Modeling the "Effective Capacitance" for the RC Interconnect of CMOS Gates

Jessica Qian, Satyamurthy Pullela, and Lawrence Pillage, *Member, IEEE*

Abstract—**With finer line widths and faster switching speeds, the resistance of on-chip metal interconnect is having a dominant impact on the timing behavior of logic gates. Specifically, the gates are switching faster and the interconnect delays are getting longer due to scaling. This results in a trend in which the RC interconnect delay is beginning to comprise a larger portion of the overall logic stage delay. This shift in relative delay dominance from the gate to the RC interconnect is increased by resistance shielding. That is, as the gate "resistance" gets smaller and the metal resistance gets larger, the gate no longer "sees" the total net capacitance and the gate delay may be significantly less than expected. This trend complicates the timing analysis of digital circuits, which relies upon simple, empirical gate delay equations for efficiency. In this paper, we develop an analytical expression for the "effective load capacitance" of an RC interconnect. In addition, when there is significant shielding, the response waveforms at the gate output may have a large exponential tail. We show that this waveform tail can strongly influence the delay of the RC interconnect. Therefore, we propose an extension of the effective capacitance equation that captures the complete waveform response accurately, with a two-piece gate-output-waveform approximation.**

I. Introduction

As integrated circuit technologies continue to improve, the feature sizes of transistors and interconnect wiring are getting smaller, thus allowing for denser chips with increased functionality. With this trend, metal interconnect resistance-per-unit-length is scaling up due to R scaling up and C not scaling down due to fringing fields. If the interconnect lengths shortened according to the scale factor, then the total metal resistance would remain relatively constant with scaling. However, since the integrated circuit densities are increasing dramatically, the average metal lengths are not scaling down as the feature sizes are reduced. This situation is evidenced by the increase in the RC-interconnect-delay portion of the overall logic-stage delay.

In addition to the gate delay getting faster while the RC delay gets longer due to scaling, the interconnect resistance also acts to reduce the gate delay due to resistance shielding. That is, as the gate's resistance becomes comparable to the total metal resistance, some of the load capacitance is shielded from the gate. This shielding tends to decrease the gate delay since the gate is driving an "effective load" that is less than the total capacitance of the net. In [1], an "effective capacitance" model is proposed that accounts for the reduction in the gate delay due to the metal shielding component but maintains compatibility with the popular (empirically derived) k-factor gate-delay equations. This effective capacitance model results in gate delay predictions that are within ±10% of a SPICE prediction. Since optimistic delay predictions are generally unacceptable, in this paper we derive an effective capacitance expression similar to the one proposed in [1] that provides a more pessimistic delay approximation.

One difficulty with the model in [1] is that while it predicts the gate delay with reasonable accuracy, it is not able to predict the transition time (rise/fall time) of the gate-output signal. Moreover, when the metal resistance is significant, the digital waveforms at the gate output begin to take on a non-digital character. That is, it is very difficult to even specify the signal transition time since the RC shielding effects give the signal a strong nonlinear character. To overcome this limitation, the effective capacitance model is augmented by a gate resistance model to approximate a complete gate-output waveform. This paper will show that given the RC interconnect parameters and the k-factor equations for the gate delay and gate-output transition time as a function of load capacitance, the gate delay and the gate output waveform can be calculated with reasonable accuracy even when there is significant resistance-shielding.

II. Background

It is recognized that the overall logic-stage delay consists of a gate delay component and an RC interconnect delay component [2]. There are two approaches to capturing the combined delay of both the gate and the interconnect which have gained considerable acceptance: 1) a switch-resistor model comprised of a linear resistor and a step function of voltage [3]–[6] and 2) empirically derived expressions for delay and output-signal transition time as a function of load capacitance and input-signal transition time (k-factor equations) [7]. Both methods are empirically based, since even the second method requires empirical fitting to approximate the resistance value as a function of input transition time and output load.

Switch resistor models have an advantage since their coupling with the RC interconnect is inherently modeled. However, it is extremely difficult to model a submicron gate in terms of a single resistor, and for this reason empirical gate

Manuscript received June 22, 1993; revised August 4, 1994. This work was supported in part by IBM Corporation, the National Science Foundation under contract MIP-9157263, and the Semiconductor Research Corporation under contract 94-DJ-343. This paper was recommended by Associate Editor R. A. Saleh.

S. Pullela and L. Pillage are with the Department of Electrical and Computer Engineering,University of Texas at Austin, Austin, TX 78712 USA.

J. Qian is with LSI Logic Corporation, Milpitas, CA USA.

IEEE Log Number 9406216.

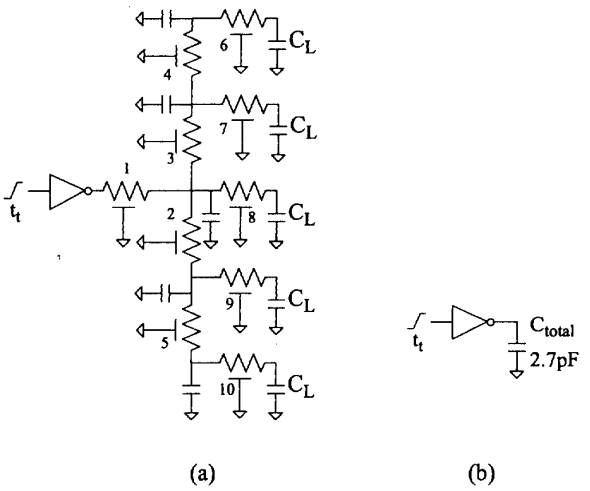

Fig. 1. (a) An inverter driving an RC interconnect with 5 loads and 10 segments of metal (URC's). (b) The same inverter driving the total capacitance of the net in (a).

delay equations are the model of choice for cell-based design styles.

With empirical gate delay models, we must decouple the gate and interconnect delay analyses, analyze them individually, and then sum the resulting delays to approximate the overall logic-stage delay. In order to decouple the interconnect and gate-delay problems, we must have a model for the loading on the gate due to the RC interconnect and fan-out gates.

For example, consider the logic-stage delay problem shown in Fig. 1(a). The fan-out for this net is 5, with each fan-out input characterized by a 10-fF load capacitance. There are 10 segments of metal, each represented by a uniform RC segment (URC) in the figure. The total R and C values for these URC segments (based upon per unit length R's and C's, and the corresponding length for each URC), are shown in Table I. There is also a capacitance associated with the interconnect vias that is modeled by a 5-fF capacitance.

For delay analysis, we are given the rise time of the inverter input-signal and we want to calculate the falling transition delays at all of the loads. For efficiency, we analyze the falling-transition gate delay (the delay of the signal just at the output of the inverter) and the RC interconnect delay (to the various fan-out points) separately, then add them together for an overall logic-stage delay. To enable such an analysis we must approximate the *driving point admittance* of the RC load. That is, we can accurately model the waveform at the output of the inverter only if we accurately approximate the load "seen by" the gate. The simplest driving point admittance load for an RC interconnect is the total capacitance of the net, as shown in Fig. 1(b). For this example, the results in Fig. 2 demonstrate that the total capacitance is a reasonably accurate model of the driving point admittance for this net.

