VLSI DESIGN
1999, Vol. 10, No. 1, pp. 21–34
Reprints available directly from the publisher
Photocopying permitted by license only

Tuning Strategies for Global Interconnects in High-Performance Deep-Submicron ICs

ANDREW B. KAHNG[a,*], SUDHAKAR MUDDU[b,†] and EGINO SARTO[b]

[a] *UCLA Computer Science Dept., 3713 Boelter Hall, Los Angeles, CA 90095-1596;*
[b] *Silicon Graphics, Inc., Mountain View, CA 94039*

(Received 7 September 1998; In final form 20 November 1998)

Interconnect tuning is an increasingly critical degree of freedom in the physical design of high-performance VLSI systems. By interconnect tuning, we refer to the selection of line thicknesses, widths and spacings in multi-layer interconnect to simultaneously optimize signal distribution, signal performance, signal integrity, and interconnect manufacturability and reliability. This is a key activity in most leading-edge design projects, but has received little attention in the literature. Our work provides the first technology-specific studies of interconnect tuning in the literature. We center on global wiring layers and interconnect tuning issues related to bus routing, repeater insertion, and choice of shielding/spacing rules for signal integrity and performance. We address four basic questions. (1) How should width and spacing be allocated to maximize performance for a given line pitch? (2) For a given line pitch, what criteria affect the optimal interval at which repeaters should be inserted into global interconnects? (3) Under what circumstances are shield wires the optimum technique for improving interconnect performance? (4) In global interconnect with repeaters, what other interconnect tuning is possible? Our study of question (4) demonstrates a new approach of offsetting repeater placements that can reduce worst-case cross-chip delays by over 30% in current technologies.

Keywords: Interconnects, tuning, shielding, scaling, repeater insertion, signal integrity, wire pitch, interconnect delay analysis

1. INTRODUCTION

With technology scaling, on-chip interconnect becomes an increasingly critical determinant of performance, manufacturability and reliability in high-end VLSI designs. Current and future designs are generally interconnect-limited, and the available routing resource must be carefully balanced among signal distribution, power/ground distribution, and clock distribution. Table I reproduces several

*Corresponding author.
†Present address: Silicon Graphics, Inc., 2011 N. Shoreline Blvd., 40L-175, Mountain View, CA 94039. Tel.: 650-933-6021, Fax: 650-932-0269, e-mail: muddu@mti.sgi.com

TABLE I Selected technology projections from the 1997 SIA NTRS

SIA national technology roadmap (1997)						
Year	1997	1999	2001	2003	2006	2009
Minimum feature size – dense lines (nm)	250	180	150	130	100	70
High-end, on-chip cross-chip clock (MHz)	750	1250	1400	1600	2000	2500
# Wiring layers	6	6–7	7	7	7–8	8–9
Minimum contacted M1 pitch (μm)	0.64	0.46	0.40	0.34	0.26	0.19
Metal height/width aspect ratio	1.8:1	1.8:1	2.0:1	2.1:1	2.4:1	2.7:1

technology projections from the 1997 SIA National Technology Roadmap for Semiconductors [15]. The implications of technology scaling – particularly for interconnects – are very complicated. Example considerations for a 7-layer metal (7LM) process might include (*cf.* [16]):

- Local interconnect layers (*e.g.*, M1 – M3) should generally remain at near-minimum dimensions and pitch to achieve acceptable routing density (an example analysis of interconnect density in 0.25 μm processes is given in [6]). For short lines (*e.g.*, several hundred microns or less), thinner metal offers less lateral coupling capacitance and driver loading, and thus locally improves circuit performance. At the same time, maximum wire width is limited by the aspect ratio upper bound. The resulting thin and narrow wires are highly resistive and also subject to reliability concerns; they are hence unsuitable for global interconnects, power distribution, *etc.*
- Layers M2 – M3 (and maybe M4) will support a mix of local and "semi-global" wiring, *e.g.*, long wires within a single block. In general, shorter wires are better routed on thinner metal. Thus, the distribution of lengths and performance goals for signals in a given design, as well as design-specific objectives (circuit robustness, guardbanding against manufacturing variation, *etc.*) will affect the interconnect tuning.
- Power distribution layers (*e.g.*, M6 – M7, maybe M5), which typically also support the top-level clock distribution (mesh or balanced -tree), should be as thick as possible for reliability. IR drop and clock skew – as well as robustness under process variations – also suggest the use of thick wire on these layers. Thick wire additionally conserves area, but can suffer from increased lateral capacitive coupling.
- Global interconnect layers (*e.g.*, M4 – M6) support inter-block signal runs with length on the order of 3000 μm – 15000 μm. To satisfy delay and signal integrity constraints, at least three degrees of freedom are available: line width and spacing, repeater insertion, and shield wiring. Repeater insertion shields downstream capacitance and is the canonical means of converting "quadratic" RC delay into "near-linear" delay; this technique also improves edge rates and hence noise immunity. When lateral coupling capacitances are large, worst-case "Miller coupling" begins to dominate noise and delay calculations; this is alleviated by increasing the line spacing and/or adding shield wiring (*i.e.*, wires connected to ground), with future techniques possibly including dedicated ground and power planes interleaved with signal layers [9].[1] Another technique to reduce the lateral coupling capacitance is to interleave signal lines which do not switch at the same signal transistion period.

