely, delay is defined to be the difference in areas covered by $u(t)$ and $f(v_i(t))$ along time. A convenient way to express delay for all nodes in the network is to use the vector notation

$$\mathrm{T}_D \equiv \int_0^\infty [\mathrm{u}(t) - f(\mathrm{v}(t))] \, dt. \tag{6}$$

By taking the Laplace transform on both sides of (3) and taking their limits as s approaches zero, we obtain

$$T_D = \lim_{s \to 0} [\, \mathbf{u}(s) - \mathbf{V}(s) \,], \tag{7}$$

where $\mathbf{u}(s)$ and $\mathbf{V}(s)$ are the Laplace transform of $\mathrm{u}(t)$ and $f(\mathrm{v}(t))$ respectively.

III. Calculating Delays

To measure the delay of a node in a pass network, it suffices to consider the normalized case where the node voltage starts from some initial value $v_i(0)$ between 0 and 1, and is driven towards some final value (maximum 1). The results obtained in this normalized case are easily adapted to both charging and discharging processes (final value 0), and to any values of supply voltage.

The first step in calculating delays is to apply the Laplace transform on both sides of (4), giving us

$$RC\, [\, s\,\mathbf{X}(s) - \mathrm{x}(0) \,] = \mathbf{1}\mathbf{e}(s) - \mathbf{V}(s), \tag{8}$$

where $\mathbf{X}(s)$ is the Laplace transform of $\mathrm{x}(t)$, $\mathrm{x}(0)$ are the initial voltages on the capacitors, and $\mathbf{e}(s)$ is the Laplace transform of $f(e(t))$. Consider the specific case where $e(t) = u(t)$, i.e., $\mathbf{e}(s) = \mathbf{1}\frac{1}{s}$. Taking the limit of (8) as s approaches zero gives us the following delay expression,

$$T_d = \lim_{s \to 0} RC\, [\, s\,\mathbf{X}(s) - \mathrm{x}(0) \,],$$

which reduces to

$$T_d = RC\, [\, \mathrm{x}(\infty) - \mathrm{x}(0) \,], \tag{9}$$

by the final value theorem.

For slow input, the delay expression is given by

$$T_d = RC\, [\, \mathrm{x}(\infty) - \mathrm{x}(0) \,] + \int_0^\infty [\, \mathrm{u}(t) - \mathbf{1}e(t) \,]^2\, dt, \tag{10.a}$$

and for a pass network with multiple *type* of excitation sources, the delay expression can be shown to be

$$T_d = RC\, [\, \mathrm{x}(\infty) - \mathrm{x}(0) \,] + RD \int_0^\infty [\, \mathrm{u}(t) - \mathrm{e}(t) \,]^2\, dt. \tag{10.b}$$

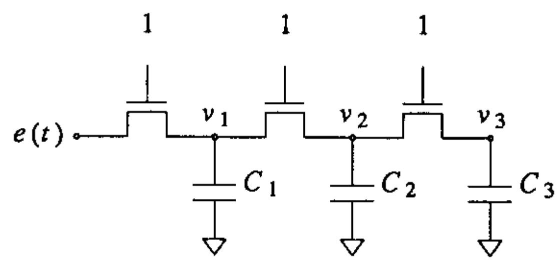
The first term is the *intrinsic delay* because it accounts for the delay due to the intrinsic characteristics (circuit topology, initial and final conditions) of the pass network. The second term is the *extrinsic delay* since it accounts for the difference between $\mathbf{u}(t)$ and $\mathbf{e}(t)$ along time. Of course, expression (10.b) is meaningful only if the integral $\int_0^\infty [\mathrm{u}(t) - \mathrm{e}(t)]^2\, dt$ converges for excitations $\mathbf{e}(t)$. The derivation of expression (10) is sufficiently general to handle pass networks which obey $\mathbf{G}^{-1}\mathbf{D1} = -\mathbf{1}$, arbitrary input, and arbitrary distribution of initial charges.

