

On the signal bounding problem in Timing analysis

Jin-Fuw Lee¹, D. L. Ostapko¹, Jeffery Soreff², C. K. Wong³

¹IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598.

²IBM Fishkill, Rt. 52, E. Fishkill, Hopewell Junction, NY 12533.

³Dept. of Computer Science and Engineering, The Chinese Univ. of Hong Kong, Shatin, N.T., Hong Kong.

{jinfuw, ostapko}@us.ibm.com
soreff@us.ibm.com

wongck@cse.cuhk.edu.hk

Abstract

In this paper, we study the propagation of slew dependent bounding signals and the corresponding slew problem in static timing analysis. The selection of slew from the latest arriving signal, a commonly used strategy, may violate the rule of monotonic delay. Several methods for generating bounding signals to overcome this difficulty are described. The accuracy and monotonicity of each method is analyzed. These methods can be easily implemented in a static timer to improve the accuracy.

1. Introduction

Timing closure is a major problem of chip design, especially in deep sub micron technology. There have been two kinds of timing analysis tools, dynamic and static timers. Dynamic timers are based on Spice or a fast circuit simulator, for which the users need to specify input patterns. The number of input patterns may become very large, if the worst case behavior of the circuit is not known a priori. Static timers [1][2] are based on the propagation of bounding signals representing the worst-case and best-case delays (arrival times) and slews (rise and fall times), and the users do not have to specify the input pattern. In general, static timers are less accurate but much faster than dynamic timers, and are usually used for analyzing a large design such as a complete chip. An alternative method to bounding signals, which employs the propagation of multiple signals, has been proposed in [3].

The bounding signal approach has been adopted in the development of many circuit optimization tools to reduce the complexity. For example, in [4], the maximum slew and maximum arrival time was used for circuit tuning. In [5], the delay model with linear slew dependence is used for optimizing interconnect sizing and buffer insertion. More complicated models for slew dependent delay were described in [6-7]. The bounding signal method may be applied to the problem of simultaneous switching [8-9]: When two transistors in series (parallel) switch simultaneously, the delay is worse (better) than the switching of any single transistor. The coupling noise [10] becomes significant in deep sub-micron technology, and the worst case delay arises when the victim net has the slowest slew while the aggressor net has the fastest slew.

In this paper, we shall study the bounding signal problem in static timing analysis. Signals arriving at a given node of a circuit may be generated from different input patterns and may travel

through different paths. In one common approach, the latest arriving signal with its slew is selected to represent the worst-case timing bound. The *latest arriving method* has two potential pitfalls. Firstly, the worst-case timing bound calculated may be invalid, since an earlier signal with a slow slew may eventually reach the primary output later. For example in Figure 1, assume that the arrival time (a_f) of the fast signal at the output of gate 1 is later than that (a_s) of the slow signal. Let the delay through gate 2 be d_f and d_s , respectively for the fast and slow signals. Since $d_f < d_s$, we may reach the situation $a_f + d_f < a_s + d_s$, and the slow signal becomes the latest one at the output of gate 2. Secondly, the monotonic property of signal propagation may be violated. By monotonic property, we mean that a speeding up (slowing down) of the arrival time of any signal must also speed up (slow down) the arrival time of the circuit output. Let us assume that in Figure 1, the fast signal is only slightly later than the slow signal at the output of gate 1. This fast signal is chosen to propagate through gate 2, and the worst-case arrival time calculated at the output of gate 2 is $a_f + d_f$. If we speed up the fast signal so that its arrival time at the output of gate 1 is $a_f < a_s$, then the slow signal is chosen to propagate, and the worst-case time at the output of gate 2 becomes $a_s + d_s$, which is greater than $a_f + d_f$. The monotonic property is thus violated.

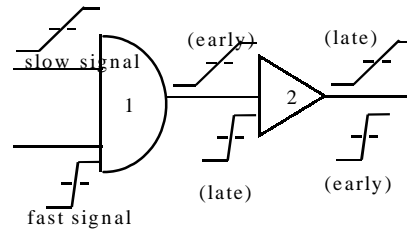


Figure 1: Dashed lines indicate the 0.5VDD points of signals.

To avoid the above problems, in some static timers, the maximum slews are derived separately, combined with the latest arrival times, and then used in the timing computation. Since the signal with the maximum slew may have a very early arrival time, the worst case time predicted is too pessimistic. In some static timers, a look-ahead algorithm is used first to search through one or more fan-out stages of gates to identify the potential worst slew which can cause the maximum delay. However, this heuristic method does not guarantee that the slew and arrival time calculated is a worst-case bound. In a recent paper [3], a pruned set of multiple worst-case candidate signals

are propagated. This can be a practical approach, if the computation time is not constrained. However, for a fast static timer or an efficient circuit tuner, the bounding signal approach is advantageous.

In this paper, we explore several new ways to improve the bounding signals. A method for comparing different bounding signals is presented. Slow sensitivities of delays are presented in Section 2. In Section 3, several construction methods for strictly bounding signals are described. In Section 4, a formal definition of bounding signal is given first and a systematic method to construct all types of bounding signals is then derived. In Section 5, a multiple signal approach is presented.

2. Characteristics of Slew dependence

Signal waveforms traveling on a VLSI chip may be approximated by the ramp function. The arrival time a is defined as the time the voltage crosses 0.5 VDD, while the slew s is typically defined as 1.25 times the duration the voltage swings between 0.1 VDD and 0.9 VDD. When there is no confusion, the signal waveform defined by $w=(s, a)$ will be called *signal*.

For the delay propagation through gates, a static timer may evaluate the delay either with some analytical delay rules or with a fast simulator on the fly. Let the arrival time and slew at the input pin j be a_j and s_j . The pin-to-pin delay and the slew at the output pin, d_{ji} and s_{ji} , are functions of the input slew:

$$d_{ji} = f_{ji}(s_j) \quad (1)$$

$$s_{ji} = g_{ji}(s_j) \quad (2)$$

where g_{ji} and f_{ji} represent respectively the slew and the delay for the path from pin j to pin i . These delay rules, Equations 1 and 2, may be a simple unit delay model, a linear delay model, to a piece-wise quadratic delay model with coefficients extracted from the curve-fitting of many simulation runs. The dependence on other circuit parameters such as capacitance loading is dropped for clarity.