[1]When two parallel neighboring lines $L1$ and $L2$ switch simultaneously in opposite directions, the driver of $L1$ sees the grounded line capacitance plus *twice* the coupling capacitance of $L1$ to $L2$. If $L2$ is quiet when $L1$ switches, then the driver of $L1$ sees the grounded line capacitance plus the coupling capacitance to $L2$. And if $L2$ switches simultaneously in the opposite direction, the driver of $L1$ sees only the grounded line capacitance. (In leading-edge processes, *each* neighbor coupling is of the same (and possibly greater) magnitude as the area coupling to ground.) The "coupling factor" or "switching factor" is often given in the range [0, 2], and since most lines have two neighbors, the total coupling factor is in the range [0, 4]. We also note that in layout synthesis, an increasingly important concept is to think of "noise-induced delay uncertainty" as "noise-induced capacitance uncertainty". The delay uncertainty is a function of slew times, voltage swings, driver strengths, and ratios of coupling to area capacitances.

The bus-dominated nature of global interconnects in building-block and high-performance designs only worsens the effects of coupling, since it results in longer parallel runs.

- All layers are subject to mutual pitch-matching, *via* sizing, *etc.*, considerations. Hence, widths and spacings on one layer cannot be chosen independently of the widths and spacings on a second layer.

The above are only a few of the applicable design considerations; the net effect is that balancing interconnect resources is now extremely difficult as designs move into and beyond the quarter-micron regime.

1.1. Interconnect Strategies

Interconnect tuning is the selection by a design team of line thicknesses, widths and spacings in multi-layer interconnect to simultaneously achieve: (i) distribution (available wiring density) for local signals, global signals, clock, power and ground; (ii) performance (signal propagation delay), particularly on global interconnects; (iii) noise immunity (signal integrity), again particularly on global interconnects; and (iv) manufacturability and reliability (*e.g.*, required margins for AC self-heat or DC electromigration on interconnects, short-circuit power in attached devices, *etc.*). Today, interconnect tuning is a key activity in most leading-edge microprocessor projects. It is clearly an option whenever the design and fabrication are owned by a single entity (in which case there is overlap with "interconnect process optimization"); however, for high-volume projects even fabless design houses exercise increasing influence on vendors' processes [6]. Nevertheless, this topic has received little attention in the literature, with only a few high-level treatments available. For example, [11] describes a characterization and analysis methodology and the need to break ideal scaling in deep submicron interconnect. [14] is another work that centers on analysis of a given multi-layer interconnect process, as opposed to the underlying interconnect tuning. [5] and [10] are examples of system-level treatments based on Rent's rule for interconnect length distribution. To our knowledge, the most notable work is the seminal paper of Rahmat *et al.* [12], which plots the constraints imposed by material, circuit performance and reliability requirements, *e.g.*, crosstalk noise, electromigration, and signal propagation delay. The paper studies such questions as: (i) maximum interconnect length that can be switched in a clock period; (ii) delay and noise envelopes for given values of horizontal and vertical pitch; (iii) coupling capacitance as a function of feature size; and (iv) maximum length of local interconnect as limited by crosstalk noise.