In particular, if there is only one type of excitation source which is equal to $\frac{t}{t+a}$, then (10) is reduced to

$$\mathrm{T}_d = \mathbf{RC}\,[\,\mathbf{x}(\infty) - \mathbf{x}(0)\,] + a\,\mathbf{1}. \qquad (11)$$

IV. Illustrated Example

Consider the following pass network with three pass transistors.



The dynamics of the system is described by

$$\begin{bmatrix} C_1 & 0 & 0 \\ 0 & C_2 & 0 \\ 0 & 0 & C_3 \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} -(G_1+G_2) & G_2 & 0 \\ G_2 & -(G_2+G_3) & G_3 \\ 0 & G_3 & -G_3 \end{bmatrix} \begin{bmatrix} f(v_1(t)) \\ f(v_2(t)) \\ f(v_3(t)) \end{bmatrix} + \begin{bmatrix} G_1 \\ 0 \\ 0 \end{bmatrix} f(e(t)),$$

or

$$\begin{bmatrix} R_1 & R_1 & R_1 \\ R_1 & R_1+R_2 & R_1+R_2 \\ R_1 & R_1+R_2 & R_1+R_2+R_3 \end{bmatrix} \begin{bmatrix} C_1 & 0 & 0 \\ 0 & C_2 & 0 \\ 0 & 0 & C_3 \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} f(e(t)) - \begin{bmatrix} f(v_1(t)) \\ f(v_2(t)) \\ f(v_3(t)) \end{bmatrix}.$$

For $e(t) = u(t)$, that gives us

$$\begin{bmatrix} T_{D1} \\ T_{D2} \\ T_{D3} \end{bmatrix} = \begin{bmatrix} R_1 & R_1 & R_1 \\ R_1 & R_1+R_2 & R_1+R_2 \\ R_1 & R_1+R_2 & R_1+R_2+R_3 \end{bmatrix} \begin{bmatrix} C_1 & 0 & 0 \\ 0 & C_2 & 0 \\ 0 & 0 & C_3 \end{bmatrix} \begin{bmatrix} x_1(\infty) - x_1(0) \\ x_2(\infty) - x_2(0) \\ x_3(\infty) - x_3(0) \end{bmatrix} .$$

V. Remarks and Conclusions

Modelling a pass transistor by a linear resistor can be conceived as the degenerated case in which the mapping $f(V) = 1 - (1 - V)^2$ becomes $f(V) \approx V$. Therefore, defining delay as the area bounded by $u(t)$ and $f(v(t))$ gives a unified notion of delay relating the linear resistor model and the nonlinear quadratic model for MOS transistors.

Equation (12) is the delay expression that we obtain with the linear resistor model.

$$\mathrm{T}_d = \mathrm{RC}\ [1 - \mathrm{x}(0)] + \lim_{s \to 0} [\mathrm{u}(s) - 1e(s)]. \tag{12}$$

Comparing (12) with (10.a), we see that they coincide when a network is excited by unit-step functions.

To conclude, we have extended Elmore's definition of delay to include a number of effects such as slow excitations which were not previously considered by others. The new definition not only accounts for both intrinsic and extrinsic delays, it also justifies the use of the linear resistor model in timing analysis.

Acknowledgment

This work has been supported in part by MICRO under Contract 4-442514-57045, and in part by the Office of Naval Research under Contract 4-482510-25801.

REFERENCES

[1] P. K. Chan, ''An Extension of Elmore's Delay,'' Computer Science Department, University of California, Los Angeles, Tech. Rep. CSD-860050, Jan. 1986.

[2] W. C. Elmore, ''The Transient Response of damped Linear Networks with particular regard to Wideband Amplifiers,'' *J. Appl. Phys.*, Vol. 19, No. 1, Jan. 1948, pp. 55-63.

[3] M. A. Horowitz, ''Timing Models for MOS Pass Networks,'' in *Proceeding on Int. Symposium Circuits and Systems*, 1983, pp. 198-201.