Performance Optimization by Wire and Buffer Sizing Under the Transmission Line Model *

Tai-Chen Chen[1], Song-Ra Pan[2], and Yao-Wen Chang[1]

[1]Department of Electrical Engineering, National Taiwan University, Taipei 106, Taiwan

{d0943008, ywchang}@cc.ee.ntu.edu.tw

[2]Department of Computer and Information Science, National Chiao Tung University, Hsinchu 300, Taiwan

gis87528@cis.nctu.edu.tw

Abstract

As the operating frequency increases to Giga Hertz and the rise time of a signal is less than or comparable to the time-of-flight delay of a line, it is necessary to consider the transmission line behavior for delay computation. We present in this paper an analytical formula for the delay computation under the transmission line model. Extensive simulations with SPICE show the high fidelity of the formula. Compared with previous works [8, 11], our model leads to smaller average errors in delay estimation. Based on this formula, we show the property that the minimum delay for a transmission line with reflection occurs when the number of round trips is minimized (i.e., equals one). Besides, we show that the delay of a circuit path is a posynomial function in wire and buffer sizes, implying that a local optimum is equal to the global optimum. Thus, we can apply any efficient search algorithm such as the well-known gradient search procedure to compute the globally optimal solution. Experimental results show that simultaneous wire and buffer sizing is very effective for performance optimization under the transmission line model.

1 Introduction

As the operating frequency increases to Giga Hertz, the rise time of a signal is less than or comparable to the time-of-flight delay of a line. Also, the die size is getting larger, resulting in longer global interconnection lines. The trends make it important to consider the transmission line behavior for delay computation [1]. Transmission line effects become significant when $t_r < 2\ t_f$, where t_r is the rise time and t_f is the time of flight determined by the wire length l divided by the velocity v [1]. There are two kinds of transmission lines. A line with negligible resistance is called a *lossless transmission line*. However, on-chip interconnections have significant resistance, and they should be treated as *lossy transmission lines* [1, 6, 19]. Obviously, it is more accurate and desirable to consider line resistance for timing estimation and optimization. In this paper, therefore, we shall focus on lossy transmission lines.

When two transmission lines on a chip are connected and these two wires have different characteristic impedance, such mismatches of wire impedance can cause reflections at the junction point [1, 13]. Since reflections may cause logic failure or increase delay, the discontinuities of impedance at junction points must be controlled in order to minimize the side effect of reflections. For a lossy transmission line, the ratio of the driving resistance to wire impedance determines the initial voltage generated on the wire. In addition to the ratio of the driving resistance to wire impedance, the number of round trips for the receiving end to reach its final value is determined by the voltage attenuation coefficient which is a function of wire resistance and impedance. Since the voltage attenuation coefficient relates to the wire impedance, we can eliminate the reflections by matching the driving resistance and the wire impedance. The driving resistance of a gate and the impedance of a wire are approximately in inverse proportion to their sizes. Hence wire and gate sizing can affect the delay, implying that sizing circuit components (wires and buffers) is applicable to delay optimization.

*This work was partially supported by the National Science Council of Taiwan ROC under Grant No. NSC-89-2215-E-009-055.

Timing is a crucial concern in high-performance circuits. Many techniques such as wire sizing and gate sizing have been proposed to optimize timing (e.g., [3, 4, 5, 12], etc); however, most of the techniques are based on the Elmore delay model [8]. Modeling and analysis techniques for simulation and timing optimization under the lossy transmission line model have been studied extensively in the literature [9, 10, 11, 14, 17, 18, 21, 24, 26, 27, 28, 29, 30]. Previous work in [17, 21] proposed precise methods for simulating waveform, but they did not present any delay estimator. The work in [18, 28] modeled the transmission line effect; they, however, did not consider delay optimization. Several works in the literature consider the minimization of delay under the transmission line model. Gao and Wong in [9, 10] applied continuous wire-sizing to minimize delay under the lossy transmission line model; however, they focused on exponentially tapered wires. Ismail and Friedman in [11] computed a uniform buffer size and the number of buffers to optimize the delay of a circuit path under the lossy transmission line model; however, their formula does not handle wire sizing. Lin and Pileggi in [14] proposed a wire sizing formulation with second order central moments, but their wire sizing formulation under the transmission line model is not always a posynomial program, and thus there is no optimality guarantee. The work in [26, 30] adopted the S-parameter macro delay model to minimize delay and skew, but the sensitivities were computed at each step using finite difference approximation which requires expensive computation. The work in [24, 27, 29] adopted higher order moments to minimize delay, but their delay models were computationally expensive.

In this paper, we focus on delay modeling and timing optimization under the transmission line model. Unlike most previous works [11, 14, 24, 26, 27, 29, 30] that are based on relatively complicated models or incur larger errors, we present a simple, yet accurate formula for the delay computation under the lossy transmission line model. Extensive simulations with SPICE show that the formula has high fidelity, with an average error of within 6.85% for lossy transmission lines. Based on this formula, we show the property that the minimum delay for a lossy transmission line with reflection occurs when the number of round trips is minimized (i.e., equals one). Besides, we show that the delay of a circuit path is a posynomial function in wire and buffer sizes, implying that a local optimum is equal to the global optimum. Thus we can apply any efficient search algorithm, such as the well-known gradient search procedure, to compute the optimal wire and buffer sizes for timing optimization for a circuit path. For a routing tree, we propose a two-stage algorithm to optimize the delay. In the first stage, we traverse the tree to determine its critical path and delay. In the second stage, we control the reflections at all branching points to prevent from falsely triggering receivers and minimize the critical path delay. We repeat the two stages until no further improvements in the delay of the tree. Experimental results show that simultaneous wire and buffer sizing is very effective in minimizing the delays of circuit paths under the transmission line model.

The remainder of this paper is organized as follows. Section 2 introduces some notation. Section 3 gives the gate and the transmission line models. Section 4 formulates the problem. Section 5 considers the simultaneous wire and buffer sizing for delay optimization. Section 6 extends the cases on a general routing tree. Section 7 shows the experimental results, and finally concluding remarks are given in Section 8.

2 Notation

We use the following notation in this paper.