We believe that our work is the first in the literature to attempt a wide-ranging study of interconnect tuning with respect to degrees of freedom (repeater insertion, choice of pitch, *etc.*) that are most applicable in the high-end design context. We center on global wiring layers (*e.g.*, M4 and M5 in a 6LM process), and interconnect tuning issues related to bus routing, repeater insertion, and choice of shielding/spacing rules for signal integrity and performance. Even though the results presented in this paper are for aluminum interconnects with SiO_2 dielectric, similar techniques can be applied for copper interconnects and low-K dielectrics. Several other parameters, notably wire tapering and choice of wire thickness, are not applicable in our design methodology and thus are not part of the present study.

We address four basic questions.

1. How should width and spacing be allocated to maximize performance for a given line pitch?
2. For a given line pitch, what criteria affect the optimal interval at which repeaters should be inserted into global interconnects?
3. Under what circumstances are shield wires the optimum technique for improving interconnect performance?
4. In global interconnect with repeaters, what other interconnect tuning is possible?

We answer these questions using technology parameters from a representative 0.25 μm CMOS process; this matches the process technology

context for many current- and next-generation microprocessors. Coupling capacitance studies are performed with the commercial QuickCap 3-D field solver, and interconnect delay and noise coupling studies are performed with the commercial HSPICE simulator. Of particular interest is our study of question (4): we demonstrate that a new methodology for offsetting repeater placements can reduce worst-case cross-chip delays by over 30% in current technologies, *versus* traditional repeater insertion methodology. All parameters used in this paper are obtained using drawn dimensions of the transistors. Actual transistor widths and interconnect length/width/spacing values correspond to a 64% shrink of drawn dimensions (of course, the 0.25 μm process itself refers to actual dimension).

2. ALLOCATION OF WIDTH AND SPACING FOR GIVEN PITCH

Our first study examines how to choose a set of pitches for wires used in routing. To choose best pitches for a given layer, we plot the decrease in pure interconnect delay against the increase in pitch, with respect to some default (or minimum) pitch. Ideally, if the decrease in delay matches the increase in pitch, it is beneficial to go for higher pitches. However, if the curve starts to flatten – *i.e.*, for every given percentage increase in pitch a lesser percentage decrease in delay results – this indicates diminishing returns. Using such delay/pitch plots we have chosen three optimal pitches for routing: (i) default, (ii) fast pitch, and (iii) super fast pitch. Figure 1 plots the decrease in delay *versus* the increase in pitch for M3 wire in a representative 0.25 μm CMOS process.

Our next study seeks to determine how width and spacing should be optimally allocated for a given line pitch. In practice, the actual line width used is considerably greater than the minimum line width achievable in lithography. Thus, there is freedom to tune the width and spacing once assumptions are in place for line thickness and target line length. We note that because very long inter-block lines will have repeaters inserted regularly (see Section 3 below), the maximum line length of interest is equal to the optimum interval between

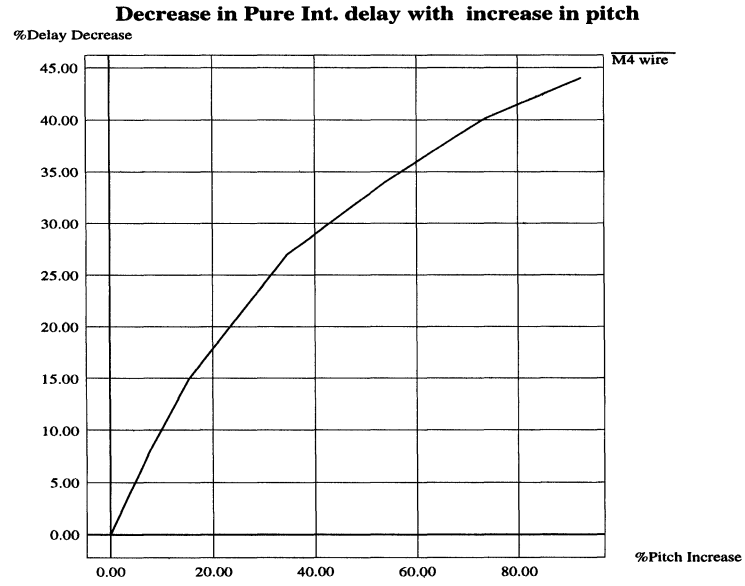

FIGURE 1 Decrease in pure interconnect delay (*i.e.*, without any load at the end of the line) as pitch for M3 wire is increased. We see that the curve starts to flatten, *i.e.*, decrease in delay saturates when pitch increase goes beyond 80% of nominal.

