Timing Analysis in Presence of Power Supply and Ground Voltage Variations*

Rubil Ahmadi
Department of ECE
University of Toronto
Toronto, Ontario, Canada
rubil@eecg.utoronto.ca

Farid N. Najm
Department of ECE
University of Toronto
Toronto, Ontario, Canada
f.najm@utoronto.ca

ABSTRACT

Given the sensitivity of circuit delay to supply and ground voltage values, static timing analysis (STA) must take into account supply voltage variations. Existing STA techniques allow one to verify the timing at different process corners which effectively only considers cases where all the supplies are low or all are high. Cases of mismatch between the supplies of driver and load are not considered. In practice, supply voltages are neither totally independent nor totally dependent. In this work, we consider the supply and ground nodes of a logic gate to be either totally independent variables, or to be directly *tied* or connected to those of some other gate(s) in the circuit. We also assume that the exact supply voltage values are not known exactly, but that only upper/lower bounds on them are known. In this framework, we propose new timing models for logic gates and identify the worst-case voltage configurations for individual gates and for simple paths. We then give an STA technique that provides the worst-case circuit delay taking supply variations into account.

1. INTRODUCTION

Deep sub-micron (DSM) CMOS requires the use of reduced supply voltages (V_{dd}). With a reduced V_{dd}, even a small drop in the local supply voltage can have a significant effect on circuit timing. As we will show below, if the supplies are allowed to vary by up to 12.5%, one can observe (by simulation) up to 2.4X increase in gate delay, in 0.13μm CMOS. Thus, for today's and future technology, the issue of circuit timing is tightly coupled to the question of supply voltage drop. To complicate matters, the sheer size of modern power grids, and the large variety of possible circuit currents, make it very difficult to do an accurate analysis of the grid in order to compute the supply voltage drop. Nevertheless, timing verification must take into account power supply variations.

We are studying the effect of variations of the grid voltages on the circuit timing, and are developing a static timing analysis (STA) approach that takes these variations into account. We assume that the exact voltage drops are not known, but that the *ranges* of voltage drops are specified. A key issue to be considered is whether the voltage drops on the grid are independent variables or not. Obviously, the voltages are neither totally independent (due to the presence of the grid) nor totally dependent (due to the presence of potentially independent circuit currents that load the grid). In this work, we start by assuming that the voltage drops are totally independent and identify the *worst-case* voltage configuration that causes a logic circuit to exhibit its worst-case delay. This question is theoretically interesting, but also practically relevant because if one cannot determine exactly how the grid voltages are dependent, then knowing the worst-case is a useful fall-back position to have available. Then, we consider certain specific types of dependencies among the supply nodes, namely that some may be tied together, and study how the worst-case delay is modified in that case. Based on all this, we then offer an STA algorithm that gives the worst-case circuit delay taking supply variations into account.

*This research was supported in part by Micronet, with funding from ATI and from Altera, and by the SRC under contract 2003-TJ-1070.

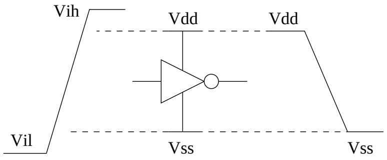


Figure 1: Modeling parameters

Consider the diagram in Fig. 1, where an inverter is shown with its input and output waveforms. The power supply nodes of the inverter are considered the reference V_{dd} and V_{ss}, and the input is assumed to rise from V_{il} to V_{ih}. The output of the inverter, as does the output of its fanout interconnect network, falls from V_{dd} to V_{ss}. It is instructive to consider what is a practical range of variations of the power supply values. In order for the circuit to function properly, the transistors must be able to turn off, which sets a limit on how large the supply variations may be. For one thing, we should have $|V_{ss} - V_{il}| < V_{tn}$ and $|V_{ih} - V_{dd}| < |V_{tp}|$. In the worst-case, if we consider opposite variations for (V_{ss}, V_{il}) and (V_{ih}, V_{dd}), then:

$$|\Delta V_{ss}| + |\Delta V_{il}| < V_{tn} \Rightarrow \text{roughly, } |\Delta V_{ss}| < V_{tn}/2 \quad (1)$$

$$|\Delta V_{dd}| + |\Delta V_{ih}| < |V_{tp}| \Rightarrow \text{roughly, } |\Delta V_{dd}| < |V_{tp}/2| \quad (2)$$

Throughout this work, we have used 0.13μm CMOS technology, with a nominal supply voltage of 1.2V, and we assumed $\pm$12.5% variation around V_{dd} and V_{ss}. This is equivalent to 0.15V fluctuation around the nominal power supply and ground. Therefore, V_{ih} and V_{dd} can vary from 1.05V to 1.35V, and V_{il} and V_{ss} can vary from -0.15V to +0.15V. In the rest of the paper, we will refer to these ranges of voltages as the *valid* or *allowed* supply voltage ranges.

2. MODELING

In order to develop a timing analysis approach in presence of power supply and ground voltage fluctuations, one needs to first develop a delay model for cells and interconnect that is dependent on these voltages. In this section, we will first define delay in a variable voltage environment and then introduce our delay models.

2.1 Delay definition

The notion of signal delay needs careful definition when the supplies are potentially different between the driver and the load. Consider the typical timing waveforms in Fig. 2. The **gate delay** is defined as $t_{d1} = t_2 - t_1$, where t_1 is the time at which the input signal reaches $(V_{ih} + V_{il})/2$ and t_2 is the time at which the output reaches $(V_{dd} + V_{ss})/2$. The **interconnect delay** is defined as $t_{d2} = t_3 - t_2$, where t_2 and t_3 are the times at which the input and the output signals of the interconnect network reach $(V_{dd} + V_{ss})/2$.

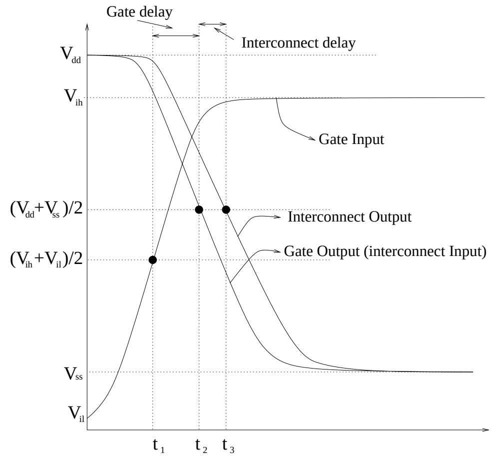


Figure 2: Gate and interconnection delay.

2.2 Gate delay model

The gate delay depends on the traditional parameters of input signal slope and output load. In addition, in this work, we model the dependence of gate delay on the four supply voltages defined above, in Fig. 1. Thus, six parameters are considered as part of the gate delay model: the input high signal level (V_{ih}), the input low signal level (V_{il}), the gate's power supply (V_{dd}), the gate's ground (V_{ss}), the input slope (S_{in}), and the gate's output load (C_l). The **input slope** is defined as the slope (dV/dt) of the input waveform at the time when it crosses $(V_{il} + V_{ih})/2$.

