

Timing Modeling and Optimization Under the Transmission Line Model

Tai-Chen Chen, Song-Ra Pan, and Yao-Wen Chang, *Member, IEEE*

Abstract—As the operating frequency increases to gigahertz and the rise time of a signal is less than or comparable to the time-of-flight delay of a wire, 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, 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.

Index Terms—Buffer sizing, delay model, inductance, interconnect, performance optimization, transmission line, wire sizing.

NOTATION

We use the following notation in this paper.

\hat{r}_b	Resistance of a gate with unit size.
r_i^b	Resistance of gate i .
\hat{c}_b	Capacitance of a gate with unit size.
c_i^b	Capacitance of gate i .
g_i	Size of gate i .
\hat{c}_w	Capacitance of a wire with unit size.
\hat{u}_w	Inductance of a wire with unit size.
\hat{r}_w	Sheet resistance of a wire.
w_i	Width of wire i .
l_i	Length of wire i .
Z_i	Characteristic impedance of wire i .
v_i	Propagation velocity of wire i .
R_s	Driver resistance.
C_L	Load capacitance.
V_{DD}	High voltage of power supply.
V_{IH}	Minimum voltage at the input of a logic gate required so that gate switches.

Manuscript received April 7, 2002; revised January 30, 2003. This work was supported in part by the National Science Council of Taiwan ROC under Grant NSC-91-2215-E-002-036.

T.-C. Chen is with the Graduate Institute of Electronics Engineering, National Taiwan University, Taipei 106, Taiwan (e-mail: d0943008@ee.ntu.edu.tw).

S.-R. Pan is with the Department of Electronic and Computer Engineering, University of California at Santa Barbara, CA 93106 USA (e-Mail: srpan@ece.ucsb.edu).

Y.-W. Chang is with the Department of Electrical Engineering and the Graduate Institute of Electronics Engineering, National Taiwan University, Taipei 106, Taiwan (e-mail: ywchang@cc.ee.ntu.edu.tw).

Digital Object Identifier 10.1109/TVLSI.2003.820529

$\alpha_{i,j}$	Transmission coefficient at point i if a signal is transmitted from point i to point j .
$\beta_{i,j}$	Reflection coefficient at point i if a reflection travels from point i to point j .
γ_i	Voltage attenuation coefficient on wire i if a signal is transmitted from its source to sink.

I. INTRODUCTION

AS THE operating frequency increases to gigahertz, the rise time of a signal is less than or comparable to the time-of-flight delay of a wire. 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 $t_r < 2t_f$, 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 wire 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], [18]. 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. On one hand, if the driving resistance is larger than the wire impedance, it requires multiple trips (a trip is a signal travels from one end of a line to the other end) to switch on the load; on the other hand, if the driving resistance is smaller than the wire impedance, the load may be falsely triggered. 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 its size and width, respectively. Hence, wire and gate sizing can affect the delay, implying that sizing circuit components (wires and buffers) is applicable to delay optimization.

A. Previous Work

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]–[5], [12], etc.); however, most of the techniques are based on the Elmore delay model [8]. Modeling and analysis techniques for simulation and

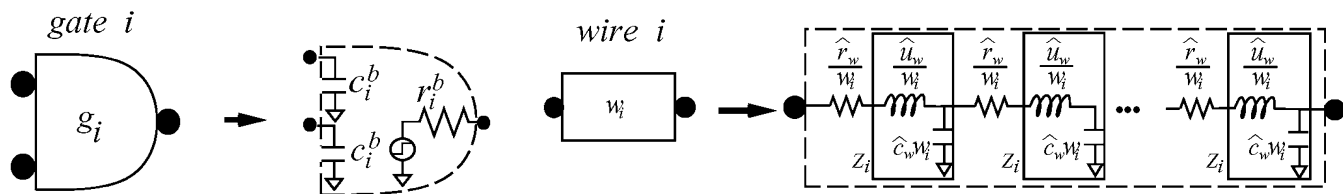


Fig. 1. A gate is the loading of its upstream, but is the driver of its downstream. A lossy transmission line is represented by a serial sections of its resistance, inductance, and capacitance, or we can merge each section of inductance and capacitance into a characteristic impedance.

timing optimization under the lossy transmission line model have been studied extensively in the literature [9]–[11], [14], [16], [17], [20], [23], [25]–[29]. Previous works in [16] and [20] proposed precise methods for simulating waveform, but they did not present any delay estimator. The works in [17] and [27] modeled the transmission line effect; however, they 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] and [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 works in [25] and [29] 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 works in [23], [26], and [28] adopted higher order moments to minimize delay, but their delay models were computationally expensive.

B. Our Contribution

In this paper, we focus on delay modeling, and timing optimization under the transmission line model. Unlike most previous works that are based on relatively complicated models (e.g., [16], [17], [26], and [27]) or incur larger errors (e.g., [8] and [11]), 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 5.61% 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 (a routing tree is a tree that interconnects all signal terminals of a net), 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 there are 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 II gives the gate and the transmission line models. Section III formulates the problem. Section IV considers the simultaneous wire and buffer sizing for delay optimization. Section V extends the cases on a general routing tree. Section VI shows the experimental results, and finally, concluding remarks are given in Section VII.

II. TRANSMISSION LINE MODEL

In this Section, we give the wire and gate models and discuss the transmission line effects which is importance when $t_r < 2t_f$, 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].

A. Gate and Wire Modeling

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

B. Reflections on a Wire

Due to the inductive and capacitive discontinuities, the resulting reflections may cause logic failure or excessively longer delay [1], [13]. As shown in Fig. 3, gate $i - 1$ drives lossy transmission line i and gate i . Inductive and capacitive discontinu-

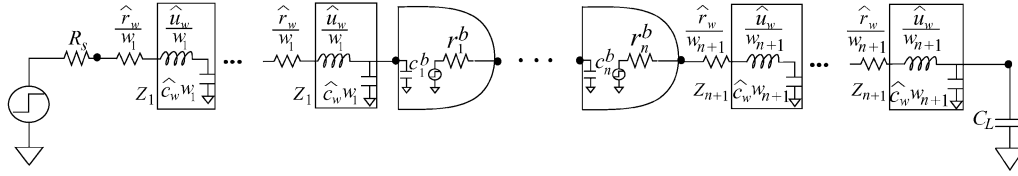


Fig. 2. A circuit path (with lossy transmission lines) is a combination of resistors, capacitors, and characteristic impedances.

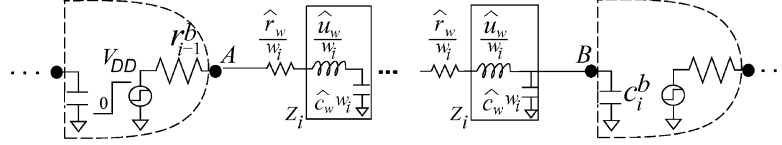


Fig. 3. The resistor with resistance r_{i-1}^b drives a lossy transmission line with characteristic impedance Z_i and a capacitor with capacitance c_i^b .

ities may occur at the points A and B . 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.

On the one hand, as shown in Fig. 4(a), if the resistance of the gate $i - 1$, r_{i-1}^b , is larger than the impedance of the wire i , Z_i , the initial voltage at point B might not reach the threshold voltage. Thus, multiple round trips along the wire may be required to correctly transmit a signal. On the other hand, as shown in Fig. 4(b), if r_{i-1}^b is smaller than Z_i , a reflection generated at point A is negative since the reflection coefficient $\beta_{i-1,i} = (r_{i-1}^b - Z_i)/(r_{i-1}^b + Z_i)$ is negative. Therefore, the voltage may oscillate at the point B , causing *overshoot* or *undershoot*. This oscillating pattern is called *ringing*. If r_{i-1}^b matches with Z_i , the source reflection coefficient $\beta_{i-1,i} = 0$. Thus, no reflections are generated at the source end A .

