Interconnect Estimation and Planning for Deep Submicron Designs

Jason Cong and David Zhigang Pan
Department of Computer Science
University of California, Los Angeles, CA 90095
Email: {cong,pan}@cs.ucla.edu *

Abstract

This paper reports two sets of important results in our exploration of an interconnect-centric design methodology for deep submicron (DSM) designs: (I) We obtain a set of efficient, accurate performance and area estimation models for optimal wire sizing (OWS) using two simple wire sizing schemes, namely single-width sizing (1-WS) and two-width sizing (2-WS). These simple, efficient estimation models enable us to explore the trade-off between delay and area of interconnect designs. They also enable high level design tools to consider interconnect layout optimization during design planning. (II) Guided by our interconnect estimation models, we study the interconnect architecture planning problem for wire-width designs. We achieve a rather surprising result which suggests that two pre-determined wire widths per metal layer are sufficient to achieve near-optimal performance for current and future technologies from $0.25\mu m$ to $0.07\mu m$ generations.. This result will greatly simplify the routing architecture and routing tools for DSM designs. We believe that our interconnect estimation and planning results will have a significant impact to guide high-performance DSM designs.

1 Introduction

As VLSI technology moves to deep submicron (DSM) dimensions and gigahertz clock frequencies, VLSI interconnects play the dominant role in determining the overall performance, power, reliability, and cost of the system. Recently, many interconnect optimization techniques, including wire sizing and spacing, buffer insertion and sizing, etc., have been proposed and shown to be very effective for interconnect performance optimization (e.g., see [1] for a recent survey). However, in the conventional VLSI design flow, interconnect optimization is usually performed at late stages in the design process. As a consequence, accurate interconnect delay and area, especially those for global interconnects are not known to higher level synthesis and design planning tools. Therefore, it is necessary to consider an interconnect-centric design flow which includes interconnect estimation and planning, optimal interconnect synthesis, and efficient interconnect layout implementation at each level of the design process. Early interconnect estimation and planning are critical to assure the proper coupling between synthesis and layout to achieve the design convergence.

So far, there has been very limited work on interconnect estimation and planning that consider interconnect layout optimization. [2] provides the first systematic study on interconnect delay estimation under interconnect optimization. It derived a set of simple delay estimation models (DEM) under various optimization techniques, e.g., optimal wire sizing, simultaneous driver and wire sizing, and simultaneous buffer insertion/sizing and wire sizing. These DEMs are shown to have 90% accuracy when compared with those

*This research is partially sponsored by Semiconductor Research Corporation under Contract 98-DJ-605.

obtained by running corresponding complex interconnect optimization algorithms (for example, those from the TRIO package in [1]) directly. These DEMs can be used in various design planning stages to provide accurate timing information without really going into the layout details.

However, [2] does not provide any estimation of wiring area used for interconnect optimization. The wiring resources must also be planned at high levels to make sure that the planned interconnect optimization is realizable at the layout level. Also, the trade-off between delay and area is not available from these models. Furthermore, the coupling effect under variable spacings is not modeled in [2], while coupling capacitance affects the delay calculation considerably in DSM designs.

In this paper, we study interconnect estimation for both delay and area, with consideration of coupling capacitance. Based on our simple but accurate estimation modeling, we propose a novel interconnect architecture planning for wire-width design. Our main contributions include the following:

- First, we show that the delay and area of the optimal wire sizing (OWS) solutions [3] can be approximated accurately by two simple wire sizing schemes, namely single-width sizing (1-WS) and two-width sizing (2-WS). When the coupling capacitance is considered explicitly, 2-WS outperforms 1-WS in terms of delay and area reduction, and achieves close to optimal solution compared to running global interconnect sizing and spacing (GISS) algorithm [4] directly.

- We study the area/delay trade-off using the closed-form interconnect estimation models from 1-WS and show that delay is not sensitive to certain variation of wire width around the optimal width. Therefore, we can achieve significant area reduction with a slight increase in delay. In particular, we show that the AT^4 (A–area, T–delay) metric works well to guide the area-efficient performance optimization and leads to more than 60% area reduction from OWS but with only about 10% delay increase.

- The delay sensitivity study further suggests that there exists some small set of "globally" optimal widths for a wide range of interconnect lengths. Therefore, we study the interconnect architecture planning for wire-width design at each metal layer.

- We derive such "globally" optimal 1-width and 2-width designs, and show rather surprisingly that using two "pre-designed" widths, we are still able to achieve close to optimal performance compared with those obtained by GISS using many possible widths. For $0.10\mu m$ technology, our 2-width design leads to optimal 2-WS solution which is only no more than 7% away from those by GISS, regardless of wire length distributions.

- We further provide the area-efficient, high-performance 2-width design recommendation for every metal layer in each of future technology generations.

The rest of the paper will be organized as follows. Section 2 states the notations and preliminaries. In Section 3, the efficient and accurate interconnect delay and area estimation models are derived. Delay versus area trade-off is also explored. In Section 4, interconnect architecture planning is studied, and the 2-width design is shown to work surprisingly well and recommended for future technologies. The conclusion follows in Section 5.

2 Notations and Preliminaries

To derive efficient and accurate interconnect estimation models, we need simple but accurate delay computation, as well as a set of key interconnect and device parameters. We model the driver as an effective resistance R_d connected to an ideal voltage source. The well-known Elmore delay model [5, 6] is used for delay computation. Although Elmore delay model is not very accurate in DSM design, especially for delay calculation of near-source critical sinks due to the resistive shielding [7], it is accurate enough for our estimation purpose to provide guidance to high-level design planning[1]. The key interconnect and device parameters for our estimation modeling are identified and listed below.

- W_{min}: the minimum wire width, in μm
- S_{min}: the minimum wire spacing in μm
- r: the sheet resistance, in $\Omega/\Box$
- c_a: the unit area capacitance, in $fF/\mu m^2$
- c_f: the unit effective-fringing capacitance[2], in $fF/\mu m$
- t_g: the intrinsic device delay in ps
- c_g: input capacitance of a minimum device, in fF
- r_g: output resistance of a minimum device, in $k\Omega$

The values of these parameters for our study are shown in Table 1. They are based on the *1997 National Technology Roadmap for Semiconductors* (NTRS'97) [8]. In the table, a tier is defined to be a pair of adjacent metal layers with the same cross-sectional dimensions [9, 10]. So from bottom to top, Tier1 refers to metal layers 1 and 2, Tier2 refers to metal layers 3 and 4, ..., and Tier4 refers to metal layers 7 and 8. NTRS'97 only provides the geometry information for Tier1. To study the effect of interconnect reverse scaling [9, 11, 10] at higher metal layers, we extract a set of RC parasitics for higher metal layers, based on the geometry information from UC Berkeley's Strawman technology [12, 13] and from SEMATECH [14]. For capacitance extraction, we use the 2.5D capacitance extraction methodology reported in [15] which uses a 3-D field solver to generate accurate capacitance values for interpolation and extrapolation.

3 Interconnect Delay and Area Estimation

Proper wire sizing has been shown to be very effective to reduce interconnect delay for DSM designs. It was first proposed in [3] and later on studied by others with various extensions [16, 17, 18, 19, 20, 21, 22]. However, these works did not consider the coupling capacitance which becomes the dominant capacitance component in DSM designs. A recent work in [4] took the coupling capacitance into consideration by performing global interconnect sizing and spacing (GISS) for multiple nets simultaneously, and provided further delay reduction than OWS. In this section, we show that the delay and area of optimal wire sizing (OWS) [3] can be estimated accurately by two simple wire sizing schemes, namely single-width sizing (1-WS) and two-width sizing (2-WS). When the coupling capacitance need to be considered explicitly, 2-WS will have further delay and area reduction than 1-WS and achieve near optimal solution compared to running complex GISS algorithm [4]. We further explore the delay/area trade-off and propose a new metric for area-efficient delay optimization.

[1] At the high-level design and planning, other sources of errors, such as interconnect estimation of coupling capacitance due to unknown neighborhood structures, will outweight the inaccuracy due to the Elmore delay model.