TABLE II Summary of M3 coupling capacitances extracted using QuickCap. Bottom M2 is a ground plane; top M4 is populated by crossover lines

Width, Space (μm)	Coupling capacitance per μm (aF)				
	Left neighbor	Right neighbor	Top plane	Bottom plane (ground)	Total
1.0, 2.2	25.20	25.61	54.79	46.84	152.66
1.2, 2.0	29.00	29.26	56.74	48.22	163.53
1.4, 1.8	33.33	33.11	57.76	51.53	177.32
1.6, 1.6	38.71	38.60	59.09	51.90	188.41
1.8, 1.4	44.75	44.12	60.22	51.52	200.92

TABLE III Delay estimates for various M3 line configurations. Driver and receiver buffer sizes: (wp = 100 μm, wn = 50 μm). Delay is computed from input of driver to input of receiver

Width, Space	50% Threshold rise delay (ps)								
	4000 μm M3 length			5000 μm M3 length			6000 μm M3 length		
(μm)	Driver load delay	Int. delay	Total delay	Driver load delay	Int. delay	Total delay	Driver load delay	Int. delay	Total delay
1.0, 2.2	106.19	113.99	220.17	132.74	168.36	301.10	159.28	233.09	392.37
1.2, 2.0	115.00	100.72	215.73	143.76	149.26	293.02	172.51	207.14	379.65
1.4, 1.8	126.61	92.80	219.41	158.27	138.04	296.31	189.92	192.10	382.02
1.6, 1.6	138.77	87.12	225.89	173.46	130.04	303.04	208.15	181.41	389.56
1.8, 1.4	151.24	82.84	234.08	189.04	124.03	313.08	226.85	173.41	400.26

repeaters; this length ranges between 2500 μm and 5000 μm for global interconnect layers in leading-edge technologies.

We have performed detailed studies of "fast" M3 interconnect with 3.2 μm pitch, assuming that M2 crossunders are dense (*i.e.*, can be approximated as a ground plane) [2] and explicitly modeling M4 crossovers. Dielectric modeling is based on actual layer data for a representative 0.25 μm CMOS process. QuickCap was used to extract coupling and area capacitances, summarized in Table II. As is typical in such analyses, we assume worst-case coupling, *i.e.*, a total coupling factor of 4.0 (worst-case coupling factor of 2.0 to each of the left and right neighbors of the (victim) line under analysis).

Table III shows HSPICE-computed line delays for M3 line lengths ranging from 4000 μm to 6000 μm. Again, dense M2 is assumed to be a ground plane, and M4 crossovers are modeled explicitly. The Table shows that (width, spacing) = (1.2, 2.0) μm gives the best performance for the given line pitch.

3. BOUNDING THE INTERVAL BETWEEN REPEATERS

A very basic study (in some sense a pre-requisite to all other interconnect tuning) asks how often repeaters should be inserted into global interconnects. This is of course a chicken-egg problem, in that the optimum repeater interval depends on the interconnect tuning, and the interconnect tuning depends on the maximum run ever made without an intervening repeater. However, the following can be noted.

- A body of study shows that repeaters should be inserted at uniform intervals. In other words, there should be a constant interconnect length (or interconnect delay) between each pair of adjacent repeaters; the first and last segments of the path are exceptions because in practice the driver and receiver sizes may not be the same as the repeater size. Actually, such theoretical results deviate from real-life practice. On any source-destination path the repeater sizes need

not be the same. It may also be better to add repeaters in parallel in order to drive larger wire lengths. This is not just for performance: repeaters locally affect device area and routing constraints. However, our studies have not yet addressed such layout issues. Using the same principle (and with certain types of methodology and chip planning constraints), it can be better to increase the size of the drivers inside the block as much as possible, which would increase the first segment length.

- Assuming that the driver size and the receiver size are the same as the size of the repeaters inserted along the path, we calculate the total delay, optimal number of repeaters and optimal distance between the repeaters.

