Calculating Worst-Case Gate Delays Due to Dominant Capacitance Coupling*

Florentin Dartu
Carnegie Mellon University
Department of ECE, Pittsburgh, PA 15213
dartu@ece.cmu.edu

Lawrence T. Pileggi
Carnegie Mellon University
Department of ECE, Pittsburgh, PA 15213
pileggi@ece.cmu.edu

ABSTRACT

In this paper we develop a gate level model that allows us to determine the best and worst case delay when there is dominant interconnect coupling. Assuming that the gate input windows of transition are known, the model can predict the worst and best case noise, as well as the worst and best case impact on delay. This is done in terms of a Ceff based gate model under general RC interconnect loading conditions.

I. INTRODUCTION

As IC dimensions scale to the deep submicron range, their multi-level interconnects are constructed such that the coupling capacitance becomes the dominant component of load capacitance. This effect is largely the result of the increased ratio between the lateral and the vertical capacitance of the line. The increased number of metal layers is the other source of coupling capacitance problems, since there is a reduced likelihood of a nearby "ground plan." The lateral capacitance is increased by the relative increase in the metal thickness with respect to line spacing that is made to control resistance.

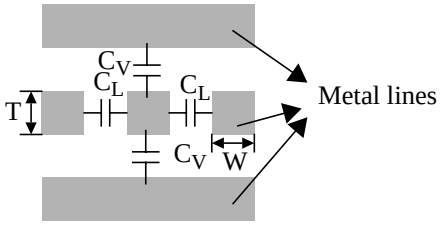

Figure 1: Cross-section into interconnect system with parasitic capacitances definition.

For example, Fig. 1 shows a cross-section of metal layers and the definition for lateral (C_L) and vertical (C_V) capacitances, metal thickness and width. As in other first-order models [1], we assume that line spacing is equal to the line width, and that the metal thickness equals to the dielectric thickness. Using the parallel plate approximation for the capacitances, we obtain:

$$\frac{C_L}{C_V} = \left(\frac{T}{W}\right)^2 = AR^2 \qquad (1)$$

where *AR* is the aspect ratio (thickness/width). Equation (1) is inexact since fringing fields are not included. Considering the fringe effects, this lateral vs. vertical capacitance ratio has a slightly weaker dependence on the aspect ratio:

$$\frac{C_L}{C_V} = AR^n \qquad n \in [1, 2] \qquad (2)$$

In [1] it was reported that the average metal aspect ratio (defined as the sum of the aspect ratio used for each layer divided by the number of metal layers) for Intel Corporation processes (which is representative of the state-of-the-art) increases by 1.22x per generation. For 0.35μm technology, the average aspect ratio is 1.3, and in [1] it is stated that aspect ratios of 2 are still desirable for RC delay benefits. From equation (2) we observe that the lateral capacitance, the main candidate for coupling between lines, can be 2-4 times bigger than the vertical capacitance. This makes the effective load capacitance of a line *strongly* dependent on the switching activity of the signals to which it is coupled -- same direction coupled-signal switching speeds up the response, while opposite direction switching slows the response [2]. If the coupling capacitance dominates the total load capacitance, then the line delay can vary by several hundreds of percent as a function of the switching activity of nearby lines.

Since interconnect modeling and RC model order reduction have advanced significantly over the past several years, it is not unfathomable to assume that we can extract the actual coupling capacitance and model coupling between lines. But since we are stuck in a pattern of computing gate delays only when linear grounded capacitors load the gate, we are immediately faced with the delicate problem of load modeling.

Sometimes the problem is approximated by modeling the coupled capacitors as elements to ground, with modified values of capacitance. For example, for opposite direction switching of two identical, perfectly symmetrical coupled lines, switching at the same instant of time, the coupling capacitance can be accurately modeled as twice the amount of capacitance to ground. While such an approximation can sometimes yield pessimistic delay approximations, for the more realistic coupling cases (as we will show through an example) this doubling-the-coupling model does not predict an upper bound on the delay in general.

To analyze noise due to capacitance coupling, one can start with a reduced order coupled interconnect model and calculate the signals on a quiet victim line by superimposing the coupled signals from all other lines. Such a model, while not exact, can render a reasonable approximation since the nonlinear CMOS gate of the non-switching victim line is behaving like a transistor in its linear region of operation, and therefore, is modeled fairly well by a linear resistor.

When we consider the impact of coupling on the delay, how-

* This work was supported in part by Intel Corporation and the Semiconductor Research Corporation under contract DC-068.

DAC 97, Anaheim, California

ever, the victim line is switching, and the problem is much more complex. As the victim line switches, the impedance of its driving gate changes by orders of magnitude, thereby influencing the amount of coupling voltage. Such effects can be accurately modeled in SPICE, but due to the circuit size, we would prefer to perform such analyses at the highest possible level of abstraction.