[2] It is defined as the sum of the fringing and coupling capacitances, as introduced in [4].

Tech. (μm)		0.25	0.18	0.13	0.10	0.07
W_{min}		0.25	0.18	0.13	0.10	0.07
S_{min}		0.34	0.24	0.17	0.14	0.10
t_g		86.6	66.4	54.4	50.1	29.8
c_g		0.282	0.234	0.135	0.072	0.066
r_g		16.2	17.1	22.1	23.4	22.1
Tier1	r	0.073	0.068	0.081	0.092	0.095
	c_a	0.059	0.060	0.046	0.053	0.056
	c_f	0.082	0.064	0.043	0.045	0.040
Tier2	r	0.016	0.011	0.018	0.022	0.030
	c_a	0.021	0.0176	0.0128	0.0136	0.0163
	c_f	0.206	0.160	0.103	0.103	0.089
Tier3	r	0.013	0.0088	0.011	0.011	0.012
	c_a	0.0125	0.0097	0.0067	0.0074	0.0077
	c_f	0.154	0.119	0.104	0.103	0.088
Tier4	r	-	-	0.0088	0.0088	0.0075
	c_a	-	-	0.0043	0.0043	0.0035
	c_f	-	-	0.0782	0.0782	0.0904

Table 1: Parameters based on NTRS'97. For Tier2 through Tier4, interconnect reverse scaling is considered.

3.1 Interconnect Estimation under Single-Width Sizing (1-WS)

3.1.1 Algorithm and Estimation Model

Given an interconnect of length l with loading capacitance C_L and driver resistance R_d, as shown in Figure 1(a), single-width sizing (1-WS) problem is to determine the best uniform width that minimizes the delay. To compute the distributed Elmore delay, the original wire is often divided into many small wire segments, and each wire segment is modeled as a π-type RC circuit. For example, Figures 1 (b) and (c) show the 1- and k-segment π-type RC circuits. In the figures, R_w is the total wire resistance, and C_w is the total wire capacitance. For uniform wire, we have the following lemma.

Lemma 1 *The Elmore delay of a uniform width wire is unchanged regardless how it is divided to shorter segments.*

Proof: First, we prove the case of two segments. Without loss of generality, we assume they have length αl and $(1-\alpha)l$. According to [6], the Elmore delay is

$$\begin{aligned} T &= R_d(C_w + C_L) + \alpha R_w \cdot \left[\frac{\alpha C_w}{2} + (1-\alpha)C_w + C_L\right] + (1-\alpha)R_w \cdot \left[\frac{(1-\alpha)C_w}{2} + C_L\right] \\ &= R_d(C_w + C_L) + R_w \cdot (\frac{C_w}{2} + C_L) \end{aligned}$$

which is independent of α. For the case of more than 2 segments, simple mathematical induction can be used to prove it. □

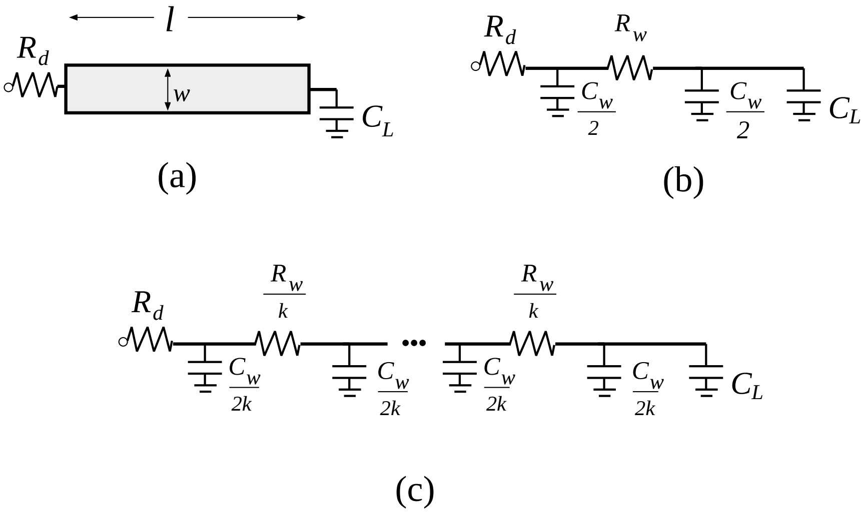


Figure 1: (a) Single-width sizing (1-WS) to determine the optimal uniform width w. (b) The 1-segment π-type RC model for the interconnect. (c) The distributed k-segment π-type RC model. It can be shown that (b) and (c) are equivalent to compute the Elmore delay.

From Lemma 1, we can just use one-segment π-model in Figure 1(b) to compute the Elmore delay for Figure 1(a).

$$\begin{aligned} T(w,l) &= R_d\left[(c_a \cdot w + c_f)\cdot l + C_L\right] + \frac{rl}{w}\cdot\left[\frac{(c_a\cdot w + c_f)\cdot l}{2} + C_L\right] \\ &= R_d c_f l + R_d C_L + \frac{1}{2} r c_a \cdot l^2 + R_d c_a l \cdot w + \left(\frac{1}{2} r c_f l^2 + r l C_L\right)\cdot\frac{1}{w} \qquad (1) \end{aligned}$$

Thus the best wire width to minimize $T(w,l)$ is

$$w^*(l) = \sqrt{\frac{r(c_f l + 2C_L)}{2R_d c_a}} \qquad (2)$$

The optimal delay for 1-WS under w^* is

$$T_{1ws}(l) = R_d C_L + R_d c_f l + \sqrt{2R_d c_a r(c_f l + 2C_L)}\cdot l + \frac{1}{2} r c_a \cdot l^2 \qquad (3)$$

The four terms at the r.h.s. of (3) are constant, linear, super-linear and quadratic terms of l, respectively and they are called Term1 through Term4.

Lemma 2 *Let $f(x) = x\sqrt{x+a}$ ($x > 0$, and $a \geq 0$). Then $f(x)$ is a convex function.* □

Theorem 1 *T_{1ws} is a quadratic convex function of the interconnect length l.*

Proof: We can easily show that Term2 and Term4 in (3) are convex functions. From Lemma 2, Term3 in (3) is also convex function of l. Since the positive linear combination of convex functions are still convex [23], we have the above theorem. □

Corollary 1 *Since T_{1ws} is convex, the equally spaced buffer insertion algorithm as in [2] can be used to perform simultaneous buffer insertion and wire sizing.* □

3.1.2 Comparison of 1-WS with OWS

Our experiments show surprisingly that the optimized delay under 1-WS is close to that from running OWS algorithm [3] for a wide range of parameters from NTRS'97. As an example, Figure 2 shows the comparison of optimized delays for an interconnect of length up to 2cm, under 1-WS and OWS for Tier1 and Tier4 using 0.10 μm technology. The 1-WS for Tier1 has only up to 10% more delay than OWS for wires shorter than 5mm. And the 1-WS for Tier4 has almost the same delay as OWS for all wire lengths up to 2cm (the chip dimension). Note that in theory, T_{1ws} is quadratic function of l, while T_{ows} from the closed-form delay estimation modeling under OWS in [2] is sub-quadratic function of l. To see under what condition 1-WS has delay close to that from OWS, we draw the delay distributions among different terms in Eqn. (3), namely Term1 through Term4, respectively. Figure 3 shows the total delay and delay distributions among these four terms. We observe that as long as the quadratic Term4 in (3), i.e., $\frac{1}{2}rc_al^2$ is smaller than both Term2 and Term3, 1-WS approximates OWS well (in general within 90% accuracy). Thus 1-WS can be used to estimate the delay for OWS provided that

$$\begin{aligned} & \frac{1}{2}rc_al^2 < R_dc_fl, \\ and \quad & \frac{1}{2}rc_al^2 < \sqrt{2R_dc_ar(c_fl + 2C_L)}. \end{aligned}$$

The two inequalities above are satisfied as long as $l < 2R_dc_f/rc_a$. For Tier1, $2R_dc_f/rc_a = 4.3mm$; for Tier4, $2R_dc_f/rc_a = 96cm$ which is much larger than the chip dimension. This explains why 1-WS and OWS delays are so close for wires shorter than 4mm in Tier1[3] and for all wires up to chip dimension in Tier4.