- $\hat{r}_b$: the resistance of a gate with unit size.
- r_i^b: the resistance of gate i.
- $\hat{c}_b$: the capacitance of a gate with unit size.
- c_i^b: the capacitance of gate i.
- g_i: the size of gate i.
- $\hat{c}_w$: the capacitance of a wire with unit size.
- $\hat{u}_w$: the inductance of a wire with unit size.
- $\hat{r}_w$: the sheet resistance of a wire.
- w_i: the width of wire i.
- l_i: the length of wire i.
- Z_i: the characteristic impedance of wire i.
- v_i: the propagation velocity of wire i.
- R_s: the resistance of source.
- C_L: the capacitance of load.
- V_{DD}: the high voltage of power supply.
- V_t: the threshold voltage.
- $\alpha_{i,j}$: the transmission coefficient at point i if a signal is transmitted from point i to point j.
- $\beta_{i,j}$: the reflection coefficient at point i if a reflection travels from point i to point j.
- γ_i: the voltage attenuation coefficient on wire i if a signal is transmitted from its source to sink.

3 Transmission Line Model

When the rise time of a signal is less than or comparable to the time-of-flight delay from one end of a wire to the other end, the wire should be modeled as a transmission line.

3.1 Gate and Wire Modeling

Figure 1 illustrates the gate and the lossy transmission line models used in this paper. For a gate i with size g_i, the gate resistance r_i^b is $\hat{r}_b/g_i$ and the gate capacitance c_i^b is $\hat{c}_b g_i$, where $\hat{r}_b$ and $\hat{c}_b$ are the unit-sized resistance and unit-sized capacitance of a gate, respectively.

A uniform lossy transmission line i of width w_i can be represented by a serial sections of unit-length resistance, $\hat{r}_w/w_i$, unit-length inductance, $\hat{u}_w/w_i$, and unit-length capacitance, $\hat{c}_w w_i$, where $\hat{r}_w$, $\hat{u}_0$, and $\hat{c}_w$ are the sheet resistance, the unit-sized inductance, and the unit-sized capacitance of a wire, respectively. The effect of inductance and capacitance can be represented by a characteristic impedance, Z_i, which equals $\sqrt{(\hat{u}_w/w_i)/(\hat{c}_w w_i)} = \sqrt{\hat{u}_w}/(w_i\sqrt{\hat{c}_w})$. The propagation velocity of a wire i, v_i, equals $1/\sqrt{\hat{u}_w \hat{c}_w}$ [1]. If the length of a wire is l_i, its total resistance, total inductance, and total capacitance are $\hat{r}_w l_i/w_i$, $\hat{u}_w l_i/w_i$, and $\hat{c}_w w_i l_i$, respectively.

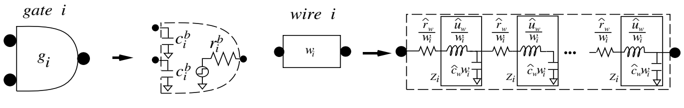

Figure 1: A gate is the loading of its upstream, but is the driver of its downstream. A lossy transmission line is represented by a serial sections of its resistance, inductance, and capacitance, or we can merge each section of inductance and capacitance into a characteristic impedance.

Therefore, with the gate and the lossy transmission line models, we can represent a circuit path by resistors, capacitors, and characteristic impedance. Figure 2 illustrates the resulting circuit modeling for a circuit path with n buffers, where R_S and C_L are the resistance of source and the capacitance of load, respectively.

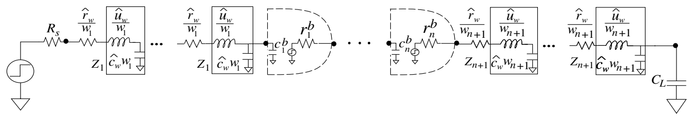

Figure 2: A circuit path (with lossy transmission lines) is a combination of resistors, capacitors, and characteristic impedances.

3.2 Reflections on a Wire

As shown in Figure 3, gate $i-1$ drives lossy transmission line i and gate i. In other words, the resistor with resistance r_{i-1}^b drives a lossy transmission line with the unit-length resistance $\hat{r}_w/w_i$, unit-length inductance $\hat{u}_w/w_i$, and unit-length capacitance $\hat{c}_w w_i$ and a capacitor with capacitance c_i^b. Inductive and capacitive discontinuities may occur at the points A and B. Due to the inductive and capacitive discontinuities, the resulting reflections may cause logic failure or excessively longer delay [1]. The initial voltage at the point B is the sum of the signal sent out from the point A and the reflection generated at the point B. When the reflection generated at the point B travels backward to the point A, a new reflection generated at the point A is transmitted toward to the point B. The new voltage at the point B is the sum of the incoming reflection, the new outgoing reflection, and the initial voltage. As shown in Figure 4(a), the initial voltage at point B does not reach the threshold voltage. Thus, multiple round trips along the line may be required to correctly transmit a signal. As shown in Figure 4(b), if the reflection generated at the point A is negative ($r_{i-1}^b < Z_i$), the voltage may oscillate at the point B, causing *overshoot* or *undershoot*. This oscillating pattern is called *ringing*.

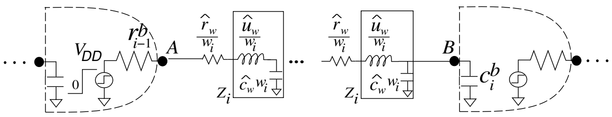

Figure 3: The resistor with resistance r_{i-1}^b drives a lossy transmission line with characteristic impedance Z_i and a capacitor with capacitance c_i^b.

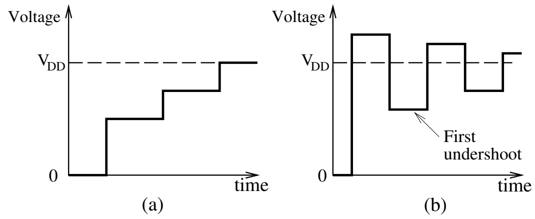

Figure 4: (a) Multiple trips are required to correctly transmit a signal. (b) Ringing may cause logic failures.