It is instructive to consider how variable the cell delays are, and how strong is their sensitivity to the supply voltages. To this end, we have built a library of cells, containing 2-input and 3-input input NAND, NOR, AND, OR and NOT gates. In our experiments, the load, transistor widths, and four voltage levels of the gates were varied across their valid ranges. Transistor width was allowed to vary from 160nm (the minimum size for 0.13μm technology) to 2400nm, and the loads from 1fF to 32.5 fF (as a comparison, the input capacitance of a minimum size gate for this technology is near 1fF). Furthermore, different combinations of consecutive gates were tested. Fig. 3 shows all possible gate type combinations along with valid parameters ranges. Normalized delay of gates in our library shows that the delay can change by up to 240% (2.4X) due to a 12.5% variation of the supply and ground around their nominal values.

Modern cell libraries represent the delay of cells using four 2-dimensional tables for each timing arc (a timing arc is an input-output node pair). In case of a falling output, one table gives the propagation delay and another gives the output slope. Another two tables correspond to the rising output case. Each table covers the range of valid input slope and output load values. Simple extension of this model to our case would require 6-dimensional tables, which would be impractical in terms of model size and cost of building the model. In order to simplify the model, we found that the delay dependence on each voltage is near-linear in the (narrow) range of valid voltages. However, to be more accurate, we have used a quadratic polynomial to represent the dependence of delay on each voltage, and made allowance for cross-product terms, by using a template expression for delay as follows:

$$t_d = \sum_k \alpha_k V_{ih}^{a_k} V_{il}^{b_k} V_{dd}^{c_k} V_{ss}^{d_k} \quad \text{where} \quad \alpha_k \in \mathcal{R}, \quad \text{and} \quad a_k, b_k, c_k, d_k \in \{0,1,2\} \tag{3}$$

The regression coefficients α_k are found by using a standard Least Mean Square (LMS) regression method [7]. The regression is performed for each grid point in the [slope, load] table, so that each cell in the [slope, load] table contains the values for a number of coefficients $\alpha_1, \alpha_2, \ldots, \alpha_m$. A similar model to this gives the output slope in terms of all four voltages and input slope and output load. We characterized (built the delay models for) all the cells in our library, then tested the error in delay between HSPICE and the library model. The results are shown in Fig. 4 for the propagation delay. It is seen that the model has very good accuracy. The output slope component of the model was also tested, and it shows an error of under 3%, which also is good.

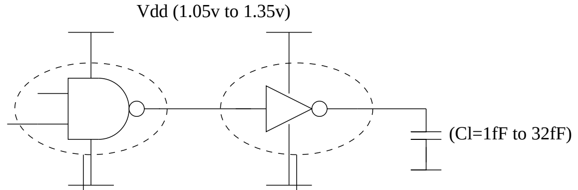


Figure 3: Possible gates and parameters combination.

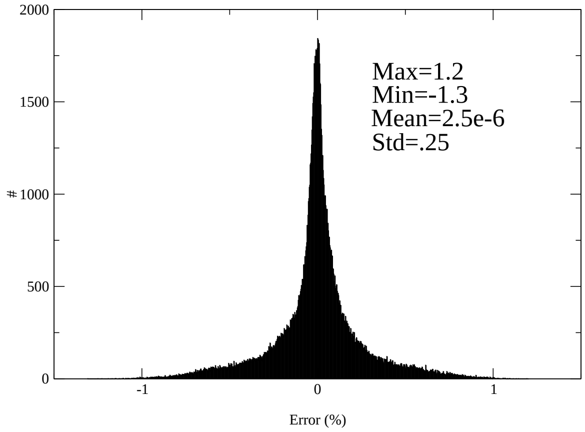


Figure 4: Modeling error.

2.3 Interconnect delay

Interconnect delay can be modeled by any of the modern ways, using either Elmore delay [6], moment matching [5], or other higher-order modeling approaches. The interconnection delay is independent of both the driver and the load gate's voltages, and it just depends on interconnect model values and the transition time of the driver gate. Therefore, the interconnect delay requires no special treatment.

3. WORST-CASE GATE DELAY

Given a logic gate with variable supplies, it is important to look for the supply configuration that gives the worst-case gate delay. The situation is complicated, due to the number of variables involved, especially for complex CMOS gates. We will first consider this in the easy special case of an inverter, where analytical expressions are possible, and then generalize to the case of arbitrary CMOS gates.

3.1 Special case: inverter

In this section, we will consider inverters, with rising and falling input signals. Simple quadratic equations are used for the NFET and PFET transistor currents and a delay expression is derived that shows, among other things, the dependence of delay on the supply and ground voltages. We then consider the sensitivity of the delay to the supply/ground variations and highlight the *sign* of the sensitivity terms, as this will turn out to be important in

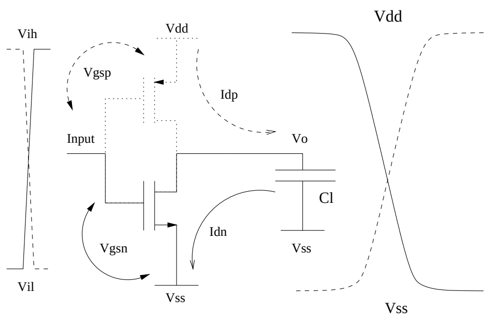


Figure 5: Inverter with falling output.

the rest of the paper. For more complex logic gates, for which analytical results are not possible, we will give empirical data to show the *sign* of the sensitivity terms.

3.1.1 Step input

Fig. 5 shows an inverter with an output load C_l. Let V_{tp} and V_{tn} be the PFET and the NFET threshold voltages, respectively. Let V_{gsp} and V_{gsn} be the gate-source voltage of the PFET and the NFET transistors.

3.1.1.1 Falling delay.

Consider a rising step signal as the input signal of inverter, as shown in Fig. 5. Initially, the input of the inverter is low, the NFET is in cutoff and the PFET in saturation. When the input becomes high, the output load is discharged through the NFET and the output voltage may be found as the solution of the following differential equation:

$$C_l \frac{\partial V_o}{\partial t} = -I_{dn} \tag{4}$$

where $V_o(0) = V_{dd}$ and where:

$$I_{dn} = \begin{cases} 0 & \text{for} \quad V_{gsn} < V_{tn} \\ \beta_n((V_{gsn} - V_{tn})V_o - \frac{V_o^2}{2}) & \text{for} \quad V_o < (V_{gsn} - V_{tn}) \\ \frac{\beta_n}{2}(V_{gsn} - V_{tn})^2 & \text{for} \quad V_o > (V_{gsn} - V_{tn}) \end{cases} \tag{5}$$

where $V_{gsn} = V_{ih} - V_{ss}$. Solving for the falling delay (the time when V_o reaches $\frac{(V_{dd}+V_{ss})}{2}$), leads to:

$$t_{df,step} = \left[\ln\left(\frac{4V_{ih} - 5V_{ss} - 4V_{tn} - V_{dd}}{V_{dd} + V_{ss}}\right) + \frac{2(V_{ss} + V_{tn} + V_{dd} - V_{ih})}{(V_{ih} - V_{ss} - V_{tn})}\right] \frac{C_l}{\beta_n(V_{ih} - V_{ss} - V_{tn})} \tag{6}$$