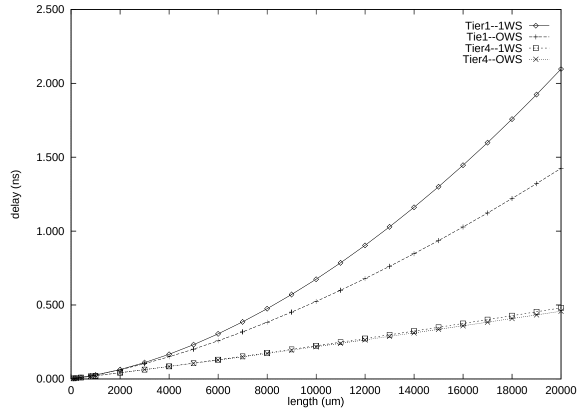
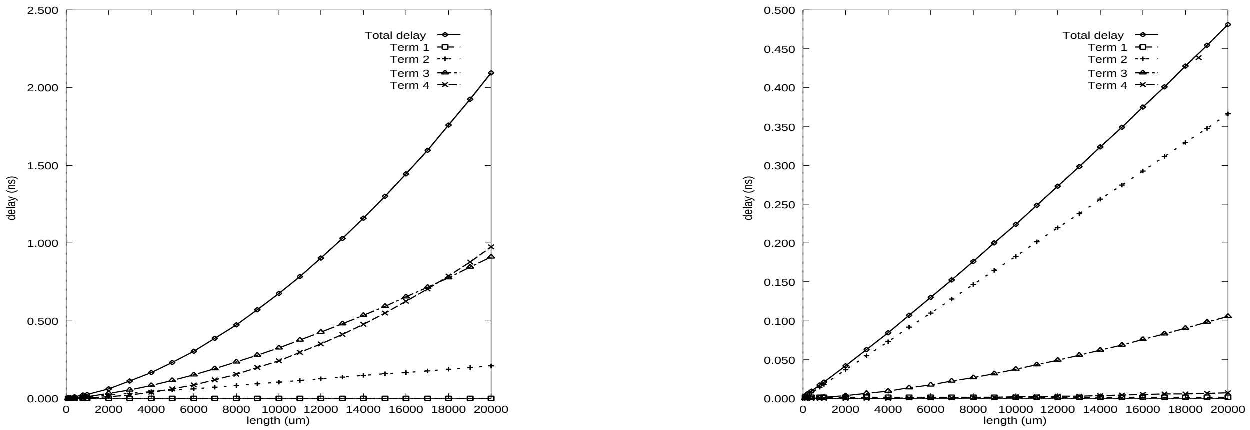


Figure 2: Comparison of 1-WS and OWS for Tier1 and Tier4 under the 0.10 μm technology. $R_d = r_g/100$, $C_L = c_g \times 100$. To run OWS algorithm, we set $W_{max} = 50 \times W_{min}$ with the width incremental to be $\frac{1}{2}W_{min}$ and the wire is segmented in every $100\mu m$ (same for other Figures).

In Figure 4, we show the comparison of average wire widths under 1-WS and OWS. To our surprise, 1-WS and OWS have very similar average wire width too. For Tier1, the average wire width under 1-WS is almost identical to that under OWS. For Tier4, the average wire width under 1-WS is just slightly larger (about 5%) than that under OWS. So we can conclude that w^* in (2) from 1-WS is a very good approximation to the average wire width (i.e., wire area) under OWS.

[3]Tier1 is mostly used for local interconnects, which are much shorter than 4mm. So in practice, 1-WS scheme works very well for all tiers compared to OWS, as we shall see in Section 4.



(a) (b)

Figure 3: Delay distribution of different terms in T_{1ws} for (a) Tier1 and (b) Tier4 under the $0.10\mu m$ technology. $R_d = r_g/100$, $C_L = c_g \times 100$.

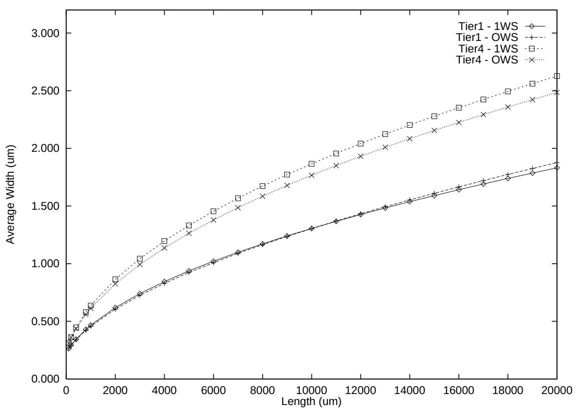


Figure 4: Comparison of 1-WS and OWS average wire width for Tier1 and Tier4 under 0.10 μm technology with $R_d = r_g/100$, $C_L = c_g \times 100$. To run OWS algorithm, we set $W_{max} = 50 \times W_{min}$.

3.1.3 Comparison of 1-WS with GISS

So far we have considered wire-sizing scenario of fixed effective-fringing capacitance, which assumes some fixed nominal spacing to neighboring nets and lumps the associated coupling capacitance with the fringing capacitance to ground to form effective-fringing capacitance. In actual routing, however, the pitch-spacings, defined as the distances between the center lines of two neighboring wires (see Figure 5), are usually fixed. So the edge-to-edge spacing will be different as wire width changes, resulting in different coupling capacitances. The 1-WS solution, in this case, is not flexible enough to take the advantage of down-sizing certain downstream (i.e., closer to sinks) wire segments (wire tapering) to reduce their coupling capacitances to the neighbors. Figure 6 shows the delay comparison of the 1-WS solution with using GISS algorithm from [4] with variable coupling capacitance. We can see that the delay from 1-WS is about 20-30% larger than that from GISS. Also 1-WS tends to use larger area since it does not allow wire tapering. However, smaller sizing in the downstream does provide more benefit in the case of fixed pitch-spacings since smaller sizing will reduce the coupling capacitance. In the next subsection, we will study 2-width sizing and show that it is sufficient to consider the variable coupling capacitances.

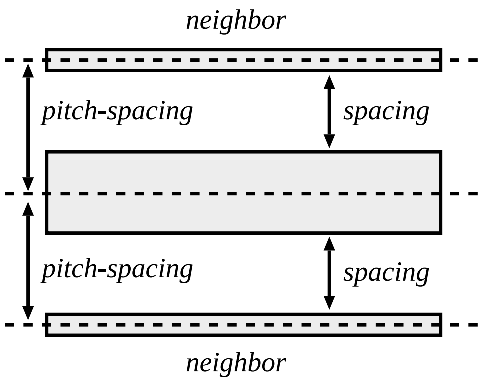


Figure 5: Illustration of the edge-to-edge spacing and the center-to-center pitch-spacing.

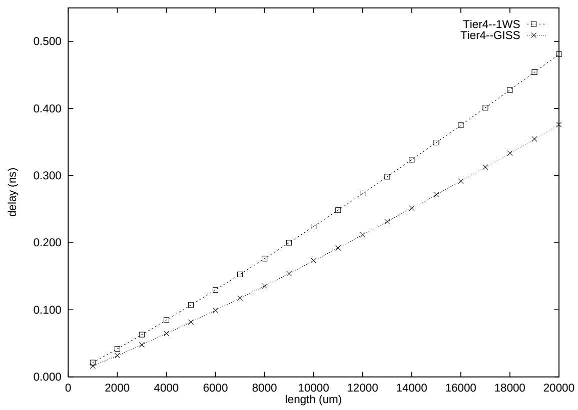


Figure 6: Comparison of 1-WS with GISS under fixed pitch-spacing. To run GISS, $W_{max} = 50 \times W_{min}$ with the width incremental to be $\frac{1}{2}W_{min}$ and 10 segments for each wire are used.