Empirical gate/cell level models remain popular for timing analysis, even for full custom designs. In [3] a gate/cell level modeling methodology was developed which achieves compatibility with RC interconnect loading through an "effective capacitance" approximation. In this paper we extend this waveform-based gate model to consider the problem of calculating the delay (and response waveshape) when there is a significant amount of coupling. The algorithms we develop in this paper for handling capacitive coupling permit two approaches for obtaining the best/worst gate delay. First, we will outline a methodology for bounding the best/worst gate delay. Second, due to the algorithm efficiency, a general optimization procedure is possible to generate accurate results.

We will begin by first reviewing some of the background of the C_{eff} model in the following section. Two cases of coupling of particular interest will be discussed in Section III: single gate switching and two gates with capacitive loads simultaneously switching. Our approach for the *n* capacitively coupled RC tree problem will be presented in Section IV. Bounding the gate delay is the subject of Section V, followed by our conclusions in Section VI.

II. WAVEFORM-BASED GATE DELAY MODELS

Gate delay modeling represents an attractive approach for timing analysis due to its simplicity, speed and accuracy. For purely capacitive loads and saturated ramp inputs, it is possible to characterize various output points (e.g. 20%, 50%, 90%, etc.) as functions of input transition time, t_{in}, and output capacitance, C_L:

$$t_\alpha = f_\alpha(t_{in}, C_L) \tag{3}$$

In (3), α is the percentage point value, t_α is the output point delay (w.r.t. the 50% point of the input signal) and f_α is the corresponding delay description function. The delay description functions can be obtained in various ways. Both analytical expressions obtained using simplified MOS models [4,5] or empirical gate delay models [6] can be used to generate (3). But for the purpose of explanation in this paper, we will assume empirical gate delay models.

Empirical gate delay models are built by running multiple SPICE simulation with the input transition time and the output capacitance sweeping a specified range. The output points of interest are selected from the SPICE results and then an algebraic function (usually polynomial) or a look-up table is fitted to the data. The gate is often characterized in terms of other parameters too, such as temperature, voltage supply, variations, etc.

As the minimum feature sizes for integrated circuits scale downward, the interconnect cannot be modeled anymore by the linear grounded capacitor assumed in the empirical gate models. The resistance of the interconnect requires the use of RC driving point and transmittance models. Generally, they are in the form of some reduced order models [7]. But even so, the simple RC loads are not single capacitors. In addition, when the resistance is significant, the waveforms are not accurately represented by just rise/fall time value [8].

Some attempts have been made to extend the empirical models to handle RC loads by increasing the number of parameters in the characterization. In [9] the RC load was modeled from the driving point admittance point of view as a π-circuit, while the input signal was represented by three parameters. This approach increases the number of SPICE runs required for characterization to very high levels (for example, from 16 to 4096 when considering every parameter sampled in 4 points). It also has difficulties establishing input parameter ranges and it is unable to give answers when the load requires higher order models than a C-R-C (π-model) circuit.

Another approache to achieve RC load compatibility employs the concept of *effective capacitance* [12,3]. Given any gate delay model (empirical or analytical) developed under purely capacitive load constraints, it maps the RC load into a single capacitance value during an iterative process to find the complete gate output waveform.

Consider a Thevenin equivalent gate model as shown in Fig. 2a [3]. The model resistance, R_d, is linear as described in [3]. It can be shown that for any actual *g*ate output waveform and R_d value, an ideal Thevenin voltage can be obtained that will allow the gate model output to perfectly match the actual output. The Ceff iterations are the process by which this voltage is obtained.

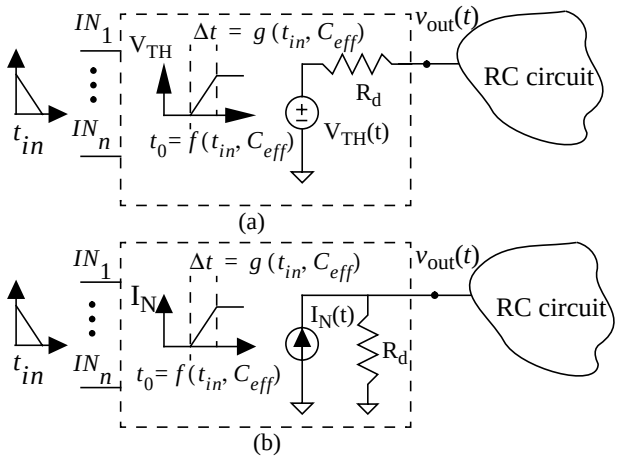

Figure 2: Time-varying CMOS gate model as described in [3]: a) Thevenin equivalent; b) Norton equivalent.