Definition: For a function $f(s)$ of slew, the average and instantaneous slew sensitivity are respectively defined as $\Delta f/\Delta s$ and df/ds . The maximum and minimum sensitivity are defined as $D_M f = \max_s df/ds$, and $D_m f = \min_s df/ds$.

By the mean-value theorem of Calculus, for any slew interval, there exists a point with the instantaneous sensitivity equal to the average slew sensitivity. Hence the average slew sensitivities are bounded in the same way as the instantaneous sensitivity:

$$\alpha_{mi} \leq \Delta f_{ji}/\Delta s \leq \alpha_{Mi} \text{ where } \alpha_{mi}=D_m f_{ji} \text{ and } \alpha_{Mi}=D_M f_{ji} \quad (3)$$

$$\beta_{mi} \leq \Delta g_{ji}/\Delta s \leq \beta_{Mi} \text{ where } \beta_{mi}=D_m g_{ji} \text{ and } \beta_{Mi}=D_M g_{ji} \quad (4)$$

Ckt/sensitivity	$\Delta f/\Delta s$	$\Delta g/\Delta s$
NAND2	0.04-0.20	0.29-0.36
NOR2	0.05-0.16	0.27-0.35
NAND3	0.01-0.22	0.20-0.35
NOR3	0.03-0.20	0.20-0.36
AOI22	0.06-0.20	0.18-0.37
OAI22	0.06-0.20	0.18-0.37

Table 1: the slow sensitivities derived from spice simulations.

The slow sensitivities are in general much less than 1 so that the slow dependence on input slew will diminish after propagating through several stages of gates [7]. From a detailed spice simulation for static CMOS gates under different slew(from 50ps to 400ps) and load conditions(from 25ff to 150 ff), the range of slow sensitivities are compiled in Table 1. The maximum sensitivities in Table 1 are $\alpha_M = \max\{\alpha_{Mi}\} \leq 0.25$ and $\beta_M = \max\{\beta_{Mi}\} \leq 0.4$. The positive minimum sensitivities, $\alpha_m = \min\{\alpha_{mi}\}$ and $\beta_m = \min\{\beta_{mi}\}$, from this table means that both f_{ji} and g_{ji} are monotonic increasing functions of slew. In the rare case when a slow signal drives a low-threshold dynamic circuit, α_m may become negative, and a slow increase might actually decrease the delay, when delay is defined as the time interval between 50% points of the input and output. We shall assume the following about β 's:

$$1 \geq \beta_{mi}, \beta_{Mi} \geq 0 \quad (5)$$

For deriving theoretical bounds on α 's, we require the following definition and assumption.

Definition: For a signal $w=(s,a)$, $b=a-s/2$ and $t=a+s/2$ are defined respectively as its 0% and 100% times. Two signals, w_i and w_j , are defined as overlapping if the transient portions of two signal waveforms overlaps. That is, the difference of the 0% times has a sign opposite to that of the 100% times.

Assumption (Causality): If two signals, w_i and w_j , present at the same input are non-overlapping in the order $a_i < a_j$, then the corresponding signals at the output will remain non-overlapping in the same order.

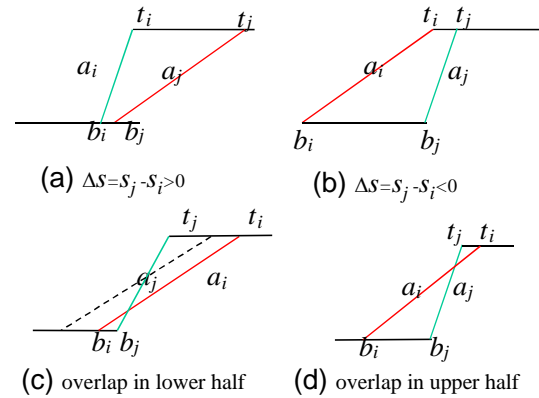


Figure 2: Comparison of two signals, w_i and w_j .

For example, signals in Figure 2(a) and (b), w_i and w_j , are non-overlapping in the order $a_i < a_j$. Algebraically, this condition is $\Delta b = b_j - b_i \geq 0$ and $\Delta t = t_j - t_i \geq 0$. In terms of $\Delta a = a_j - a_i$ and $\Delta s = s_j - s_i$, this yields $\Delta(a-s/2) \geq 0$ and $\Delta(a+s/2) \geq 0$, or $\Delta a \geq |\Delta s/2|$ in short. According to the causality assumption, the signals at the gate output, w'_i and w'_j , will be non-overlapping in the same order $a'_i < a'_j$; $\Delta a' \geq |\Delta s'/2|$. From the gate delay rules, $\Delta a' = \Delta a + \Delta f$ and $\Delta s' = \Delta g$ where $\Delta f = f(s_j) - f(s_i)$ and $\Delta g = g(s_j) - g(s_i)$. The sufficient condition for $\Delta a' \geq |\Delta s'/2|$ can be rewritten as

$$2(\Delta a + \Delta f) \geq |\Delta g| \quad (6)$$

For the case $\Delta s = s_j - s_i > 0$ as shown in Figure 2(a), divide Equation 6 with Δs and we have $2\Delta a/\Delta s + 2\Delta f/\Delta s \geq |\Delta g/\Delta s|$. This inequality is true, if the minimum of the left hand side is not smaller than the maximum of the right hand side. The minimum

of $\Delta a/|\Delta s|$ for non-overlapping signals is 1/2. The minimum of $\Delta f/|\Delta s|$ is α_m , while the maximum of the right side is β_M . Hence

$$1+2\alpha_m \geq \beta_M \quad (7)$$

For the case $\Delta s < 0$ as shown in Figure 2(b), divide Equation 6 with $-\Delta s$ (a positive number) and we have $2\Delta a/|\Delta s| \geq 2\Delta f/|\Delta s| + |\Delta g/|\Delta s|$. This inequality is true, if the minimum of the left hand side is not smaller than the maximum of the right hand side. The minimum of $\Delta a/|\Delta s|$ for non-overlapping signals is 1/2. The maximum of $\Delta f/|\Delta s|$ is α_M , while the maximum of the right side is β_M . This means that

$$1 \geq 2\alpha_M + \beta_M \quad (8)$$

Theorem 1: Under the causality assumption, the bounds on α 's are $(1 - \beta_M)/2 \geq \alpha_M$, $\alpha_m \geq (\beta_M - 1)/2$. If $1 \geq \beta_M \geq 0$, then $0.5 \geq \alpha_M$, $\alpha_m \geq -0.5$.