Since we can use such a simple model for the driving point admittance, the gate's delay can be pre-characterized in terms of the input-signal transition time, t_t, and the total load capacitance, C_L. That is, for efficiency, the gate delay for a falling output transition, t_d, and the waveform fall time, t_f,

TABLE I
URC VALUES FOR THE RC INTERCONNECT IN FIG. 1(a)

URC#	R (Ω)	C (fF)
1	1.20	11.1
2	1.33	365
3	3.55	223
4	30.6	275
5	1.20	11.1
6	50.7	224
7	50.5	432
8	64.8	583
9	70.6	194
10	80.4	306

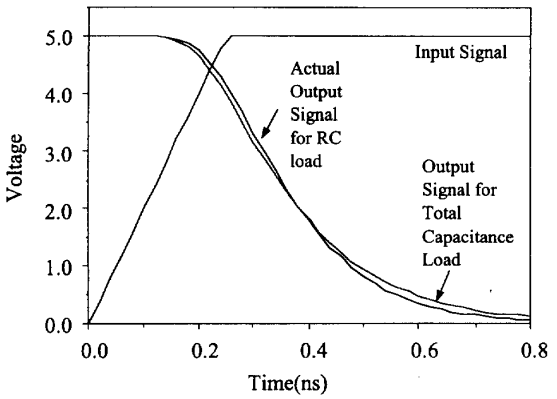

Fig. 2. The total capacitance is a fairly accurate model of the driving point admittance for this example.

can be characterized in terms of empirically derived k-factor equations such as [7]:

$$t_d = (k_1 + k_2 C_L)t_t + k_3 C_L^3 + k_4 C_L + k_5 \quad (2.1)$$

$$t_f = (k_1' + k_2' C_L)t_t + k_3' C_L^2 + k_4' C_L + k_5' \quad (2.2)$$

where the k's are empirical fitting parameters and the falling output delay (50% point delay), t_d, and the fall time, t_f, are defined as shown in Fig. 3. Notice that we have defined t_f to be the 100% to 0% (or 0 to 100% for t_r) time determined by fitting a straight line through the 20 and 80% points of the output voltage waveform. It is apparent from Fig. 3 that such an approximation is reasonable for this waveform.

For timing analysis, once the delay and the output transition time are efficiently evaluated using (2.1) and (2.2), the interconnect delay is calculated by driving the RC circuit model with a voltage defined by a transition time, t_f. The RC interconnect can be efficiently precharacterized [8] in terms of a reduced-order transfer function. In this case, Asymptotic Waveform Evaluation (AWE) [9] is used to determine the reduced order model. The overall logic stage delay is the gate delay, t_d, plus the RC delay, t_{RC}.

A delay analysis procedure such as the one outlined above can be extremely efficient and accurate for delay analysis of digital gates and interconnect. However, as the logic gates switch faster with lower "resistance," and as the interconnect resistance per-unit-length increases with technology advances,

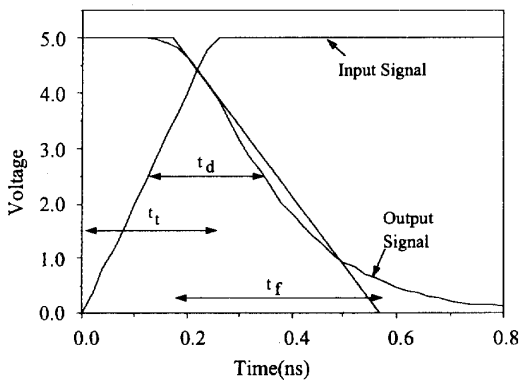

Fig. 3. Defining the transition time and delay for an output signal waveform.

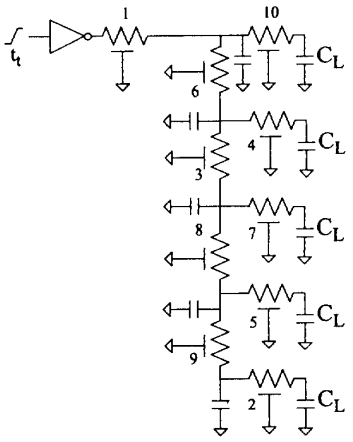

Fig. 4. The inverter example from Fig. 1 with a different interconnect topology.

some of the interconnect capacitance can be shielded from the logic gate. That is, a total capacitance model is guaranteed to be a pessimistic approximation for the driving point admittance. In some cases, the delay prediction can be extremely pessimistic.

As an example, consider once again the inverter in Fig. 1, still driving five load capacitors and 10 URC's of metal. If we change the interconnect topology of this net, as shown in Fig. 4, the total capacitance is unchanged, however the "effective capacitance" is changed significantly. When we compare the signal waveform for the output of the inverter in Fig. 4 with that for the inverter in Fig. 1(b), the delay and transition time are dramatically different, as shown in Fig. 5. Notice that the actual waveform has a smaller delay due to the metal resistance shielding some of the load capacitance. Also, it is apparent that the waveshape does not appear as "digital" as the total capacitance waveshape, due to the RC effects.

To better capture the delay and the output signal transition time for the circuit in Fig. 4 requires a better approximation for the driving point admittance of the RC interconnect. One can synthesize higher-order equivalent circuit models for the driving point admittance of the RC interconnect. The procedure for doing this using AWE is described in [10].

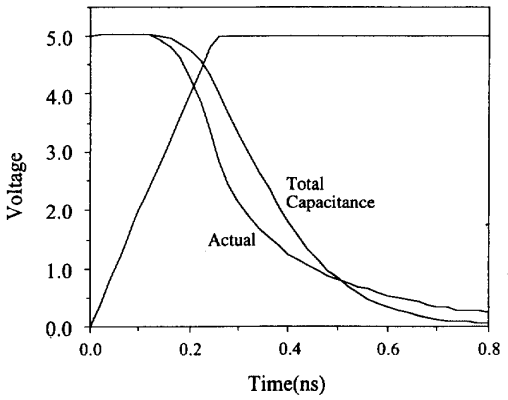

Fig. 5. The inverter-output response waveform for the circuit in Fig. 4.

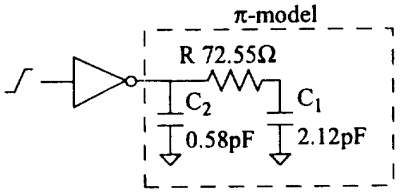

Fig. 6. A π-model of the driving point admittance for the RC interconnect in Fig. 4.

A π-model load is easily synthesized using the first three moments of the driving point admittance [11], [10]. The π-model synthesis for the RC interconnect in Fig. 4 is shown in Fig. 6. Note that the sum of capacitors C_1 and C_2 is equal the total capacitance of this net [11]. The resistance R represents the amount of capacitance shielding. When the π-model resistance R is comparable to the "resistance" of the switching inverter, we can expect significant shielding and a waveshape with an RC tail.