We define the **sensitivity** of this delay to V_{dd} to be given by $\partial t_{df,step}/\partial V_{dd}$, and likewise for the other voltage variables. These sensitivities can be found analytically by differentiation; it is found that, for the whole range of allowable voltage variations, the sensitivity of this delay to V_{dd} and to V_{ss} is *positive*, and its sensitivity to V_{ih} is *negative*, so that:

$$\frac{\partial t_{df,step}}{\partial V_{dd}} \geq 0 \quad \frac{\partial t_{df,step}}{\partial V_{ss}} \geq 0 \quad \frac{\partial t_{df,step}}{\partial V_{ih}} \leq 0 \quad \frac{\partial t_{df,step}}{\partial V_{il}} = 0 \tag{7}$$

Therefore, the worst-case inverter falling delay may be found by setting $V_{dd} = H$, $V_{ss} = H$ and $V_{ih} = L$ (H stands for the highest allowable value, and L stands for the lowest allowable value), which may be represented by the mnemonic:

$$\begin{pmatrix} L \\ * \end{pmatrix} \begin{pmatrix} H \\ H \end{pmatrix} \tag{8}$$

3.1.1.2 Rising delay.

In the case when the input is a falling step, similar results can be found as follows. While the input signal is initially high, the PFET is in the cutoff mode and the NFET is in saturation. When the input falls, the output load will be charged through PFET, as shown in Fig. 5. The output voltage may be found as the solution of the following differential equation:

$$C_l \frac{\partial V_o}{\partial t} = I_{dp} \tag{9}$$

where $V_o(0) = V_{ss}$ and where:

$$I_{dp} = \begin{cases} 0 & \text{for} \quad V_{gsp} > V_{tp} \\ \beta_p((V_{gsp} - V_{tp})V_{dsp} - \frac{V_{dsp}^2}{2}) & \text{for} \quad V_{dsp} > (V_{gsp} - V_{tp}) \\ \frac{\beta_p}{2}(V_{gsp} - V_{tp})^2 & \text{for} \quad V_{dsp} < (V_{gsp} - V_{tp}) \end{cases} \tag{10}$$

where $V_{gsp} = V_{il} - V_{dd}$ and $V_{dsp} = V_o - V_{dd}$. Solving for the rising delay (the time when V_o reaches $\frac{(V_{dd}+V_{ss})}{2}$) leads to:

$$t_{dr,step} = \frac{C_l}{\beta_p(V_{il} - V_{dd} - V_{tp})} \left[\frac{2(V_{il} - V_{tp} - V_{ss})}{(V_{il} - V_{dd} - V_{tp})} + \ln\left(\frac{(V_{dd} - V_{ss})}{(-4V_{il} + 3V_{dd} + 4V_{tp} + V_{ss})}\right)\right] \tag{11}$$

The sensitivities of this delay to the various voltages can be found analytically. It is seen that $t_{dr,step}$ is independent of V_{ih}, and that the sensitivities to V_{dd} and V_{ss} are both *negative* while the sensitivity to V_{il} is *positive*:

$$\frac{\partial t_{dr,step}}{\partial V_{dd}} \leq 0 \quad \frac{\partial t_{dr,step}}{\partial V_{ss}} \leq 0 \quad \frac{\partial t_{dr,step}}{\partial V_{ih}} = 0 \quad \frac{\partial t_{dr,step}}{\partial V_{il}} \geq 0 \tag{12}$$

Therefore, the worst-case inverter rising delay may be found by setting $V_{dd} = L$, $V_{ss} = L$ and $V_{il} = H$, which may be represented by the mnemonic:

$$\begin{pmatrix} * \\ H \end{pmatrix} \begin{pmatrix} L \\ L \end{pmatrix} \tag{13}$$

3.1.2 Ramp input

The previous two sections were based on an assumption of a *step input*. In order to obtain more realistic results, we consider a saturated ramp input. In this case, analytical results are possible, based on a case analysis [2] in which the input slope value is used to select one of two cases: 1) the input is fast, fast enough that it reaches its final value before the transistor (NFET for rising input, PFET for falling input) exits the saturation region, and 2) the input is slow, slow enough that the transistor (NFET for rising input, PFET for falling input) exits the saturation region before the input reaches its final value.

3.1.2.1 Fast input case.

For the fast input case, new differential equations can be formulated, and the falling and rising delays are given by:

$$t_{df} = t_{df,step} + \left[\frac{V_{ih} + 2V_{tn} + 2V_{ss} - 3V_{il}}{6S}\right] \tag{14}$$

$$t_{dr} = t_{dr,step} + \left[\frac{3V_{ih} - V_{il} - 2V_{dd} - 2V_{tp}}{6S}\right] \tag{15}$$

where S is the slope of input signal. It is helpful to rewrite these equations in the following form:

$$t_{df} = C_l g_f(\mathbf{V}) + \frac{h_f(\mathbf{V})}{S} \tag{16}$$

$$t_{dr} = C_l g_r(\mathbf{V}) + \frac{h_r(\mathbf{V})}{S} \tag{17}$$

where g_f, g_r, h_f and h_r are functions of the four voltages ($\mathbf{V}$ is a vector of the four voltages V_{dd}, V_{ss}, V_{ih}, and V_{il}) whose analytical

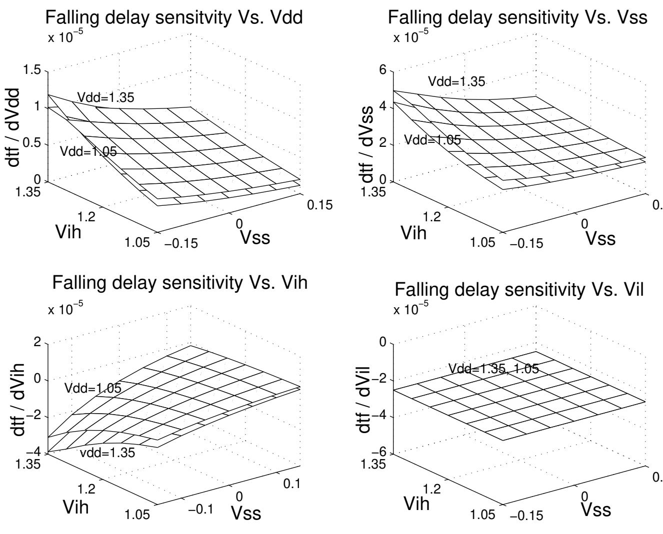


Figure 6: Falling delay sensitivities for ramp input, with V_{il} set at nominal value.