In this paper, we shall also assume that the wiring delay is either negligible, or can be modeled by Equations 3 to 8.

3. Constructing bounding signals for positive α_m

Given a set of signals $W = \{...w_i...\}$ at the same gate input pin, the strictly bounding signal is defined as a representative signal which will arrive later than all the signals in W at the gate output.

Definition: For a pair of signals, $w_i = (s_i, a_i)$ and $w_j = (s_j, a_j)$, w_i strictly dominates w_j if $a_j \leq a_i$ and $t_j \leq t_i$ (100% times).

Theorem 2: Let w_i, w_j be signals at gate input and w'_i, w'_j be the corresponding signals at gate output. If $\alpha_m > 0$ and w_i strictly dominates w_j , then w'_i strictly dominates w'_j .

Proof: If w_i and w_j do not overlap, then from the causality assumption, w'_i will be later, and strictly dominates w'_j . If w_i and w_j overlap, they only overlap in the lower half voltage interval $(0, VDD/2)$. Let $w = (s, a)$ be the signal shown as the dashed line in Figure 2(c). At the gate output, w'_j due to a faster slew is earlier than w' and w' is earlier than w'_i by the amount $a_i - a_j$. So w'_i strictly dominates w'_j .

Hence if signals in W do not overlap in the upper half voltage interval $(VDD/2, VDD)$ as shown in Figure 2(d), the signal which strictly dominates will be the bounding signal. In the general case, the bounding signal may not be found inside W , and a new one must be constructed. The methods of constructing bounding signals are presented below along with the latest arriving method:

1. Latest arriving method: Pick the signal $w_{late} = (s_{late}, a_{late})$ with the maximum arrival time. See the solid line through a_j in Figure 3.
2. Maximum slew method [4]: Construct a signal, $w_{slew} = (s_{slew}, a_{slew})$, with the maximum slew and arrival time. See the dash-dotted line marked "S" in Figure 3.
3. Full envelope method[11]: Construct a signal, $w_{full} = (s_{full}, a_{full})$ which covers all the signals from the right side. See the dotted line marked "F" in Figure 3.
4. Half envelope method: Construct a signal, $w_{half} = (s_{half}, a_{half})$ which covers all the signals in the upper half voltage range

$(VDD/2, VDD)$ from the right side. See the long dashed line marked "H" in Figure 3.

In summary, the representative signals are :

$$a_{late} = \max_i a_i, \quad s_{late} = s_{late}$$

$$a_{slew} = \max_i a_i, \quad s_{slew} = \max_i s_i$$

$$a_{full} = (\max_i b_i + \max_i t_i)/2, \quad s_{full} = \max_i t_i - \max_i b_i \quad (9)$$

$$a_{half} = \max_i a_i, \quad s_{half} = 2(\max_i t_i - \max_i a_i)$$

It can be shown from Theorem 2 that w_{slew} , w_{full} , and w_{half} are indeed bounding signals, while w_{late} may not be so. Since only the upper half voltage range is relevant in this positive α_m case, w_{half} yields the tightest upper bound among these methods.

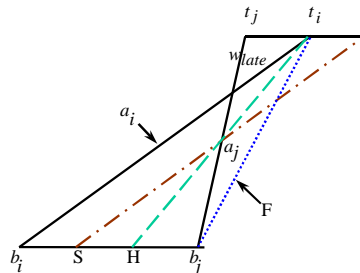


Figure 3: $W = \{w_i, w_j\}$.

There is some accuracy penalty paid when W is replaced by any of the above bounding signals. Assume that at gate input W contains two signals w_i and w_j , with $s_i > s_j$. Let w'_i and w'_j be the corresponding signals at the gate output. Let $a = \max(a_i, a_j)$ and $\Delta a = a_i - a_j$. Then the worst-case arrival time at the gate output is $a' = \max(a'_i, a'_j)$. For comparison, a' and the arrival times at output for methods 1-4 are plotted in Figure 4 as functions of Δa with a fixed at 0. The half envelope method (dashed line) has less error than both max slew (dash-dotted line) and full envelope method (dotted line), while the latest arriving method (dash-dot-dotted line) may underestimate the worst arrival time.

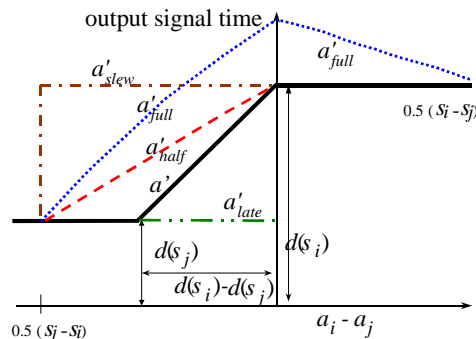


Figure 4: The arrival times at gate output as functions of $a_i - a_j$.

4. Generating constraints for bounding signals

Methods 2-4 in the previous section require that both the 100% and 50% times be bounded during the propagation through all paths. The criterion for the bounding signal may be relaxed as follows. Assume that $w_i = (s_i, a_i)$ and $w_j = (s_j, a_j)$, present at the same gate input pin, propagate through fan-out paths to the

latches. If one signal, say w_i , arrives at all the catching latches in the fan-out cone always with a later 50% time, w_i will impose a stronger timing constraint against the clock edges and shall be considered as the worst-case signal. With this new criterion, the 50% time is not bounded in the intermediate stages.