C. Voltage Attenuation on a Wire

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

$$\gamma_i = e^{-(\hat{r}_w l_i / w_i) / 2Z_i} = e^{-\hat{r}_w l_i \sqrt{\hat{c}_w} / 2\sqrt{\hat{u}_w}}. \quad (1)$$

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

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

D. When to use Transmission Line Analysis

According to [1], [15], and [21], the transmission line behavior is significant when

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

and

$$Rl \leq 2Z_0 \quad (4)$$

where $t_r = 2.2r_{i-1}^b(\hat{c}_w w_i l_i + c_i^b)$ 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

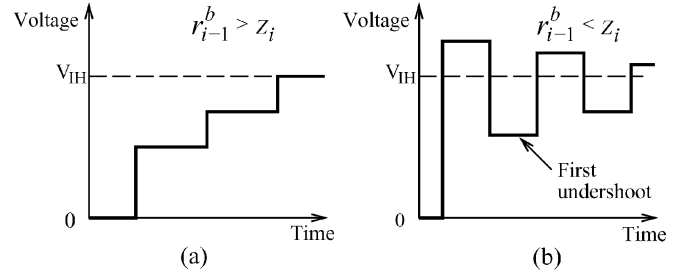


Fig. 4. (a) Multiple round trips are required to correctly transmit a signal. (b) Ringing may cause logic failures.

resistance, and $Z_0 = Z_i$ is the characteristic impedance. As illustrated in Fig. 3, we can rewrite (3) and (4) as Inequalities (5) and (6) as follows:

$$2.2r_{i-1}^b(\hat{c}_w w_i l_i + c_i^b) < 2\frac{l_i}{v_i} \quad (5)$$

and

$$\frac{\hat{r}_w l_i}{w_i} \leq 2Z_i. \quad (6)$$

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_{IH} . In other words, the following inequality must be satisfied.

$$\begin{aligned} V_B &= 2\alpha_{i-1,i}\gamma_i(1 + \gamma_i^2\beta_{i-1,i} + \gamma_i^4\beta_{i-1,i}^2 + \dots)V_{DD} \\ &= \frac{2\alpha_{i-1,i}\gamma_i V_{DD}}{1 - \gamma_i^2\beta_{i-1,i}} \\ &\geq V_{IH} \end{aligned} \quad (7)$$

where

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

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

TABLE I
RC PARAMETERS OF THE 0.13- μm TECHNOLOGY IN SIA'99

\hat{c}_w (fF/ μm^2)	\hat{u}_w (pH/ \square)	\hat{r}_w (Ω/\square)	\hat{c}_b (fF/ μm)	\hat{r}_b (k $\Omega \cdot \mu\text{m}$)	A_b (μm^2)	R_s (Ω)	C_L (fF)
0.06	1.667	0.043	1.17	3.6	6.76	250	23.4

Note that, (5) can be reduced as follows by discarding c_i^b :

$$r_{i-1}^b < \frac{2}{2.2\hat{c}_w w_i v_i} < Z_i. \quad (8)$$

Since $r_{i-1}^b < Z_i$, ringing occurs [see Fig. 4(b)]. If ringing occurs, we may need to model a wire as a transmission line; it should be modeled as a distributed RC line, otherwise.

E. Delay Model

In this Section, we introduce our delay model. Our delay model is based on the RC model. However, since the resistive loss causes voltage attenuation on a wire and the discontinuity of impedance at a junction point incurs reflection, we need to modify the original RC model. First, as given in (1), the voltage attenuation coefficient γ_i along a lossy transmission line is less than 1 because the exponent of e is negative. Therefore, the *effective* resistance is not equal to the total resistance of the wire, implying that the pull-up resistance needs to be modified. Second, due to voltage attenuation and reflection, the final voltage may not equal V_{DD} , implying that we need to use an approximate method to correct the delay model. Third, due to reflection, multiple round trips may be required to correctly transmit a signal, implying that the number of round trips be considered in delay model. We describe in detail how to modify the original RC model in the following.

The time 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 pull-up resistance and C_L is the total capacitive load [18], [19], [24]. 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 *effective* resistance supplies the current. If the total resistance of a wire causes voltage attenuation, the voltage at the receiving end becomes zero and the *effective* resistance equals the total resistance. In Section II-C, 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 and supplying current.

Consequently, the pull-up 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 pull-up resistance R_p for the line as follows:

$$R_p = Z_i + (1 - \gamma_i) \frac{\hat{r}_w l_i}{w_i}. \quad (9)$$

Hence, the time t_c for charging the capacitive load (at 50% of the voltage of the first overshoot) of a transmission line is given by

$$t_c = \ln 2 \left(Z_i + (1 - \gamma_i) \frac{\hat{r}_w l_i}{w_i} \right) c_i^b. \quad (10)$$

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

$$\begin{aligned} t'_c &= \frac{\ln 2 \left(Z_i + (1 - \gamma_i) \frac{\hat{r}_w l_i}{w_i} \right) c_i^b}{\left(\frac{2\gamma_i Z_i}{r_{i-1}^b + Z_i} \right)} \\ &= \eta_i \left(r_{i-1}^b + \frac{\sqrt{\hat{u}_w}}{w_i \sqrt{\hat{c}_w}} \right) c_i^b, \end{aligned} \quad (11)$$

where

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

Since 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], [18]. Thus, the propagation delay $\Delta(g_{i-1}, g_i)$ from the gate g_{i-1} to the next gate g_i in Fig. 3 is given by

$$\Delta(g_{i-1}, g_i) = (2n - 1) \frac{l_i}{v_i} + \eta_i \left(r_{i-1}^b + \frac{\sqrt{\hat{u}_w}}{w_i \sqrt{\hat{c}_w}} \right) c_i^b \quad (12)$$

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

F. 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 I, where \hat{c}_w , \hat{u}_w , \hat{r}_w , \hat{c}_b , \hat{r}_b , A_b , R_s , and C_L are the unit capacitance of a wire, the unit inductance of a wire, the sheet resistance of a wire, the unit capacitance and resistance of a gate, the area of minimum-size buffer, the driver resistance, and the load capacitance, respectively. This set of parameters is based on the 0.13- μm technology of the SIA'99 roadmap [22].

In the first and second experiments, we used fixed wire lengths (2.5 and 5 mm) with a variety of wire widths. The wire widths for all experiments satisfy (5)–(7). Therefore, the wire widths ranged from 130 to 480 nm for the first experiment, and ranged from 130 to 530 nm for the second experiment. In

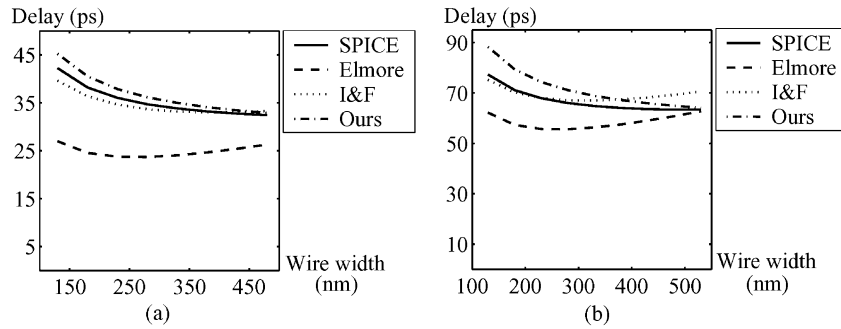


Fig. 5. Comparison of the delays calculated by SPICE, Elmore, I&F, and our delay models for lossy transmission lines; (a) wire length = 2.5 mm; (b) wire length = 5 mm.