3.2 Interconnect Estimation under Two-Width Sizing (2-WS)

3.2.1 Algorithm and Estimation Model

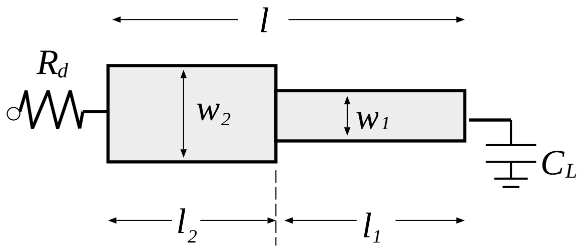


Figure 7: Two-width sizing (2-WS). We need to determine the best w_2, w_1, l_2, and l_1 with the constraint of $l_1 + l_2 = l$.

In this subsection, we study two-width sizing (2-WS). As shown in Figure 7, given R_d, l and C_L, we shall determine the best w_2, w_1, l_2 and l_1, with $l_2 + l_1 = l$. Using Lemma 1, the distributed Elmore delay can be written as

$$\begin{aligned} T(w_1, w_2, l_1, l_2) &= R_d \cdot (c_f l + c_a w_2 l_2 + c_a w_1 l_1 + C_L) \\ &\quad + \frac{r l_2}{w_2} \cdot [(c_a w_2 + c_f) l_2/2 + (c_a w_1 + c_f) l_1 + C_L] \\ &\quad + \frac{r l_1}{w_1} \cdot [(c_a w_1 + c_f) l_1/2 + C_L] \\ &= R_d(c_f l + C_L) + \frac{1}{2} r c_a (l_2^2 + l_1^2) + R_d c_a (w_2 l_2 + w_1 l_1) \\ &\quad + r c_a l_1 l_2 \frac{w_1}{w_2} + \frac{r c_f l_1 l_2}{w_2} + \frac{r c_f l_2^2}{2 w_2} + \frac{r c_f l_1^2}{2 w_1} + r C_L (\frac{l_2}{w_2} + \frac{l_1}{w_1}) \end{aligned} \tag{4}$$

Substituting the constraint $l_1 = l - l_2$ into the above equation, it is not difficult to conclude that $T(w_1, w_2, l_2)$ is not a posynomial [24] or a convex function. Therefore, multiple local optimal solutions may exist. However, if we assume that l_2 is fixed (which implies $l_1 = l - l_2$ is also fixed), then T becomes a function of w_1 and w_2. It is easy to see that T is now a posynomial function of w_1 and w_2 in the following form.

$$T(w_1, w_2) = K_1 w_1 + \frac{K_2}{w_1} + K_3 w_2 + \frac{K_4}{w_2} + K_5 \frac{w_1}{w_2}. \tag{5}$$

where K_1, K_2, K_3, K_4, K_5 are some constant coefficients.

It is well known that a posynomial function can be transformed into a convex function [24] so that a local optimal solution is the global optimal solution. We can obtain the optimal w_1 and w_2 in (5) by adapting the local refinement (LR) operation[4] first introduced in [3] for discrete wire-sizing. The LR procedure is shown in Figure 8. Our experiments show that for a precision of $\delta_w = 0.001\mu m$, and $\delta_T = 1ps$, only 3-5 iterations are needed in step 2 to get the optimal w_1 and w_2[5]. Since for each given length l_2 (or equivalently l_1), we can compute the best w_1, w_2 and T in just a few steps, we can use a simple linear search to get the best l_2 (and l_1). We denote the best delay under 2-WS to be $T_{2ws}(R_d, l, C_L)$, and the best widths to be w_2^* and w_1^*.

[4] Directly setting $\partial T/\partial w_1 = 0$ and $\partial T/\partial w_2 = 0$ from (5) will result in a 5-th order equation which does not have closed-form solution. So we use the fast numerical technique of LR.

[5] In fact, the results in [25] shows that LR converges to optimal solution at the exponential rate, i.e., $|x^{[k]} - x^*| \leq \alpha^k |x^{[0]} - x^*|$ ($0 < \alpha < 1$ is the convergence rate).

Local refinement to compute 2-WS
Input: K_1, K_2, K_3, K_4, K_5 as in Eqn.(5)

1. Initialize $w_1, w_2 \leftarrow W_{min}$, and compute T from Eqn.(5);
2. repeat {
    $w_2' \leftarrow w_2$;
    $w_1' \leftarrow w_1$;
    $T' \leftarrow T$;
    $w_2 \leftarrow \sqrt{(K_4 + K_5 w_1')/K_3}$;
    $w_1 \leftarrow \sqrt{K_2/(K_1 + K_5/w_2')}$;
    Compute T from Eqn.(5);
  } until ($\lvert w_2' - w_2 \rvert < \delta_w$ and $\lvert w_1' - w_1 \rvert < \delta_w$ and $\lvert T' - T \rvert < \delta_T$);
3. return w_2 and w_1;

Figure 8: The local refinement procedure for 2-WS.

Note that a 2-WS solution can handle different coupling capacitances and thus different effective-fringing capacitances, c_{f1} and c_{f2} (in general, $c_{f1} < c_{f2}$ as $w_1 < w_2$), as shown in Figure 9. The nice feature of using 2-width is that it can determine the best length l_2 (or equivalently l_1) to adjust to the variable coupling capacitance effect and minimize the delay. Similar to (4), we can write the delay T as a function of l_2 in the following form.

$$T(w_1, w_2, l) = A \cdot l_2^2 + B \cdot l_2 + C \tag{6}$$

where

$$A = rc_a(1 - \frac{w_1}{w_2}) + \frac{1}{2} r(\frac{c_{f1}}{w_1} - \frac{c_{f2}}{w_2}) \tag{7}$$

$$B = R_d(c_{f2} - c_{f1}) + R_d c_a(w_2 - w_1) + rc_a l(\frac{w_1}{w_2} - 1) + rl(\frac{c_{f2}}{w_2} - \frac{c_{f1}}{w_1}) + r(\frac{1}{w_2} - \frac{1}{w_1}) \tag{8}$$

$$C = R_d(c_{f1} l + c_a w_1 l) + \frac{1}{2} r(c_a + \frac{c_{f1}}{w_1}) l^2 + \frac{rC_L l}{w_1} \tag{9}$$

Then the optimal length for l_2 is

$$l_2^* = \begin{cases} 0, & -\frac{B}{2A} < 0 \\ -\frac{B}{2A}, & 0 \le -\frac{B}{2A} \le l \\ l, & -\frac{B}{2A} > l \end{cases} \tag{10}$$

The optimal delay is thus

$$T^*(w_1, w_2, l) = A \cdot {l_2^*}^2 + B \cdot l_2^* + C \tag{11}$$

and the area is

$$A^*(w_1, w_2, l) = w_2 l_2^* + w_1(l - l_2^*) \tag{12}$$

3.2.2 Comparison of 2-WS with OWS and GISS

Figure 10 shows the optimized delay comparison of 1-WS, 2-WS and OWS for Tier1 and Tier4 under $0.10\mu m$ technology. For Tier1, 2-WS and 1-WS have very similar delay up to interconnect length of 5mm. For a 2cm

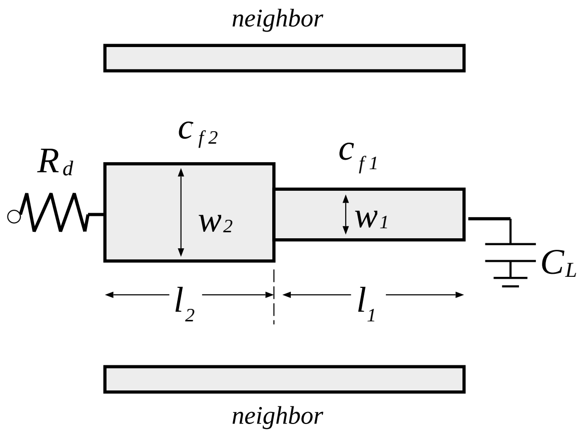


Figure 9: 2-width sizing with different effective-fringing capacitances.

global interconnect, the delay using 2-WS is 1.76ns, which is 16% less than the 2.09ns delay under 1-WS. The average wire width under 2-WS is still almost the same as OWS, similar to Figure 4. The optimal widths w_2^* and w_1^* are shown in Figure 11. It is interesting that for all the wire lengths, the ratio of w_2^*/w_1^* is about 2-3. This observation will be useful in Section 4 to guide interconnect planning. Note that we have a fixed ratio of $w_2/w_1 = \rho$, we can solve (5) directly to get the best w_1^*, without going through LR iterations.