Definition: For a pair of signals, $w_i=(s_i, a_i)$ and $w_j=(s_j, a_j)$, at the output pin of gate k , let a_{in} and a_{jn} be the 50% arrival time at latch n on the fan-out cone. $w_i \leq w_j$ is defined by the criterion that $a_{in} \leq a_{jn}$ for every latch n on the fan-out cone of gate k . We shall say that w_i is **dominated** by w_j , or w_i **dominates** w_j .

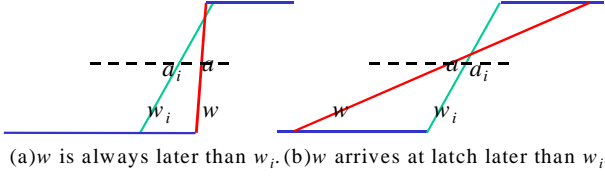


Figure 5: Two cases where w dominates w_i

To explore the sufficient conditions for $w_i \leq w$, consider two extreme situations: In the case shown in Figure 5(a), a faster signal w is initially later ($a \geq a_i$ and $s \leq s_i$), and the excess delay gain by w_i over w through the fan-out propagation is still small enough that the arrival time of w is always later. In the case shown in Figure 5(b), a slower signal w is initially slightly earlier ($a \leq a_i$ and $s \geq s_i$), but picks up enough delay gain over w_i during the propagation and eventually appears at latches with a later arrival time.

Let us examine more closely the new criterion for a dominating signal. First a path, which consists of multiple stages of gates, can be regarded as one composite delay element with its own delay rules. The cumulative path delay and the output signal slew at the end of path p are functions of the input signal slew s :

$$d_p = f_p(s) \text{ and } s_p = g_p(s) \quad (10)$$

Let the minimum and maximum slew sensitivity of f_p (g_p) be α_{mp} (β_{mp}) and α_{Mp} (β_{Mp}). In Section 4a, we'll describe a method to compute these sensitivities. Assume that signals $w=(s, a)$ and $w_i=(s_i, a_i)$ propagate from the same gate k through a fan-out path p to the capturing latch n . Then the difference in path delays, $d_p - d_{ip}$, of two signals is bounded:

$$\alpha_{mp} \leq (d_p - d_{ip}) / (s - s_i) \leq \alpha_{Mp} \text{ where } \alpha_{mp} = D_{mf_p} \text{ and } \alpha_{Mp} = D_{Mf_p} \quad (11)$$

Let the arrival times of two signals at latch n be a_n and a_{in} . Since $d_p - d_{ip} = (a_n - a) - (a_{in} - a_i) = (a_n - a_{in}) - (a - a_i)$, this leads to

$$(a - a_i) / (s - s_i) + \alpha_{mp} \leq (a_n - a_{in}) / (s - s_i) \leq (a - a_i) / (s - s_i) + \alpha_{Mp}. \quad (12)$$

For the smaller slew case $s_i \geq s$ in Figure 5(a), a sufficient condition to assure $(a_n - a_{in}) / (s - s_i) \leq 0$ and hence $a_{in} \leq a_n$ is that the upper bound on the right-hand side of Equation 12 is negative or zero. This condition, which is equivalent to $\alpha_{Mp} \leq -(a - a_i) / (s - s_i)$, must hold for every fan-out path from gate k to latches. Hence

$$-(a - a_i) / (s - s_i) \geq r_M \text{ or } a + r_M s \geq a_i + r_M s_i \text{ where } r_M = \max_p \alpha_{Mp} \quad (13)$$

Consider the path with maximum slew sensitivity: $\alpha_{Mp} = r_M$. Equation 13 implies that the signal with a larger $a + r_M s$ will reach

the end of the path with a later arrival time a_n , and hence a_n is a monotonic increasing function of $a + r_M s$. We shall call $y = a + r_M s$ the **most slew-sensitive time** of w .

For the larger slew case $s_i \leq s$ in Figure 5(b), a sufficient condition to assure $(a_n - a_{in}) / (s - s_i) \geq 0$ and hence $a_{in} \leq a_n$ is that the lower bound on the left-hand side of Equation 12 is positive or zero. This condition, which is equivalent to $-(a - a_i) / (s - s_i) \leq \alpha_{mp}$, must hold for every fan-out path from gate k to latches. Hence

$$-(a - a_i) / (s - s_i) \leq r_m \text{ or } a + r_m s \geq a_i + r_m s_i \text{ where } r_m = \min_p \alpha_{mp}. \quad (14)$$

Consider the path with minimum slew sensitivity: $\alpha_{mp} = r_m$. Equation 14 implies that the signal with a larger $a + r_m s$ will reach the end of this path with a later arrival time a_n , and hence a_n is a monotonic increasing function of $a + r_m s$. We shall call $x = a + r_m s$ the **least slew-sensitive time** of w .

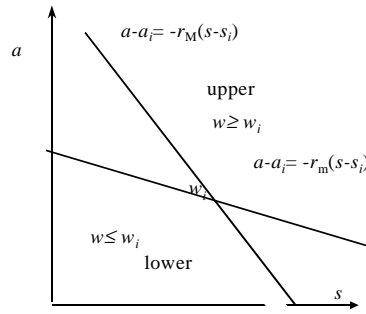


Figure 6: The bound quadrants of signal w_i .

Signal w dominates w_i , if both Equations 13 and 14, are satisfied. Hereafter, they will be called as **max** and **min constraint** equations on w_i , and r_M and r_m as the slew sensitivity of max and min constraints. These constraint equations define a feasible region for $w_i \leq w$, which is marked as the upper quadrant in Figure 6. Extension of the two lines representing constraint equations will divide the $a-s$ plane into four quadrants. For signal w in the upper quadrant, $w_i \leq w$. For signal w in the lower quadrant, w_i will be in the region bounded by the constraint equations on w and hence $w \leq w_i$. In either cases, we shall say that two signals are **comparable**. Signals outside these two quadrants are **non-comparable**.