One way of thinking about the Thevenin equivalent of a gate is to see it through the eyes of a simulator. In a transient analysis, at every time point the simulator solves a nonlinear system, usually through a modified Newton-Raphson algorithm [10]. At every iteration, the standard Newton-Raphson algorithm linearizes every nonlinear element (represented by its I-V characteristic) as shown in Fig. 3a. At the last iteration the true operating point is obtained and the corresponding linearization can be used to obtain a Thevenin or Norton equivalent of the gate. *The gate output resistance obtained this way will have different values at each time point and represents the actual small signal output resistance*.

The successive chord method [10] is another algorithm that can be used to solve nonlinear systems. Its linearization method is shown in Fig. 3b and implies a *constant resistor value*. The Thevenin or Norton gate equivalents will have the same output resistance for every time point. Because the successive chord method is as accurate as Newton-Raphson, its gate equivalents are also accurate. The successive chord method has a slower rate of nonlinear convergence in general, which is why N-R is more widely used for circuit simulation.

The gate model presented in Fig. 2a is based on a linear approximation of this ideal Thevenin voltage waveform. The ramp-like shape of the Thevenin waveform was confirmed experi-

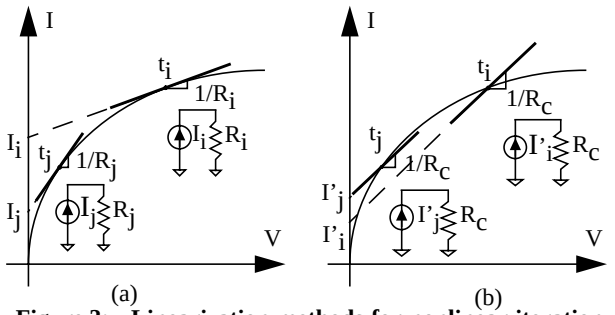

Figure 3: Linearization methods for nonlinear iteration algorithms: a) standard Newton-Raphson and b) successive chord method.

mentally in [11]. It is interesting to remark that the errors of the ideal Thevenin waveform linearization are attenuated at the model output due to the low-pass nature of the system.

The Norton form of the Ceff model, shown in Fig. 2b, will be used in this paper. The I_N steady state values are 0 and V_{DD}/R_d. For a purely capacitive load, the model current source (MCS) parameters, t_0 and Δt, are determined by forcing the model output voltage to be equal to the actual gate output voltage for two output points (e.g. 20% and 50% points):

$$\begin{cases} v_{out}(t_{20}(t_{in}, C_L)) = \begin{cases} 0.2V_{DD} & \text{output rising} \\ 0.8V_{DD} & \text{output falling} \end{cases} \\ v_{out}(t_{50}(t_{in}, C_L)) = 0.5V_{DD} \end{cases} \quad (4)$$

where $v_{out}(t)$ is the model response for a capacitive load, that depends on the model unknown parameters t_0 and Δt, C_L is the load capacitance, t_{in} is the gate input transition time and $t_\alpha(t_{in},C_L)$ is a gate delay model as in equation (3).

To achieve compatibility between the RC load model and the gate equations in (4), the effective capacitance principle is applied: *For every given gate, input transition time and RC load, there exists a C_{eff} that will force the same linearized Norton equivalent for the gate*. In order to find this C_{eff} value we force the equality of the average output voltage for $t \in (t_0, t_0 + \Delta t)$ for both model driving C_{eff} and the actual RC load:

$$\frac{1}{\Delta t}\int_{t_0}^{t_0+\Delta t} v_{out}^{C_{eff}}(t)\,dt = \frac{1}{\Delta t}\int_{t_0}^{t_0+\Delta t} v_{out}^{RC}(t)\,dt \quad (5)$$

where $v_{out}^{C_{eff}}(t)$ is the model response for C_{eff} load and $v_{out}^{RC}(t)$ is the model response for the actual load. It can be shown that the average voltage principle described in equation (5) is mathematically equivalent to the average current principle described in [3]. This model is conceptually extended in the following sections to model the coupling problem.

III. GATE MODELS WITH DOMINANT COUPLING CAPACITANCE

III.1 Single Gate Switching

We will first consider the "simplest" problem that can be caused by coupling capacitance. The situation is described in Fig. 4a for two inverters, but extends to any other type of gate (buffers, NANDs, etc.). The two drivers, one switching and one quiet, are loaded by interconnect lines that are capacitively coupled. The coupling capacitance is modeled by C_c while C_{g1} and C_{g2} model the capacitance to ground.