3.3 Voltage Attenuation on a Wire

In a lossy transmission line, the resistance of a line causes voltage attenuation, and the voltage attenuation coefficient γ_i along a lossy transmission line i is derived in [1] as follows:

$$\gamma_i = e^{-\frac{\left(\frac{\hat{r}_w l_i}{w_i}\right)}{2Z_i}} = e^{-\frac{\hat{r}_w l_i \sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}}}. \qquad (1)$$

Therefore, in Figure 3, the voltage at the point B before reflection is given by

$$V_B \quad = \quad \gamma_i V_A. \qquad (2)$$

3.4 When to Use Transmission Line Analysis

According to [1, 15, 22], the transmission line behavior is significant when

$$t_r \quad < \quad 2\, t_f, \qquad (3)$$

and

$$Rl \quad \leq \quad 2\; Z_0, \tag{4}$$

where $t_r = 2.2\; r^b_{i-1}(\hat{c}_w w_i l_i + c^b_i)$ is the rise time of wire i, $t_f = l_i/v_i$ is the time-of-flight delay, $Rl = \hat{r}_w l_i/w_i$ is the total resistance, and $Z_0 = Z_i$ is the characteristic impedance. As illustrated in Figure 3, we can rewrite Inequalities (3) and (4) as Inequalities (5) and (6) as follows:

$$2.2\; r^b_{i-1}(\hat{c}_w w_i l_i + c^b_i) \quad < \quad 2\; \frac{l_i}{v_i}, \tag{5}$$

and

$$\frac{\hat{r}_w l_i}{w_i} \leq 2\; Z_i. \tag{6}$$

Besides, to make the voltage at the point B correctly drive the gate i, the voltage at the point B after infinite reflections should be greater than or equal to V_t. In other words, the following inequality must be satisfied.

$$\begin{aligned} V_B &= 2\,\alpha_{i-1,i}\gamma_i(1 + \gamma_i^2\beta_{i-1,i} + \gamma_i^4\beta^2_{i-1,i} + \ldots)V_{DD} \\ &= \frac{2\,\alpha_{i-1,i}\gamma_i V_{DD}}{1 - \gamma_i^2\beta_{i-1,i}} \\ &\geq V_t, \end{aligned} \tag{7}$$

where

$$\begin{aligned} \alpha_{i-1,i} &= \frac{Z_i}{r^b_{i-1} + Z_i} \\ \beta_{i-1,i} &= \frac{r^b_{i-1} - Z_i}{r^b_{i-1} + Z_i} \\ \gamma_i &= e^{-\frac{\hat{r}_w l_i\sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}}}. \end{aligned}$$

Therefore, we should model a line as a lossy transmission line if Inequalities (5)–(7) are satisfied; it should be modeled as a distributed RC line, otherwise.

3.5 Delay Model

The time t_c for charging the capacitive load (defined at 50% of the final value) of the lumped network equals $\ln 2 R_p C_L$, where R_p is the pullup resistance and C_L is the total capacitive load [19, 20, 25]. According to [1], the current that a lossless transmission line can supply is limited by its characteristic impedance. As a result, looking from the receiving end, the line behaves like a resistor with a value Z_0. In a lossy transmission line, not only its characteristic impedance, but also its partial resistance of the line that causes voltage attenuation supply the current. If the total resistance of a line causes voltage attenuation, the voltage at the receiving end becomes zero. In Section 3.3, we know that the voltage at the receiving end V_B equals $\gamma_i V_A$. This implies that there is only $(1 - \gamma_i)$ percentage of the total resistance for the line between nodes A and B, $\hat{r}_w l_i/w_i$, causing voltage attenuation.

Consequently, the pullup resistance R_p for the transmission line is equal to the sum of the characteristic impedance of the line, and partial resistance of the wire which causes voltage attenuation. We have the pullup resistance R_p for the line as follows:

$$R_p \quad = \quad Z_i + (1 - \gamma_i)\,\frac{\hat{r}_w l_i}{w_i}. \tag{8}$$

Hence, the time t_c for charging the capacitive load of a transmission line is given by

$$t_c \quad = \quad \ln 2\; \left(Z_i + (1 - \gamma_i)\,\frac{\hat{r}_w l_i}{w_i}\right) c^b_i. \tag{9}$$

With $\alpha_{i-1,i} = Z_i/(r^b_{i-1} + Z_i)$, $\gamma_i = e^{-\frac{\hat{r}_w l_i\sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}}}$, and the effect of reflection, the final voltage, $V^{'}_{DD}$, at the receiving end after reflection equals $V'_{DD} = 2\gamma_i Z_i/(r^b_{i-1} + Z_i)$, which may not equal V_{DD}. Thus, we can use an approximate method that divides t_c by the final voltage, V'_{DD}, to obtain the charging time, $t^{'}_c$, for which the voltage equals $0.5V_{DD}$. Therefore, we have

$\hat{c}_w$ $(fF/\mu m^2)$	$\hat{u}_w$ $(pH/\square)$	$\hat{r}_w$ $(\Omega/\square)$	$\hat{c}_b$ $(fF/\mu m)$	$\hat{r}_b$ $(k\Omega \cdot \mu m)$	R_s (Ω)	C_L (fF)
0.06	1.667	0.043	1.17	3.6	250	23.4

Table 1: RC parameters of the 0.13 μm technology in SIA'99.

$$\begin{aligned} t^{'}_c &= \ln 2\; \left(Z_i + (1 - \gamma_i)\,\frac{\hat{r}_w l_i}{w_i}\right) c^b_i / \left(\frac{2\gamma_i Z_i}{r^b_{i-1} + Z_i}\right) \\ &= \eta_i \left(r^b_{i-1} + \frac{\sqrt{\hat{u}_w}}{w_i\sqrt{\hat{c}_w}}\right) c^b_i, \end{aligned} \tag{10}$$

where

$$\begin{aligned} \eta_i &= \frac{\ln 2\,\left(e^{\theta_i} + 2\theta_i\,\left(e^{\theta_i} - 1\right)\right)}{2} \\ \theta_i &= \frac{\hat{r}_w l_i\sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}}. \end{aligned}$$

Because transmission line analysis always gives the correct answer independent of the rise time of the driver, delay is the sum of the time-of-flight t_f along the wire and the time $t^{'}_c$ for charging the capacitive load [1, 19]. Thus, the propagation delay $\Delta(g_{i-1}, g_i)$ from the gate g_{i-1} to the next gate g_i in Figure 3 is given by

$$\Delta(g_{i-1}, g_i) \quad = \quad (2n - 1)\,\frac{l_i}{v_i} + \eta_i\,\left(r^b_{i-1} + \frac{\sqrt{\hat{u}_w}}{w_i\sqrt{\hat{c}_w}}\right) c^b_i \tag{11}$$

where n is the number of required round trips to correctly transmit a signal.

3.6 Accuracy

We used SPICE to verify the accuracy of our delay model. The experiments were performed on a signal wire with no buffers. The parameters we used are listed in Table 1. The unit capacitance $\hat{c}_w$ of a wire, the unit inductance $\hat{u}_w$ of a wire, the sheet resistance $\hat{r}_w$ of a wire, the unit capacitance $\hat{c}_b$ and resistance $\hat{r}_b$ of a gate, the resistance of source R_s, and the load capacitance C_L are 0.06 $fF/\mu m^2$, 1.667 $pH/\square$, 0.043 $\Omega/\square$, 1.17 $fF/\mu m$, 3.6 $k\Omega \cdot \mu m$, 250 Ω, and 23.4 fF, respectively. This set of parameters is based on the 0.13 μm technology of the SIA'99 roadmap [23].