Bounded regions of two or more signals

Assume that two signals, w_i and w_j , are present at the same output pin of a gate. We would like to construct an upper-bound signal $w=(s, a)$ which dominates both w_i and w_j . According to the definition of dominance, w must satisfy the max and min constraints for both w_i and w_j , and thus fall into the intersection of upper quadrants of w_i and w_j . If the two signals are comparable, the intersection is equal to the upper quadrant from the dominant signal. Any signal in this quadrant will dominate both signals. The dominant signal, which also resides in this quadrant, is the least upper bounding signal. In Figure 7(b), the fast signal w_j is the dominant signal and w must fall into the upper quadrant of w_j . In Figure 7(a), the slow signal w_i is the

dominant signal and w must fall into the upper quadrant of w_i . For the non-comparable case in Figure 7(c), the intersection of two upper quadrants can be obtained as follows: First take the min constraint equation on w_j and the max constraint equation on w_i . Then find the intersection, U, of the two lines. Any signal in the upper quadrant of U will dominate both w_i and w_j , and U is the least upper-bound signal. In the case that $r_m < 0$, the slopes of lines representing the min constraints become positive as shown in Figure 7(d). Note that for non-comparable signals, the least upper-bound signal differs from either of the original signals.

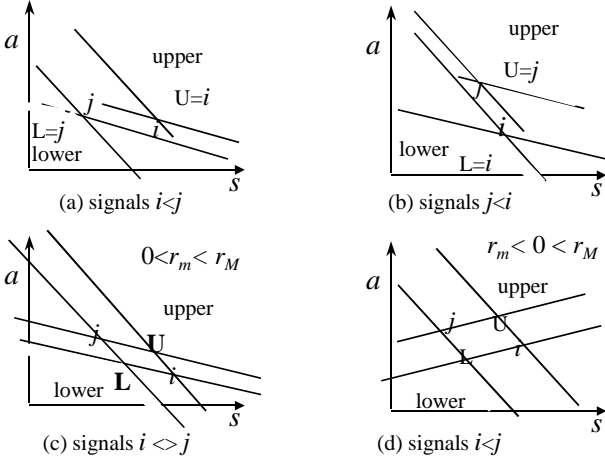


Figure 7: The upper-bound and lower-bound of two signals.

For the general bounding problem involving a set of signals $W = \{w_1, w_2, \dots, w_n\}$, let the slew-sensitive times of w_i be $x_i = a_i + r_m s_{i1}$ and $y_i = a_i + r_M s_{i2}$. The slew-sensitive times of an upper-bound signal (s, a) must satisfy $x = a + r_m s \geq x_i$ and $y = a + r_M s \geq y_i$ for every i . Hence $x \geq \max(x_1, x_2, \dots, x_n)$ and $y \geq \max(y_1, y_2, \dots, y_n)$. Let the slew sensitive times of $U = (s_U, a_U)$ be $x_U = \max(x_1, x_2, \dots, x_n)$ and $y_U = \max(y_1, y_2, \dots, y_n)$. Then $x_U \geq x_i$ and $y_U \geq y_i$. U dominates w_i and is an upper-bound signal. On the other hand, any bounding signal satisfies $x \geq x_U$ and $y \geq y_U$. U is indeed the least upper-bound signal which can be used in the worst case timing. By similar arguments, the greatest lower-bound signal L can be derived for use in the best-case timing. See Figure 8 for graphic illustration.

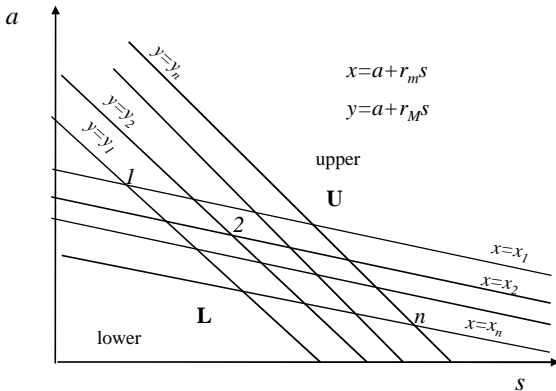


Figure 8: The bounding region for a set of signals.

Theorem 3: For a set of signals, $W = \{w_1, w_2, \dots, w_n\}$, let $x_i = a_i + r_m s_{i1}$ and $y_i = a_i + r_M s_{i2}$. The slew-sensitive times of the least upper-bound signal $U = (s_U, a_U)$ are $x_U = a_U + r_m s_U = \max(x_1, x_2, \dots, x_n)$ and $y_U = a_U + r_M s_U = \max(y_1, y_2, \dots, y_n)$, whereas those times for the greatest lower bound signal $L = (s_L, a_L)$ are $x_L = a_L + r_m s_L = \min(x_1, x_2, \dots, x_n)$ and $y_L = a_L + r_M s_L = \min(y_1, y_2, \dots, y_n)$.

The bounds described in Theorem 3 and Figure 8 can be interpreted in the signal plane. By multiplying VDD/slew, the most and least sensitive times are translated into the most and least sensitive voltage level. Since $a + 0.5s$ and $a - 0.5s$ are respectively the 100% and 0% time, the percentage in VDD for the most and least slew-sensitive times are $50\% + r_M$ and $50\% + r_m$. (For example, if $r_M = 0.25$, the most slew sensitive time is at 75% VDD voltage level). Thus a bounding signal must bound the times at which all signals cross the most and least slew-sensitive voltages as shown in Figure 9 for the case of $W = \{w_i, w_j\}$.

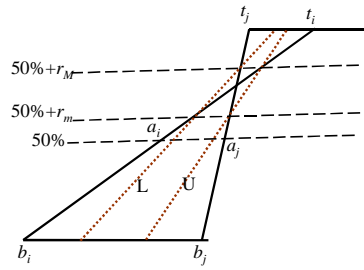


Figure 9: The construction of bounding signals, U and L.