TABLE II
EXPERIMENTAL RESULTS FOR THE ACCURACY OF ELMORE, I&F, AND OUR DELAY MODELS FOR LOSSY TRANSMISSION LINES; WIRE LENGTH = 2.5 mm

Width (nm)	SPICE (ps)	Elmore (ps)	E_Err (%)	I&F (ps)	I&F_Err (%)	Ours (ps)	O_Err (%)
130	42.23	26.97	-36.13	39.60	-6.23	45.20	6.58
180	38.20	24.55	-35.74	36.35	-4.85	40.51	5.69
230	36.02	23.74	-34.09	34.63	-3.86	37.85	4.84
280	34.69	23.69	-31.72	33.68	-2.92	36.15	4.03
330	33.81	24.04	-28.89	33.18	-1.86	34.96	3.28
380	33.19	24.65	-25.74	33.00	-0.58	34.08	2.61
430	32.74	25.41	-22.38	33.04	0.93	33.41	2.00
480	32.42	26.29	-18.90	33.27	2.61	32.88	1.39
Absolute average			29.20		2.98		3.80

TABLE III
EXPERIMENTAL RESULTS FOR THE ACCURACY OF ELMORE, I&F, AND OUR DELAY MODELS FOR LOSSY TRANSMISSION LINES; WIRE LENGTH = 5 mm

Width (nm)	SPICE (ps)	Elmore (ps)	E_Err (%)	I&F (ps)	I&F_Err (%)	Ours (ps)	O_Err (%)
130	77.28	62.15	-19.58	75.18	-2.72	88.20	12.38
180	71.12	57.30	-19.44	70.14	-1.38	79.32	10.34
230	67.89	55.68	-17.98	67.93	0.06	74.30	8.63
280	65.98	55.58	-15.77	67.07	1.65	71.08	7.17
330	64.78	56.29	-13.11	66.98	3.40	68.83	5.88
380	64.03	57.50	-10.20	67.40	5.27	67.17	4.68
430	63.59	59.03	-7.17	68.19	7.23	65.90	3.50
480	63.38	60.78	-4.10	69.26	9.28	64.89	2.33
530	63.37	62.70	-1.06	70.56	11.34	64.08	1.10
Absolute average			12.05		4.70		6.22

Fig. 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]. Tables II and III show the experimental results, where width denotes the wire width, *SPICE* denotes the delay calculated by SPICE, *Elmore* denotes the delay calculated by the Elmore delay model, *E_Err* denotes the percentage of the error between SPICE and the Elmore delay model, *I&F* denotes the delay calculated by the I&F delay model, *I&F_Err* denotes the percentage of the error between SPICE and the I&F delay models, *Ours* denotes the delay calculated by our delay model, and *O_Err* denotes the percentage of the error between SPICE and our delay models. The percentage of the error is calculated by $X - SPICE / SPICE \times 100\%$, where X denotes Elmore, I&F, or Ours. 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 and 130 nm) with a variety of wire lengths. As mentioned earlier, the wire lengths for all experiments satisfy (5)–(7). Therefore, the wire lengths ranged from 3.7 to 6.2 mm for the third experiment, and ranged from 0.82 to 7 mm for the fourth experiment. In Fig. 6, the delays are plotted as functions of the wire lengths for SPICE, Elmore, I&F, and our delay models. Tables IV and V show the experimental results, where Length denotes the wire length. 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% (30.55%), the maximum error calculated by the I&F delay model is 10.99% (-14.19%) and the average error is 9.52% (5.95%), and the maximum error calculated by our delay model is 1.99% (20.31%) and the average error is 1.49% (10.94%).

According to the above four experiments, the average error of our delay model is 5.61% . Besides, based on the observa-

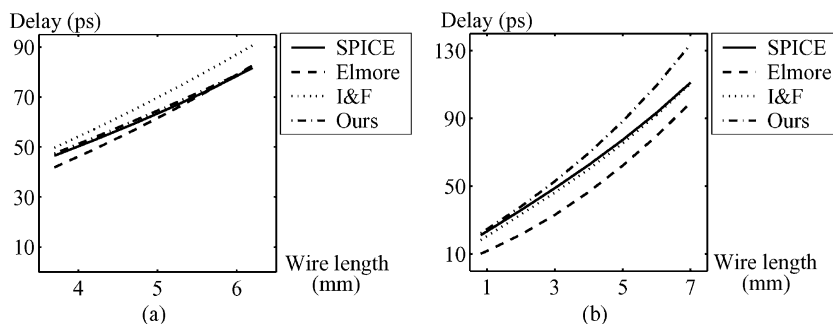


Fig. 6. Comparison of the delays calculated by SPICE, Elmore, I&F, and our delay models for lossy transmission lines; (a) wire width = 500 nm; (b) wire width = 130 nm.

TABLE IV

EXPERIMENTAL RESULTS FOR THE ACCURACY OF SPICE, ELMORE, I&F, AND OUR DELAY MODELS FOR LOSSY TRANSMISSION LINES; WIRE WIDTH = 500 nm

Length (mm)	SPICE (ps)	Elmore (ps)	E.Err (%)	I&F (ps)	I&F.Err (%)	Ours (ps)	O.Err (%)
3.7	46.53	41.87	-10.01	49.75	6.92	47.39	1.84
4.2	52.77	49.04	-7.06	57.14	8.29	53.82	1.99
4.7	59.28	56.70	-4.35	64.91	9.49	60.46	1.98
5.2	66.16	64.85	-1.98	73.07	10.44	67.32	1.75
5.7	73.55	73.49	-0.08	81.63	10.99	74.42	1.19
6.2	81.65	82.62	1.19	90.63	10.99	81.80	0.18
Absolute average			4.11		9.52		1.49

TABLE V

EXPERIMENTAL RESULTS FOR THE ACCURACY OF SPICE, ELMORE, I&F, AND OUR DELAY MODELS FOR LOSSY TRANSMISSION LINES; WIRE WIDTH = 130 nm

Length (mm)	SPICE (ps)	Elmore (ps)	E.Err (%)	I&F (ps)	I&F.Err (%)	Ours (ps)	O.Err (%)
0.82	21.18	10.22	-51.74	18.17	-14.19	22.30	5.31
1	23.37	11.75	-49.71	20.58	-11.96	24.58	5.16
2	35.81	21.41	-40.21	33.25	-7.16	37.97	6.03
3	48.81	33.03	-32.33	46.13	-5.50	52.84	8.26
4	62.56	46.61	-25.50	59.98	-4.13	69.47	11.05
5	77.28	62.15	-19.58	75.18	-2.72	88.20	14.13
6	93.30	79.65	-14.63	91.95	-1.45	109.41	17.27
7	111.02	99.11	-10.73	110.46	-0.50	133.57	20.31
Absolute average			30.55		5.95		10.94

from the simulations, the delays computed from our model are 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. Notice that the maximum inaccuracy of our delay model occurs at the minimum wire size and the maximum wire length. The reason for this phenomenon is that the total resistance is comparable to the impedance. According to Section II-D, the transmission line behavior is insignificant for this situation.

III. 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 V.

IV. OPTIMAL WIRE AND BUFFER SIZING FOR A PATH

A. 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: Considering reflections, the minimum delay based on our model for a circuit section occurs when the number of round trips equals one.