Comparing the actual response waveform to that obtained using the π-model load for this example shows that the waveforms are nearly identical (Fig. 7). The only difficulty with this approach is that the π-model load is incompatible with the k-factor Equations (2.1) and (2.2). In order to generate empirical k-factor equations for all possible input transition times and π-model loads would require fitting a 4D table of empirical data. Generating such models does not seem practical from a storage or run-time point of view.

Instead of a 4D table, an "effective capacitance" model was proposed in [1] that would maintain compatibility with the k-factor equations while also modeling the resistance shielding effect. In [1], an effective capacitance could be generated that captures the delay to within $\pm 10\%$ of the actual delay. For the example in Fig. 4, an effective capacitance value was obtained (as shown in Fig. 8) that adequately approximated the loading behavior of the π-model. The results are shown in Fig. 9.

The effective capacitance from [1] is reasonably accurate, as shown by the example in Fig. 9. However, it is occasionally optimistic, which is not always acceptable for applications such as worst-case timing analysis. It is also apparent from the waveforms in Fig. 9 that while a single capacitance value

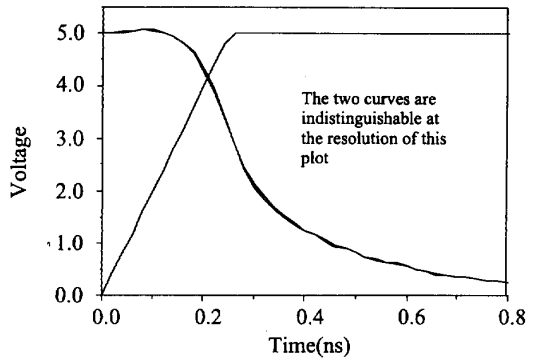

Fig. 7. The excellent agreement between the actual response and the approximate responses using the π-model for the driving point admittance.

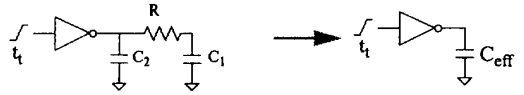

Fig. 8. An effective capacitance that captures the effects of the π-model load.

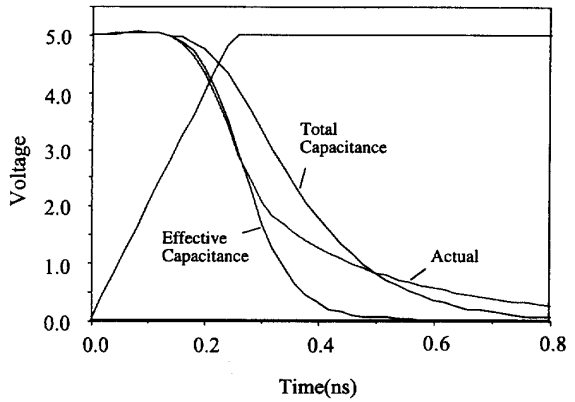

Fig. 9. A comparison of the actual waveform with the results when using a total capacitance and an effective capacitance to model the driving point admittance.

seems to result in a delay close to that obtained using the π-model load, it is impossible to capture both the delay and the output waveshape with a single capacitance value. In Sections III and IV we consider these limitations and propose models which capture the effective delay as well as the RC waveform tail.

III. The "Effective" Capacitance

To establish an expression for an effective capacitance that considers the resistance shielding of the interconnect, we attempt to find a single capacitor that will result in the same 50% point delay as a π-model load. The approach taken in [1] is to determine the capacitance load that has the same average current (therefore the same total charge transfer) as the π-model load. Referring to Fig. 10, we equate the average currents for the waveforms of $V_{\text{out}}(t)$ up to the 50% delay

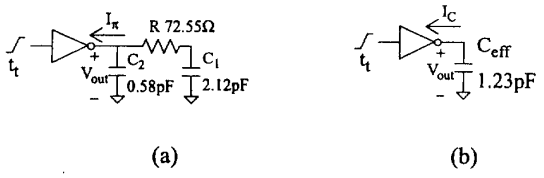

Fig. 10. Equating the average currents for the (a) $-pi$-model load and (b) the effective capacitance load.

point, t_D.

$$\frac{1}{t_D}\int_0^{t_D} I_\pi(t)dt = \frac{1}{t_D}\int_0^{t_D} I_C(t)dt \tag{3.1}$$

where

$$I_\pi(s) = Y_\pi(s)V_{\text{out}}(s) \tag{3.2}$$

and

$$I_C(s) = sC_{\text{eff}}V_{\text{out}}(s) \tag{3.3}$$

and $Y_\pi(s)$ is the admittance of the π-model. Note that we make a distinction between the time at which the 50% point is reached, t_D, and the 50% point delay value from (2.1), t_d, since the latter represents the time difference between the 50% points of the input and output waveforms.

Since the delay point, t_D, is what we seek, we assume a waveshape for $V_{\text{out}}(t)$ and equate the integrals in (3.1) using (3.2) and (3.3). Almost any waveshape can be used to equate this averaging of currents, however the more realistic the waveshape assumption, the more accurate the delay approximation. In [12], CMOS gates are modeled using a combination of quadratic and linear functions that correspond to the operating regions of the MOSFET's. Following this reasoning, we use the following waveshape assumption:

$$V_{\text{out}}(t) = \begin{cases} V_i - ct^2 & 0 \le t \le t_x \\ a + b(t - t_x) & t_x \le t \le t_D \end{cases} \tag{3.4}$$

Starting at an initial voltage, V_i, the waveshape is quadratic to the 20% point, t_x. (V_i is equal to V_{DD} for a falling waveform, and equal to zero for a rising waveform.) Then, from the 20% point to the midpoint, the transistors are in saturation and the voltage is assumed to be linear up to the 50% point, t_D. The constants, s, a, b, and c are determined by the factors we must solve for in order to determine the delay. One simplifying assumption is that the voltage waveform and its first derivative are continuous at t_x, therefore,

$$\begin{aligned} a &= V_i - ct_x^2 \\ b &= -2ct_x \end{aligned} \tag{3.5}$$

Using this waveshape assumption, the average current in the capacitance load C_{eff} is

$$\begin{aligned} \overline{I}_C(t) &= \frac{1}{t_D}\left[\int_0^{t_x} C_{\text{eff}} \cdot (-2ct)dt \right. \\ &\quad \left. + \int_{tx}^{t_D} C_{\text{eff}} \cdot (-2ct_x)dt\right] \\ &= \frac{-2C_{\text{eff}} \cdot c \cdot t_x}{t_D}\left[t_D - \frac{t_x}{2}\right] \end{aligned} \tag{3.6}$$

Similarly, the average current in capacitor C_2 of the π-model is given by

$$\bar{I}_{C_2}(t) = \frac{-2C_2 \cdot c \cdot t_x}{t_D}\left[t_D - \frac{t_x}{2}\right] \tag{3.7}$$