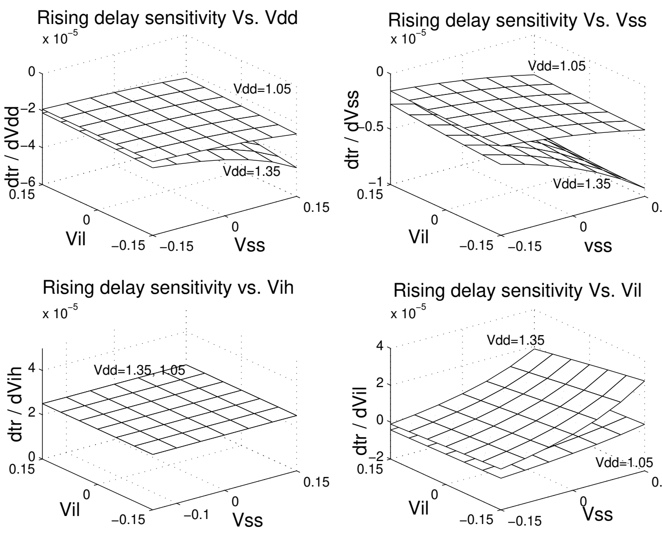


Figure 7: Rising delay sensitivities for ramp input, with V_{ih} set at nominal value.

expressions are clear from (6), (11), (14), and (15). Sensitivities can again be obtained by differentiation, leading to:

$$\frac{\partial t_{df}}{\partial V_*} = C_l \frac{\partial g_f}{\partial V_*} + \frac{1}{S}\frac{\partial h_f}{\partial V_*} \tag{18}$$

$$\frac{\partial t_{dr}}{\partial V_*} = C_l \frac{\partial g_r}{\partial V_*} + \frac{1}{S}\frac{\partial h_r}{\partial V_*} \tag{19}$$

where V_* is any of the four voltages V_{dd}, V_{ss}, V_{ih}, or V_{il}.

Notice that $\partial g_f/\partial V_*$ has the same sign as $\partial t_{df,step}/\partial V_*$, due to (14) and (16) and $\partial g_r/\partial V_*$ has the same sign as $\partial t_{dr,step}/\partial V_*$, due to (15) and (17). Therefore, for a falling output, notice that, whenever $\partial g_f/\partial V_*$ has the same sign as $\partial h_f/\partial V_*$, then $\partial t_{df}/\partial V_*$ has the same sign as $\partial t_{df,step}/\partial V_*$. Thus, the only case when $\partial t_{df}/\partial V_*$ and $\partial t_{df,step}/\partial V_*$ *may* have different signs is when $\partial g_f/\partial V_*$ has a different sign from $\partial h_f/\partial V_*$, which one can easily show occurs only when V_* corresponds to V_{ih} (in which case $\partial g_f/\partial V_{ih}$ is negative and $\partial h_f/\partial V_{ih}$ is positive) and for small values of S and C_l. Since both S and C_l are bounded, we have set them at their minimum values and computed sensitivities for different voltage combinations. Across the whole range of allowed voltages, it was found that $\partial t_{df}/\partial V_*$ has the same sign as $\partial t_{df,step}/\partial V_*$, as can be seen in Fig. 6, which was generated for the case when S and C_l are at their respective minimum values. Therefore, an important conclusion is that, in the fast input case, with the output falling, the sensitivities have the *same* signs as was found in the step input case, for all possible values of input slope and output load:

$$\frac{\partial t_{df}}{\partial V_{dd}} \geq 0 \quad \frac{\partial t_{df}}{\partial V_{ss}} \geq 0 \quad \frac{\partial t_{df}}{\partial V_{ih}} \leq 0 \quad \frac{\partial t_{df}}{\partial V_{il}} \leq 0 \tag{20}$$

A similar analysis applies to t_{dr}. The only case where $\partial g_r/\partial V_*$ has a different sign from $\partial h_r/\partial V_*$ is when V_* corresponds to V_{il}, in which case $\partial g_r/\partial V_{il}$ is positive and $\partial h_r/\partial V_{il}$ is negative. Again, setting both S and C_l to their minima, it was found that, for all voltages in the allowed range, $\partial t_{dr}/\partial V_*$ has the same sign as $\partial t_{dr,step}/\partial V_*$, as can be seen in Fig. 7, which was generated for the case when S and C_l are at their respective minimum values. Therefore, an important conclusion is that, in the fast input case, with the output rising, the sensitivities have the *same* signs as was found in the step input case, for all possible values of input slope and output load:

$$\frac{\partial t_{dr}}{\partial V_{dd}} \leq 0 \quad \frac{\partial t_{dr}}{\partial V_{ss}} \leq 0 \quad \frac{\partial t_{dr}}{\partial V_{ih}} \geq 0 \quad \frac{\partial t_{dr}}{\partial V_{il}} \geq 0 \tag{21}$$

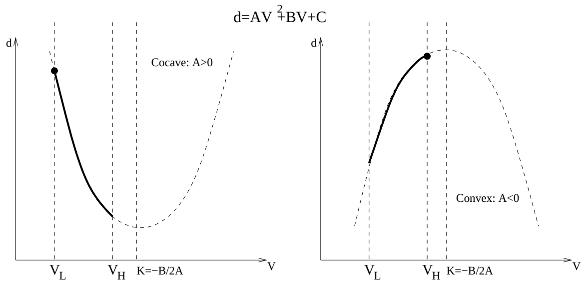


Figure 8: Characteristics of quadratic curves.

3.1.2.2 Slow input case.

For the slow input case, the analysis becomes much more complicated. It is possible to obtain expressions for the rising and falling delays, but the sensitivities were then obtained by numerical differentiation (finite difference approximation). The same results are found, as (20) and (21), for the signs of the sensitivities.

3.1.3 Summary: inverter case

Sensitivities of inverter delay to supply voltage variations have the signs given in (20) and (21), for all possible voltages, slopes, and loads, in the allowed ranges. Correspondingly, the worst-case voltage configuration is given by:

$$\text{for falling output: } \begin{pmatrix} L \\ L \end{pmatrix}\begin{pmatrix} H \\ H \end{pmatrix} \tag{22}$$

$$\text{for rising output: } \begin{pmatrix} H \\ H \end{pmatrix}\begin{pmatrix} L \\ L \end{pmatrix} \tag{23}$$

3.2 General case: arbitrary gates

As we have seen, the delay of a cell depends approximately quadratically on the four voltages. In order to study sensitivities, it is instructive to consider separately the quadratic dependence on each voltage variable. As shown in Fig. 8, if a quadratic function, $Y = AV^2 + BV + C$, is "concave up" (i.e., if $A > 0$), its maximum occurs at a corner of the V-domain. If it is "concave down" (i.e. if $A < 0$), its maximum occurs at $K = \frac{-B}{2A}$ if K is in the V-domain, otherwise the maximum occurs again at a corner of the V-domain.

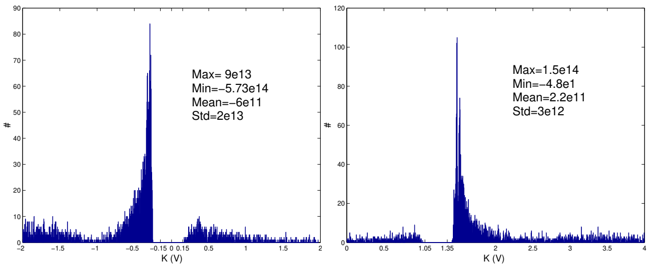
Fig. 9(a) shows data for the two quadratic expressions for V_{ss} and V_{il}. The V-domain in this case is the interval from -0.15V to $+0.15$V. It is seen that K is out of V-domain for all the cells and

Table 1: Inverting gate delay sensitivity.