  The total delay for a path with K repeaters is

$$T_{\text{tot}}^{K} = T_{\text{first\_stage}} + (K-1) * T_{\text{Rep\_stage}} + T_{\text{Final\_stage}}$$

  The delay of the first stage is the total delay from the output of driver to the input of the first repeater, *i.e.*, $T_{\text{first\_stage}} = T_{\text{gd}} + T_{\text{int}}$, where gate load delay is $T_{\text{gd}} = R_{\text{rep}}(C_{\text{int}}^{\text{eff}} + C_{\text{rep}})$, interconnect delay is $T_{\text{int}} = R_{\text{int}}(C_{\text{int}}/2 + C_{\text{rep}})$, and R_{rep}, C_{rep} are repeater output resistance and input gate capacitance. The effective capacitance at the gate output can be approximated as $C_{\text{int}}^{\text{eff}} = \alpha C_{\text{int}}$ where α is a constant between 1/6 and 1 [8]. Let L_p be the interconnect path length between driver and receiver. Then for optimal placement of repeaters the interconnect length between repeaters is $L_p/K+1$. Therefore, the total delay for the path is

$$\begin{aligned} T_{\text{tot}}^{K} &= (K+1) * (T_{\text{gd}} + T_{\text{int}}) \\ &= (K+1) * R_{\text{rep}}\left(\alpha * c * \frac{L_p}{K+1} + C_{\text{rep}}\right) \\ &\quad + r * L_p\left(c * \frac{L_p}{2(K+1)} + C_{\text{rep}}\right) \end{aligned} \tag{1}$$

  where r, c are resistance and capacitance per unit length of the interconnect line. We compute the optimal number of repeaters that minimizes total delay by setting $\partial T_{\text{tot}}/\partial K = 0$, and obtain

$$K = \sqrt{\frac{rcL_p^2}{2R_{\text{rep}}C_{\text{rep}}}} - 1 \tag{2}$$

  To minimize total delay, gate load delay and interconnect delay should be equal. If effective capacitance is not considered in the gate load delay computation, and with current technology trends, gate load delay will always be greater than interconnect delay. Under these conditions, to minimize total delay one can increase the time of flight (or wire length) between repeaters until slew time constraints become tight. In the current range of 0.35 μm and 0.25 μm process generations, global interconnects have repeaters inserted with periods ranging from 2500 μm to 10000 μm.
- Repeater insertion is also driven by pure interconnect delay, since larger time of flight implies larger slew time on the transition seen at the receiver. Edges with large slew times cause much larger gate delays, are more susceptible to noise, are more susceptible to process-distribution influenced delay variations, and also increase the short-circuit power dissipation. Even in today's designs, slew times above 600 – 700 ps cannot be tolerated. Thus, even without the delay minimization objective, edge rate control will force insertion of repeaters. In fact, some of the functionality of "post-layout optimization" tools for gate sizing and repeater insertion is driven by edge rate checks as opposed to signal delay reduction.
- In practice, repeaters will be implemented using inverters whenever possible, due to performance and area efficiency.

Table IV summarizes M3 interconnect slew times for line width 1.0 μm and line spacing 1.2 μm (corresponding to a "dense" M3 routing pitch), and input slew time of 400 ps. All capacitance extractions were performed with Quick Cap, and correspond to M4 and M1 as the top and bottom ground planes, respectively. Switching

TABLE IV Summary of M3 interconnect slew times. M4 is top layer; M1 is bottom layer. Two combinations of width/spacing are shown, along with three different coupling factor assumptions. The input slew time is 400 ps and the output slew times are computed as 10%–90% for rise time and 90%–10% for fall time