Estimating the average current in capacitor C_1 of the π-model is not quite as simple. What we must consider is the current in capacitor C_1 as a function of the voltage waveshape at C_2. First, the average current in C_1 due to the quadratic voltage waveform at C_2, can be shown to be

$$\bar{I}_{C_1}(t) = \frac{-2C_1 \cdot c}{t_x}\left[\frac{t_x^2}{2} - \mathrm{RC}_1 t_x + (\mathrm{RC}_1)^2\left(1 - e^{\frac{-t_x}{\mathrm{RC}_1}}\right)\right] \tag{3.8}$$

Next, we consider the average current in C_1 due to a linear voltage at C_2 from t_x to t_D, which is given by

$$\bar{I}_{C_1}(t) = \frac{a - V_{C1i}}{t_D - t_x} C_1\left(1 - e^{\frac{-(t_D - t_x)}{\mathrm{RC}_1}}\right) + bC_1\left[1 - \frac{\mathrm{RC}_1}{t_D - t_x}\left(1 - e^{\frac{-(t_D - t_x)}{\mathrm{RC}_1}}\right)\right] \tag{3.9}$$

Note that we have also considered that there is an initial voltage on capacitor C_1 at time t_x, V_{C1i}, which is different than the initial voltage at C_2. We approximate this initial voltage by integrating the average C_1 current from $t = 0$ to t_x:

$$V_{C1i} = V_i - c\left[t_x^2 - 2\mathrm{RC}_1 t_x + 2(\mathrm{RC}_1)^2\left(1 - e^{\frac{-t_x}{\mathrm{RC}_1}}\right)\right] \tag{3.10}$$

Using (3.5) and (3.10), (3.9) evaluates to

$$\bar{I}_{C_1}(t) = \frac{-C_1 c}{t_D - t_x}\left[2\mathrm{RC}_1 t_x - 2(\mathrm{RC}_1)^2\left(1 - e^{\frac{-t_x}{\mathrm{RC}_1}}\right)\right] \times \left(1 - e^{\frac{-(t_D - t_x)}{\mathrm{RC}_1}}\right) - 2ct_x C_1\left[1 - \frac{\mathrm{RC}_1}{t_D - t_x}\left(1 - e^{\frac{-(t_D - t_x)}{\mathrm{RC}_1}}\right)\right] \tag{3.11}$$

Then, from (3.8) and (3.11), the total average current in C_1 for the interval $(0, t_D)$ is

$$\bar{I}_{C_1}(t) = \frac{-2cC_1}{t_D}\left[\frac{t_x^2}{2} + t_x(t_D - t_x - \mathrm{RC}_1) + (\mathrm{RC}_1)^2 \times \left(e^{\frac{-(t_D - t_x)}{\mathrm{RC}_1}} - e^{\frac{-t_D}{\mathrm{RC}_1}}\right)\right] \tag{3.12}$$

Finally, we equate the average currents in the π-model ((3.7) and (3.12)) to those in the C_{eff} model ((3.6)) and solve for the "effective capacitance value":

$$C_{\text{eff}} = C_2 + C_1\left[1 - \frac{\mathrm{RC}_1}{t_D - \frac{t_x}{2}} + \frac{(\mathrm{RC}_1)^2}{t_x\left(t_D - \frac{t_x}{2}\right)} e^{\frac{-(t_D - t_x)}{\mathrm{RC}_1}}\left(1 - r^{\frac{-t_x}{\mathrm{RC}_1}}\right)\right] \tag{3.13}$$

As expected, the effective capacitance value lies between C_2 (the first capacitance of the π-model) and $C_1 + C_2$ (the total capacitance) and is determined by the values of R, t_x, and t_D. As expected, in the limit as R goes to zero in (3.13), the effective capacitance is equal to the total capacitance. And, in the limit as R tends toward infinity, the effective capacitance is C_2.

The values t_x and t_D are determined in part by the characteristics of the driving gate and the input signal to the gate. That is, we could calculate these quantities from the π-model parameters, the input-signal slope information, and the k-factor Equations (2.1) and (2.2):

$$t_D = t_d + \frac{t_t}{2} \tag{3.14}$$

$$t_x = t_d + \frac{t_t}{2} - 0.5 \cdot t_f \tag{3.15}$$

From (3.14) and (3.15), we can see that the effective capacitance is a function of the delay and the fall (or rise) time of the output voltage waveform. Of course, the delay and the fall time of the waveform are not known *a priori*, since they are the values that we seek. Therefore, the effective capacitance must be calculated iteratively using the k-factors and (3.13), (3.14), and (3.15).

The iteration procedure is as follows:

1) Set the load capacitance value equal to the total capacitance.
2) Use the load capacitance value to obtain a delay and an output-signal transition time using (2.1) and (2.2).
3) Using the t_d and t_f obtained in step 2, calculate a C_{eff} using (3.13), (3.14), and (3.15).
4) If the value of C_{eff} is still changing, set the load capacitance value equal to C_{eff} and go to step 2.

This iteration procedure can be used with k-factor equations to calculate the effective capacitance. k-factor equations like the ones shown in Fig. 11 are usually generated using regression techniques to fit the data from thousands of SPICE runs to such equations as (2.1) and (2.2) (for various capacitance loads and input transition times). To avoid the error associated with the k-factor fit, we instead used SPICE to extract the fall time and the delay at each effective capacitance iteration in order to emulate "perfect" k-factor Equations ((2.1) and (2.2)). Therefore, the differences in the waveforms in Fig. 9, which compares the effective capacitance in Fig. 10(b) with the π-model in Fig. 10(a), is due only to the effective capacitance approximation, and there is no error attributable to the k-factor equations. In practice, however, the k-factor errors can influence this approximation depending on the regression fit error.

For the example in Fig. 10, the effective capacitance converged to a value of 1.23 pF in 3 iterations. Empirically, we have found that the effective capacitance value converges in 3 to 4 iterations. The speed of convergence and the guarantee of convergence (a unique solution for C_{eff}) are most easily explained using Fig. 11. Referring to the figure, (3.13) is shown plotted as a function of t_d for the three iterations of the aforementioned example. Note that each iteration (each curve) corresponds to a different delay, and therefore, a different value of t_f. Shown on the plot is the k-factor Equation (2.1) for the 50× inverter used in the previous

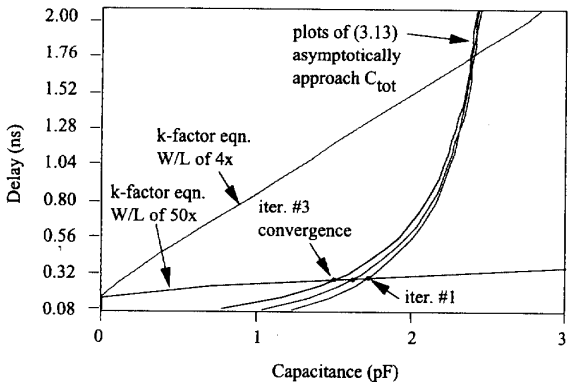

Fig. 11. Plots of the effective capacitance value as a function of delay for three iterations (three different values of fall time) of the example. Also shown on the plot are two k-factor equations for the delay ((Equation 2.1)) of a 50× and a 4× inverter.