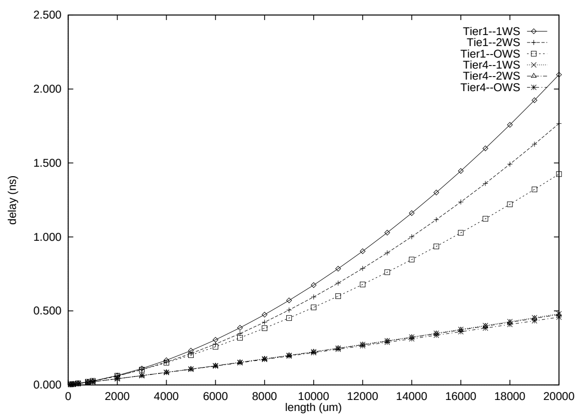


Figure 10: Comparison of 1-WS, 2WS and OWS for Tier1 and Tier4 using the 0.10 μm technology. $R_d = r_g/100$, $C_L = c_g \times 100$.

In the case of fixed pitch spacing (i.e., variable coupling capacitance), the 2-WS scheme is more flexible than 1-WS. Figure 12 shows the delay comparison of using 1-WS, 2-WS and GISS. 2-WS has up to 15% delay reduction from 1-WS and has close to optimal solution from that by running GISS.

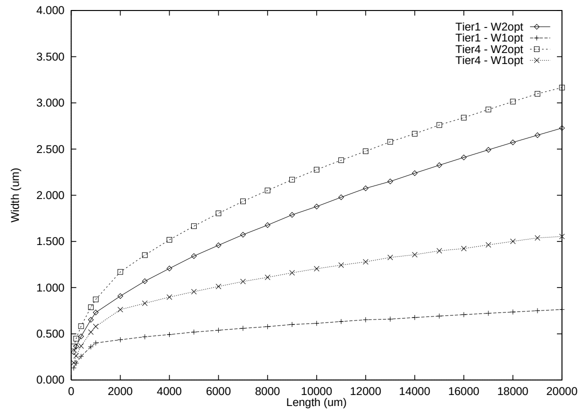


Figure 11: The two optimal widths w_2^* and w_1^* for Tier1 and Tier4 under the 0.10 μm technology. $R_d = r_g/100$, $C_L = c_g \times 100$.

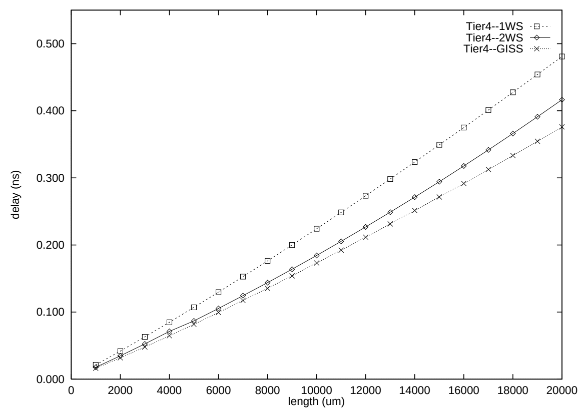


Figure 12: Comparison of 1-WS, 2-WS and GISS with variable coupling capacitance for Tier4 using the 0.10 μm technology. $R_d = r_g/100$, $C_L = c_g \times 100$.

3.3 Delay/Area Trade-off and Sensitivity Study

The simple closed-form delay formula of 1-WS enables us to study the sensitivity of delay versus wire width. From Eqn. (1), we can compute the differential

$$\frac{dT}{dw} = R_d c_a - \frac{\frac{1}{2} r c_f l^2 + r l C_L}{w^2}.$$

As shown in Figure 13, delay decreases sharply as width increases from minimum wire width (i.e., 0.10 μm) since $\frac{dT}{dw} << 0$ when $w \approx W_{min}$, then flattens as $\frac{dT}{dw}$ slowly achieves 0 where the delay is the minimum, and after that delay increases slowly as $\frac{dT}{dw} > 0$. The optimal width w^* is about 2.6μm for a 2cm global interconnect in Tier4 under 0.10μm. However, it is not difficult to see in order to achieve the minimum delay, the cost, in terms of wire area, is high. For example, using wire width of 1μm has only 10% more delay than the optimal OWS, but saves 62% area. Therefore, delay minimization only could lead to significantly larger area!

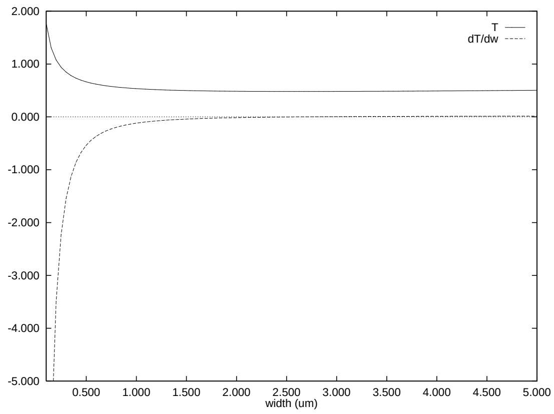


Figure 13: The delay T and its sensitivity to w, $\frac{dT}{dw}$, using different uniform wire width for a 2cm global interconnect using the 0.10 μm technology. $R_d = r_g/100$, $C_L = c_g \times 100$.

To obtain a good metric for area efficient performance optimization, we have performed extensive experiments on different area-delay metrics, including T (delay only), AT (area-delay product), AT^2 (area-delay-square product), AT^3, AT^4, AT^5, etc. Our study concludes that AT^4 is a metric that is suited for area-efficient performance optimization, with in general of only about 10% delay slack from OWS, but with significant area reduction. Figure 14 shows an example. The optimal widths of a 2cm interconnect for T, AT, AT^2, AT^3, AT^4, AT^5 are 2.6μm, 0.10μm, 0.30μm, 0.60μm, 1.0μm, and 1.15μm, respectively, with delays of 0.48ns, 1.77ns, 0.84ns, 0.62ns, 0.53ns, and 0.52ns respectively. The optimal 1-WS solution under the AT^4 metric uses 62% smaller wiring area compared to OWS (20,000μm^2 vs. 52,000μm^2) with only 10% increase of delay. The performance-driven but area-efficient metric AT^4 will be used in Section 4 for interconnect architecture planning.

3.4 Other Applications of Interconnect Estimation Models

Our simple interconnect delay and area estimation models can be used in a wide spectrum of applications to guide high level design planning. Here are some examples:

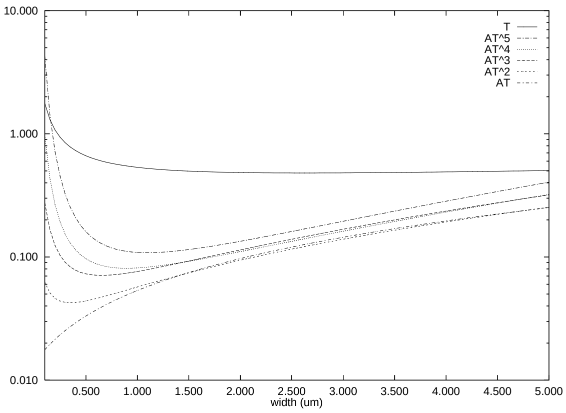


Figure 14: Different optimization metrics for a 2cm interconnect in Tier4 under the 0.10 μm technology. $R_d = r_g/100$, $C_L = c_g \times 100$. The y-axis is scaled to compare all metrics in one figure.