Driver/Receiver (wp, wn)(μm)	Width (μm)	Space (μm)	Length (μm)	SF	Delay (ps)	Rise time (ps)	Fall time (ps)
(130, 65)/(130, 65)	1	1.1	10000	4	589	1679	1510
(130, 65)/(130, 65)	1	1.1	9000	4	486	1421	1265
(130, 65)/(130, 65)	1	1.1	8000	4	393	1187	1044
(130, 65)/(130, 65)	1	1.1	7000	4	310	975	847
(130, 65)/(130, 65)	1	1.1	5000	4	172	623	525
(130, 65)/(130, 65)	1	1.1	10000	3	488	1405	1267
(130, 65)/(130, 65)	1	1.1	9000	3	404	1193	1066
(130, 65)/(130, 65)	1	1.1	8000	3	327	1001	885
(130, 65)/(130, 65)	1	1.1	7000	3	259	828	723
(130, 65)/(130, 65)	1	1:1	5000	3	147	538	458
(130, 65)/(130, 65)	1	1.1	10000	2	388	1131	1026
(130, 65)/(130, 65)	1	1.1	9000	2	323	966	869
(130, 65)/(130, 65)	1	1:1	8000	2	263	817	728
(130, 65)/(130, 65)	1	1.1	7000	2	209	682	601
(130, 65)/(130, 65)	1	1.1	5000	2	120	456	393
(130, 65)/(130, 65)	1.4	1.6	10000	4	366	1123	980
(130, 65)/(130, 65)	1.4	1.6	9000	4	303	963	832
(130, 65)/(130, 65)	1.4	1.6	8000	4	246	818	698
(130, 65)/(130, 65)	1.4	1.6	7000	4	195	686	578
(130, 65)/(130, 65)	1.4	1.6	5000	4	111	465	384
(130, 65)/(130, 65)	1.4	1.6	10000	3	320	992	869
(130, 65)/(130, 65)	1.4	1.6	9000	3	266	854	740
(130, 65)/(130, 65)	1.4	1.6	8000	3	217	729	625
(130, 65)/(130, 65)	1.4	1.6	7000	3	172	615	522
(130, 65)/(130, 65)	1.4	1.6	5000	3	99	422	352
(130, 65)/(130, 65)	1.4	1.6	10000	2	275	862	759
(130, 65)/(130, 65)	1.4	1.6	9000	2	229	746	650
(130, 65)/(130, 65)	1.4	1.6	8000	2	188	640	553
(130, 65)/(130, 65)	1.4	1.6	7000	2	150	543	465
(130, 65)/(130, 65)	1.4	1.6	5000	2	87	382	322

factors range from 4 (both neighbors switching in the opposite direction from the victim) to 2 (both neighbors quiet, or one neighbor switching in the opposite direction and one neighbor switching in the same direction with respect to the victim). We see that the M3 distance between repeaters has an upper bound of 5000 μm due to edge rate considerations alone. Separate studies show that this upper bound on distance between repeaters is essentially unaffected by changes to the driver/receiver sizing or the input slew time.

4. BENEFITS OF SHIELD WIRING

Our third study addresses the question of whether shield wiring is an effective means of improving delay and signal integrity performance of long global interconnects. We consider various width-spacing rules for M3 interconnect, in order to evaluate the utility of spacing *vs.* shielding techniques. Our evaluations are with respect to delay only; for all of the configurations, the assumed slew time upper bounds of approximately 600 ps imply that noise coupling will not be problematic. Figure 2 contrasts five pitch-matched width-spacing rules:

- **Rule 1:** 1.2 μm width, 1.0 μm spacing
- Single-V_{SS}: 1.2 μm width, 1.0 μm spacing, with every third line grounded (*i.e.*, every signal line has one grounded neighbor to shield it)
- **Rule 2:** 1.2 μm width, 2.1 μm spacing
- **Rule 3:** 2.2 μm width, 2.2 μm spacing
- Double-V_{SS}: 1.2 μm width, 2.1 μm spacing, with every other line grounded (*i.e.*, every signal line has two grounded neighbors to shield it)