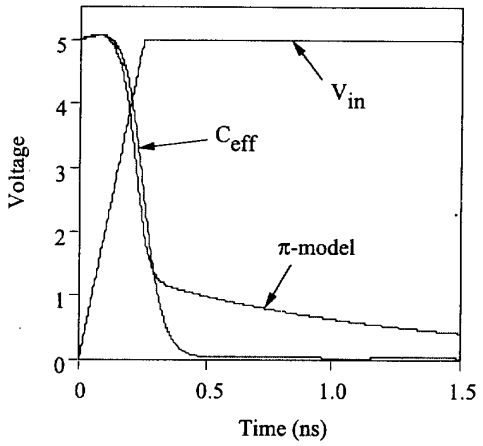

Fig. 12. Comparison of C_{eff} and π-model load for a large RC tree.

examples. The intersection of the C_{eff} equation and the k-factor equation represents one iteration point solution. As the iterations proceed, the C_{eff} curves move toward smaller delay values. Convergence is reached when the k-factor delay agrees with the delay value for the C_{eff} curve that it intersects, as shown in Fig. 11.

Also shown in Fig. 11 is a k-factor equation for an inverter with a W/L ratio of $4x$. This corresponds to significantly smaller W/L ratios for the p- and n-channel transistors that comprise this inverter. As shown in Fig. 11, the k-factor equation is significantly weaker for the 4× inverter as compared to the 50× inverter. Notice that for this somewhat weak inverter, the k-factor equation intersects the C_{eff} curves at a point very close to the total capacitance value. This is expected, since the inverter's "resistance" is now significantly larger than the resistance of the π-model load, hence, the inverter sees the total capacitance of the net. For the 4 × inverter example, the C_{eff} value would converge in one iteration.

As a second example we consider an inverter and RC interconnect problem that is characterized by a 30× inverter driving a π-model with R =232 Ω, C_1 =3.849 pF, and C_2 =0.283 pF. After 3 iterations, a C_{eff} value of 0.668 pF is obtained. The response for the C_{eff} load is compared with that for the π-model load in Fig. 12. Once again, there is good agreement between the two waveforms up to the 50% point.

It is apparent, however, from the examples in Figs. 9 and 12 that an effective capacitance model is able to capture the delay with reasonable accuracy, but that the overall waveshape is not captured beyond the 50% point. It is not unexpected that a single capacitance value is unable to capture the complete behavior of a π-model load. Moreover, it is the resistance shielding effect that gives these response waveforms such long exponential tails, which are significantly different than the more "digital" waveshapes for the capacitance loads. If we were concerned only with obtaining the delay of the gate, the C_{eff} model would be sufficient. But since we are planning to use the output waveform (at the driving point of the RC interconnect) to calculate the RC interconnect delay, this C_{eff} waveform may be unacceptable.

IV. Approximating the Driving Point Waveform

As observed in Fig. 9 and Fig. 12, the slow decaying tail portion of the response waveform is not accurately captured using the effective capacitance model. One explanation for the tail portion of this waveform is that the CMOS gate is behaving like a resistor, and its interaction with a π-model load is described by a second-order exponential function. This line of reasoning follows from the piecewise CMOS inverter models described in [12].

In [12], the transient response of a CMOS inverter is analyzed as four separate regions of operation based upon the operating regions of the p- and n-channels. We will assume throughout this section that the input signal is a rising transition, therefore an output falling transition is considered for this discussion. One can argue that when the rising input-signal to an inverter is greater than the output voltage by more than the threshold voltage, the n-channel goes from saturation to linear and the p-channel is off or barely conducting. The inverter, or any logic gate, can be accurately modeled by a resistor to ground for this region of operation. This resistance value can be approximated by the large-signal output resistance of the gate. A similar argument for a two-region gate model was made previously in [13].

Assuming that the gate is behaving like a resistance to ground, it is apparent from Fig. 13 that the C_{eff} model is going to yield a vastly different response than the π-model load. Therefore, the C_{eff} model is accurate only up to the point at which the gate begins to behave like a resistance. And, it is shown in [4]–[6] that a single resistance model accurately captures the latter portion of a CMOS response waveform, however the initial delay and the initial portion of the response waveform are more difficult to capture. Therefore, we propose to use the C_{eff} model to capture the initial delay and a resistance model to capture the remaining portion of the response.

Given the k-factor equations for a gate and the input transition time value, we would begin by iteratively calculating the effective capacitance as described in Section III. With our assumption of a falling output transition, and given the values

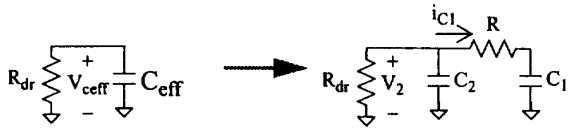

Fig. 13. An effective capacitance model is not an accurate representation of the π-model load when the gate is behaving like a resistor.

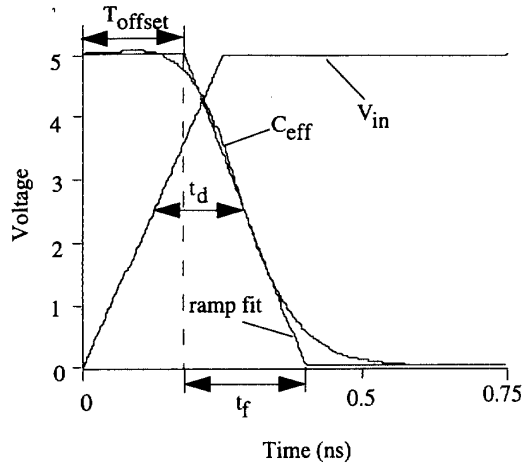

Fig. 14. Fitting a ramp response to the initial portion of the effective capacitance response waveform.

for the gate delay, t_d, and gate-output-waveform transition time, t_f, we can model the initial portion of the response waveform by a single ramp as shown in Fig. 14. This ramp is a valid approximation since it simplifies the driving-point waveform description, and it pessimistically models the response waveform as it will impact the RC interconnect delay at the fan-out points.

The intrinsic delay of the gate, or the waveform offset T_{offset}, is given by

$$T_{\text{offset}} = t_d + \frac{t_t}{2} - \frac{t_f}{2} \tag{4.1}$$

which is simply obtained by solving for the point at which a straight line with slope t_f passing through the 50% point time, t_D, will intersect the initial voltage point of the waveform. Using (4.1), the initial portion of the falling waveform in Fig. 14 is approximated by:

$$V(t) = \frac{V_t\left(t_d + \frac{t_t}{2}\right)}{t_f} + \frac{V_i}{2} - \frac{V_i}{t_f} \cdot t \tag{4.2}$$

where V_i is the initial voltage (V_{DD} for a falling waveform and zero for a rising waveform) as defined previously.