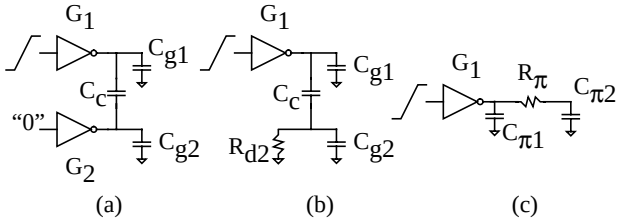

Figure 4: Transforming the coupling capacitance problem with one line quiet into an RC load problem: a) two drivers, one switching and one quiet, with their loads coupled; b) the quiet driver is replaced by its output resistance; c) the load of gate G_1 is mapped into a π-circuit.

The problem in Fig. 4 is a simplified one for two reasons: First, due to the uncertainty of the input signal arrival times, there is a non-zero probability of overlapping transitions for G_1 and G_2. If we are certain that gate G_1 will always switch while G_2 is in steady state, then we don't speak about the best/worst case delay for G_1 (at least from the coupling point of view) but rather the exact value. We have also neglected the interconnect resistance, which we know is not possible in general. But we will remove these simplifications in later sections.

For this example we can safely replace gate G_2 by a linear resistor to ground [3] as shown in Fig. 4b, whenever the noise amplitude at the output of G_2 is small (up to 20-25% from V_{DD}) so that the nonlinear variations are negligeable. The error incurred in the noise amplitude due to this assumption will be shown later. The RC load in Fig. 4b can be exactly translated into a π-model load (Fig. 4c).

The equivalent circuit shown in Fig. 4c can be evaluated by the *effective capacitance* algorithm described in [3]. First of all, it should be noted that the C_{eff} value is bounded below and above by $C_{\pi 1}$ and ($C_{\pi 1}$+$C_{\pi 2}$=C_{g1}+C_c), respectively. If C_c dominates C_{g1} and C_{g2}, these bounds are not directly useful. The Ceff gate model approximation for the circuit in Fig. 4c is shown in Fig. 5.

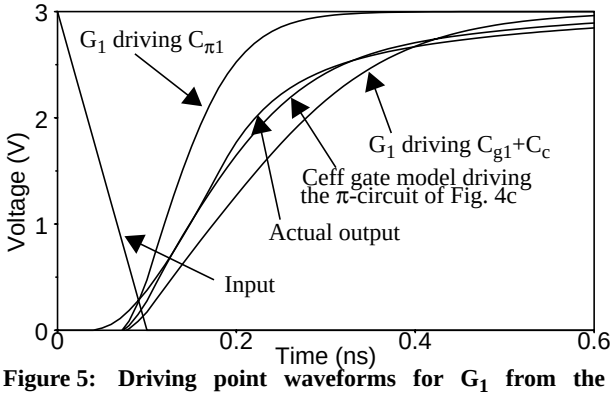

Figure 5: Driving point waveforms for G_1 from the circuit shown in Fig. 4.

Note that while this model is simplified, it still captures the effect of the quiet line gate resistance. A simple charge sharing model does not capture this effect.

$$V_{noise} = V_{DD}\frac{C_c}{C_c + C_{g2}} \quad (6)$$

Where V_{noise} is the amplitude of the coupling noise at the output of G_2 generated by a transition of G_1. For this example, equation (6) will predict a noise amplitude of 62.5% from V_{DD}, whereas the

simplified analysis in Fig. 5 is much more accurate.

Once the Ceff model (Fig. 2) for the example in Fig. 4 reaches convergence, the now linear circuit represents a two-pole system. The noise amplitude, therefore, can be computed *analytically* from the circuit parameters (C_{g1}, C_{g2}, C_c) and the gate parameters (R_{d2} -- G_2 output resistance, R_{d1} and Δt of the G_1 model as described in Fig. 2). This computation is beyond the scope of this paper but we will mention that it accurately predicts the noise amplitude for our example: 25% from V_{DD}.

The single gate switching case can be complicated by taking into consideration the line resistance and the coupling to more than one other line. These modifications will only result in a more complicated driving point admittance and transfer function models, so the Ceff gate model approach still applies.

III.2 Two Switching Gates Coupled

A much more interesting, realistic, case is presented in Fig. 6a. Both drivers can switch simultaneously and their input signals are described as "windows" of arrival times. For this situation we are interested in best/worst cases at the outputs in order to generate the "window" of arrival times for those signals, and so on.

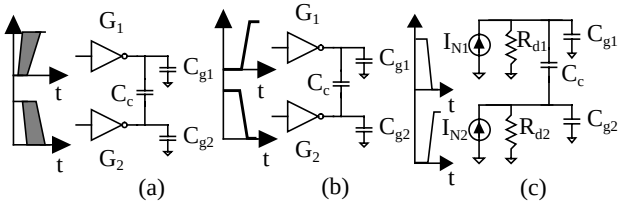

Figure 6: Two coupled gates switching simultaneously. a) general case; b) worst case scenario; c) Norton equivalent gate model applied to this problem.