Input Signal	$\frac{\partial t_d}{\partial V_{dd}}$	$\frac{\partial t_d}{\partial V_{ss}}$	$\frac{\partial t_d}{\partial V_{ih}}$	$\frac{\partial t_d}{\partial V_{il}}$
Rising	$+$	$+$	$-$	$-$
Falling	$-$	$-$	$+$	$+$

all the gate combinations in the library. As a result, in all cases, the maximum of these quadratic terms occurs at a corner of the V-domain. Similar results were obtained, and similar conclusions found, for the quadratic expressions for V_{dd} and V_{ih}, as shown in Fig. 9(b). A corollary conclusion is that *the sensitivity of the delay to a given voltage variable does not change sign as that voltage is varied across its whole range.* It remains always either positive or negative. This is consistent with what was found for inverters, above.

Finally, we considered a cascade of two gates, as in Fig. 3, where the supplies of the "driver gate" are V_{ih} and V_{il} and the supplies of the "load gate" are V_{dd} and V_{ss}, for the following reason. In practice, every CMOS gate is driven by another CMOS gate, so that a variation of the supply and ground of the driver gate would affect its output slope, and hence the input slope of the load gate. Thus, it is important that the sensitivities and the worst-case settings of V_{ih}, V_{il}, V_{dd}, and V_{ss} be made in a realistic situation where changes of V_{ih} and V_{il} have an effect on the input slope of the load gate. Analytical study of this situation is not possible. Instead, all combinations of gates, of varying sizes, were simulated in the configuration of Fig. 3. All inverting gates in our library show the same sign pattern that was found analytically for the inverter, summarized in Table 1, and which leads to the worst-case voltage settings in (22) and (23).



(a) **(b)**

Figure 9: a) K factors for V_{ss} and V_{il}, and b) K factors for V_{dd} and V_{ih}.

3.3 Gates with connected supplies (blocks)

We now extend the analysis to handle combinations of gates whose supplies are not independent. Especially interesting is the special case when several consecutive inverting gates on a path share a common power supply and ground; we call this structure a *block*. Thus, a block may be a simple AND cell from a cell library, or a general path of consecutive inverting gates with connected supplies. The case of a block consisting of a single gate will be considered a degenerate or trivial case, and will be referred to as a *trivial block*. In general, the term "block" will refer to a non-trivial block. For block analysis, analytical methods are not available, and we will use empirical data to study the delay sensitivities.

Recall that, for a case such as in Fig. 10(a), where the output of gate 1 is rising, the worst-case delay of gate 2 corresponds to $\binom{L}{L}\binom{H}{H}$. If the two supplies of the driver and load gates are connected, such as in Fig. 10(b), then the worst-case setting for the delay of gate 2 is simply $\binom{L}{H}$, irrespective of signal polarity in fact. This is a commonly known fact, and can easily be shown analytically by replacing V_{ih} by V_{dd} and V_{il} by V_{ss} in (14) and (15) and differentiating both equations. Indeed, it is not hard to see that irrespective of the type of gates, and the length of the path, for an arrangement such as in Fig. 10(c), the worst-case delay of the block identified in the figure corresponds to $\binom{L}{H}$.

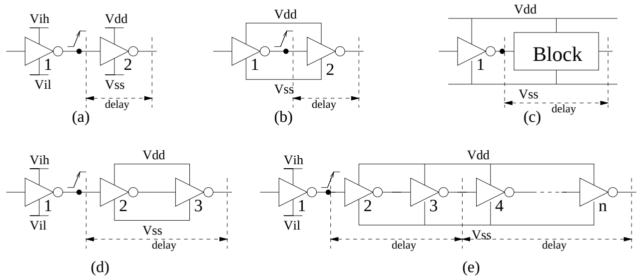


Figure 10: Various test cases.

Table 2: Voltage configurations for worst-case delay

Gate	Input Falling				Input Rising			
	V_{dd}	V_{ss}	V_{ih}	V_{il}	V_{dd}	V_{ss}	V_{ih}	V_{il}
Inverting gate	L	L	H	H	H	H	L	L
Block	L	H	H	H	L	H	L	L

Consider now the case in Fig. 10(d), where, for a rising input to gate 2, we are interested in the delay of the block composed of gates 2 and 3. In this case, and according to the preceding discussion, the worst-case delay of gate 2 is achieved for $V_{dd} = H$, while the worst-case delay of gate 3 requires $V_{dd} = L$. How is this conflict to be resolved? We have found, empirically, that under all conditions of slopes and loads, the sensitivity of gate 3 is dominant, so that the worst-case combination turns out to be $\binom{L}{L}\binom{L}{H}$. This happens because the delay of a logic gate whose output is being pulled high (such as gate 3) is more dependent on the value of V_{dd} than the delay of a gate whose output is being pulled low (such as gate 2). This conclusion was also found to apply for all cases where the gates 1, 2, and 3 are any other inverting gate from our library.

Finally, consider Fig. 10(e). Since it has the same supplies as its driver, the worst-case delay for the block composed of gates 4, 5, $\cdots$, n corresponds to $V_{dd} = L$, $V_{ss} = H$. Since the worst-case delay for the block composed of gates 2 and 3 is *also* $\binom{L}{H}$, then the general conclusion (we have similarly analyzed the falling input case) is that for any (non-trivial) block, the worst-case block delay corresponds to the following:

$$\text{for a rising input: } \begin{pmatrix} L \\ L \end{pmatrix} \begin{pmatrix} L \\ H \end{pmatrix} \tag{24}$$

$$\text{for a falling input: } \begin{pmatrix} H \\ H \end{pmatrix} \begin{pmatrix} L \\ H \end{pmatrix} \tag{25}$$

A summary for the worst-case block delay configurations, for both inverting gates (trivial blocks) and for general (non-trivial) blocks, which includes non-inverting cells, is given in Table 2.

4. WORST-CASE PATH DELAY

The total delay of a signal along a path of gates (specifically, along a path of timing arcs) is the sum of the individual delays of all the timing arcs on the path. Consider all the supply voltages of the gates on a path. If these voltages are viewed as independent variables, then what is the combination of supply values that gives the worst-case path delay? The path delay corresponding to this setting is the absolute worst-case in practice and is therefore worth studying. We first consider this question, and then consider the case when the path includes both gates with independent supplies *and* blocks.

4.1 Gates with independent supplies

Consider the simple 2-gate path shown in Fig. 11, with a falling input. In the following, we will use the following simplified notation so as to simplify the presentation. For a gate "i", we will denote its delay sensitivity to "its" supply voltage as $\partial t_{dri}/\partial V_{dd}$ (for the rising output case), even though that supply node may be labeled differently on the diagram. For instance, in Fig. 11, the sensitivity of gate 1 to its supply voltage will be denoted

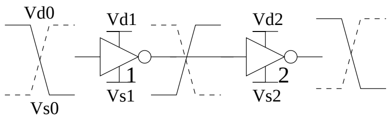


Figure 11: Connected inverting gates with independent supplies and grounds.

$\partial t_{dr1}/\partial V_{dd}$ and the sensitivity of the delay of gate 2 to its supply voltage will be denoted $\partial t_{df2}/\partial V_{dd}$, even though the two supply nodes are labeled V_{d1} and V_{d2} in the figure. Likewise, for the other voltages.