Proof: With the gate and the wire models described in Section II-A, we can divide a circuit path into sections, and the sections can be handled one by one. Consider a section shown in Fig. 3. Suppose on the contrary that the minimum delay for a circuit section occurs when the number of round trips is larger than one. Let r_{i-1}^b and Z_i be the values that result in a (local) minimal delay for the circuit section when the number of round trips equals one, and r_{i-1}^{b*} and Z_i^* be the values that result in the globally minimum delay (the number of round trips is larger than one). According to (12), we have the following:

$$D_1 = \frac{l_i}{v_i} + \eta_i (r_{i-1}^b + Z_i) c_i^b \quad (13)$$

and

$$D_k = (2k-1) \frac{l_i}{v_i} + \eta_i (r_{i-1}^{b*} + Z_i^*) c_i^b \quad (14)$$

where

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

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

D_1 is the (local) minimal delay when the number of round trips equals one, D_k is the globally minimum delay, and k is the number of round trips. Here, $k > 1$. Since $k > 1$ and a wire may need to be modeled as a transmission line if ringing occurs (see Section II-D), the first undershoot for D_k is smaller than V_{IH} , implying that the first undershoot for D_k is smaller than that for D_1 . The first undershoot can be calculated as follows:

$$V_u^1(x) = \frac{2x}{1+x} \gamma_i \left(1 + \frac{1-x}{1+x} \gamma_i^2 \right) \quad (15)$$

where $x = x_i$ or x_i^* , $x_i = Z_i/r_{i-1}^b$, and $x_i^* = Z_i^*/r_{i-1}^{b*}$. On one hand, if $V_u^1(x)$ increases as x increases, $x_i > x_i^*$, implying that $r_{i-1}^{b*} Z_i > r_{i-1}^b Z_i^*$. On the other hand, if $V_u^1(x)$ decreases as x increases, $x_i < x_i^*$, implying that $r_{i-1}^{b*} Z_i < r_{i-1}^b Z_i^*$. To show that $D_k > D_1$, we need to discuss the following two cases:

Case 1) $r_{i-1}^{b*} Z_i > r_{i-1}^b Z_i^*$:

- 1) $r_{i-1}^{b*} > r_{i-1}^b$ and $Z_i^* = Z_i$: The first and second terms of (14) are always larger than those of (13). Thus, $D_k > D_1$.
- 2) $Z_i^* < Z_i$ and $r_{i-1}^{b*} = r_{i-1}^b$: Subtracting (14) from (13), we have

$$D_k - D_1 = (2k-2) \frac{l_i}{v_i} + \eta_i (Z_i^* - Z_i) c_i^b. \quad (16)$$

By (7), we have

$$x < \frac{1 - \gamma_i^2}{4\gamma_i - \gamma_i^2 - 1}. \quad (17)$$

When $\gamma_i = 1/e$ (≈ 0.37), x can be as large as $e^2 - 1/4e - 1 - e^2$ (≈ 2.57). By (8) and (17), we have the range of x as follows:

$$1 < x < \frac{e^2 - 1}{4e - 1 - e^2}. \quad (18)$$

Therefore, by (18), (16) can be rewritten as follows:

$$\begin{aligned} D_k - D_1 &> (2k-2) \frac{l_i}{v_i} + \eta_i \left(r_{i-1}^{b*} - \frac{e^2 - 1}{4e - 1 - e^2} r_{i-1}^b \right) c_i^b \\ &> (2k-2) \frac{l_i}{v_i} + \eta_i (r_{i-1}^b - 3r_{i-1}^b) c_i^b \\ &> (2k-2) \frac{l_i}{v_i} - 2\eta_i r_{i-1}^b (c_i^b + \hat{c}_w w_i l_i) \\ &> (2k-2) \frac{l_i}{v_i} - \frac{2}{1.1} \eta_i \frac{l_i}{v_i}. \end{aligned} \quad (19)$$

According to (6), the minimum of the right-hand side of (19) occurs when $\theta_i = 1$, resulting in the minimum value $0.12 l_i/v_i$.

Thus, we have $D_k > D_1$.

- 3) $r_{i-1}^{b*} > r_{i-1}^b$ and $Z_i^* < Z_i$: According to Case 1.1 and Case 1.2, $D_k > D_1$.
- 4) $r_{i-1}^{b*} < r_{i-1}^b$ and $Z_i^* < Z_i$: Let $r_{i-1}^b = y_1 r_{i-1}^{b*}$ and $Z_i = y_2 Z_i^*$, where $y_1, y_2 > 1$. Since $r_{i-1}^{b*} Z_i > r_{i-1}^b Z_i^*$, $y_1 < y_2$. The first undershoot caused by r_{i-1}^{b*} and Z_i^* is the same as that caused by r_{i-1}^b and Z_i , where $r_{i-1}^{b'} = r_{i-1}^b/y_1$ and $Z_i' = Z_i/y_1$. Substituting $r_{i-1}^{b'}$ for r_{i-1}^b and Z_i' for Z_i , the second term of (13) becomes smaller. Therefore, when the number of round trips equals one, $r_{i-1}^{b'}$ and Z_i' lead to a (local) minimal delay for the circuit section, contradicting the assumption that r_{i-1}^b and Z_i give a (local) minimal delay for the circuit section. Thus, the case that $r_{i-1}^{b*} < r_{i-1}^b$ and $Z_i^* < Z_i$ will never happen.
- 5) $r_{i-1}^{b*} > r_{i-1}^b$ and $Z_i^* > Z_i$: Let $r_{i-1}^{b*} = y_3 r_{i-1}^b$ and $Z_i^* = y_4 Z_i$, where $y_3, y_4 > 1$. Since $r_{i-1}^{b*} Z_i > r_{i-1}^b Z_i^*$, $y_3 > y_4$. Similar to Case 1.4, r_{i-1}^{b*} and Z_i^* lead to the globally minimum delay, where $r_{i-1}^{b*'} = r_{i-1}^{b*}/y_4$ and $Z_i^{*'} = Z_i^*/y_4$, contradicting the assumption that r_{i-1}^{b*} and Z_i^* give the globally minimum delay. Thus, the case that $r_{i-1}^{b*} > r_{i-1}^b$ and $Z_i^* > Z_i$ will never happen.

Case 2) $r_{i-1}^{b*} Z_i < r_{i-1}^b Z_i^*$:

- 1) $r_{i-1}^{b*} < r_{i-1}^b$ and $Z_i^* = Z_i$: Subtracting (14) from (13), we have

$$\begin{aligned} D_k - D_1 &= (2k-2) \frac{l_i}{v_i} + \eta_i (r_{i-1}^{b*} - r_{i-1}^b) c_i^b \\ &> (2k-2) \frac{l_i}{v_i} - \eta_i r_{i-1}^b c_i^b \\ &> (2k-2) \frac{l_i}{v_i} - \eta_i r_{i-1}^b (\hat{c}_w w_i l_i + c_i^b) \\ &> (2k-2) \frac{l_i}{v_i} - \frac{1}{1.1} \eta_i \frac{l_i}{v_i}. \end{aligned} \quad (20)$$

According to (6), the minimum of the right-hand side of (20) occurs when $\theta_i = 1$, resulting in the minimum value $2.06 l_i/v_i$.

Thus, we have $D_k > D_1$.