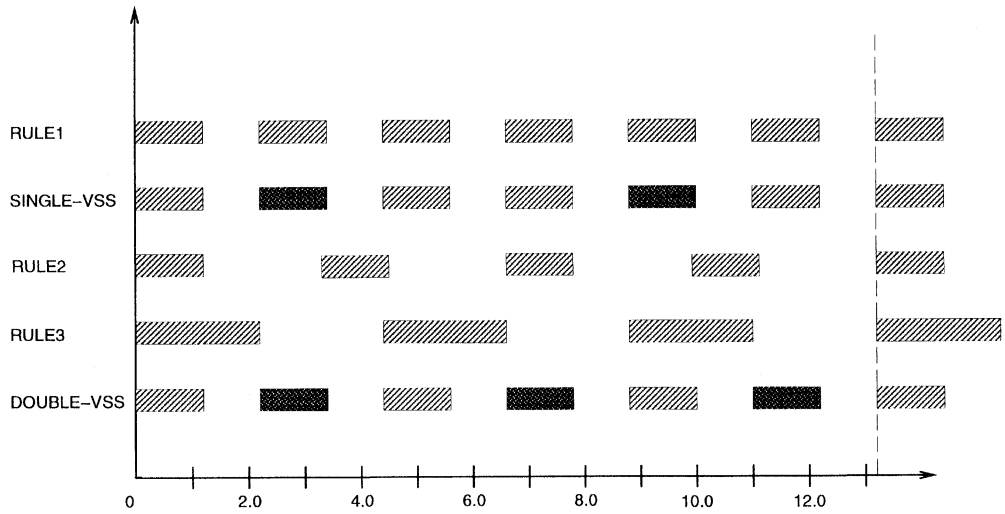

FIGURE 2 Pitch-matched width-spacing rules. Rule 1 allows six lines per 13.2 μm; Rule 2 and the Single-V_{SS} rule (Rule 1 width/spacing, but every third line grounded) both allow four signal lines per 13.2 μm; and Rule 3 and the Double-V_{SS} rule (Rule 1 width/spacing, but every other line grounded) both allow three signal lines per 13.2 μm.

Again, QuickCap was used to extract capacitive couplings of a given victim line to its neighbor lines and the neighboring top/bottom layers; these results are shown in Table V. Notice that the Rule 1, Rule 2 and Rule 3 rules have worst-case coupling factors = 4. On the other hand, the Single-V_{SS} rule has worst-case coupling factor = 3, and the Double-V_{SS} rule has worst-case coupling factor = 2. Table VI shows the delay performance for a 4000 μm M3 line, under various bottom ground and top plane configurations. We observe:

- The Rule 3 rule provides 37% decrease in total delay, but since C_{eff} was not used in the gate load delay computation, actual delay reductions could be even greater.
- The Single-V_{SS} rule is less effective than the Rule 2 rule; note that the two rules are equivalent in terms of effective routing density. Our studies have not yet addressed the routing interactions that can potentially affect this analysis. In particular, shield lines may be added to bring power and ground connections to repeater blocks.
- The Double-V_{SS} rule gives improved total delays compared with the Rule 3 rule, with the rules being equivalent in terms of effective routing

TABLE V M3 coupling capacitances extracted using QuickCap for various interconnect tuning rules and combinations of bottom and top planes

M3 Rules	Width, Space (μm)	Ground, Top planes	Coupling capacitance per μm (aF)				
			Left neighbor	Right neighbor	Top plane	Bottom plane (ground)	Total
Rule1	1.2, 1.0	Substrate, M4 Line	68.23	68.15	43.68	14.79	195.03
Rule1	1.2, 1.0	M2, M4 Line	60.30	60.92	43.96	34.88	202.37
Rule1	1.2, 1.0	M2, –	74.67	74.23	–	42.99	192.44
Rule2	1.2, 2.1	Substrate, M4 Line	36.87	34.37	58.58	18.07	148.29
Rule2	1.2, 2.1	M2, M4 Line	26.96	27.10	58.51	48.72	160.41
Rule2	1.2, 2.1	M2, –	42.17	42.43	–	59.15	143.96
Rule3	2.2, 2.2	Substrate, M4 Line	35.09	36.50	77.61	22.14	171.52
Rule3	2.2, 2.2	M2, M4 Line	26.18	25.61	77.51	67.92	198.82
Rule3	2.2, 2.2	M2, –	44.33	43.86	–	73.23	162.14

TABLE VI Delay estimates for a 4000 μm M3 line, under various interconnect tuning configurations. Driver and receiver buffer sizes: (wp = 100 μm, wn = 50 μm). Delay is computed from input of driver to input of receiver