If $t_p = t_{dr1} + t_{df2}$ is the total path delay, then $\Delta t_p = \Delta t_{dr1} + \Delta t_{df2}$ leads to:

$$\Delta t_p = \frac{\partial t_{dr1}}{\partial V_{ih}}\Delta V_{d0} + \frac{\partial t_{dr1}}{\partial V_{il}}\Delta V_{s0} + \frac{\partial t_{dr1}}{\partial V_{dd}}\Delta V_{d1} + \frac{\partial t_{dr1}}{\partial V_{ss}}\Delta V_{s1} + \frac{\partial t_{df2}}{\partial V_{ih}}\Delta V_{d1} + \frac{\partial t_{df2}}{\partial V_{il}}\Delta V_{s1} + \frac{\partial t_{df2}}{\partial V_{dd}}\Delta V_{d2} + \frac{\partial t_{df2}}{\partial V_{ss}}\Delta V_{s2} \quad (26)$$

and, collecting terms, this leads to:

$$\Delta t_p = (\frac{\partial t_{dr1}}{\partial V_{dd}} + \frac{\partial t_{df2}}{\partial V_{ih}})\Delta V_{d1} + (\frac{\partial t_{dr1}}{\partial V_{ss}} + \frac{\partial t_{df2}}{\partial V_{il}})\Delta V_{s1} + (\frac{\partial t_{df2}}{\partial V_{dd}})\Delta V_{d2} + (\frac{\partial t_{df2}}{\partial V_{ss}})\Delta V_{s2} + (\frac{\partial t_{dr1}}{\partial V_{ih}})\Delta V_{d0} + (\frac{\partial t_{dr1}}{\partial V_{il}})\Delta V_{s0} \quad (27)$$

Considering Table 1, it is clear that the coefficient of ΔV_{d1} in (27) is negative. Therefore, in order to have the maximum delay, one should set $V_{d1} = L$. Likewise, for the other voltages, $V_{d0} = H$, $V_{d2} = H$, $V_{s0} = H$, $V_{s1} = L$, and $V_{s2} = H$. For a rising input signal, we have the same expression with different signs, leading to the following worst-case voltage setting: $V_{d0} = L$, $V_{d1} = H$, $V_{d2} = L$, $V_{s0} = L$, $V_{s1} = H$, and $V_{s2} = L$. Since Table 1 is valid for all inverting gates not only inverters, then this result is general and applies to arbitrary inverting gates.

It is interesting that the worst-case delay is so easily identifiable and corresponds to a setting of:

$$\begin{pmatrix} H \\ H \end{pmatrix} \begin{pmatrix} L \\ L \end{pmatrix} \begin{pmatrix} H \\ H \end{pmatrix} \quad (28)$$

for the falling input case, and the opposite setting for the rising input case. The reason this works so well is that the individual worst-case assignments of the gates match exactly due to the reversed polarity of the transitions at the outputs of consecutive gates.

Indeed, it is clear that this result extends naturally to paths of arbitrary length, by induction. Therefore, for a multi-gate path composed of all inverting gates with independent supplies, the worst-case voltage setting for a falling input is given the staggered arrangement:

$$\begin{pmatrix} H \\ H \end{pmatrix} \begin{pmatrix} L \\ L \end{pmatrix} \begin{pmatrix} H \\ H \end{pmatrix} \begin{pmatrix} L \\ L \end{pmatrix} \begin{pmatrix} H \\ H \end{pmatrix} \cdots \quad (29)$$

and, for a rising input, it is given by the alternate staggered arrangement:

$$\begin{pmatrix} L \\ L \end{pmatrix} \begin{pmatrix} H \\ H \end{pmatrix} \begin{pmatrix} L \\ L \end{pmatrix} \begin{pmatrix} H \\ H \end{pmatrix} \begin{pmatrix} L \\ L \end{pmatrix} \cdots \quad (30)$$

4.2 Mix of independent gates and blocks

When considering a path that mixes non-inverting gates and general blocks, then it is possible to observe a *conflict* between the sensitivities to the supplies, so that the solution is not necessarily the nice staggered arrangements seen above. In theory, a conflict in the supply voltage assignment can always be resolved during timing analysis (as will be described below) by generating and following various *alternatives*. The mechanism for doing this will be seen to require the generation of additional "signals" to be propagated during the timing analysis. However, in order to reduce the overhead due to these signals, we will describe ways in which certain conflicts can be resolved easily without the need for additional signals during timing analysis.

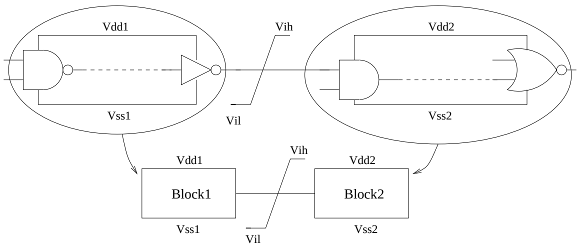
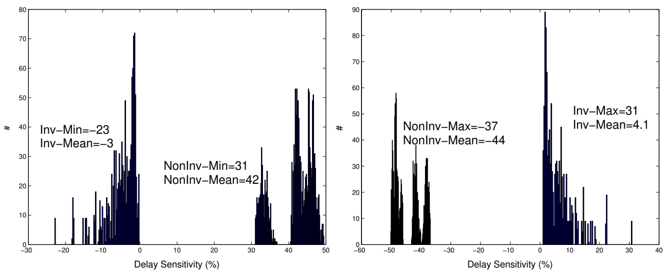


Figure 12: Consecutive blocks with independent supplies

Conflicts can be resolved easily in case of a series connection of two blocks. If the first block is a trivial block (a single inverting gate), then there is actually no conflict. To see this, consider the case when the signal at the intermediate node (input of the 2nd block) is rising. Then, based on the preceding analysis, the worst-case delay of both the gate and the block are achieved when the gate's supplies are set to $\binom{L}{L}$. If that signal is falling then the gate's supplies should be $\binom{H}{H}$ in order to maximize both the gate delay and block delay, and there is no conflict. Conflict arises when the first block is non-trivial, as follows.

Consider two consecutive blocks with independent supplies, as shown in Fig. 12, and consider the case when the output of the first block is rising. The first (non-trivial) block requires a setting of $\binom{L}{H}$ for its supplies. The second block requires a setting of $\binom{L}{L}\binom{H}{H}$. Thus, there is a conflict in the setting of the ground value of the first block. We have found, empirically, that the sensitivity of t_{d2} (delay of block 2) to V_{ss1} is always *smaller* (in magnitude; recall, this sensitivity is negative) than the (positive) sensitivity of t_{d1} (delay of block 1) to V_{ss1}, leading to the conclusion that V_{ss1} must be set to H in order to maximize the path delay. Basically, the sensitivity of the delay of a logic gate to its supply voltage turns out to be larger than its sensitivity to the input signal level. When the intermediate signal is falling, the conflict has to do with the value of V_{dd1}, and we have found that the worst-case corresponds to setting V_{dd1} to L. We now show some empirical justification for these conclusions - further data are available but are not shown due to lack of space.