The important problem of how to find the pair of input gate signals that will generate the best/worst case for the output signals will be treated later. It is not difficult to observe that the worst case scenarios can be different for the two gates. Without loss in generality, we will study G_1's behavior under noise generated by G_2. G_1 will be referred to as the *victim*, while G_2 will be the *aggressor*. For the moment we will assume that we know the worst case switching conditions for the output of G_1, as shown in Fig. 6b.

With the problem converted to a deterministic one, we attempt to solve it with the best accuracy versus computation time possible. We begin by viewing this circuit from superposition-like standpoint. Of course our circuit is highly nonlinear, so superposition doesn't actually apply, but this view of the problem allows us to consider the aggressor as acting on the victim. Starting with this argument we consider the following facts:

- the "passive" gate output resistance (R_{d2} in Fig. 4c) is nonlinear, spanning orders of magnitude during a transition. It has a low value only if there is a path of transistors working in the linear region from the output to ground or V_{DD}. The higher the resistance value at a given time point, the higher the noise amplitude will be.
- the noise injected by G_1 on R_{d2} depends on G_1's output waveform. But the true output signal of G_1 includes the noise injected by G_2 on G_1's output resistance. This convoluted argument is basically another way of saying that superposition doesn't apply to nonlinear systems.

In solving this nonlinear, time-dependent system, we have the option to use the successive chord method for the nonlinear solver. Such a solution scheme yields a Norton equivalent gate model with constant output resistance upon convergence for all time points. The Norton current source is a function of time but can be linearized with reasonable accuracy so that the Ceff gate responses can be approximated. The Ceff model shown for the coupled system in Fig. 6c can be interpreted as the result of such a procedure.

The time varying gate model described in [3] successfully models single gates switching for a variety of loads. It computes the charge exchanged by the gate with the load for a specific period during the transition and finds the *effective capacitance* that would ask for the same charge. Observing the currents through the coupling capacitance of Fig. 6 it is obvious that there is no qualitative difference between these currents and the current delivered by a gate to a capacitive load. *The noise current is just another load for the gate*.

Based on the above observation we extended the effective capacitance algorithm to the problem described in Fig. 6. Replacing the gates by their time varying Norton equivalent, Fig. 6c, we form a system where the parameters of the current sources, t_0's and Δt's, are unknown. In order to incorporate the information from the empirical gate delay models we also introduce an effective capacitance for each gate. We maintain the same principles to solve for the unknowns:

- the charge delivered by the gate while its Norton current source is in transition should be the same for the actual load (including noise current) and for C_{eff} load;
- the same Norton equivalent for the actual load and for C_{eff}.
- the Norton equivalent parameters allow the model to fit two points of the actual gate response for capacitive load.

The above principles translate in the following equations:.

$$\begin{cases} \dfrac{1}{\Delta t_1}\displaystyle\int_{t_{01}}^{t_{01}+\Delta t_1} v^1_{actual}(t)\,dt = \dfrac{1}{\Delta t_1}\displaystyle\int_{t_{01}}^{t_{01}+\Delta t_1} v^1_{Ceff}(t)\,dt \\ \dfrac{1}{\Delta t_2}\displaystyle\int_{t_{02}}^{t_{02}+\Delta t_2} v^2_{actual}(t)\,dt = \dfrac{1}{\Delta t_2}\displaystyle\int_{t_{02}}^{t_{02}+\Delta t_2} v^2_{Ceff}(t)\,dt \\ v^1_{Ceff}(t_{20}(t_{in1}, C_{eff1})) = \begin{cases} 0.2V_{DD} & \text{output rising} \\ 0.8V_{DD} & \text{output falling} \end{cases} \\ v^2_{Ceff}(t_{20}(t_{in2}, C_{eff2})) = \begin{cases} 0.2V_{DD} & \text{output rising} \\ 0.8V_{DD} & \text{output falling} \end{cases} \\ v^1_{Ceff}(t_{50}(t_{in1}, C_{eff1})) = 0.5V_{DD} \\ v^2_{Ceff}(t_{50}(t_{in2}, C_{eff2})) = 0.5V_{DD} \end{cases} \quad (7)$$

where $v^1_{Ceff}(t)$ is the G_1 model voltage response for C_{eff1} load, $v^2_{Ceff}(t)$ is the G_2 model voltage response for C_{eff2} load, $v^1_{actual}(t)$ and $v^1_{actual}(t)$ are the G_1 and G_2 model voltage responses, respectively, for the actual load (including noise contribution). The model parameters, t_0's and Δt's, are hidden inside these model voltage responses. The model responses to the actual loads can be written as:

$$\begin{cases} V^1_{actual}(s) = I_{N1}(s)\,Z_{11}(s) + I_{N2}(s)\,Z_{21}(s) \\ V^2_{actual}(s) = I_{N1}(s)\,Z_{12}(s) + I_{N2}(s)\,Z_{22}(s) \end{cases} \quad (8)$$

where $I_{N1}(s)$ and $I_{N2}(s)$ are the Laplace transforms of the Norton equivalent current sources and $Z_{ij}(s)$ are the z-parameters for the two-port seen by the two current sources.