In the first and second experiments, we used fixed wire lengths (2.5 mm & 5 mm) with a variety of wire widths. The wire widths for all experiments satisfy Inequalities (5)–(7). Therefore, the wire widths ranged from 130 nm to 480 nm for the first experiment, and ranged from 130 nm to 530 nm for the second experiment. In Figure 5, the delays are plotted as functions of the wire widths for SPICE, Elmore, I&F, and our delay models, where I&F denotes the delay model presented in [11]. Compared to SPICE and based on the lossy transmission line of 2.5 mm (5 mm) long, the maximum error calculated by the Elmore delay model is -36.13% (-19.58%) and the average error is 29.20% (12.05%), the maximum error calculated by the I&F delay model is -6.23% (11.34%) and the average error is 2.98% (4.70%), and the maximum error calculated by our delay model is 6.58% (12.38%) and the average error is 3.80% (6.22%).

In the third and fourth experiments, we used fixed wire widths (500 nm & 130 nm) with a variety of wire lengths. As mentioned earlier, the wire lengths for all experiments satisfy Inequalities (5)–(7). Therefore, the wire lengths ranged from 3.7 mm to 6.2 mm for the third experiment, and ranged from 0.82 mm to 7.75 mm for the fourth experiment. In Figure 6, the delays are plotted as functions of the wire lengths for SPICE, Elmore, I&F, and our delay models. Compared to SPICE and based on the lossy transmission line of 500 nm (130 nm) wide, the maximum error calculated by the Elmore delay model is -10.01% (-51.74%) and the average error is 4.11% (28.11%), the maximum error calculated by the I&F delay model is 10.99% (-14.19%) and the average error is 9.52% (5.30%), and the maximum error calculated by our delay model is 1.99% (22.47%) and the average error is 1.49% (12.22%).

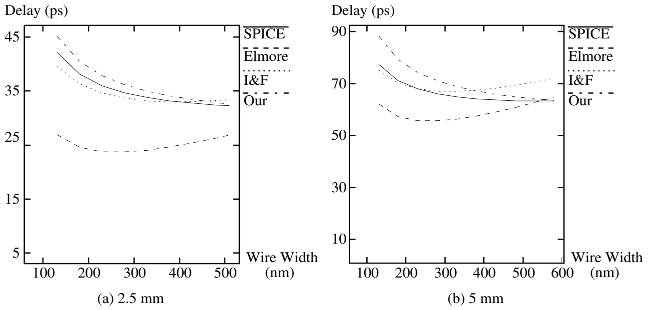

Figure 5: Comparison of the delays calculated by SPICE, Elmore, I&F, and our delay models for lossy transmission lines; (a) wire length = 2.5 mm; (b) wire length = 5 mm.

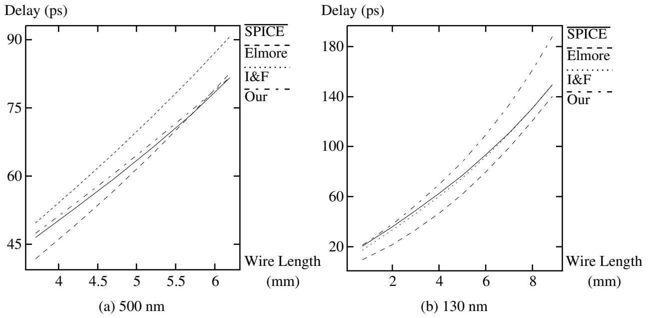

Figure 6: Comparison of the delays calculated by SPICE, Elmore, I&F, and our delay models for lossy transmission lines; (a) wire width = 500 nm; (b) wire width = 130 nm.

According to the above four experiments, our delay model has the maximum inaccuracy at the minimum wire size and the average error is 6.85% under the lossy transmission line model. In particular, the delays computed from our model are always upper bounds of those obtained by SPICE, which makes our model a reliable delay estimator under the lossy transmission line model. The Elmore delay model, however, has a significant negative percentage of errors. Therefore, the Elmore delay model is not a suitable delay estimator for the lossy transmission line model. Also, the I&F delay model incurs positive as well as negative errors for different wire widths of the same length. Hence, although the I&F delay model may be more accurate in some corner cases, it is less suitable for delay estimation under the lossy transmission line model when we apply wire sizing to optimize a circuit. Often circuit designers prefer overestimating delay to underestimate, since an over-optimistic estimation of delay may lead to timing violations. Therefore, our delay model should be more suitable than the Elmore and I&F delay models for practical applications.

4 Problem Formulation

This paper targets at minimizing delay by sizing circuit components. We formulate this problem as follows:

- **Input:** A circuit path and the lower and upper bounds for wire and buffer sizes.
- **Objective:** Determine the optimal wire and buffer sizes for each segment in a circuit path, so that delay is minimized.

We will reformulate this problem for a routing tree in Section 6.

5 Optimal Wire and Buffer Sizing for a Path

5.1 Reflection Considerations

In practice, designers typically desire to optimize performance without generating undesirable reflections and transmit a signal correctly within a limited number of round trips. As the VLSI technology advances, the wire length is increasing and the capacitance of a gate is decreasing, making the time-of-flight delay dominate the delay. Therefore, we have the following theorem for the optimal number of round trips for delay optimization.