This ramp approximation using C_{eff} is valid up to some timepoint t_s, which we define (for a falling output signal) as the time at which the n-channel enters its linear region of operation as discussed above. We estimate t_s as follows: if the input reaches its final value before the logic gate output reaches its 20% of falling transition, that is $t_t < t_{20}$, then t_s is assigned t_{20}; or, if the input transition is greater than t_{20}, then t_s is assigned the value t_t. The case of really slow input transition, that is, the logic gate output completes 80% of its falling transition before the input reaches its final value, is not considered since the effective capacitance waveform is a reasonable approximation for such a case. That is, the 20 to 80% point time range is accurately approximated by a straight line in such cases, and the effective capacitive model is therefore an accurate depiction of the load for this period of time.

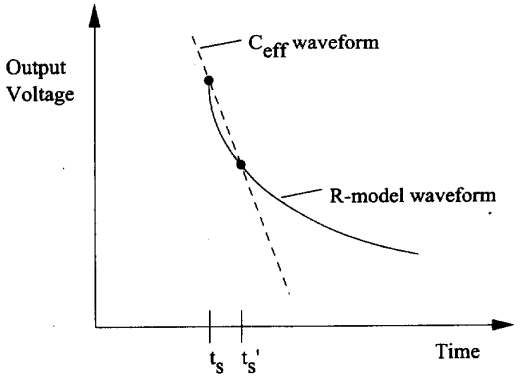

Fig. 15. The intersection of the C_{eff} and the resistance-model curves.

The time points t_{20} and t_{80}, which denote the times when the output reaches 20 and 80% of falling transition, respectively, are approximated by the ramp equation in (4.2). From these time points, we determine t_s and use t_s to indicate when to switch from the C_{eff} model to the situation of a linear resistance driving the π-model load (as shown in Fig. 13). The intersection of these two models is shown in Fig. 15.

Notice in Fig. 15 that the derivatives of the R-model and the C_{eff} approximation are not continuous at t_s. Forcing these derivatives to be equal at this point automatically specifies the gate resistance value, which fails to capture the overall waveshape and the proper RC exponential tail. Instead, we force only the voltage on C_2 to be continuous at time t_s. The voltage V_2 on C_2 of π-model should equal the voltage V_{ceff} of the effective capacitance model at $t = t_s$:

$$V_2(t_s) = V_{\text{ceff}}(t_s) \tag{4.3}$$

Then, we calculate the effective driving resistance R_{dr} in Fig. 13 from the k-factor equations that characterize the gate. For the case of a linear load capacitance, the k-factors contain the information that describe the gate's behavior for all possible output voltage values. We would like to estimate the gate's pull-down resistance from time t_s to the end of the waveform. Since the k-factors were generated by fitting the 20%, 50%, and 80% points for a purely capacitive load, we use them to estimate an effective pull-down resistance for the gate between the time point t_s and the time at which the 80% point value would be reached for the effective capacitance model.

If the gate is behaving like a linear resistance for $t > t_s$, for the case of a linear capacitance load (which is what the k-factors describe), this portion of the response waveform would be described by a single exponential:

$$V(t) = V(t_s)e^{-\frac{t}{R_{\text{dr}}C_{\text{eff}}}} \tag{4.4}$$

This single exponential approximation is similar to the one described in [5], however instead of approximating the complete transition as a single exponential, we calculate the resistance of the gate for a smaller portion of the response waveform (t_s to t_{80}). The value $R_{\rm dr}$ for our approach is calculated as:

$$R_{\rm dr} = \frac{(t_{80} - t_s)}{\ln\left[\frac{V(t_s)}{V(t_{80})}\right] C_{\rm eff}} \tag{4.5}$$

Similar arguments and expressions result for a rising output signal.

From Fig. 13, the voltage $V_2(t)$ for time greater than t_s can be expressed in terms of the initial conditions on the π-model and the π-model parameters. The voltage response of this two RC circuit for time $t > t_s$ is:

$$V_2(t) = a_1 e^{p_1(t-t_s)} + a_2 e^{p_2(t-t_s)} \tag{4.6}$$

And the poles, p_1 and p_2, can be symbolically analyzed by (4.7), shown at the bottom of this page.

The coefficients a_1 and a_2 in (4.6) are solved for using the initial conditions at $t = t_s$:

$$V_2(t_s) = a_1 + a_2 \tag{4.8}$$

$$\frac{V_2(t_s)}{R_{\rm dr}} + i_{c_1}(t_s) = -C_2(a_1 p_1 + a_2 p_2) \tag{4.9}$$

Solving (4.8) and (4.9), the values a_1 and a_2 are given by:

$$a_2 = \frac{\left[V_2(t_s)\left(p_1 + \frac{1}{C_2 R_{\rm dr}}\right) + \frac{i_{C_1}(t_s)}{C_2}\right]}{p_1 - p_2}$$

$$a_1 = V_2(t_s) - a_2 \tag{4.10}$$

To calculate a_1 and a_2, it is apparent from (4.10) that current through C_1 at t_s must be approximated since it is part of the initial conditions on the π-model. Considering continuity with the $C_{\rm eff}$ model, we approximate this current by:

$$i_{c_1}(t_s) = i_{\rm ceff}(t_s) - i_{c_2}(t_s) = k(C_{\rm eff} - C_2) \tag{4.11}$$

where k represents the slope of the ramp and the capacitor current is simply kC at $t = t_s$. For a falling transition:

$$k = \frac{-V_{\rm DD}}{t_f} \tag{4.12}$$

Using the initial conditions above, the double exponential approximation in (4.6) starting at $t = t_s$ will intersect the ramp approximation in (4.2) at some time t'_s, as shown in Fig. 15. To characterize the driving point waveform, one could use the linear approximation from the $C_{\rm eff}$ estimate up to the timepoint t_s, and then the resistance approximation from t_s to infinity. However, to ensure a pessimistic waveform approximation, one can also choose to characterize the response using (4.2) up to t'_s, the point at which the resistance-model becomes more pessimistic than $C_{\rm eff}$. Using t'_s instead of t_s guarantees a pessimistic result, however it does require solving a transcendental equation to obtain the value of t'_s. We should also mention that it can be shown that t'_s is guaranteed to exist (meaning the resistor model always intersects with the straight-line $C_{\rm eff}$ approximation as shown in Fig. 15) using the equations and the initial conditions above [14].

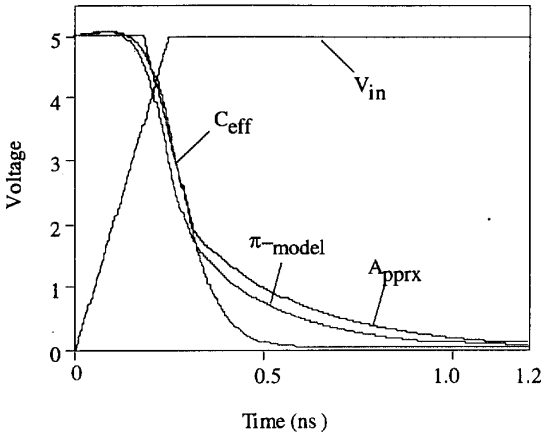

Fig. 16. Comparison of waveforms using π-model load (π-model), effective capacitance (C_{eff}), and the complete waveform approximation (Apprx) for Fig. 4.