4a. Computations for r_m and r_M

Assume that signal, $w = (s, a)$ is present at the output pin of gate i . Consider a path p from this output pin to a latch n . Let j be a gate in the fan-out of gate i , and q be the sub-path from the output pin of j to n . Let the signal at the output pin of gate j and the input pin of latch n be $w_j = (s_j, a_j)$ and $w_n = (s_n, a_n)$. Then from the delay rules of gate j , we have $d_j \equiv a_j - a = f_j(s)$ and $s_j = g_j(s)$. From the delay rules of path q , we have $d_q \equiv a_n - a_j = f_q(s_j)$ and $s_n = g_q(s_j)$. The delay rules for path p can be defined by the following recursive formula:

$$g_p(s) = (g_q \circ g_j)(s) \quad (15)$$

$$f_p(s) = a_n - a = (f_j + f_q \circ g_j)(s) \quad (16)$$

where \circ is the chaining operator: $f \circ g(s) = f(g(s))$. The following bounds on the slew sensitivity can be easily proved:

Theorem 4: Assume that $D_M g \geq D_m g \geq 0$. We have

$$(a) D_M (f+g) \leq D_M f + D_M g \text{ and } D_m (f+g) \geq D_m f + D_m g$$

$$(b) D_M (f \circ g) \leq \max\{D_M f D_M g, D_M f D_m g\} \text{ and } D_m (f \circ g) \geq \min\{D_m f D_m g, D_m f D_M g\}.$$

Applying Theorem 4 to Equation (16), we obtain $D_M f_p \leq \max\{\alpha_{Mj} + D_M f_q \beta_{Mj}, \alpha_{Mj} + D_M f_q \beta_{mj}\}$ where $\alpha_{mj} = D_m f_j$, $\alpha_{Mj} = D_M f_j$, $\beta_{Mj} = D_M g_j$, and $\beta_{mj} = D_m g_j$. Since the set of fan-out paths from gate i is the union of fan-out paths from all its immediate successor gate j , we have

$$r_{Mi} = \max_p D_M f_p \leq \max_j \{ \alpha_{Mj} + \max \{ \max_q (D_M f_q) \beta_{Mj}, \max_q (D_M f_q) \beta_{mj} \} \}$$

The same argument applies to r_{mi} . We obtain the recursive relations:

$$r_{Mi} \leq \max_j \{ \alpha_{Mj} + \max \{ r_{Mj} \beta_{Mj}, r_{mj} \beta_{mj} \} \} \quad (17).$$

$$r_{mi} \geq \min_j \{ \alpha_{mj} + \min \{ r_{mj} \beta_{Mj}, r_{mj} \beta_{mj} \} \} \quad (18).$$

This leads to the following algorithm for computing the slew sensitivity of constraints:

- 1) Sort the gate in the reverse order. That is, if there is a path from gate i to gate j , then place j before i .
- 2) Initialize r_{mi} and r_{Mi} to 0 at latches.
- 3) For each gate i , compute r_{Mi} and r_{mi} as follows:
 - a) If r_{Mj} is positive, set $r_{Mij} = \alpha_{Mj} + r_{Mj} \beta_{Mj}$. Otherwise, set $r_{Mij} = \alpha_{Mj} + r_{mj} \beta_{mj}$.
 - b) If r_{mj} is positive, set $r_{mij} = \alpha_{mj} + r_{mj} \beta_{mj}$. Otherwise, set $r_{mij} = \alpha_{mj} + r_{mj} \beta_{mj}$.
 - c) Compute r_{Mi} as the maximum of $\{r_{Mij}\}$ and r_{mi} as the minimum of $\{r_{mij}\}$, where j is taken from the set of successor gates.

The cost of this backward sweep algorithm is linear in the number of gates in the circuit.

The case of positive α

For the case that α 's are all positive, the slew sensitivities of constraints are also positive, and we have

$$r_{Mi} \leq \max_j \{ \alpha_{Mj} + r_{Mj} \beta_{Mj} \} \text{ and}$$

$$r_{mi} \geq \min_j \{ \alpha_{mj} + r_{mj} \beta_{mj} \} \quad (19)$$

Since $\beta_{Mk} \leq \beta_M$ and $\alpha_{Mk} \leq \alpha_M$ for all gates, r_{Mi} is bounded above by the series $\sum_{k=1, N} \alpha_M (\beta_M)^{k-1}$:

$$r_{Mi} \leq \alpha_M (1 - \beta_M^N) / (1 - \beta_M) \quad (20)$$

where N is the maximum number of stages in the fan-out cone. Since the geometric series converges very fast, we can use the infinite series as an upper bound:

$$r_M \leq \alpha_M / (1 - \beta_M) \quad (21)$$

Since $\beta_{mk} \geq \beta_m$ and $\alpha_{mk} \geq \alpha_m$ for all gates, r_{mi} is bounded below by the series $\sum_{k=1, n} \alpha_m (\beta_m)^{k-1}$:

$$r_{mi} \geq \alpha_m (1 - \beta_m^n) / (1 - \beta_m) \quad (22)$$

where n is the minimum number of stages in the fan-out cone. For signals at latches, $n=0$ and $r_m=0$. For signals on path with at least one gate,

$$r_m \geq \alpha_m \quad (23)$$

For incomplete designs, the maximum (minimum) number of stages are unknown, Equations 21 and 23 should be used for the max and min constraint.