Theorem 1 *The minimum delay for a circuit path with reflection occurs when the number of round trips equals one.*

Thus, we can rewrite Equation (11) as follows:

$$\Delta(g_{i-1}, g_i) = l_i\sqrt{\hat{u}_w\hat{c}_w} + \eta_i\left(r_{i-1}^b + \frac{\sqrt{\hat{u}_w}}{w_i\sqrt{\hat{c}_w}}\right)c_i^b, \quad (12)$$

where

$$\eta_i = \frac{\ln 2\left(e^{\theta_i} + 2\theta_i\left(e^{\theta_i}-1\right)\right)}{2}$$

$$\theta_i = \frac{\hat{r}_w l_i\sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}}.$$

5.2 Optimal Wire Sizing

In this section, we minimize the delay of a circuit path by wire sizing. If all buffer sizes and locations are fixed, the delay function of a circuit path from the source s to sink t with $n+1$ segments ($w_1, \ldots, w_{n+1}$) can be calculated as follows:

$$\begin{aligned}\Delta(s,t) &= \sum_{i=1}^{n+1} l_i\sqrt{\hat{u}_w\hat{c}_w} + \eta_1\left(R_S + \frac{\sqrt{\hat{u}_w}}{w_1\sqrt{\hat{c}_w}}\right)c_1^b \\ &+ \sum_{i=2}^{n}\eta_i\left(r_{i-1}^b + \frac{\sqrt{\hat{u}_w}}{w_i\sqrt{\hat{c}_w}}\right)c_i^b \\ &+ \eta_{n+1}\left(r_n^b + \frac{\sqrt{\hat{u}_w}}{w_{n+1}\sqrt{\hat{c}_w}}\right)C_L, \quad (13)\end{aligned}$$

where

$$\eta_i = \frac{\ln 2\left(e^{\theta_i} + 2\theta_i\left(e^{\theta_i}-1\right)\right)}{2}$$

$$\theta_i = \frac{\hat{r}_w l_i\sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}}.$$

Notice that Equation (13) is a posynomial function in $w_1, \ldots, w_{n+1}$, implying that the wire-sizing problem has a unique global minimum [2, 7]. Thus, we can apply any efficient search algorithm, such as the well-known gradient search procedure, to find a locally optimal solution and thus the globally optimal solution.

Theorem 2 *With fixed buffer sizes and locations, the delay of a circuit path is a posynomial function in wire sizes.*

5.3 Optimal Buffer Sizing

In this section, we minimize the delay of a circuit path by buffer sizing. If all wire sizes and buffer locations are fixed, the delay function of a circuit path from the source s to sink t with $n+1$ segments ($g_1, \ldots, g_n$) can be calculated as follows:

$$\begin{aligned}\Delta(s,t) &= \sum_{i=1}^{n+1} l_i\sqrt{\hat{u}_w\hat{c}_w} + \eta_1\left(R_S + \frac{\sqrt{\hat{u}_w}}{w_1\sqrt{\hat{c}_w}}\right)\hat{c}_b g_1 \\ &+ \sum_{i=2}^{n}\eta_i\left(\frac{\hat{r}_b}{g_{i-1}} + \frac{\sqrt{\hat{u}_w}}{w_i\sqrt{\hat{c}_w}}\right)\hat{c}_b g_i \\ &+ \eta_{n+1}\left(\frac{\hat{r}_b}{g_n} + \frac{\sqrt{\hat{u}_w}}{w_{n+1}\sqrt{\hat{c}_w}}\right)C_L, \quad (14)\end{aligned}$$

where

$$\eta_i = \frac{\ln 2\left(e^{\theta_i} + 2\theta_i\left(e^{\theta_i} - 1\right)\right)}{2}$$
$$\theta_i = \frac{\hat{r}_w l_i \sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}}.$$

Notice that Equation (14) is also a posynomial function in $g_1, \ldots, g_n$, implying that the buffer-sizing problem has a unique global minimum [2, 7]. Thus, we can apply any efficient search algorithm, such as the well-known gradient search procedure, to find a locally optimal solution and thus the globally optimal solution.

Theorem 3 *With fixed wire sizes and buffer locations, the delay of a circuit path is a posynomial function in buffer sizes.*

5.4 Optimal Simultaneous Wire and Buffer Sizing

In this section, we minimize the delay of a circuit path by simultaneous wire and buffer sizing. If all buffer locations are fixed, the delay function of a circuit path from the source s to sink t with $2n+1$ segments $(w_1, \ldots, w_{n+1}, g_1, \ldots, g_n)$ is the same as Equation (14).

Notice that Equation (14) is also a posynomial function in $w_1, \ldots, w_{n+1}, g_1, \ldots, g_n$, implying that the simultaneous wire- and buffer-sizing problem has a unique global minimum [2, 7]. Thus, we can apply any efficient search algorithm, such as the well-known gradient search procedure, to find a locally optimal solution and thus the globally optimal solution.

Theorem 4 *With fixed buffer locations, the delay of a circuit path is a posynomial function in wire and buffer sizes.*

6 Extensions to Wire and Buffer Sizing for a Routing Tree

Given a routing tree, our objective is to minimize the critical path delay under the constraints that the first undershoot at each branching point is within the same signal level, and the number of round trips required for correctly transmitting a signal from the root to each load is at most one. We formulate this problem as follows:

- **Input:** A routing tree and the lower and upper bounds for wire and buffer sizes.
- **Output:** Determine the optimal wire and buffer sizes of the tree, so that the critical path delay is minimized under the constraints that the first undershoot is within the same signal level, and the number of round trips is at most one.

We shall first discuss the problem on binary routing trees, and then apply the technique to general routing trees.

6.1 Reflection Constraints

As shown in Figure 7, when a signal is sent out from the source R_S and passes through the point 1 to the point 2, a reflection may be generated at the point 2 and travels backward to the point 1. When the reflection reaches the point 1, the voltage at the point 1 will be interfered. What is worst, if a reflection propagated down to one load is large enough, it could cause logic failure at the load. To prevent from

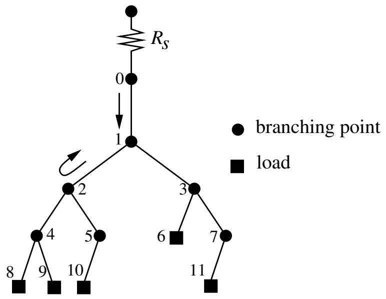

Figure 7: A signal is sent out from the point 0 and then passes through the point 1 to the point 2.

falsely triggering the load, the reflection coefficient at each node must be large enough. For the example shown in Figure 7, if the reflection coefficient $\beta_{1,0}$ at the point 1 is larger, the reflections generated at the points 2 and 3 have smaller impact on the point 1. The reflection coefficient $\beta_{1,0}$ is given by

$$\beta_{1,0} = \frac{Z_2 Z_3 - Z_1 Z_2 - Z_1 Z_3}{Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3}. \tag{15}$$

By Equation (15), $\beta_{1,0}$ becomes larger when $Z_1 Z_2$ and $Z_1 Z_3$ are smaller. If $\beta_{1,0}$ becomes larger, the transmission coefficient $\alpha_{2,1} = \frac{2 Z_1 Z_3}{Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3}$ is smaller. When a reflection generated at the point 2 travels backward to the point 1, the impact of the reflection may be negligible if $\alpha_{2,1}$ is small enough. Similarly, the impact of the reflections generated at the point 3 on other points can also be negligible. For each point i of a routing tree, if the reflections generated at the point i have little interference at other points, a signal can be correctly transmitted from the source to the loads. In order to correctly trans-