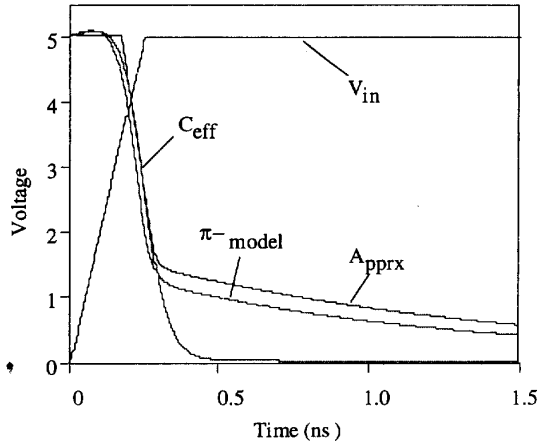

Fig. 17. Comparison of waveforms using π-model load (π-model), effective capacitance (C_{eff}), and the complete waveform approximation (Apprx) for a large RC tree.

To demonstrate the accuracy of our approach, consider once again the RC interconnect example in Fig. 4. Comparing our approximate waveform model with the effective capacitance waveform and the π-model waveform in Fig. 16, the approximate waveform provides a pessimistic estimate as expected. Similarly, for the circuit example that was characterized by the response waveform with an extremely large RC tail in Fig. 12, the waveform approximation described above accurately captures the overall shape, pessimistically, as shown in Fig. 17.

$$p_1, p_2 = \frac{-[(c_1 + C_2)R_{\rm dr} + \mathrm{RC}_1] \pm \sqrt{[(C_1 + C_2)R_{\rm dr} + \mathrm{RC}_1]^2 - 4R_{\rm dr}\mathrm{RC}_1C_2}}{2R_{\rm dr}\mathrm{RC}_1C_2} \tag{4.7}$$

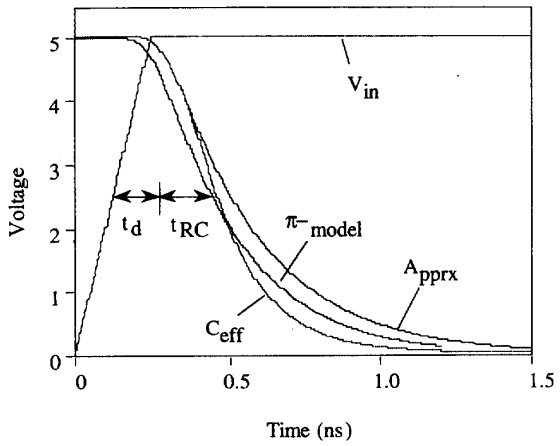

Fig. 18. Comparison of output waveforms at a fan-out point in the RC tree in Fig. 4. The driving point waveforms (from Fig. 16) were calculated using a π-model load (π), an effective capacitance (C_{eff}), and the complete waveform approximation (Apprx) described in this section.

Of course, in general we are not interested in the gate output waveform, but merely the delay and the waveforms at the fan-out points (since these waveforms are used to determine the delay of subsequent logic stages). Even though the C_{eff} model approximates the gate delay rather accurately, the RC interconnect acts as a low-pass filter on the driving point waveform. Therefore, if we use an incorrect driving point waveform to determine the delays and waveforms at the fan-out points, we can end up with an erroneous waveshape and delay.

To illustrate the importance of capturing the complete driving point waveform, consider the farthest fan-out node for the RC interconnect in Fig. 4. Fig. 18 contains the response waveforms at this fan-out node when using the various driving point waveforms from Fig. 16. That is, the time domain plots for a C_{eff} load, our two-piece approximate waveform, and a π-model load were numerically convolved with an accurate, yet approximate, RC tree transfer function from the driving point of the interconnect to this fan-out point of interest. In practice, we do not use numerical convolution, but we convolve the analytical waveshape expressions for the C_{eff} load response and the resistance model response directly with the approximate analytical transfer function from AWE for a symbolic expression for the fan-out waveform. Numerical convolution is used here so that we can avoid any waveshape fitting errors for the C_{tot} and C_{eff} driving point waveforms.

Notice that even though the effective capacitance model starts out pessimistically in Fig. 16, the delay starts becoming optimistic at this fan-out point since the tail portion of this driving point waveform is optimistic. That is, the delay time at the fan-out point occurs when the C_{eff} driving point waveform has practically reached its final value. We would expect that this optimistic error would increase in absolute size as the resistance of the RC interconnect increases and the speed of the driving point waveforms increases. Moreover, C_{eff} will always tend to predict a faster signal transition at the fan-out points than is actually the case, and this optimistic transition time will be propagated as an optimistic delay and an optimistic output transition time for the next logic stage.

TABLE II
SPICE MODEL PARAMETERS

	P-channel	N-channel		P-channel	N-channel
VTO	-1.0	1.0	CGSO	0.22E-9	0.22E-9
TOX	0.2E-7	0.2E-7	CGDO	0.22E-9	0.22E-9
NSUB	4E16	2E16	CGBO	0.17E-7	0.17E-7
PB	0.37	0.4	CJSW	0.68E-3	0.47E-3
JS	0.1E-2	0.1E-2	MJ	0.4	0.4
NSS	0.0	0.0	MJSW	0.17	0.11
NFS	0.2E12	0.2E12	TPG	1	1
XJ	0.3E-06	0.3E-06	LD	0.3E-07	0.3E-07
VMAX	3.40E+04	3.0E+04	UCRIT	1.0E+04	1.2E+04
NEFF	1.0	1.0	UEXP	0.4	0.4
FC	0.5	0.5	DELTA	0.0	0.0
XQC	0.499	0.499	UO	300	700

TABLE III
THE C_{eff} VALUES FOR SEVERAL EXAMPLE CIRCUITS

Inverter Size	π-model parameters			C_{eff}	# of
	C2 (pF)	R (Ω)	C1(pF)	(pF)	iterations
30x	0.245	89.5	1.46	0.938	4
30x	0.283	232	3.85	0.668	3
30x	0.454	106	2.50	1.29	3
50x	0.575	72.6	2.13	1.48	3
200x	125	1620	750	126	2

V. RESULTS

All of the examples in this paper were generated for a 1-μm CMOS technology. The SPICE model parameters are shown in Table II. Five circuit examples were tested using this approach, as shown in Table III. Shown for each example is the size of the driving inverter and the π-model parameters for an actual RC interconnect. Also shown are the C_{eff} values and the number of iterations required to reach convergence.