(a) **(b)**

Figure 13: a) V_{ss} and V_{il} sensitivity for falling output b) V_{dd} and V_{ih} sensitivity for rising output

Fig. 13(a) shows $\partial t_{df2}/\partial V_{il}$ and $\partial t_{dr1}/\partial V_{ss}$ in the same histogram when the first block (Fig. 12) is a cascade of two inverting gates and the second block is a single inverting gate. It is seen that the former sensitivity is negative and the latter is positive, but the minimum value of the latter is greater than the absolute value of the minimum value of the former. Therefore, the sensitivity of the path delay to V_{ss1} is positive and V_{ss1} must be set to H. Fig. 13(b) shows $\partial t_{dr2}/\partial V_{ih}$ and $\partial t_{df1}/\partial V_{dd}$ for the same circuit when the intermediate node is falling. Here, the former sensitivity is positive and the latter is negative, and the maximum value of the former is less than the absolute value of the maximum value of the latter, therefore the overall delay sensitivity of the path to V_{dd1} is negative and V_{dd1} must be set to L. The above data was obtained for all combinations of gates in our library. If

Table 3: STA and HSPICE worst-case delay for the random-range case.

Circuit	STA(ns)			SPICE(ns)		
	Nominal	Min Supply	Worst-Case	Nominal	Min Supply	Worst-Case
C1355	2.48	3.24	3.5 (7%)	2	2.89 (45%)	3.24 (63%)
C1908	3.2	3.97	4.35 (8.75%)	2.7	3.62 (34%)	4 (48%)
C2670	3.4	4.3	4.4 (10%)	3	3.86 (29%)	3.98 (32%)
C3540	4.26	5.4	5.6 (6.7%)	3.7	4.87 (32%)	5.25 (42%)
C432	4.33	5.3	5.5 (5%)	3.9	5 (28%)	5.24 (34%)
C499	1.97	2.6	2.8 (1.5%)	1.77	2.36 (33%)	2.76 (56%)
C5315	3.7	4.75	5.1 (9.6%)	3.24	4.26 (31%)	4.65 (43%)
C7552	2.89	3.66	3.98 (2%)	2.58	3.5 (35%)	3.9 (51%)
C880	2.14	2.6	3 (10%)	1.8	2.3 (26%)	2.73 (50%)
S1494	1.78	2.23	2.3 (0%)	1.67	2.2 (32%)	2.3 (38%)
S420	1.23	1.52	1.61 (14%)	.98	1.32 (35%)	1.41 (43%)
S444	1.13	1.4	1.61 (3%)	.96	1.25 (30%)	1.55 (61%)
S510	1	1.2	1.34 (11%)	.84	1 (19%)	1.2 (43%)
$C1355\_20\_70$	2.4	3.21	3.51 (8%)	2	3 (50%)	3.25 (63%)
$C1908\_10\_50$	3.25	3.81	4 (12%)	2.73	3.28 (20%)	3.54 (30%)
$C432\_3\_80$	4.42	5.4	5.63 (1.2%)	3.93	5.35 (36%)	5.56 (41%)
$C499\_7\_10$	1.87	2.4	2.74 (5.4%)	1.7	2.38 (34%)	2.6 (46%)
$C7552\_11\_75$	2.91	3.4	3.63 (6%)	2.61	3.3 (26%)	3.42 (31%)
$S420\_20\_10$	1.2	1.5	1.58 (12%)	.98	1.27 (30%)	1.4 (42%)
$S444\_3\_60$	1.1	1.45	1.5 (10%)	.94	1.3 (38%)	1.36 (44%)
$S510\_4\_50$	1	1.2	1.26 (14%)	.84	1 (19%)	1.1 (30%)

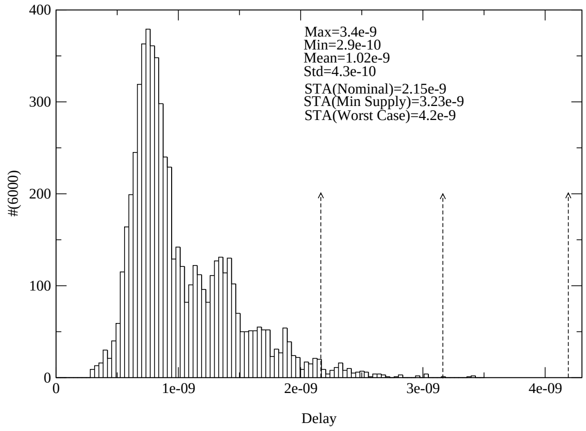


Figure 14: C880 Delay with falling outputs.

the first block is longer than simply two gates, its sensitivity to its supply or ground will only increase (in magnitude) so that the same conclusions hold. If the second block is non-trivial, then it has more delay and its sensitivity to V_{ss1} or V_{dd1} increases in a way which could, in theory, negate our conclusion. However, since the input signal mainly affects the delay of the first one or two gates in the path (again, we have established this empirically but it is not hard to see why it is true), this does not happen, and the conclusion is intact.

5. TIMING ANALYSIS

Static Timing Analysis (STA) gives the maximum delay of a combinational circuit. The available techniques range from the early work of Kirkpatrick [4] and Hitchcock [3], to recent work by Blaauw [1], which is significant in that it carefully takes into account the effect of the input slope on path delay during signal propagation. Our implementation of STA is based on [1]. We consider that supply nodes of the logic gates in a circuit are either tied together, in arbitrary combinations, or are independent.

For each primary input, two *signals* are created, one rising and one falling, each with an arrival time of 0. For each logic gate, we propagate the signals at its inputs to its output, and then we *prune* the signal set at that output node. If the supply nodes of that gate are tied nowhere else, then each signal at a gate's input node yields one signal at the gate's output node, which has arrival time and slope as determined by our timing model, using the worst-case supply settings for that gate and for that polarity of transition. This supply setting becomes part of the signal description, and is carried along. Once all the signals at the gate's inputs have been propagated thus to its output, the signal set at the output is pruned as in [1]. At the circuit primary outputs, the signal with the latest arrival time determines the circuit delay, and the voltage assignment tagged to that signal is the worst-case voltage assignment for that circuit.

If, however, the supply nodes of the logic gate *are* tied elsewhere, meaning that this gate is either part of a block or that this gate's supplies are tied to some other gate's supply elsewhere in the circuit, then a *conflict* may arise in that the voltage assignment that one would like to make for this gate's supplies may conflict with other assignments that are already part of this signal's description, or may conflict with future assignments that one may want to make for these supplies in connection with another gate downstream. Conflict resolution is done by generating extra signals. Each signal at an input of this gate is propagated as *two* new signals at the gate output, each with a different setting of the supplies. Since there are two supplies (V_{dd} and V_{ss}), one would think that four signals would be required. However, we actually use only two signals, which are chosen in a conservative way, meaning that we may err slightly but pessimistically on the delay. At the gate output, signals whose voltage assignments do not conflict are pruned separately, as a sub-group.

6. EXPERIMENTAL RESULTS

The above STA technique was implemented and tested on the ISCAS85 and the combinational parts of the ISCAS89 benchmarks. Experiments were run on a 1 GHz Sun machine with 4 GB memory. The execution time of STA was very fast, under 12 seconds for every circuit that we tested, and less than 1 second for most of them. The results of the analysis of circuit C880 with independent power supplies and grounds, shown in Fig. 14, illustrate a key point. The figure shows a histogram of the circuit delay (using HSPICE) for 6000 different input vector pairs, with a worst-case setting of the supply voltages (within their allowable ranges), as identified by our STA for that circuit. This does not exhaustively cover all vector pairs for this circuit, but will help illustrate the point. The figure also shows the circuit delay as measured by our STA, using three different settings for the supplies. The first setting (Nominal) gives the circuit delay when all supplies are set at their nominal (ideal, no voltage drop) values.