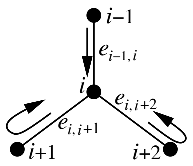

Figure 8: A signal is sent from the point $i-1$ to the point i, and then $i+1$ and $i+2$. The impact of the reflections generated at the points $i+1$ and $i+2$ on the point i may be negligible if $\alpha_{i+1,i}$ and $\alpha_{i+2,i}$ are small enough.

mit a signal from the source of a routing tree to each load, the voltage at each branching point must be larger than or equal to the threshold voltage V_t within one round trip. As shown in Figure 8, the following constraint must be satisfied for the point i:

$$\alpha_{i-1,i}\gamma_{e_{i-1,i}}(1 + \beta_{i,i-1}) \geq V_t. \tag{16}$$

Based on Inequality (16), the initial voltage at the point i will be greater than or equal to the threshold voltage when a signal from the point $i-1$ arrives at the point i. Let $e_{i,j}$ denote the edge between the points i and j, and $\tau_{i,j}$ represent the length of $e_{i,j}$. Since $\tau_{i-1,i}$, $\tau_{i,i+1}$, and $\tau_{i,i+2}$ in Figure 8 could be different, the reflections generated at those points will arrive at the point i at different times. Without loss of generality, assume that $\tau_{i,i+1} \leq \tau_{i,i+2} \leq \tau_{i-1,i}$. The first reflection arrives at the point i is sent out from the point $i+1$, next is from the point $i+2$, and the last is generated from the point $i-1$. In order to prevent the reflections from changing the signal level at the point i, we have the following constraints:

$$\alpha_{i-1,i}\gamma_{e_{i-1,i}}(1 + \beta_{i,i-1} + \alpha_{i,i+1}\gamma^2_{e_{i,i+1}}\beta_{i+1,i}\alpha_{i,i-1}) \geq V_t \tag{17}$$

$$\text{left-hand side of } (17) + \alpha_{i-1,i}\gamma_{e_{i-1,i}}\alpha_{i,i+2}\gamma^2_{e_{i,i+2}}\beta_{i+2,i}\alpha_{i,i-1} \geq V_t \tag{18}$$

$$\text{left-hand side of } (18) + \alpha_{i-1,i}\gamma^3_{e_{i-1,i}}\beta_{i,i-1}\beta_{i-1,i}(1 + \beta_{i,i-1}) \geq V_t. \tag{19}$$

If all constraints are satisfied, the reflection coefficient at each point will be large enough, implying that the reflections generated at the point i have little interference at other points. As a result, a signal can be correctly transmitted from the source to the loads in a tree.

6.2 Delay Calculation

Given a routing tree, we number its nodes level by level, and from left to right on each level (see Figure 7). Let m, L, and P denote the number of edges in the tree, the set of loads, and the critical path, respectively. Similar to Equtaion (11), the critical path delay of a routing tree from the source s to a load t is given by

$$\Delta(s,t) = \sum_{e_{i,j} \in P} (2n_{i,j} - 1)\frac{\tau_{i,j}}{v_{i,j}} + \sum_{e_{i,j} \in P, j \in L} \xi_{i,j}(R_s + Z_{i,j})c_j, \tag{20}$$

Algorithm: *Find-Critical-Path*(T)
Input: T–a routing tree
Output: the critical path
begin
1 **for** each edge $e_{i,j}$ in T
2 Determine the number of round trips $n_{i,j}$ on $e_{i,j}$.
3 Label $e_{i,j}$ with the weight $n_{i,j}\tau_{i,j}$.
4 Call a depth-first traversal to find the longest path.
5 **return** the longest path
end

Figure 9: The Algorithm for determining the critical path of a tree.

Algorithm: *Minimize-Tree-Delay*(T)
Input: T–a routing tree.
Output: wire sizes $\vec{w} = (w_1, w_2, \cdots, w_n)$
begin
repeat
critical-path $\leftarrow$ *Find-Critical-Path*(T)
$w \leftarrow$ *Gradient-Search-Procedure*(*critical-path*)
until no improvement on the delay of T
end

Figure 11: The Alogrithm for minimizing the delay of a tree.

where

$$\begin{aligned} \xi_{i,j} &= \frac{\ln 2\left(e^{\vartheta_{i,j}} + 2\vartheta_{i,j}\left(e^{\vartheta_{i,j}} - 1\right)\right)}{2} \\ \vartheta_{i,j} &= \frac{\hat{r}_w \xi_{i,j}\sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}}, \end{aligned}$$

c_i denotes the capacitance of node i, $v_{i,j}$ and $Z_{i,j}$ denote the propagation velocity and the impedance of edge $e_{i,j}$, respectively.

We propose Algorithm *Find-Critical-Path* (summarized in Figure 9) to find the critical path of a routing tree T. First, we determine the number of round trips $n_{i,j}$ along edge $e_{i,j}$ required to correctly transmit a signal (Line 2). The number of round trips is the minimum $n_{i,j}$ that satisfies the following constraint:

$$\sum_{t=1}^{n_{i,j}} \alpha_{i,j}\gamma_{e_{i,j}}^{2t-1}\beta_{i,j}^{t-1}\beta_{j,i}^{t} \ge V_t.$$

After determining the number of round trips on each edge, we label each edge with the weight $n_{i,j}\tau_{i,j}$ (Line 3). The critical path delay is the sum of edge weights along the longest path. We then apply the depth first traversal to compute the longest path in $O(p)$ time, where p is the number of nodes (Line 4).

6.3 General Routing Tree

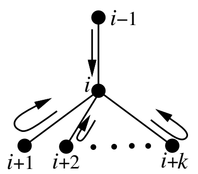

Figure 10: The point i has k children, and the signal is sent from the point $i-1$ to other points.