Table IV shows the accuracy of this approach for these five eclectic examples. In all cases, the 20%, 50%, and 80% points are pessimistic, but reasonably accurate, approximations to the actual (nominal results using SPICE and the π-model load) percentage points. We should also point out that the waveform approximation following the procedure in Section IV will sometimes yield an 80% estimate that is more pessimistic than the total capacitance estimate. However, it is important to recognize that the overall waveshape is more accurately captured using the waveform approximation in Section IV than it is with the total capacitance. Moreover, the waveshape at the driving point of the interconnect affects the RC delay more so than the 80% point value since we ultimately convolve this waveform with the approximate transfer function to estimate the delays and waveshapes at the fan-out points.

This model has been applied to thousands of real circuit examples, mainly from high-speed microprocessor chips at IBM, Austin TX, and these five shown above represent a sampling of some of the larger, more resistive loads. The

TABLE IV
COMPARISON OF CRITICAL WAVEFORM POINT PREDICTIONS USING TOTAL CAPACITANCE (C_{tot}), π-MODEL LOAD (π), THE EFFECTIVE CAPACITANCE (C_{eff}), AND THE COMPLETE WAVEFORM APPROXIMATION IN SECTION IV (APPR.)

Inv.	20% point (pS) using:				50% point (pS) using:				80% point (pS) using:			
Size	C_{tot}	π	C_{eff}	appr	C_{tot}	π	C_{eff}	appr	C_{tot}	π	C_{eff}	appr
30	257	203	225	216	348	264	282	282	481	450	358	501
30	337	192	208	206	551	240	257	257	865	470	311	739
30	299	215	238	224	452	287	308	308	677	614	406	708
50	262	211	230	221	350	269	287	287	477	434	361	497
200	4540	894	896	897	11260	1870	1869	2095	21117	3313	3295	3321

results from this testing indicate errors and improvements similar to those in the tables shown above.

VI. CONCLUSION

We have shown a complete scheme for modeling the delays of CMOS logic gates when the resistance of the interconnect significantly shields some of the load capacitance. First, an effective capacitance value is used to capture the initial portion of the response waveform. Then, a resistor-model is used to capture the remaining portion of the waveform, which may include a long exponential tail due to the RC interconnect effects. The approach is completely compatible with the popular k-factor modeling equations, thereby making it suitable for incorporating circuit-level effects such as resistance shielding into higher-level tools such as timing analyzers. This concept of effective capacitance has also been used recently to generate a linear, time-varying Thevenin equivalent gate model in [15].

ACKNOWLEDGMENT

The authors would like to thank Jean-Paul Watson of IBM for incorporating and testing this model.

REFERENCES

[1] C. Ratzlaff, S. Pullela, and L. T. Pillage, "Modeling the RC interconnect effects in a hierarchical timing analyzer," in *Proc. Custom Integrated Circuits Conf.*, May 1992.
[2] R. Putatunda, "Auto-delay: a program for automatic calculation of delay in LSI/VLSI chips," in *Proc. 19th Design Automation Conf.*, June 1981, pp. 616–621.
[3] J. Rubenstein, P. Penfield, Jr., and M. A. Horowitz, "Signal delay in RC tree networks," *IEEE Trans.Computer-Aided Design*, vol. CAD-2, pp. 202–211, 1983.
[4] J. K. Ousterhout, "Switch-level delay models for digital MOS VLSI," in *Proc. 21st Design Automation Conf.*, 1984.
[5] L. M. Brocco, S. P. McCormick, and J. Allen, "Macromodeling CMOS circuits for timing simulation," *IEEE Trans. Computer-Aided Design*, Dec. 1988.
[6] J. L. Wyatt, *Circuit Analysis, Simulation and Design*. The Netherlands: North-Holland, ch. on "Signal Propagation Delay in RC Models of Interconnect," 1987.
[7] N. H. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*. New York: Addison Wesley, 2nd ed., pp. 221–223, 1993.
[8] C. Ratzlaff, N. Gopal, and L. T. Pillage, "RICE: Rapid interconnect circuit evaluator," in *Proc. 28th ACM/IEEE Design Automation Conf.*, June 1991.
[9] L. T. Pillage and R. A. Rohrer, "Asymptotic wavform evaluation for timing analysis," *IEEE Trans. Computer-Aided Design*, Apr. 1990.
[10] N. Gopal, D. P. Neikirk, and L. T. Pillage, "Evaluation of on-chip interconnect using moment matching," in *Proc. Int. Conf. Computer-Aided Design*, Nov. 1991.
[11] P. R. O'Brien and T. L. Savarino, "Modeling the driving point characteristic of resistive interconnect for accurate delay estimation," in *Proc. Int. Conf. Computer-Aided Design*, Nov. 1989.
[12] D. Holberg, S. Dutta, and L. T. Pillage, "DC parametrized piecewise function transistor models for bipolar and MOS logic stage delay evaluation," in *Proc. IEEE Int. Conf. Computer-Aided Design*, Nov. 1990.
[13] M. D. Matson and L. A. Glasser, "Macromodeling and optimization of digital MOS VLSI circuits," *IEEE Trans. Computer-Aided Design*, Oct. 1986.
[14] J. Qian, "Modeling the 'effective capacitance' of RC interconnect," M. S. thesis, Univ. of Texas at Austin, TX, May 1994.
[15] F. Dartu, N. Menezes, J. Qian, and L. T. Pillage, "A gate delay model for high performance CMOS," in *Proc. Design Automation Conf.*, June 1994.

Jessica Wen-qian Qian received her B.S. and M.S. in electrical and computer engineering from the University of Texas at Austin in August 1992 and May 1994, respectively. Her research interests were delay modeling and timing analysis for high-performance VLSI design.

She joined LSI Logic Corporation, Milpitas, CA. in January 1994. Currently, she is working on test program generation for manufacturing testing generation for manufacturing testing.

Satyamurthy Pullela received his Master of Engineering from Indian Institute of Science, Bangalore in 1991 and Bachelor of Technology from Indian Institute of Technology, Madras, in 1989. Currently, he is a Ph.D student at the University of Texas, Austin. His research interests include timing analyses for vlsi, circuit simulation, delay modeling, and synthesis of on chip interconnect.

Lawrence T. Pillage (M'85) received his Ph.D. in electrical and computer Engineering from Carnegie Mellon University (CMU) in 1989. Prior to joining CMU, from 1984 through 1986 he worked as an integrated circuit and system designer at Westinghouse Research and Development. During that time he acquired three patents and in 1986 was recognized with the corporation's highest engineering achievement award.

Presently, he is an associate professor at the University of Texas at Austin where he holds a Temple Endowed Faculty Fellowship in Engineering. In 1991 he received the IEEE Transactions on Computer-Aided Design Best Paper Award. Also in 1991, he received a Presidential Young Investigator Award from the National Science Foundation. In 1992 he received a Technical Excellence Award from the Semiconductor Research Corporation (SRC). In 1993 he received a Technical Invention Award from the SRC. He has published over 40 papers in the area of VLSI design and circuit-level design automation.