- 2) $Z_i^* > Z_i$ and $r_{i-1}^{b*} = r_{i-1}^b$: The first and second terms of (14) are always larger than those of (13). Thus, $D_k > D_1$.
- 3) $r_{i-1}^{b*} < r_{i-1}^b$ and $Z_i^* > Z_i$: According to Case 2.1 and Case 2.2, $D_k > D_1$.
- 4) $r_{i-1}^{b*} < r_{i-1}^b$ and $Z_i^* < Z_i$: Let $r_{i-1}^{b*} = y_5 r_{i-1}^{b*}$ and $Z_i = y_6 Z_i^*$, where $y_5, y_6 > 1$. Since $r_{i-1}^{b*} Z_i < r_{i-1}^b Z_i^*$, $y_5 > y_6$. Similar to Case 1.4, $r_{i-1}^{b'}$ and Z_i' lead to a (local) minimum delay, where $r_{i-1}^{b'} = r_{i-1}^b/y_6$ and $Z_i' = Z_i/y_6$, contradicting the assumption that r_{i-1}^b and Z_i give a (local) minimal delay for the circuit section. Thus, the case that $r_{i-1}^{b*} < r_{i-1}^b$ and $Z_i^* < Z_i$ will never happen.
- 5) $r_{i-1}^{b*} > r_{i-1}^b$ and $Z_i^* > Z_i$: Let $r_{i-1}^{b*} = y_7 r_{i-1}^b$ and $Z_i^* = y_8 Z_i$, where $y_7, y_8 > 1$. Since $r_{i-1}^{b*} Z_i < r_{i-1}^b Z_i^*$, $y_7 < y_8$. Similar to Case 1.4, $r_{i-1}^{b*'}$ and $Z_i^{*'}$ lead to the globally minimum delay, where $r_{i-1}^{b*' } = r_{i-1}^{b*}/y_7$ and $Z_i^{*' } = Z_i^*/y_7$, contradicting the assumption that r_{i-1}^{b*} and Z_i^* give the globally minimum delay. Thus, the case that $r_{i-1}^{b*} > r_{i-1}^b$ and $Z_i^* > Z_i$ will never happen.

Therefore, the globally minimum delay occurs when the number of round trips equals one. ■

According to Theorem 1, we can rewrite (12) as follows:

$$\Delta(g_{i-1}, g_i) = l_i \sqrt{\hat{u}_w \hat{c}_w} + \eta_i \left(r_{i-1}^b + \frac{\sqrt{\hat{u}_w}}{w_i \sqrt{\hat{c}_w}} \right) c_i^b \quad (21)$$

where

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

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

B. 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, \dots, w_{n+1}) can be calculated as follows:

$$\begin{aligned} \Delta(s, t) = & \sum_{i=1}^{n+1} l_i \sqrt{\hat{u}_w \hat{c}_w} + \eta_1 \left(R_S + \frac{\sqrt{\hat{u}_w}}{w_1 \sqrt{\hat{c}_w}} \right) c_1^b \\ & + \sum_{i=2}^n \eta_i \left(r_{i-1}^b + \frac{\sqrt{\hat{u}_w}}{w_i \sqrt{\hat{c}_w}} \right) c_i^b \\ & + \eta_{n+1} \left(r_n^b + \frac{\sqrt{\hat{u}_w}}{w_{n+1} \sqrt{\hat{c}_w}} \right) C_L \end{aligned} \quad (22)$$

where

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

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

Notice that (22) is a posynomial function in w_1, \dots, 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 based on our model is a posynomial function in wire sizes.

C. 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 segments (g_1, \dots, g_n) can be calculated as follows:

$$\begin{aligned} \Delta(s, t) = & \sum_{i=1}^{n+1} l_i \sqrt{\hat{u}_w \hat{c}_w} + \eta_1 \left(R_S + \frac{\sqrt{\hat{u}_w}}{w_1 \sqrt{\hat{c}_w}} \right) \hat{c}_b g_1 \\ & + \sum_{i=2}^n \eta_i \left(\frac{\hat{r}_b}{g_{i-1}} + \frac{\sqrt{\hat{u}_w}}{w_i \sqrt{\hat{c}_w}} \right) \hat{c}_b g_i \\ & + \eta_{n+1} \left(\frac{\hat{r}_b}{g_n} + \frac{\sqrt{\hat{u}_w}}{w_{n+1} \sqrt{\hat{c}_w}} \right) C_L \end{aligned} \quad (23)$$

where

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

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

Notice that (23) is also a posynomial function in g_1, \dots, 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 based on our model is a posynomial function in buffer sizes.

D. 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 $(w_1, \dots, w_{n+1}, g_1, \dots, g_n)$ is the same as (23).

Notice that (23) is also a posynomial function in $w_1, \dots, w_{n+1}, g_1, \dots, g_n$, 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.

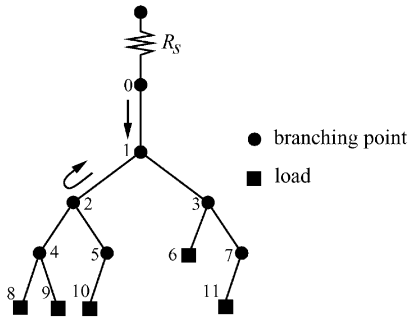


Fig. 7. A signal is sent out from the point 0 and then passes through the point 1 to the point 2.

Theorem 4: With fixed buffer locations, the delay of a circuit path based on our model is a posynomial function in wire and buffer sizes.

V. 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.

A. Reflection Constraints

As shown in Fig. 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. Further, if a reflection propagated down to one load is large enough, it could cause logic failure at the load. To prevent from falsely triggering the load, the reflection coefficient at each node must be large enough. For the example shown in Fig. 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. According to [1], the reflection coefficient $\beta_{1,0}$ is given by

$$\beta_{1,0} = \frac{Z_2 Z_3 - Z_1 Z_2 - Z_1 Z_3}{Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3}. \quad (24)$$

By (24), $\beta_{1,0}$ becomes larger when $Z_1 Z_2$ and $Z_1 Z_3$ are smaller. If $\beta_{1,0}$ becomes larger, the transmission coefficient $\alpha_{2,1} = 2Z_1 Z_3 / (Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3)$ 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

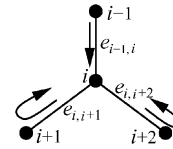


Fig. 8. A signal is sent from the point $i - 1$ to the point i , and then $i + 1$ and $i + 2$. The impact of the reflections generated at the points $i + 1$ and $i + 2$ on the point i may be negligible if $\alpha_{i+1,i}$ and $\alpha_{i+2,i}$ are small enough.

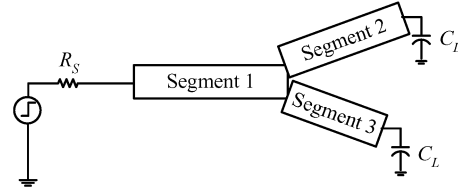


Fig. 9. A binary routing tree without buffers.

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 transmit 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_{IH} within one round trip. As shown in Fig. 8, the following constraint must be satisfied for the point i :

$$\alpha_{i-1,i} \gamma_{e_{i-1,i}} (1 + \beta_{i,i-1}) \geq V_{IH}. \quad (25)$$

Based on (25), 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 Fig. 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:

$$\alpha_{i-1,i} \gamma_{e_{i-1,i}} (1 + \beta_{i,i-1} + \alpha_{i,i+1} \gamma_{e_{i,i+1}}^2 \beta_{i+1,i} \alpha_{i,i-1}) \geq V_{IH} \quad (26)$$

left-hand side of (26)

$$+ \alpha_{i-1,i} \gamma_{e_{i-1,i}} \alpha_{i,i+2} \gamma_{e_{i,i+2}}^2 \beta_{i+2,i} \alpha_{i,i-1} \geq V_{IH} \quad (27)$$

left-hand side of (27)

$$+ \alpha_{i-1,i} \gamma_{e_{i-1,i}}^3 \beta_{i,i-1} \beta_{i-1,i} (1 + \beta_{i,i-1}) \geq V_{IH}. \quad (28)$$

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 routing tree.