- Physical and RTL level floorplan: During the placement and sizing of functional blocks, one can use our models to accurately predict the delay and area of global interconnects.
- Placement-driven synthesis and mapping: One may keep a companion placement during synthesis and technology mapping [26]. For every logic synthesis operation, the companion placement will be updated. Once the cell positions are known, one can use our models to accurately predict interconnect delay and area for the synthesis engine with consideration of wire sizing.
- Interconnect architecture planning: Interconnect parameters (e.g., metal width, aspect ratio, spacing, etc.) may be tuned to optimize the delays predicted by our models for global, average and local interconnects under certain wire-length distributions. In next section, we will study the interconnect architecture planning problem for wire-width design to achieve area-efficient performance optimization.

4 Interconnect Architecture Planning

4.1 Motivation

From our study of 1-WS and 2-WS in the previous section, a very interesting observation is that the delay is not sensitive to certain degree wire width variations around the optimal solution (e.g., see Figure 13). This not only suggests that we can achieve close to optimal performance with significant area saving (as we show in Section 3.3), but also suggests that there may exist a small set of "globally" optimal widths for a range of interconnect lengths, so that by just using such a small set of pre-determined "fixed" widths for all the lengths within a reasonably wide range, we are still able to get close to optimal performance for all interconnects in the length range! In Figure 15, we draw the delay sensitivity versus wire width for three interconnects of length 0.5cm, 1cm and 2cm. The optimal widths for them are about 1.0 μm, 1.4μm, and 2.6μm. However, any 1-WS with width from 1.0 μm to 2.0 μm will have less than 10% delay from that of OWS for all three lengths.

This crucial observation motivates our study on the interconnect architecture planning for optimal wire-width design. In particular, we want to determine certain "optimal" and *fixed* one width for 1-WS or a pair of fixed widths for 2-WS during design planning such that by using just these pre-determined widths,

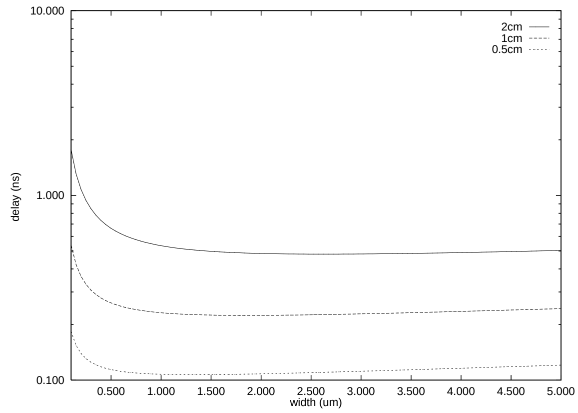


Figure 15: Delay sensitivity of using different width for a 0.5cm, 1cm and 2cm lines for Tier4 of 0.10 μm technology. $R_d = r_g/100$, $C_L = c_g \times 100$.

we can achieve close to optimal performance for a wide range of interconnect lengths (not just one length!) compared to the wire sizing solution with many different widths obtained from running complicated wire sizing (and/or spacing) algorithms. This optimal wire-width design, on the one hand, still guarantees close to optimal performance; on the other hand, greatly simplifies the routing problem and the interaction of layout optimization with other higher level design planning tools and lower level routing tools. In particular, if only one or two fixed widths are used for every metal layer, a full-blown gridless router may not be necessary. This will significantly simplify many problems, including RC extraction, detailed routing and layout verification.

4.2 Problem Formulation

Our wire-width planning is tier-based, i.e., we will determine the best 1-W and 2-W designs for each tier. In general, local interconnects are routed in the lowest tier (Tier1), while global interconnects are routed in the highest tier (Tier3 or Tier4, depending on technology). The wire length distribution on different tiers usually varies from design to design, and also depends on the layout tools and optimization objectives. In our study, we assume that the maximum wire length (l_{max}) in Tier1 is $10,000\times$ feature size, and l_{max} in the top tier is L_{edge}, i.e, the chip dimension [14]. The l_{max} in the intermediate tiers will be determined by a geometric sequence such that for any tier i, $l_{max}(i+1)/l_{max}(i) = l_{max}(i)/l_{max}(i-1)$. For example, in 0.10 μm technology, $l_{max}(1) = 1000\mu m$, $l_{max}(4) = 22800\mu m$. Since $22.8^{1/4} = 2.84$, we have $l_{max}(2) = 2840\mu m$, and $l_{max}(3) = 8040\mu m$. The minimum wire length for tier i is the maximum length for tier $i-1$, i.e., $l_{min}(i) = l_{max}(i-1)$. Table 2 shows the wire length range of each tier for NTRS'97 technologies. We also take a representative driver for each metal tier for our wire width planning. The drivers for Tier1 through Tier4 are $10\times$, $40\times$, $100\times$, and $250\times$ of the minimum gate in the given technology, respectively.

Note that our interconnect planning tool is very flexible. If the designer specifies a different wire length distribution scheme for each layer, we can easily determine the optimal width accordingly. Given the wire length range for each tier, the wire-width design problem is to find the best width vector $\vec{W}$ such that the following objective function

$$\Phi(\vec{W}, l_{min}, l_{max}) = \int_{l_{min}}^{l_{max}} \lambda(l) \cdot f(\vec{W}, l) dl \tag{13}$$

is minimized, where $\lambda(l)$ is the distribution function of l, and $f(\vec{W}, l)$ is the objective function to be minimized

Tech. (μm)	0.25	0.18	0.13	0.10	0.07
Tier 1	0–2.50	0–1.80	0–1.30	0–1.00	0–0.70
Tier 2	2.50–6.50	1.80–5.85	1.30–3.27	1.00–2.84	0.70–2.30
Tier 3	6.50–17.3	5.85–19.0	3.27–8.23	2.84–8.04	2.30–7.57
Tier 4	-	-	8.23–20.7	8.04–22.8	7.57–24.9

Table 2: Wire length ranges (in mm) that are assigned to each tier.

by the design. In this study we choose $f(l) = A^j(\vec{W}, l) \cdot T^k(\vec{W}, l)$, where $A(l) = w_{avg} \cdot l$ is the area for l with average wire width of w_{avg}, and $T(\vec{W}, l)$ is the optimized delay using 1-WS or 2-WS. For our 1-width design, $\vec{W}$ has only one component W. For 2-width design, $\vec{W}$ has two components W_1 and W_2.

For our current study, we assume $\lambda(l)$ is a uniform distribution function. Other distribution functions such as wire length distribution function in [27], can also be used. Yet, our results in Section 4.4 show that our 2-width design is so robust that it can be applied to *any length distribution function*, with predictable small amount of errors compared with the optimal design objective using many possible widths. For $j = 0$ and $k = 1$, the objective is performance optimization only. However, as we observe in Section 3, delay minimization only tends to use large wire width with very marginal performance gain, since the delay/width curve becomes very flat while approaching optimal delay. Again, we use the AT^4 (i.e., $j = 1$ and $k = 4$) metric as suggested in Section 3.3 which was shown to be a good metric for area-efficient performance-driven design. For comparison, however, we will show the wire width designs under both metrics T and AT^4.

4.3 Overall Approaches

The overall approach of the wire-width planning is straightforward. Basically, we want to find the best 1-width or 2-width pair to minimize the objective function in (13). We take 1-width planning with metric T as an example to illustrate how the wire-width planning works. For 1-width planning, we need to determine the best width W^* to minimize

$$\int_{l_{min}}^{l_{max}} T(w, l) dl \tag{14}$$

where $T(w, l)$ is from Eqn. (1). So the "globally" optimal width W^* is thus

$$\begin{aligned} W^* &= \sqrt{\frac{\int_{l_{min}}^{l_{max}} r(\frac{1}{2} r c_f l + r C_L) l dl}{\int_{l_{min}}^{l_{max}} R_d c_a l dl}} \\ &= \sqrt{\frac{\frac{1}{3} r c_f (l_{max}^3 - l_{min}^3) + r C_L (l_{max}^2 - l_{min}^2)}{R_d c_a (l_{max}^2 - l_{min}^2)}} \end{aligned} \tag{15}$$

If $l_{max}^2 >> l_{min}^2$, which is the case for our length range for each tier, then W^* can be approximated as

$$W^* \approx \sqrt{\frac{\frac{1}{3} r c_f l_{max} + r C_L}{R_d c_a}}, \tag{16}$$

which is about $\sqrt{\frac{2}{3}} \cdot w^*(l_{max})$ from (2) provided that $C_L << c_f l_{max}$.