4b. Methods characterized by two slopes

The shaded area in Figure 10 shows the feasible quadrant for the upper bound of two non-comparable signals, w_i and w_j ($s_i \geq s_j$). For any point $w=(s,a)$ in the s - a plane, we can draw two lines, one through w_j and the other one through w_i . Let the slopes of these lines be $-r_1$ and $-r_2$. Then

$$a + r_1 s = u_j = a_j + r_1 s_j \text{ and } a + r_2 s = v_i = a_i + r_2 s_i \quad (24)$$

On the other hand, given r_1 and r_2 , Equation (24) can be solved for a and s :

$$\begin{aligned} s &= (v_i - u_j) / (r_2 - r_1) = (a_i - a_j + r_2 s_i - r_1 s_j) / (r_2 - r_1) \\ a &= (r_2 u_j - r_1 v_i) / (r_2 - r_1) \\ &= (-r_1 a_i + r_2 a_j - r_1 r_2 (s_i - s_j)) / (r_2 - r_1) \end{aligned} \quad (25)$$

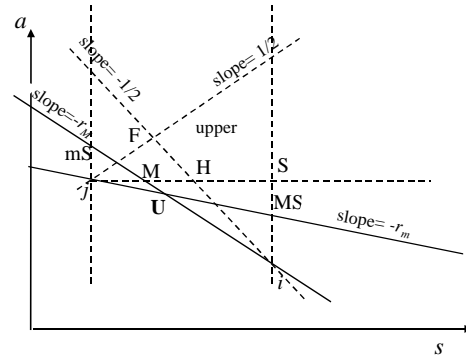


Figure 10: Bounding signals in s - a plane.

Given r_1 and r_2 , Equation 25 provides a formula for constructing a representative signal from two signals. Not all values are good for r_1 and r_2 . If the signal needs to reside in the feasible quadrant within the slew range $[s_j, s_i]$, then the slopes have to satisfy the following bounds:

$$r_2 \geq r_M \geq r_m \geq r_1. \quad (26)$$

The construction methods with two slopes can be extended to the general case:

Theorem 5: In the general case with a set of signals $\{w_1 \dots w_i \dots\}$, let $u_i = a_i + r_1 s_i$, and $v_i = a_i + r_2 s_i$. If $r_2 \geq r_M \geq r_m \geq r_1$, then the signal (s, a) defined by the intersection of

$$a + r_1 s = \max(u_i) \text{ and } a + r_2 s = \max(v_i) \quad (27)$$

is inside the feasible quadrant formed by constraints r_M and r_m .

Methods from Section 3 for positive α can all fit in this characterization scheme. The corresponding values for r_1 and r_2 are listed in Table 2. Point S (the maximum slew method) with $r_1=0$ and $r_2=\infty$ is inside the feasible quadrant. The latest arriving

method picks the fast-rise later signal w_j , which is outside the feasible quadrant. H and F fall in the feasible quadrant, if $r_2=0.5 \geq r_M$. By using Equation 21, this is equivalent to the bound $(1-\beta_M)/2 \geq \alpha_M$ given in Theorem 1.

Four new points, M, MS, mS and U, are added to the table. M has the same arrival time as S, but with slew less than both H and S. MS has the maximum slew and an arrival time earlier than S. mS has the minimum slew and an arrival time later than S. U is the optimal upper bound signal.

It can be proved that Equations 25 and 27 reduce to Equation 9 when appropriate r_1 and r_2 are used. For example, for the half envelope method ($r_1=0, r_2=0.5$), $u_i=a_i, v_i=a_i+0.5s_i=t_i$, and the signal with maximum a_i and maximum t_i is selected.

If r_m becomes negative due to negative α 's, then points M, H, and S become infeasible, while U, F, MS and mS are still valid methods for worst-case timing analysis.

Methods	r_1	r_2
j: Latest arriving	0	$(a_i-a_j)/(s_i-s_j)$
S: Max slew	0	∞
H: Half envelope	0	0.5
F: Full envelope	-0.5	0.5
M: Modified half envelope	0	r_M
MS: Modified max slew	r_m	∞
mS: Modified min slew	$-\infty$	r_M
U: Least upper bound	r_m	r_M

Table 2: Representative points for various methods

4c. Monotonicity

Monotonic delay is often a desired property which can be used to speed up the circuit timing and tuning process.

Definition (Monotonic): For a set of signals, $W = \{w_1, w_2, \dots, w_n\}$, let w_i be the one with maximum $a+r_1s$ and w_j be the one with maximum $a+r_2s$. Then s and a of the representative signal from Equations 25 and 27 are functions of four variables a_i, a_j, s_i, s_j , and so are the slew-sensitive times: $x = a+r_1s$ and $y = a+r_2s$. Since the arrival times at latches are monotonic increasing functions of the slew-sensitive times, we shall say the representative signal construction method preserves the monotonic property, if x and y are monotonic increasing functions of variables a_i and a_j .

Therefore, to satisfy the monotonic property, the partial derivatives of slew-sensitive times with respect to a_i and a_j need to be either zero or positive. In other words, the 2-dimensional gradient vectors $\nabla x = (\partial x / \partial a_i, \partial x / \partial a_j)$ and $\nabla y = (\partial y / \partial a_i, \partial y / \partial a_j)$ must not contain any negative derivative. From Equation 25, we have $\nabla s = (1, -1)/(r_2-r_1)$ and $\nabla a = (-r_1, r_2)/(r_2-r_1)$ for the case of constant r_1 and r_2 . This leads to

$$\nabla x = (r_m-r_1, r_2-r_m)/(r_2-r_1)$$

$$\nabla y = (r_M-r_1, r_2-r_M)/(r_2-r_1)$$

Under the assumptions that $r_2 \geq r_M \geq r_m \geq r_1$, the derivatives of x and y with respect to a_i and a_j are never negative. Hence we have

Theorem 6: The signal bounding methods using two constant slopes r_1 and r_2 , such that $r_2 \geq r_M \geq r_m \geq r_1$, preserve the monotonic property.

The latest arriving method is not monotonic as described in the introduction, while all the bounding methods are.

5. The multiple signal approach

When further accuracy is required, the single signal approach is not adequate. For example, in Figure 11, three fan-out gates have different slew-dependent rules, d_1, d_2, d_3 . The arrival times at the outputs, a'_1, a'_2, a'_3 , for three signals w_1, w_2, w_3 , are listed in Table 3. At output a'_1 , w_1 produces the latest signal. At output a'_2 , w_2 produces the latest signal. At output a'_3 , w_3 produces the latest signal. No single signal can be chosen to generate the latest signals at all outputs. The arrival time from several bound methods are also listed in Table 3. No bound method gives the correct latest arrival time for all three outputs. Therefore, to be able to compute the true latest arrival times at all latches, we need to store and propagate a set of signals.