B. Delay Calculation

Given a routing tree, we number its nodes level by level, and from left to right on each level (see Fig. 7). Let m , L , and P denote the number of edges in the tree, the set of loads, and the

TABLE VI
PARAMETERS AND EXPERIMENTAL RESULTS FOR THE ACCURACY OF ELMORE AND OUR DELAY MODELS ON A BINARY ROUTING TREE WITHOUT BUFFERS

Experimental ID	Segment ID	Parameter				Experimental Result				
		R_S (Ω)	Length (mm)	Width (μm)	Load (fF)	SPICE (ps)	Elmore (ps)	E.Err (%)	Ours (ps)	O.Err (%)
1	1	10	1	1	–	–	–	–	–	–
	2	–	1	1	23.4	22.80	13.03	-42.87	21.68	-4.91
	3	–	1	1	23.4	22.80	13.03	-42.87	21.68	-4.91
2	1	10	1	1	–	–	–	–	–	–
	2	–	0.9	1	23.4	21.80	11.41	-47.67	20.65	-5.27
	3	–	0.7	1	23.4	19.70	10.79	-45.20	18.59	-5.62
3	1	10	1	1	–	–	–	–	–	–
	2	–	0.9	1.2	23.4	21.30	11.39	-46.55	20.39	-4.27
	3	–	0.7	0.8	23.4	20.40	11.10	-45.60	18.97	-7.02
4	1	10	1.5	1	–	–	–	–	–	–
	2	–	1.3	1.2	23.4	30.80	21.11	-31.47	29.50	-4.21
	3	–	1	0.8	23.4	29.00	20.39	-29.70	27.08	-6.63
5	1	10	2	1	–	–	–	–	–	–
	2	–	1.7	1.2	23.4	41.10	33.75	-17.89	38.63	-6.00
	3	–	1.3	0.8	23.4	36.70	32.41	-11.69	35.20	-4.08
Absolute average								36.15		5.29

critical path, respectively. Similar to (12), the critical path delay of a routing tree from the source s to a load t is given by

$$\Delta(s, t) = \sum_{e_{i,j} \in P} (2n_{i,j} - 1) \frac{\tau_{i,j}}{v_{i,j}} + \sum_{e_{i,j} \in P, j \in L} \xi_{i,j} (R_s + Z_{i,j}) c_j \quad (29)$$

where

$$\xi_{i,j} = \frac{\ln 2 (e^{\vartheta_{i,j}} + 2\vartheta_{i,j} (e^{\vartheta_{i,j}} - 1))}{2} \quad (30)$$

$$\vartheta_{i,j} = \frac{\hat{r}_w \tau_{i,j} \sqrt{\hat{c}_w}}{2\sqrt{\hat{u}_w}} \quad (31)$$

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 used SPICE to verify the accuracy of our delay model. The experiments were performed on a binary routing tree with no buffers (as shown in Fig. 9). Table VI shows the parameters and the experimental results, where R_S denotes the driver resistance, *Length* and *Width* denote the length and width of each segment, *Load* denotes the load capacitance of segments 2 and 3, *SPICE* denotes the delay calculated by SPICE, *Elmore* denotes the delay calculated by the Elmore delay model, *E.Err* denotes the percentage of the error between SPICE and the Elmore delay model, *Ours* denotes the delay calculated by our delay model, and *O.Err* denotes the percentage of the error between SPICE and our delay model. The percentage of the error is calculated by $(X - SPICE)/SPICE \times 100\%$, where X denotes Elmore or Ours.

We propose Algorithm *Find-Critical-Path* (summarized in Fig. 10) 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:

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

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

Fig. 10. The Algorithm for determining the critical path of a routing tree.

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).

C. General Routing Tree

We extend the technique discussed in Section IV-A and B to general routing trees. As shown in Fig. 11, 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 $\tau_{i,i+1} \leq \tau_{i,i+2} \leq \dots \leq \tau_{i,i+k} \leq \tau_{i-1,i}$. To prevent the reflections generated at the children from changing the signal level at the point i , we have the following constraints:

$$\alpha_{i-1,i} \gamma_{e_{i-1,i}} (1 + \beta_{i,i-1} + \alpha_{i,i+1} \gamma_{e_{i,i+1}}^2 \beta_{i+1,i} \alpha_{i,i-1}) \geq V_{IH}$$

left-hand side of the preceding row

$$+ \alpha_{i-1,i} \gamma_{e_{i-1,i}} \alpha_{i,i+2} \gamma_{e_{i,i+2}}^2 \beta_{i+2,i} \alpha_{i,i-1} \geq V_{IH}$$

\vdots

left-hand side of the preceding row

$$+ \alpha_{i-1,i} \gamma_{e_{i-1,i}} \alpha_{i,i+k} \gamma_{e_{i,i+k}}^2 \beta_{i+k,i} \alpha_{i,i-1} \geq V_{IH}$$

left-hand side of the preceding row

$$+ \alpha_{i-1,i} \gamma_{e_{i-1,i}}^3 \beta_{i,i-1} \beta_{i-1,i} (1 + \beta_{i,i-1}) \geq V_{IH}.$$

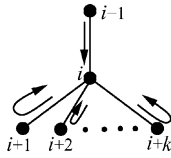


Fig. 11. The point i has k children, and the signal is sent from the point $i - 1$ to other points.

Algorithm: *Minimize-Tree-Delay*(T)
Input: T – a routing tree
Output: wire sizes $\vec{w} = (w_1, w_2, \dots, w_n)$
begin
 1 **repeat**
 2 $\text{critical-path} \leftarrow \text{Find-Critical-Path}(T)$
 3 $w \leftarrow \text{Gradient-Search-Procedure}(\text{critical-path})$
 4 **until** no improvement on the delay of T
end

Fig. 12. The Algorithm for minimizing the delay of a routing tree.

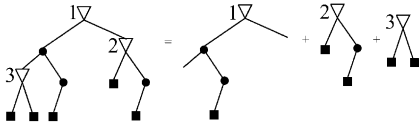


Fig. 13. A routing tree with buffers.

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.

D. 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 routing tree is dominated by the critical path delay, our problem is to find the wire sizes $\vec{w} = (w_1, w_2, \dots, w_n)$ that minimize the critical path delay $\Delta(s, t)$ of a routing tree subject to the constraints listed in (25)–(28). 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 Fig. 12). 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.

E. Simultaneous Wire and Buffer Sizing for a Routing Tree

Based on the gate and wire models presented in Section II, we can divide a buffered routing tree into subtrees. In Fig. 13, 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.

VI. EXPERIMENTAL RESULTS

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 s. The parameters used are listed in Table I.

Given four lines of the lengths 2.5, 5, 10, and 15 mm, we inserted a specified number of buffers at equidistance. Then, we applied wire and/or buffer sizing to minimize delay. Listed in Tables VII–X, Column D1 gives the delays and areas by sizing wires and buffers simultaneously (denoted by SWBS); Column D2 (D3) gives the delays and areas by sizing wires alone (denoted by WS), with the resistance of each gate equal to 90Ω (60Ω); and Column D4 (D5) lists the delays and areas by sizing buffers alone (denoted by BS) with the fixed wire width of $0.3 \mu\text{m}$ ($0.13 \mu\text{m}$), where the area is the sum of wire area (the product of width and length) and buffer area (the product of buffer size and the area of minimum-size buffer). In Fig. 14(a)–(d), the path delays are plotted as functions of the number of buffers for the five optimization techniques D1, D2, D3, D4, and D5.