The convergence properties for the Newton-Raphson procedure solving this 6x6 system are good. Only 5-7 iterations are required to achieve .0001 relative accuracy. On a IBM PowerPC (Power Series 850 @133MHz) the algorithm solves 300 bidimensional cases per second. Some results giving an idea about the model accuracy are given in Fig. 7.

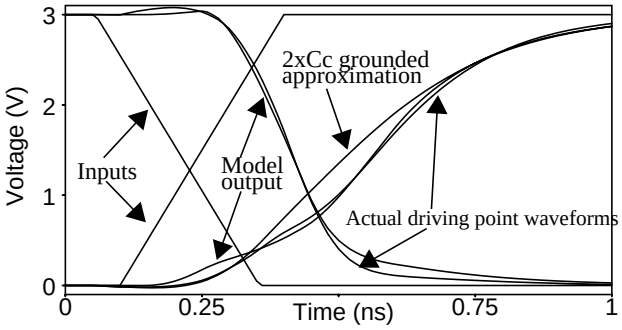

Figure 7: Results for a circuit like the one in Fig. 6a. It shows that grounding twice the coupling capacitance does not provide an upper bound for the output signal.

IV. "CEFF" APPROACH FOR CAPACITIVELY COUPLED RC INTERCONNECT

The general problem we want to solve is described in Fig. 8. Each of the *n* gates drives an RC tree that is capacitively coupled to other RC trees (there is a dc path from node x_i to node $y_{j,k}$ if and only if $i=j$). Without reducing the generality of the problem we arranged the drivers in the order of their arrival time.

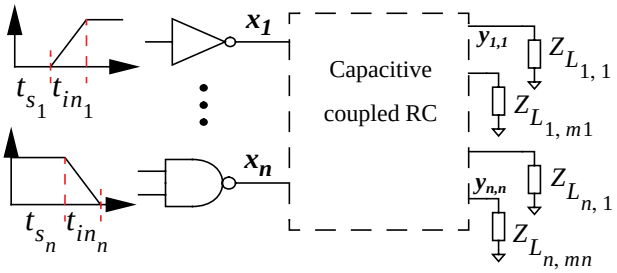

Figure 8: The problem statement: *n* capacitive coupled RC trees with their drivers.

Once we choose the Norton form for the gate model it is advisable to use a reduced order z-parameter representation for the N-port. We extend the effective capacitance algorithm for this N-port problem by computing a C_{eff} value for each port, together with the gate model parameters, i.e. t_0 and Δt, for each driver (as in Section III.2). The unknowns are coupled together, requiring the solution of a (3N)x(3N) system.

One third of the system of equations results from the averaging principle: the average voltage at port *i* should equal the average voltage delivered on the effective capacitance of node *i* for a specified period of time (in our case the time period for which the *i*-th model current source is in transition, Δt_i). Therefore,

$$\frac{1}{\Delta t_i}\int_{t_{0i}}^{t_{0i}+\Delta t_i} L^{-1}\left(\sum_{j=1}^{n} I_{N_j} s \cdot Z_{ji} s\right) dt = \frac{1}{\Delta t_i}\int_{t_{0i}}^{t_{0i}+\Delta t_i} \hat{v}\,(t, C_{eff_i})\,dt \quad (9)$$

where $Z_{ji}(s)$ is the transimpedance from node *j* to node *i* ($Z_{ji}(s)=Z_{ij}(s)$ since the RC circuits are reciprocal). It should be noted that the average voltage principle is mathematically equivalent to the current averaging principle in [3].

The other 2N equations are obtained by forcing the voltage response of the C_{eff} circuit to satisfy the *k*-factor equations for two percentage points: the 50% point and another point before it (we use the 20% point) [3]. Accuracy results for a bidimensional coupling problem with significant line resistance are given in Fig. 9.