We extend the technique discussed in previous subsections to general routing trees. As shown in Figure 10, assume that the point i has k children, and a signal is sent out from the point $i-1$ and then propagates down to the children of the point i. Without loss of generality, assume that $\tau_{i,i+1} \le \tau_{i,i+2} \le \cdots \le \tau_{i,i+k} \le \tau_{i-1,i}$. To prevent the reflections generated at the children from changing the signal level at the point i, we have the following constraints:

$\alpha_{i-1,i}\gamma_{e_{i-1,i}}(1 + \beta_{i,i-1} + \alpha_{i,i+1}\gamma^2_{e_{i,i+1}}\beta_{i+1,i}\alpha_{i,i-1} \ge V_t$

left-hand side of the preceding row $+ \alpha_{i-1,i}\gamma_{e_{i-1,i}}\alpha_{i,i+2}\gamma^2_{e_{i,i+2}}\beta_{i+2,i}\alpha_{i,i-1} \ge V_t$

$\vdots$

left-hand side of the preceding row $+ \alpha_{i-1,i}\gamma_{e_{i-1,i}}\alpha_{i,i+k}\gamma^2_{e_{i,i+k}}\beta_{i+k,i}\alpha_{i,i-1} \ge V_t.$

left-hand side of the preceding row $+ \alpha_{i-1,i}\gamma^3_{e_{i-1,i}}\beta_{i,i-1}\beta_{i-1,i}(1 + \beta_{i,i-1}) \ge V_t.$

If all constraints are satisfied, the reflection coefficient at each point will be large enough; thus, a signal can be correctly transmitted from the source to the loads in a general routing tree.

6.4 Our Algorithm

Our objective is to minimize the critical path delay of a routing tree under the constraints that a signal can be correctly transmitted within one round trip and the reflection is sufficiently small to prevent from falsely triggering loads. Since the delay of a tree is dominated by the critical path delay, our problem is to find the wire sizes $\vec{w} = (w_1, w_2, \cdots, w_n)$ that minimize the critical path delay $\Delta(s,t)$ of a tree subject to the constraints listed in Inequalities (16)–(19). We can apply any search algorithm such as the well-known gradient search procedure to find a solution. Algorithm *Minimize-Tree-Delay* computes the minimum delay of a routing tree (see Figure 11). It consists of two stages. The first stage applies the procedure *Find-Critical-Path* to compute the critical path of a routing tree. The second stage applies the gradient search procedure to determine the wire sizes that minimize the critical path delay. We repeat the two stages until no improvements on the delay of the tree.

6.5 Simultaneous Wire and Buffer Sizing for a Routing Tree

Based on the gate and wire models presented in Section 3, we can divide a buffered routing tree into subtrees. In Figure 12, the routing tree is divided into three subtrees. We can treat each subtree as a routing tree with no buffers, and then obtain the reflection constraints for each subtree. Thus, we can minimize the delay of a buffered routing tree under the constraints that a signal can be correctly transmitted within one round trip, and the first undershoot is controlled to prevent from changing the signal level if the reflection constraints for each subtree are satisfied.

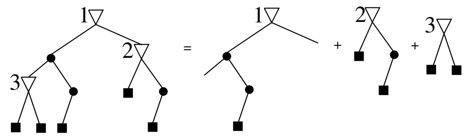

Figure 12: A routing tree with buffers.

7 Experimental Results

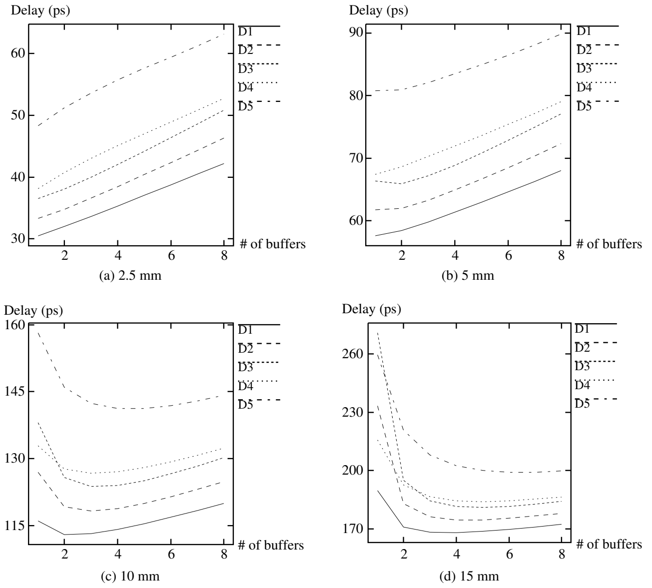

Figure 13: Comparison of different optimization techniques; D1: simultaneous wire and buffer sizing; D2 & D3: wire sizing alone, and the gate resistances are 90 Ω and 60 Ω, respectively; D4 & D5: buffer sizing alone, and the wire widths are 0.3 μm and 0.13 μm, respectively.

We used the nonlinear programming solver, the LINGO 6.0 system, on an Intel Pentium II 400 MHz PC to compute the optimal wire and buffer sizes in a circuit path. All computations are less than 1 sec. The parameters used are listed in Table 1.

Given four lines of the lengths 2.5 mm, 5 mm, 10 mm, and 15 mm, we inserted a specified number of buffers at equidistance. Then, we applied wire and/or buffer sizing to minimize delay. In Figures 13(a), (b), (c), and (d), the path delays are plotted as functions of the number of buffers for the five optimization techniques D1, D2, D3, D4, and D5. D1 gives the delays by sizing wires and buffers simultaneously (denoted by SWBS). D2 (D3) gives the delays that is optimized by sizing wires alone (denoted by WS) with the resistance of each gate equal to 90 Ω (60 Ω). D4 (D5) gives the delays by sizing buffers alone (denoted by BS) with the fixed wire width 0.3 μm (0.13 μm).

As shown in Figure 13, the ranking of those techniques for optimizing circuit performance, from the most effective to the least, is given by SWBS $\succ$ WS $\succ$ BS. These phenomena show the effectiveness of simultaneous wire and buffer sizing under the transmission line model. Further, the number of buffers required for performance optimization is quite small for simultaneous wire and buffer sizing. Because the delay is inversely proportional to the voltage at the receiving end, and voltage attenuation increases as wire length increases. Therefore, inserting buffers can partition a wire into sections of smaller length, which decreases the voltage attenuation and also the path delay.

8 Conclusions

In this paper, we have presented an analytical model for computing the delay of a wire under the transmission line model. Extensive simulations have shown the high fidelity of our model. Compared with previous works [8, 11], our model leads to smaller average errors in delay estimation. Based on our model, we have shown the property that the minimum delay for a transmission line with reflection occurs when the number of round trips is minimized (i.e., equals one). Besides, we have shown that the delay of a circuit path is a posynomial function in wire and buffer sizes under the transmission line model, implying that a local optimum is equal to the global optimum. Thus, we can determine the optimal wire and buffer sizes for performance optimization by applying an efficient algorithm, such as the gradient search procedure. Experimental results have shown the effectiveness of simultaneous wire and buffer sizing in performance optimization under the transmission line model.