As shown in Fig. 14, 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. Since the delay is inversely proportional to the voltage at the receiving end, and voltage attenuation increases as wire length increases, inserting buffers can partition a wire into sections of smaller length, which decreases the voltage attenuation and also the path delay.

VII. 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.

TABLE VII

EXPERIMENTAL RESULTS. D1: SIMULTANEOUS WIRE AND BUFFER SIZING. D2 and D3: WIRE SIZING ALONE AND THE GATE RESISTANCES ARE 90 AND 60 Ω, RESPECTIVELY. D4 and D5: BUFFER SIZING ALONE AND THE WIRE WIDTHS ARE 0.3 AND 0.13 μm, RESPECTIVELY. PATH LENGTH = 2.5 mm

# buffers	D1		D2		D3		D4		D5	
	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)
1	31.00	4.48	33.30	8.52	36.54	8.66	38.15	0.81	48.29	0.37
2	32.00	7.76	34.81	10.31	38.04	10.58	40.77	0.84	51.20	0.39
3	33.63	8.36	36.59	11.31	39.97	11.71	43.06	0.86	53.60	0.39
4	35.32	8.76	38.47	12.00	42.06	12.54	45.12	0.87	55.69	0.40
5	37.03	9.05	40.40	12.54	44.21	13.22	47.03	0.87	57.60	0.40
6	38.75	9.29	42.34	13.01	46.40	13.82	48.91	0.89	59.41	0.40
7	40.48	9.50	44.31	13.42	48.61	14.37	50.80	0.90	61.24	0.41
8	42.21	9.69	46.28	13.80	50.83	14.89	52.71	0.92	63.10	0.42

TABLE VIII

EXPERIMENTAL RESULTS. D1: SIMULTANEOUS WIRE AND BUFFER SIZING. D2 and D3: WIRE SIZING ALONE, AND THE GATE RESISTANCES ARE 90 AND 60 Ω, RESPECTIVELY. D4 and D5: BUFFER SIZING ALONE AND THE WIRE WIDTHS ARE 0.3 AND 0.13 μm, RESPECTIVELY. PATH LENGTH = 5 mm

# buffers	D1		D2		D3		D4		D5	
	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)
1	57.57	11.98	61.69	15.94	66.33	16.07	67.37	1.56	80.76	0.70
2	58.39	14.64	61.92	19.68	65.90	19.95	68.62	1.59	80.93	0.71
3	59.76	15.95	63.23	21.56	67.13	21.97	70.27	1.61	82.09	0.72
4	61.31	16.75	64.86	22.76	68.85	23.30	71.96	1.62	83.49	0.72
5	62.94	17.30	66.64	23.62	70.79	24.30	73.63	1.63	84.96	0.73
6	64.61	17.73	68.48	24.31	72.83	25.12	75.38	1.64	86.49	0.73
7	66.29	18.07	70.37	24.89	74.93	25.84	77.18	1.66	88.13	0.74
8	68.00	18.34	72.29	25.39	77.09	26.48	79.02	1.67	89.85	0.74

TABLE IX

EXPERIMENTAL RESULTS. D1: SIMULTANEOUS WIRE AND BUFFER SIZING. D2 and D3: WIRE SIZING ALONE, AND THE GATE RESISTANCES ARE 90 AND 60 Ω, RESPECTIVELY. D4 and D5: BUFFER SIZING ALONE, AND THE WIRE WIDTHS ARE 0.3 AND 0.13 μm, RESPECTIVELY. PATH LENGTH = 10 mm

# buffers	D1		D2		D3		D4		D5	
	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)
1	116.01	19.33	126.88	28.94	138.03	29.07	132.84	3.06	158.16	1.35
2	112.97	26.38	119.17	37.43	125.81	37.70	127.72	3.09	146.04	1.36
3	113.14	29.77	118.20	41.48	123.71	41.88	126.76	3.11	142.37	1.37
4	114.10	31.76	118.78	43.88	123.93	44.42	127.07	3.12	141.28	1.37
5	115.39	33.01	119.98	45.49	125.05	46.17	127.97	3.14	141.27	1.38
6	116.83	33.93	121.45	46.70	126.57	47.51	129.23	3.15	141.90	1.38
7	118.36	34.63	123.08	47.65	128.31	48.59	130.69	3.17	142.92	1.39
8	119.94	35.19	124.81	48.43	130.20	49.52	132.28	3.18	144.19	1.40

TABLE X

EXPERIMENTAL RESULTS. D1: SIMULTANEOUS WIRE AND BUFFER SIZING. D2 and D3: WIRE SIZING ALONE, AND THE GATE RESISTANCES ARE 90 AND 60 Ω, RESPECTIVELY. D4 and D5: BUFFER SIZING ALONE, AND THE WIRE WIDTHS ARE 0.3 AND 0.13 μm, RESPECTIVELY. PATH LENGTH = 15 mm

# buffers	D1		D2		D3		D4		D5	
	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)	Delay (ps)	Area (kμm ²)
1	189.68	19.98	233.42	39.94	270.83	40.07	215.65	4.60	259.48	2.00
2	171.05	35.27	183.03	54.21	195.01	54.48	192.62	4.59	220.81	2.01
3	168.41	41.63	176.29	60.73	184.54	61.13	186.61	4.61	207.95	2.02
4	168.17	45.28	174.66	64.52	181.60	65.06	184.54	4.63	202.47	2.02
5	168.79	47.61	174.70	67.02	181.10	67.69	184.09	4.65	200.05	2.03
6	169.81	49.23	175.48	68.81	181.66	69.63	184.49	4.66	199.21	2.04
7	171.05	50.45	176.65	70.19	182.77	71.14	185.35	4.68	199.24	2.05
8	172.43	51.39	178.04	71.30	184.20	72.38	186.50	4.69	199.79	2.05

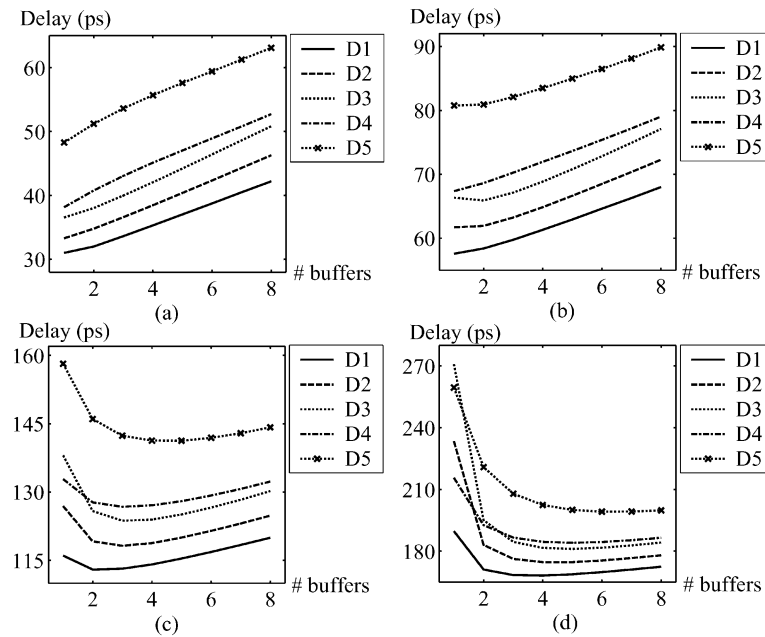


Fig. 14. Comparison of different optimization techniques D1: simultaneous wire and buffer sizing. D2 and D3: wire sizing alone, and the gate resistances are 90 and 60 Ω , respectively. D4 and D5: buffer sizing alone, and the wire widths are 0.3 and 0.13 μm , respectively. (a) Path length = 2.5 mm. (b) Path length = 5 mm. (c) Path length = 10 mm. (d) Path length = 15 mm.