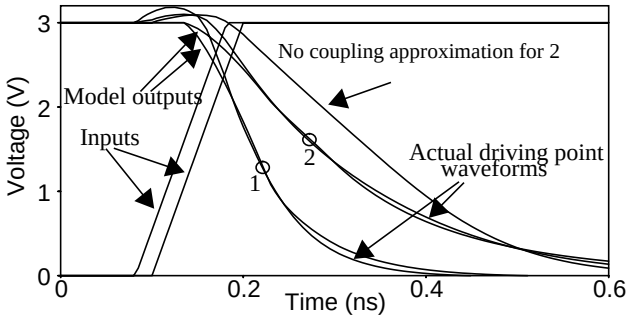

Figure 9: Two drivers loaded by coupled RC lines. This example shows that ignoring the coupling capacitance does not offer a secure lower bound.

V. BEST/WORST GATE DELAY CASE

In a timing analysis environment, we only know min-max ranges for the input transition time of each gate and the time slack between every pair of drivers. To find the set of these variables that generates the best/worst case for the output signals is a difficult problem. Traditionally, this problem has been avoided by expressing the best/worst case in terms of a modified coupling capacitance value. For example, to achieve the worst case, a grounded capacitor equal to twice of the coupling capacitance is added to mimic the conditions of opposite direction switching. But as proved by the results in Fig. 7 and Fig. 9, this approach can give significant errors.

From the output signal transition times point of view it is easy to observe that the victim should have the slowest output while the aggressors should be as fast as possible (for the biggest noise amplitude). This situation can be directly translated in terms of input transition time. The conditions for the best case are reversed.

We will first consider the formulation of the problem which finds the worst case delay using linear driver models. Then we will extend the approach to the Ceff models.

V.1 The Linear Case

We are interested in the worst case delay for the sum of two signals, named "original signal" and "noise" in Fig. 10. The only variable in this case is the time shift between the two responses, t_s, defined from the beginning of the "victim signal" to the peak of the "noise." This is apparently a difficult optimization problem where we have to worry about local optima. Moreover, we have only an implicit formula for the optimization goal, t_{df}. This will tend to make its second derivative w.r.t. t_s overly complicated.

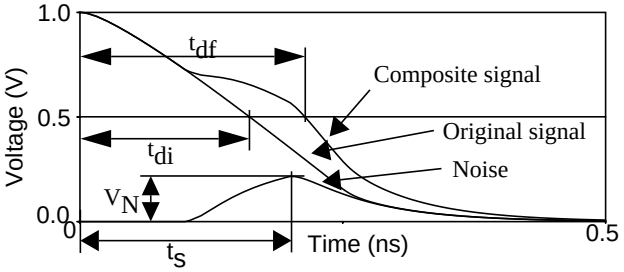

Figure 10: Defining the variables involved in this problem.

However, there is a surprisingly simple solution: the worst case delay for the composite signal corresponds to the time, t_M in Fig. 11, when the "original signal" crosses (0.5-V_N). Therefore, we first compute V_N, which is a simpler optimization goal than calculating t_{df}. For example, if this waveform is modeled as a two-pole response (see Section III.1) we have an explicit formula for the "noise" amplitude. It's understood that this noise is not so large that it would exceed the noise margins.

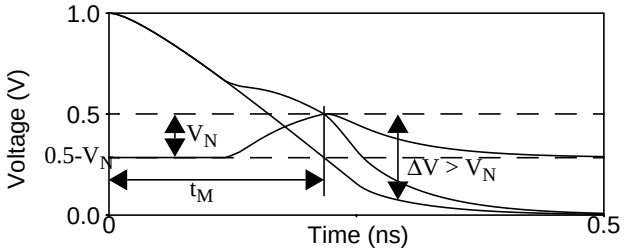

Figure 11: The worst case delay for the problem presented in Fig. 10.

Secondly, we solve for the time when the "original signal" crosses (0.5-V_N). These two steps are significantly easier than a problem formulation in terms of t_{df} directly, since they require two root finding solutions instead of a nonlinear optimization. Moreover, given that there are plenty of intelligent initial solutions, and that an analytical root finding solution is available for a two-pole system, the two step solution approach is further simplified.

V.2 The Actual Problem

Of course the actual problem is not linear. Thinking in terms of a Ceff model, both the "original signal" and the "noise" waveforms are modified by their relative position which greatly complicates finding the worst case. The observation made on the linear problem can be used to upper bound the gate delay. The composite signal delay cannot be bigger than what results by using the maximum noise amplitude, max(V_N), and the slowest "original signal."