For the 1-width design under metric AT^4, a simple analytical formula like (15) or (16) cannot be obtained as we need to solve an 8-th order equation for w (it is not difficult to verify), which does not have analytical solutions. But since the complexity of our delay and area modeling is so low that we can just use a exhaustive search from available wire width (provided by given technology) to find the the best width design.

Similarly for the two-width design, we can obtain the "globally" optimal width pair W_1^* and W_2^* in a exhaustive search manner. Without loss of generality, we assume that $W_2^* = \alpha W_1^*$. To enable the grid-based routing, it is best to set α to be an integer. From Section 3.2, we find that α is usually between 2 to 3. Given each α, we can easily search the best W_1^*. In practice, we just need to search two α's[6], $\alpha = 2$ and $\alpha = 3$. Since according to Section 3.2, we have closed-form delay formula given some w_1 and $w_2 = \alpha w_1$, the complexity of our search for the best W_1^* and W_1^* is still very low. Indeed we shall point out that the complexity for our wire-width planning is not a major concern, since we just need to run it *once*.

4.4 Detailed Study of Wire-Width Planning for 0.10μm Technology

In this subsection, we present our detailed study and comparison of using 1-width and 2-width designs under both T and AT^4 metrics. Our study concludes that the 2-width design under AT^4 metric has both area efficiency and also near-optimal performance.

Table 3 shows the 1-width design the W^*'s for metric T in different tiers of 0.10μm technology. It takes about minimum wire width for Tier1, and sizes up in a factor from 2.5 to 5, with 3.82μm in Tier4. The average delays for Tier1 through Tier4 are about 70ps to 167ps, which are only a few percent larger than those obtained by running OWS algorithm with many different wire widths, listed at the last row of the table. Table 3 also shows the 1-width design W'^* under the optimization metric of AT^4. The W'^* for Tier2 to Tier4 is only 1/2 to 1/4 of corresponding W^* shown in Table 3, meaning a area reduction of 50% to 75%. But the delay under W'^* is still just 10-15% larger.

Tier	1	2	3	4
Length Range (mm)	0-1.00	1.00-2.84	2.84-8.04	8.04-22.8
W^* for T (μm)	0.11	0.55	1.40	3.82
$T_{avg}(W^*)$ (ps)	69.2	134.8	160.5	166.8
W'^* for AT^4 (μm)	0.10	0.13	0.43	1.83
$T_{avg}(W'^*)$ (ps)	69.3	155.5	181.1	180.2
T_{avg} by OWS (ps)	69.2	132.2	156.2	158.9

Table 3: 1-Width planning for 0.10μm technology.

Table 4 shows the 2-width design under T and AT^4 metrics in different tiers of 0.10μm technology. Again, it suggests that $W_1'^*$ and $W_2'^*$ are area efficient but still close to OWS.

However, when we assume fixed pitch-spacing and consider variable coupling capacitance when performing wire sizing, the 2-width design shows much more flexibility than the 1-width design. Table 5 shows the comparison of the average delay, the maximum delay difference (in percentage) compared with GISS (ΔT_{max}), and the average widths by using the pre-determined 1-width design, 2-width design (with metric AT^4) and by using GISS algorithm [4] for Tier4 under different pitch-spacings (pitch-sp). For pitch-spacing of 2.0 μm, 1-width design has average delay about 14% and 20% larger than those from 2-width design and GISS. Moreover, it has average wire width (thus area) about 1.83$\times$ and 1.92$\times$ of those from 2-WS and GISS. The 2-width design, however, has close to optimal delay compared to the solution obtained from running GISS algorithm (just 3-6% larger) and uses only slightly bigger area (less than 5%) than that of GISS.

[6]In fact, we try many different α's (not just integers) in our experiments and it turns out that $\alpha = 2$ or 3 is good enough.

Tier	1	2	3	4
W_1^* for T (μm)	0.10	0.33	0.84	2.32
W_2^* for T (μm)	0.20	0.66	1.68	4.64
$T_{avg}(W_1^*, W_2^*)$ (ps)	69.2	134.0	159.2	163.9
$W_1'^*$ for AT^4 (μm)	0.10	0.10	0.22	1.00
$W_2'^*$ for AT^4 (μm)	0.10	0.20	0.44	2.00
$T_{avg}(W_1'^*, W_2'^*)$ (ps)	69.3	144.1	180.2	176.6
T_{avg} by OWS (ps)	69.2	132.2	156.2	158.9

Table 4: 2-Width planning for $0.10\mu m$ technology.

Note that when the pitch-spacing becomes larger, the difference between 1-width design, 2-width design and GISS will get smaller. In Table 5, we also list the maximum delay difference(ΔT_{max}) from GISS. This is an important metric which bounds our estimation error under *any length distribution function* $\lambda(l)$ in our objective function based on the following theorem.

Scheme	pitch-sp=2.0 μm			pitch-sp=2.9 μm			pitch-sp=3.8 μm		
	T_{avg}	ΔT_{max}	avg-w	T_{avg}	ΔT_{max}	avg-w	T_{avg}	ΔT_{max}	avg-w
1-width	0.245	28.2%	1.98	0.177	15.7%	1.83	0.143	5.9%	1.63
2-width	0.215	7.0%	1.08	0.167	5.9%	1.23	0.140	3.9%	1.41
GISS [4]	0.204	-	1.03	0.159	-	1.19	0.136	-	1.38

Table 5: Comparison of the average delay (in ns), the maximum delay difference (in percentage) compared with GISS, and the average wire width (in μm) of using 1-width design, 2-width design and running GISS algorithm with consideration of variable coupling capacitance under different pitch-spacings. Tier4 $0.10\mu m$ technology is used.

Theorem 2 *If* $\left|\frac{f(\vec{W},l)-f(\vec{W}^*,l)}{f(\vec{W}^*,l)}\right| \le \delta_{max}$ *for any* $l \in (l_{min}, l_{max})$*, then for any distribution function* $\lambda(l)$*, we have*

$$\left|\frac{\Phi(\vec{W}, l_{min}, l_{max}) - \Phi(\vec{W}^*, l_{min}, l_{max})}{\Phi(\vec{W}^*, l_{min}, l_{max})}\right| \le \delta_{max} \tag{17}$$

Proof: The left hand side of (17) can be written as

$$\begin{aligned} l.h.s. &= \left|\frac{\int_{l_{min}}^{l_{max}} \lambda(l) \cdot \left[f(\vec{W},l) - f(\vec{W}^*,l)\right] dl}{\int_{l_{min}}^{l_{max}} \lambda(l) \cdot f(\vec{W}^*,l) dl}\right| \\ &\le \left|\frac{\int_{l_{min}}^{l_{max}} \lambda(l) \cdot \delta_{max} \cdot f(\vec{W}^*,l) dl}{\int_{l_{min}}^{l_{max}} \lambda(l) \cdot f(\vec{W}^*,l) dl}\right| \\ &= \delta_{max} \end{aligned}$$

Now the significance of ΔT_{max} in Table 5 is clear. Although we derive the optimal 2-width design using the uniform distribution $\lambda(l) \equiv 1$, our maximum delay difference ΔT_{max} using 2-width design is only 3.9–7% under different pitch spacings. Therefore, from Theorem 2, our 2-width design differs from many-width design by at most 3.9–7% for *any distribution function* $\lambda(l)$.