REFERENCES

- [1] H. B. Bakoglu, *Circuit, Interconnections and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.
- [2] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*. New York: Wiley, 1993.
- [3] C. P. Chen, Y. P. Chen, and D. F. Wong, "Optimal wire-sizing formula under the elmore delay model," in *Proc. Design Automation Conf. (DAC)*, 1996, pp. 487–490.
- [4] C. P. Chen, C. C. N. Chu, and D. F. Wong, "Fast and exact simultaneous gate and wire sizing by lagrangian relaxation," in *Proc. Int. Conf. Computer-Aided Design (ICCAD)*, 1998, pp. 617–624.
- [5] C. C. N. Chu and D. F. Wong, "A polynomial time optimal algorithm for simultaneous buffer and wire sizing," in *Proc. Design Automation and Test Europe (DATE)*, 1998, pp. 479–485.
- [6] A. Deutsch *et al.*, "When are transmission-line effects important for on-chip interconnections?," *IEEE Trans. Microwave Theory Tech.*, vol. 45, pp. 1836–1846, Oct. 1997.
- [7] R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric Programming: Theory and Application*. New York: Wiley, 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," in *Proc. Int. Conf. Computer-Aided Design (ICCAD)*, 1998, pp. 611–616.
- [10] —, "Wire-sizing for delay minimization and ringing control using transmission line model," in *Proc. Design Automation and Test Europe (DATE)*, 2000, pp. 512–516.
- [11] Y. I. Ismail and E. G. Friedman, "Effects of inductance on the propagation delay and repeater insertion in VLSI circuits," *IEEE Trans. VLSI Syst.*, vol. 8, pp. 195–206, Apr. 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," in *Proc. Design Automation Conf. (DAC)*, 1999, pp. 90–95.
- [13] J. Lee and E. Shragowitz, "Overshoot and undershoot control for transmission line interconnects," in *Proc. Electronic Components and Technology Conf.*, 1999, pp. 879–884.
- [14] T. Lin and L. T. Pileggi, "RC(L) interconnect sizing with second order considerations via posynomial programming," in *Proc. Int. Symp. Physical Design (ISPD)*, 2001, pp. 16–21.
- [15] F. Moll, M. Roca, and A. Rubio, "Inductance in VLSI interconnection modeling," *Inst. Elect. Eng. Circuits, Devices Systems*, vol. 145, no. 3, pp. 175–179, June 1998.
- [16] L. T. Pillage and R. A. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. Computer-Aided Design*, vol. 9, pp. 352–366, Apr. 1990.
- [17] R. Gupta and L. Pileggi, "Modeling lossy transmission lines using the method of characteristics," *IEEE Trans. Circuits Syst. I*, vol. 43, pp. 580–582, July 1996.
- [18] M. J. Rabaey, *Digital Integrated Circuits: A Design Perspective*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [19] J. Rebinstein, P. Penfield Jr, and M. A. Horowitz, "Signal delay in RC tree networks," *IEEE Trans. Computer-Aided Design*, vol. CAD-2, pp. 202–211, July 1983.
- [20] J. S. Roychowdhury, A. R. Newton, and D. O. Pederson, "Algorithms for the transient simulation of lossy interconnect," *IEEE Trans. Computer-Aided Design*, vol. 13, pp. 96–104, Jan. 1994.
- [21] K. L. Shepard, D. Sitaram, and Y. Zheng, "Full-chip, three-dimensional, shapes-based RLC extraction," in *Proc. Int. Conf. Computer-Aided Design (ICCAD)*, 2000, pp. 142–149.
- [22] *International Technology Roadmap for Semiconductors 1999 Edition*, Semiconductor Industry Association, 1999.
- [23] Y. Sugiuchi, B. Katz, and R. A. Rohrer, "Interconnect optimization using asymptotic waveform Evaluation(AWE)," in *Proc. Multi-Chip Module Conf.*, 1994, pp. 120–125.
- [24] W. Wolf, *Modern VLSI Design: Systems on Silicon*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [25] J. S. H. Wang and W. W. M. Dai, "Optimal design of self-damped lossy transmission lines for multichip modules," in *Proc. Int. Conf. Computer-Aided Design (ICCAD)*, 1994, pp. 594–598.
- [26] T. Xu, E. S. Kuh, and Q. Yu, "A sensitivity-based wiresizing approach to interconnect optimization of lossy transmission line topologies," in *Proc. Multi-Chip Module Conf.*, 1996, pp. 117–122.
- [27] Q. Yu and E. S. Kuh, "Exact moment matching model of transmission lines and application to interconnect delay estimation," *IEEE Trans. VLSI Syst.*, vol. 3, pp. 311–322, June 1995.
- [28] Q. Yu, E. S. Kuh, and T. Xu, "Moment models of general transmission lines with application to interconnect analysis and optimization," *IEEE Trans. VLSI Syst.*, vol. 4, pp. 477–494, Dec. 1996.
- [29] Q. Zhu and W. M. Dai, "High-speed clock network sizing optimization based on distributed RC and lossy RLC interconnect models," *IEEE Trans. Computer-Aided Design*, vol. 15, pp. 1106–1118, Sept. 1996.



Tai-Chen Chen received the B.S. and M.S. degrees in computer and information science from the National Chiao Tung University, Hsinchu, Taiwan, in 1999 and 2001, respectively. He is currently working toward the Ph.D. degree in the Graduate Institute of Electronics Engineering, National Taiwan University, Taipei. His current research interests include computer-aided design and interconnect optimization for deep submicron technology.

Mr. Chen received the Best Master's Thesis Award from the National Science Council of the Republic of

China in 2002.



Song-Ra Pan received the B.S. and M.S. degrees in computer and information science from National Chiao Tung University, Hsinchu, Taiwan, in 1998 and 2000, respectively. He is currently working toward the Ph.D. degree at the University of California, Santa Barbara.

From 2002 to 2003, he was with the Taiwan Semiconductor Manufacturing Company, Ltd., Hsinchu Science- Based Industrial Park, Taiwan, R.O.C. His current research interests include computer-aided design, testing, and verification.



Yao-Wen Chang (S'94-M'96) received the B.S. degree from National Taiwan University, Taipei, in 1988, and the M.S. and the Ph.D. degrees, all in computer science, from the University of Texas at Austin in 1993 and 1996, respectively.

Currently, he is an Associate Professor in the Department of Electrical Engineering and the Graduate Institute of Electronics Engineering, National Taiwan University. In summer 1994, he was with the VLSI design group of IBM T. J. Watson Research Center, Yorktown Heights, New York. From 1996 to 2001,

he was on the faculty of the Department of Computer and Information Science, National Chiao Tung University, Taiwan, where he received an inaugural all-university Excellent Teaching Award in 2000. His research interests include physical design automation, architectures, and systems for VLSI and combinatorial optimization.

Dr. Chang received the Best Paper Award at the 1995 IEEE International Conference on Computer Design (ICCD-95) for his work on FPGA routing, reviewers' Best Paper nominations at the 2000 ACM/IEEE Design Automation Conference (DAC) for his work on the B*-tree floorplan representation, and Best Paper nomination at the 2002 IEEE/ACM International Conference on Computer-Aided Design (ICCAD) for his work on multilevel routing. He has served on the Technical Program Committees of several international conferences on VLSI Design Automation. He is a Member of IEEE Circuits and Systems Society, Association for Computing Machinery (ACM), and ACM/SIGDA.