Table 4: **STA and HSPICE worst-case delay for the full-range case.**

Circuit	STA(ns)			SPICE(ns)		
	Nominal	Min Supply	Worst-Case	Nominal	Min Supply	Worst-Case
C1355	2.48	3.83	4.54 (4.6%)	2	3.41 (70%)	4.34 (117%)
C1908	3.15	4.96	5.9 (1%)	2.8	4.71 (68%)	5.84 (108%)
C2670	3.39	5.44	5.82 (8.5%)	3	5 (66%)	5.36 (78%)
C3540	4.26	6.88	7.43 (13%)	3.5	5.9 (68%)	6.55 (97%)
C432	4.33	6.67	7.22 (1.7%)	3.6	6.27 (74%)	7.1 (97%)
C499	1.97	3.13	3.67 (-4.2%)	1.8	3 (66%)	3.8 (111%)
C5315	3.7	5.91	6.79 (4.3%)	3.3	5.45 (65%)	6.51 (97%)
C7552	2.89	4.56	5.48 (-1%)	2.7	4.41 (63%)	5.54 (105%)
C880	2.15	3.23	4.2 (0.6%)	1.85	2.97 (60%)	4.16 (125%)
S1494	1.78	2.86	3 (1.7%)	1.68	2.75 (63%)	2.95 (76%)
S420	1.23	1.88	2.13 (6.5%)	1	1.8 (80%)	2 (100%)
S444	1.13	1.67	2.16 (-6.5%)	.96	1.6 (67%)	2.31 (140%)
S510	.99	1.48	1.85 (2.8%)	.83	1.32 (59%)	1.8 (116%)
$C1355\_20\_70$	2.48	3.85	4.43 (3%)	2	3.5 (75%)	4.27 (113%)
$C1908\_10\_50$	3.15	4.97	5.58 (-2%)	2.81	4.7 (67%)	5.72 (103%)
$C432\_3\_80$	4.39	6.7	6.8 (6.25%)	3.75	6.35 (69%)	6.4 (70%)
$C499\_7\_10$	2	3.14	3.64 (-4%)	1.77	3 (69%)	3.8 (114%)
$C7552\_11\_75$	2.58	4.52	4.8 (-2%)	2.51	4.23 (68%)	4.9 (95%)
$S420\_20\_10$	1.23	1.88	2.137 (6.85%)	1	1.8 (80%)	2 (100%)
$S444\_3\_60$	1.11	1.63	1.72 (1.8%)	.94	1.54 (63%)	1.69 (80%)
$S510\_4\_50$	.91	1.35	1.61 (12%)	.79	1.27 (60%)	1.44 (82%)

It is clear from the figure that this significantly under-estimates the circuit delay. The second (Min Supply) setting corresponds to the case when all V_{dd} supplies are set to low and all grounds to high, within their allowable ranges. This case corresponds to what one is able to do today with existing STA tools. Here too, it is clear that this analysis is not adequate because there are paths with longer delay than that given by the Min Supply setting. Finally, the third setting corresponds to the case where our STA considers all possible mismatches between the supply nodes and finds the maximum delay, in this case assuming that all supplies are independent. Note that there are no vector pairs that violate our estimate of worst-case delay.

Further results on all the benchmarks are presented in Tables 3 and 4. We considered cases of independent supplies and cases of connected supplies. The name of each circuit indicates what was done in each case. For example $C1355\_20\_70$ is a variant of circuit C1355 which has 20 power supply and ground nodes that are shared by 70% of the gates, according to some randomly chosen assignment, while the remainder of the gates have independent supplies. The two tables differ in the way that the allowable supply ranges were assigned. In Table 3, we assigned to each supply node an arbitrarily chosen voltage range that is within the allowable $\pm 12.5\%$ range. We simply used a random number generator to choose these less-than-full ranges. In Table 4, we gave each supply node the maximum allowable voltage range, i.e., $\pm 12.5\%$ of nominal.

Each table gives the delay values measured by our STA and by HSPICE in the three cases of Nominal, Min Supply, and Worst-Case, explained above. The percentage values given in parentheses represent the relative increase of delay over the Nominal case. Getting the exact delay using HSPICE is not possible because of the large number of possible vector pairs. Therefore, for each circuit, once the critical path is identified by our STA, we extract that path and simulate it with HSPICE. Notice that the critical path may be different in the Nominal, Min Supply, and Worst-Case scenarios.

Notice that the delays under the SPICE Min Supply column are higher than the delays of the Nominal case. The advantage of our technique, and the need for it, are evident from the last column (SPICE, Worst-Case). The *significant* increase of delay over Nominal and over Min Supply underscores the fact that allowing mismatch between the supplies leads to a higher worst-case delay. Finally, notice that the delay comparisons between the corresponding columns of STA and HSPICE are very good, and show that the gate delay model works well in this case.

7. CONCLUSION

Motivated by the sensitivity of circuit delay to supply voltage variations, we are working on new techniques for static timing analysis (STA) that are cognizant of the dependence of delay on supply voltage. Specifically, we have explored the case where the supply voltage values of a driver gate and a load gate can be mismatched. This can happen due to the fact that circuit currents may be independent. It was found that this mismatch leads to potentially large increase in delay, up to 2.4X in one case, due to only $\pm 12.5\%$ variation in supply voltage. We developed a model for gate delay in terms of the four voltages of a gate and its driver. If the supplies are independent, we showed that there is an easy-to-find configuration of supply and ground voltages along a path that leads to the worst-case delay, depending on the polarity of the transition at the input. We also considered specific power supply and ground dependencies, namely that some supply nodes are tied together. We then used this as the basis for an STA technique which gives the circuit delay as well as the worst-case voltage configuration. With the voltage configuration in-hand, we simulated the critical paths with HSPICE and showed how ignoring the possible mismatch between the supplies can lead to significant under-estimation of circuit delay.

8. REFERENCES

[1] D. Blaauw, V. Zolotov, and S. Sundareswaran. Slope propagation in static timing analysis. *IEEE Transactions on Computer-Aided Design*, 21(10):1180–1195, Oct. 2002.

[2] N. Hedenstierna and K. O. Jeppson. CMOS circuit speed and buffer optimization. *IEEE Trans. on Computer-Aided Design*, 6(2):270–281, March 1987.

[3] R. B. Hitchcock, S. G. L. Smith, and D. D. Cheng. Timing analysis of computer hardware. *IBM Journal of Research and Development*, 26(1):100–105, Jan. 1982.

[4] T. I. Kirkpatrick and N. R. Clark. PERT as an aid to logic design. *IBM Journal of Research and Development*, 10(2):135–141, March 1966.

[5] C. L. Ratzlaff and L. T. Pillage. RICE: rapid interconnect circuit evaluation using AWE. *IEEE Trans. on Computer-Aided Design*, 13(6):763–776, June 1994.

[6] J. Rubenstein, P. Penfield, Jr., and M. A. Horowitz. Signal delay in RC tree networks. *IEEE Transactions on Computer-Aided Design*, CAD-2(3):202–211, July 1983.

[7] B. Widrow. *Adaptive signal processing*. Prentice-Hall, 1st edition, 1985.