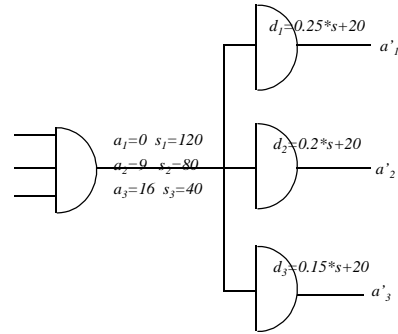


Figure 11: Three signals propagate through gates

	a'_1	a'_2	a'_3
w_1	50	44	38
w_2	49	45	41
w_3	46	44	42
U	50	46	42
MS	50	48	46
MS	54	48	42
H	58	53.6	49.2
F	64	60.8	49.2
S	66	60	54
M	50	47.2	44.4

Table 3: arrival times at output.

However, the signal set may grow very large in size, as it propagates through a large circuit network. Fortunately, we can use the dominance relation to trim out those signals, which are dominated by other signals. In other words, we need to consider only maximal signals. A *maximal* signal is defined as one which no other signal dominates it, and a *maximal signal set* is defined as the subset consisting of maximal signals. Since the dominance order of signals will not change during the propagation through the circuit, the worst-case timing can be computed using the maximal signal set. The maximal signal set can be obtained by a simple sorting, according to Theorem 7. An example of a sorted maximal signal set is shown in Figure 8. The proposed multiple signal method differs from that in reference [3], in that the pruning is done based solely on the causality assumption.

Theorem 7: For a maximal signal set, $W = \{w_1, w_2, \dots, w_n\}$, let $x_i = a_i + r_m s_i$ and $y_i = a_i + r_M s_i$. Then

1. All x_i 's and y_i 's are distinct.
2. If $\{x_i\}$ is sorted in a strictly descending sequence, then $\{y_i\}$ is simultaneously sorted in a strictly ascending sequence.

6. Experimental results and conclusion

To assess the accuracy of these different methods, the following experiments are performed on ISCAS benchmark circuits. Linear delay rules are used: gate delay = 100ps + 0.25(Tx-200ps) + 100ps(gain-1), and gate output slew = 200ps + 0.4(Tx-200ps) + 200ps(gain-1), where Tx is the input slew and gain is the ratio of capacitance load over the input capacitance. For simplicity, we assume that all gates have the same size, and the gain is equal to the number of fan-outs. Sharp signals with zero slew are fed into primary inputs and latches at time 0. Because of the slew-dependent delay rule, the signal slew will gradually increase and spread into a range between 0 and some maximum slew value after propagating through gates.

Six timing runs are done: the exact multiple signal method (E), the latest arriving method (late), the max slew method (S), the full envelope method (F), the half envelope method (H), and the upper bound method (U). Table 4 list the maximum difference of worst-case delay at primary outputs between the last five methods and the exact method. The latest arriving method gives too optimistic timing in the case of circuit c432. The maximum slew method is the most pessimistic method among the five. The full envelope method sometimes yields modest error. The half envelope method performs better than the full envelope method, while the upper bound method produces the best result among the five. It turns out that in this experiment the number of signals in the maximal signal set never exceeds 10, because of the trimming. Hence the exact method with multiple signals may also be a viable solution.

References

[1] R. B. Hitchcock, Sr., "Timing verification and the timing analysis program," 19th Design Automation Conference. June 1982, pp. 594-604.

[2] J. K. Ousterhout, "A Switch-Level Timing Verifier for Digital MOS VLSI," in IEEE Trans. on CAD, vol. 3, July 1985, pp. 336-349.

[3] D. Blaauw, V. Zototov, S. Sundareswaren, C. Oh, R. Panda, "Slope Propagation in Static Timing Analysis," in Proc. ICCAD, Nov., 2000, pp. 338-341

[4] C. Visweswariah, A. R. Conn, "Formulation of Static Optimization with Reduced Size, Degeneracy and Redundancy by Timing Graph Manipulation," in Proc. ICCAD 1999, pp. 244-251.

[5] J. Lillis, C. K. Cheng and T. T. Y. Lin, "Optimal Wire Sizing and Buffer Insertion for low power and a Generalized Delay Model," in Proc. ICCAD 1995, pp. 138-143.

[6] N. Hedenstierna and K. O. Jeppson, "CMOS Circuit Speed and Buffer Optimization," IEEE Trans. on CAD, vol. 6, March 1987, pp. 270-281.

[7] A. I. Kayssi, K. A. Sakallah, T. N. Mudge, "The impact of Signal Transition Time on Path Delay Computation," IEEE transaction on circuit and systems-II: Analog and digital signal processing, Vol. 40, No. 5, May 1993.

[8] K. L. Shepard, V. Narayanan, P. C. Elmendorf, G. Zeng, "Global Harmony: Coupled Noise Analysis for Full Chip RC Interconnect Networks," in Proc. ICCAD, Nov. 1997, pp. 139-146.

[9] C. T. Gray, W. Liu and R. K. Cavin, "Wave Pipelining: Theory and CMOS Implementation," pp. 69, by Kluwer Academic Publisher, 1994

[10] T. M. Burks, J. F. Lee and D. L. Ostapko, "Method to Analyze Worst-Case Simultaneous Switching Delay," pp. 33-41, Vol. 40, No. 3, 1997

[11] D. Hathaway, private communications.

	Latest arriving	max slew	Full envelope	Half envelope	Upper bound envelope
Ckt	late- E	S - E	F - E	H - E	U - E
C17	0	0	0	0	0
C432	-61	60	35	18	6
C499	0	0	0	0	0
C880	0	40	18	7	0
C1355	0	0	0	0	0
C1908	0	0	0	0	0
C2670	0	20	8	3	0
C3540	0	0	0	0	0
C5315	0	132	0	0	0
C6288	0	20	4	4	0
C7552	0	2692	402	165	0

Table 4: maximum difference of worst delay.