References

[1] H. B. Bakoglu, *Circuit, Interconnections and Packaging for VLSI*, Addison-Wesley Publishing Company, Inc., 1990.

[2] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley and Sons, Inc., NY, 1993.

[3] C. P. Chen, Y. P. Chen, and D. F. Wong, “Optimal Wire-Sizing Formula Under the Elmore Delay Model,” *Proc. DAC*, pp. 487–490, June 1996.

[4] C. P. Chen, C. C. N. Chu, and D. F. Wong, “Fast and Exact Simultaneous Gate and Wire Sizing by Lagrangian Relaxation,” *Proc. ICCAD*, pp. 617–624, Nov. 1998.

[5] C. C. N. Chu and D. F. Wong, “A Polynomial Time Optimal Algorithm for Simultaneous Buffer and Wire Sizing,” *Proc. DATE*, pp. 479–485, 1998.

[6] A. Deutsch, Gerard V. Kopcsay, Phillip J. Restle, Howard H. Smith, G. Katopis, Wiren D. Becker, Paul W. Coteus, Christopher W. Surovic, Barry J. Rubin, Richard P. Dunne, Jr., T. Gallo, Keith A. Jenkins, Lewis M. Terman, Robert H. Dennard, George A. Sai-Halasz, Byron L. Krauter, and Daniel R. Knebel, “When are Transmission-Line Effects Important for On-Chip Interconnections?”, *IEEE Trans. on Microwave Theory and Techniques*, vol. 45, pp. 1836–1846, Oct. 1997.

[7] R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric Programming: Theory and Application*, John Wiley & Sons, Inc., NY, 1967.

[8] W. C. Elmore, “The Transient Response of Damped Linear Networks with Particular Regard to Wide Band Amplifiers,” *J. Applied Physics*, Vol. 19, No. 1, 1948.

[9] Y. Gao and D. F. Wong, “Shaping a VLSI Wire to Minimize Delay Using Transmission Line Model,” *Proc. ICCAD*, pp. 611–616, Nov. 1998.

[10] Y. Gao and D. F. Wong “Wire-sizing for Delay Minimization and Ringing Control Using Transmission Line Model,” *Proc. DATE*, pp. 512–516, 2000.

[11] Y. I. Ismail and E. G. Friedman, “Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits,” *IEEE TVLSI*, vol. 8, no. 2, April 2000.

[12] H. R. Jiang, J. Y. Jou, and Y. W. Chang, “Noise-Constrained Performance Optimization by Simultaneous Gate and Wire Sizing Based on Lagrangian Relaxation,” *Proc. DAC*, pp. 90–95, June 1999.

[13] J. Lee and E. Shragowitz, “Overshoot and Undershoot Control for Transmission Line Interconnects,” *Proc. ECTC*, pp. 879–884, 1999.

[14] T. Lin and L. T. Pileggi, “RC(L) Interconnect Sizing with Second Order Considerations via Posynomial Programming,” *Proc. ISPD*, pp. 16–21, April 2001.

[15] F. Moll, M, Roca, A. Rubio, “Inductance in VLSI interconnection modeling,“ *Proc. of IEE Circuits, Devices and Systems*, vol. 145, issue 3, pp. 175–179, June 1998.

[16] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, Inc., 1996.

[17] L. T. Pillage and R. A. Rohrer, “Asymptotic Waveform Evaluation for Timing Analysis,” *IEEE TCAD*, Vol. 9, No. 4, pp. 352–366, April 1990.

[18] R. Gupta, L. Pileggi, “Modeling Lossy Transmission Lines Using the Method of Characteristics,” *IEEE TCAS-1*, Vol. 43, No. 7, pp. 580–582, July 1996.

[19] Jan M. Rabaey, *Digital Integrated Circuits: A Design Perspective*, Prentice Hall, Inc., 1996.

[20] J. Rebinstein, P. Penfield, Jr., and M. A. Horowitz, “Signal Delay in RC Tree Networks,” *IEEE TCAD*, Vol. CAD-2, pp. 202–211, July 1983.

[21] J. S. Roychowdhury, A. R. Newton, and D. O. Pederson, “Algorithms for the transient simulation of lossy interconnect,” *IEEE TCAD*, vol. 13, pp. 96–104, Jan. 1994.

[22] K. L. Shepard, D. Sitaram, and Yu Zheng, “Full-chip, three-dimensional, shapes-based RLC extraction,” *Proc. ICCAD*, pp. 142–149, 2000.

[23] Semiconductor Industry Association, *International Technology Roadmap for Semiconductors 1999 Edition*, Nov. 1999.

[24] Y. Sugiuchi, B. Katz, and R. A. Rohrer, “Interconnect Optimization using Asymptotic Waveform Evaluation(AWE),” *Proc. MCMC*, pp. 120–125, 1994.

[25] Wayne Wolf, *Modern VLSI Design: Systems on Silicon*, 2nd Editon Prentice Hall, Inc., 1996.

[26] J. S. H. Wang and W. W. M. Dai, “Optimal Design of Self-Damped Lossy Transmission Lines for Multichip Modules,” *Proc. ICCAD*, pp. 594–598, 1994.

[27] T. Xu, E. S. Kuh, and Q. Yu, “A Sensitivity-Based Wiresizing Approach to Interconnect Optimization of Lossy Transmission Line Topologies,” *Proc. MCMC*, pp. 117–122, 1996.

[28] Q. Yu and E. S. Kuh, “Exact Moment Matching Model of Transmission Lines and Application to Interconnect Delay Estimation,” *IEEE TVLSI*, Vol. 3, No. 2, pp. 311–322, June 1995.

[29] Q. Yu, E. S. Kuh, and T. Xu, “Moment Models of General Transmission Lines with Application to Interconnect Analysis and Optimization,” *IEEE TVLSI*, Vol. 4, No. 4, pp. 477–494, Dec. 1996.

[30] Q. Zhu, W. W. M. Dai, and J. G. Xi, “Optimal Sizing of High-Speed Clock Networks Based on Distributed RC and Lossy Transmission Line Models,” *Proc. ICCAD*, pp. 628–633, 1993.

[31] Q. Zhu and W. M. Dai, “High-speed clock network sizing optimization based on distributed RC and Lossy RLC interconnect models,” *IEEE TCAD*, vol. 15, no 9, pp. 1106–1118, Sep. 1996.