The maximum noise amplitude is obtained by the fastest aggressor. Because the victim and the aggressor are on opposite transitions, the victim also slows down the aggressor. Consequently, the quiet victim situation will lower bound the aggressor output transition time and will upper bound the noise amplitude. Once we have determined the maximum noise waveform, we pessimistically assume that there is no influence of the victim to the aggressor. For every time shift of the aggressor Norton equivalent waveform, the victim's Ceff is different (because the average noise voltage while the Norton current source is in transition will be different as shown by equations (7) and (8)). As for purely capacitive loads, the gate response is slower for a bigger Ceff. An upper bound for the gate delay can be found by solving the problem described in Section V.1, where the model response corresponding to maximum Ceff and the noise corresponding to the quiet victim are used. For more than one aggressor a similar procedure is applicable. Every aggressor is fastest when all other gates are quiet. The worst case noise waveforms will be summed such that the resulting signal has the maximum amplitude.

If the above bounds are not tight enough, a better approximation for the worst case delay can be generated by another procedure based on the algorithm presented in Section V.1. In this case we model the effects of shifting the noise on the "original response" by computing the effective capacitance corresponding to each time shift. However, we ignore the effects of the time shift on the noise waveshape. Although the solution given in Section V.1 doesn't upper bound the delay for this problem, it still gives a good approximation of the worst case. We start by considering the biggest noise amplitude (for the quite victim) and compute the victim's maximum Ceff for this situation. This effective capacitance is used to compute the new noise at the victim's output and so on. For the examples shown in Fig. 7 and Fig. 9 this procedure converged within 5% of the worst/best case in 3 iterations.

VI. CONCLUSIONS

In this paper we discussed some of the problems generated by the coupling capacitance. We have shown that even for the simplest case, single gate switching and negligeable line resistance, the classical methods of computing noise amplitude and gate delay can generate large errors. In order to solve these problems we extended a Norton equivalent gate model based on the effective capacitance algorithm to handle the general problem of *n* drivers loaded by coupled RC loads. We presented multiple examples to prove the accuracy of what we found to be a very fast algorithm.

All of these results are obtained for a deterministic set of input signals. In reality, the best we know about the input signals arrival times is their window of uncertainty, an early and a late arrival time. Using the Ceff gate model it is possible to find bounds for the best/worst case delay by viewing the coupled Ceff problem as analogous to the linear superposition of two waveforms.

BIBLIOGRAPHY

[1] M. Bohr, "Interconnect scaling - the real limiter to high performance ULSI," *Intl. Electronic Device Meeting*, pp. 241-244, 1995.

[2] H.B. Bakoglu, "Circuits, interconnections, and packaging for VLSI," Addison-Wesley, 1990.

[3] F. Dartu, N. Menezes, L.T. Pileggi, "Performance computation for precharacterized CMOS gates with RC-loads," *IEEE Transactions on CAD*, vol. 15, pp. 544-553, May 1996.

[4] T. Sakurai, A.R. Newton, "Alpha-power model, and its application to CMOS inverter delay and other formulas," *IEEE Journal of Solid State Circuits*, vol. 25, pp. 584-594, April 1990.

[5] A.I. Kayssi, K.A. Sakallah, T.M. Burks, "Analytical transient response of CMOS inverters," *IEEE Transactions on CAS-I, vol. 39*, pp 42-45, January 1992.

[6] N.H.E. Weste, K. Eshragian, "Principles of CMOS VLSI Design," Addison-Wesley, 1990.

[7] L.T. Pillage, R.A. Rohrer "Asymptotic waveform evaluation for timing analysis," *IEEE Transactions on CAD*, vol. 9, pp. 352-366, 1990.

[8] F. Dartu, L.T. Pileggi, "Modeling signal waveshapes for empirical CMOS gate delay models," *6th Intl. Workshop PATMOS '96*, pp. 57-66, Bologna, Italy, 1996

[9] Y. Miki, M. Abe, Y. Ogawa, "PCHECK: A delay tool for high performance LSI design," *IEEE Custom Integrated Circuits Conference*, pp. 267-270, 1995.

[10] W.J McCalla, "Fundamentals of computer-aided circuit simulation," Kluwer Academic Publishers, 1988.

[11] F. Dartu, N. Menezes, J. Qian, L.T. Pillage "A gate-delay model for high speed CMOS circuits," *Proc. 31st ACM/IEEE Design Automation Conference*, pp. 576-580, 1994.

[12] J. Qian, S. Pullela, L.T. Pillage, "Modeling the 'effective capacitance' of the RC interconnect," *IEEE Transactions on CAD*, vol. 13, pp. 1526-1535, December 1994.

[13] P.R O'Brian, T.L. Savarino, "Modeling the driving-point characteristic of resistive interconnect for accurate delay estimation," *Proc. IEEE Intl. Conference Computer-Aided Design*, pp. 512-515, 1989.

[14] S.A. Kuhn, M.B. Kleiner, P. Ramm, W. Weber "Interconnect Capacitances, Crosstalk, and Signal Delay in Vertical Integrated Circuits," *Intl. Electronic Device Meeting*, pp. 249-252, 1995.