4.5 Recommendation for Future Technologies

To complete our study, we have performed interconnect architecture planning for all major technology generations listed in NTRS'97 from $0.25\mu m$ to $0.07\mu m$. Our recommendation is based on the optimal 2-width design considering the area-efficient performance optimization metric AT^4. The results are shown in Table 6. It suggests the minimum widths for local interconnects in Tier1. For Tier2 to Tier4, there are two different pre-determined wire widths with 1:2 ratio. Therefore, we have a wiring hierarchy on different metal layers such that Tier2 is about 1-2 times wider than Tier1, Tier3 is about 2-3 times wider than Tier2, and Tier4 (if available) is about 4-5 times wider than Tier3. Such a wiring hierarchy can effectively minimize the interconnect delays for all local, semi-global and global interconnects while ensuring high routing density and highly simplified routing solutions. Our experiments show that by just using these pre-determined width-pairs for each tier, we can still achieve close to optimal delays compared to those obtained from running complex wire sizing and spacing algorithms [3, 4] with many different width selections.

Tech. (μm)		0.25	0.18	0.13	0.10	0.07
Tier1	$W_1'^*$	0.25	0.18	0.13	0.10	0.07
	$W_2'^*$	0.25	0.18	0.13	0.10	0.07
Tier2	$W_1'^*$	0.25	0.18	0.13	0.10	0.08
	$W_2'^*$	0.50	0.36	0.26	0.20	0.16
Tier3	$W_1'^*$	0.65	0.47	0.24	0.22	0.23
	$W_2'^*$	1.30	0.94	0.48	0.44	0.46
Tier4	$W_1'^*$	-	-	0.98	1.00	1.06
	$W_2'^*$	-	-	1.96	2.00	2.12

Table 6: 2-width design (in μm) for area-efficient performance optimization.

5 Conclusion

To summarize, we have presented in this paper two sets of important results on interconnect-centric design methodology. First, we obtain efficient and accurate interconnect delay and area estimation models of optimal wire sizing by using two simple wire sizing schemes, 1-WS and 2-WS. Based on our simple but accurate estimation models, we study the delay and area sensitivity for interconnect designs and propose an area-efficient performance optimization metric AT^4. Our models can also be used in many other applications such as to provide interconnect performance estimation to high-level and logic-level design tools, and to study interconnect-centric design planning.

Based on our interconnect estimation study, we achieve very interesting results on area-efficient, high-performance interconnect architecture planning for wire-width design. We obtain a rather surprising result which shows that by just using two pre-determined wire widths for each metal layer, we can achieve close to optimal performance compared to that obtained by running complex wire sizing and spacing algorithms with many different wire width choices. This result will greatly simplify the routing architecture and routing tools for DSM designs.

Acknowledgments

The authors would like to thank Prof. Robert Brayton and Wilsin Gosti at UC Berkeley for providing the Strawman technology, and Lei He from UCLA to provide the 2.5D capacitance extraction methodology. We also thank Dr. Lukas van Ginneken from Magma Design Automation, Prof. Cheng-Koh Koh from Purdue Uni., Dr. Kei-Yong Khoo and Dr. Dongmin Xu from UCLA for their helpful discussions.

References

[1] J. Cong, L. He, K.-Y. Khoo, C.-K. Koh, and Z. Pan, "Interconnect design for deep submicron ICs," in *Proc. Int. Conf. on Computer Aided Design*, pp. 478–485, 1997.

[2] J. Cong and Z. Pan, "Interconnect performance estimation models for synthesis and design planning," in *IEEE/ACM Int. Workshop on Logic Synthesis*, June, 1998 (also to appear in ASP-DAC'99).

[3] J. Cong and K. S. Leung, "Optimal wiresizing under the distributed Elmore delay model," in *Proc. Int. Conf. on Computer Aided Design*, pp. 634–639, 1993.

[4] J. Cong, L. He, C.-K. Koh, and Z. Pan, "Global interconnect sizing and spacing with consideration of coupling capacitance," in *Proc. Int. Conf. on Computer Aided Design*, pp. 628–633, 1997.

[5] W. C. Elmore, "The transient response of damped linear networks with particular regard to wide-band amplifiers," *Journal of Applied Physics*, vol. 19, pp. 55–63, Jan. 1948.

[6] J. Rubinstein, P. Penfield, Jr., and M. A. Horowitz, "Signal delay in RC tree networks," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. CAD-2, pp. 202–211, July 1983.

[7] L. Pileggi, "Timing metrics for physical design of deep submicron technologies," in *Proc. Int. Symp. on Physical Design*, pp. 28–33, 1998.

[8] Semiconductor Industry Association, *National Technology Roadmap for Semiconductors*, 1997.

[9] G. Sai-Halasz, "Performance trends in high-end processors," *Proceedings of the IEEE*, vol. 83, pp. 20–36, Jan. 1995.

[10] J. Davis and J. Meindl, "Is interconnect the weak link?," *IEEE Circuits and Devices Magazine*, vol. 14, no. 2, pp. 30–36, 1998.

[11] J. Davis, V. De, and J. Meindl, "A stochastic wire length distribution for gigascale integration (gsi)," in *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 140–145, 1996.

[12] R. Otten and R. K. Brayton, "Planning for performance," in *Proc. Design Automation Conf*, pp. 122–127, June 1998.

[13] W. Gosti, private communication, 1998.

[14] P. Fisher and R. Nesbitt, "The test of time. clock-cycle estimation and test challenges for future microprocessors," *IEEE Circuits and Devices Magazine*, vol. 14, pp. 37–44, March 1998.

[15] J. Cong, L. He, A. B. Kahng, D. Noice, N. Shirali, and S. H.-C. Yen, "Analysis and justification of a simple, practical 2 1/2-d capacitance extraction methodology," in *Proc. ACM/IEEE Design Automation Conf.*, pp. 40.1.1–40.1.6, June, 1997.

[16] P. K. Sancheti and S. S. Sapatnekar, "Interconnect design using convex optimization," in *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 549–552, 1994.

[17] J. Cong and L. He, "Optimal wiresizing for interconnects with multiple sources," *ACM Trans. on Design Automation of Electronics Systems*, vol. 1, pp. 478–511, Oct. 1996.

[18] J. Lillis, C. K. Cheng, and T. T. Y. Lin, "Optimal wire sizing and buffer insertion for low power and a generalized delay model," in *Proc. Int. Conf. on Computer Aided Design*, pp. 138–143, Nov. 1995.

[19] C. P. Chen, Y. W. Chang, and D. F. Wong, "Fast performance-driven optimization for buffered clock trees based on Lagrangian relaxation," in *Proc. Design Automation Conf*, pp. 405–408, 1996.

[20] J. P. Fishburn and C. A. Schevon, "Shaping a distributed-RC line to minimize Elmore delay," *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, vol. 42, pp. 1020–1022, Dec. 1995.

[21] C. P. Chen, Y. P. Chen, and D. F. Wong, "Optimal wire-sizing formula under the Elmore delay model," in *Proc. Design Automation Conf*, pp. 487–490, 1996.

[22] R. Kay and L. Pileggi, "EWA: efficient wiring-sizing algorithm for signal nets and clock nets," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, pp. 40–49, Jan. 1998.

[23] D. G. Luenberger, *Linear and Nonlinear Programming.* Addison-Wesley, 1984.

[24] J. G. Ecker, "Geometric programming: Methods, computations and applications," *SIAM Review*, vol. 22, pp. 338–362, July 1980.

[25] C. C. N. Chu and D. F. Wong, "Greedy wire-sizing is linear time," in *Proc. Int. Symp. on Physical Design*, pp. 39–44, April 1998.

[26] M. Pedram, N. Bhat, and E. Kuh, "Combining technology mapping and layout," *The VLSI Design: An Int'l Journal of Custom-Chip Design, Simulation and Testing*, vol. 5, no. 2, pp. 111–124, 1997.

[27] J. Davis, V. De, and J. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI) i. derivation and validation," *IEEE Transactions on Electron Devices*, vol. 45, no. 3, pp. 580–9, 1998.