M3 Rules	Width, Space (μm)	Ground, Top planes	50% threshold rise delay (ps)			% Gain w.r.t. Rule1
			Driver load delay	Interconnect delay	Total delay	
Rule1	1.2, 1.0	Substrate, M4 Line	173.04	116.88	289.92	–
Rule1	1.2, 1.0	M2, M4 Line	167.84	114.03	281.87	–
Rule1	1.2, 1.0	M2, –	178.03	119.62	297.65	–
Rule2	1.2, 2.1	Substrate, M4 Line	114.47	84.75	199.22	29
Rule2	1.2, 2.1	M2, M4 Line	112.50	83.66	196.16	30
Rule1 with Single VSS	1.2, 1.0	Substrate, M4 Line	137.41	97.34	234.75	17
Rule1 with Single VSS	1.2, 1.0	M2, M4 Line	136.17	96.66	232.83	17
Rule1 with Single VSS	1.2, 1.0	M2, –	139.14	98.28	237.42	16
Rule2	1.2, 2.1	M2, –	119.29	87.39	206.68	27
Rule3	2.2, 2.2	Substrate, M4 Line	126.91	49.95	176.85	37
Rule3	2.2, 2.2	M2, M4 Line	130.08	50.90	180.98	36
Rule3	2.2, 2.2	M2, –	130.40	50.99	181.39	36
Rule1 with Double VSS	1.2, 1.0	Substrate, M4 Line	99.74	78.11	177.85	37
Rule1 with Double VSS	1.2, 1.0	M2, M4 Line	104.34	80.83	185.17	34
Rule1 with Double VSS	1.2, 1.0	M2, –	121.14	78.53	199.67	29

density. However, the Rule 3 rule yields smaller interconnect delays, so that driver size reductions have greater potential for delay improvement. Thus, the Rule 3 rule seems preferable. When two buses have activity patterns such that each is quiet when the other is active, then their lines can be interleaved such that they effectively follow the Double-V_{SS} rule. In such a case, interleaving is clearly superior to the Rule 3 rule, since the effective routing density is doubled.

- Gate load delays are larger than interconnect delays, suggesting that it is preferable to decrease line widths and increase line spacings. We also note that a dense M4 top layer decreases total delay, and a dense M2 bottom (ground plane) layer decreases total delay for smaller line widths only.

5. NEW REPEATER OFFSET METHODOLOGY FOR GLOBAL BUSES

Finally, we study another form of tuning that is possible for global interconnects. Our motivations are three-fold: (i) global interconnect is increasingly dominated by wide buses; (ii) present methodology designs global interconnects for *worst-case* Miller coupling; and (iii) present methodology routes long global buses using repeater *blocks*, *i.e.*, blocks of co-located inverters spaced every, say, 4000 μm.

We have proposed a simple method to improve global interconnect performance. The idea is to reduce the worst-case Miller coupling by offsetting the inverters on adjacent lines (see Fig. 3). In the previous methodology (Fig. 3(a)), the worst-case switching of a neighbor line (*i.e.*, simultaneously and in the opposite direction to the switching of the victim line) persists through the entire chain of inverters. However, with offset inverter locations (Fig. 3(b)), any worst-case simultaneous switching on a neighbor line persists only for half of each period between consecutive inverters, *and furthermore becomes best-case simultaneous switching for the other half of the period*!.

To confirm the advantages of this method, the following experimental methodology was used.

- We study systems of three parallel interconnect lines, with lengths either 10000 μm or 14000 μm. These lines are stimulated by a waveform with risetime = falltime = 200 ps. The middle line is considered the "victim" for analysis purposes.
- We model two "technologies" representative of M3 and M4 in an 0.25 μm CMOS process. In each technology, line resistance is 50 Ω per

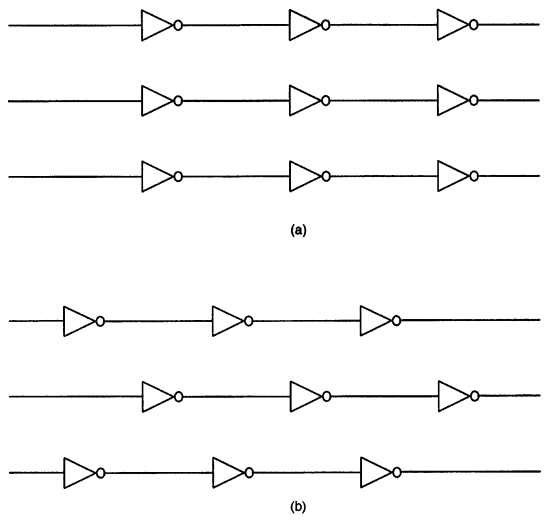

FIGURE 3 2 Reduction of worst-case Miller coupling by offsetting inverters. In (a), inverters on the left and right neighbor lines are at phase = 0 with respect to the inverters on the middle line. In (b), inverters on the left and right neighbors are at phase = 0.5.

1000 μm. In Technology I, capacitive couplings to left neighbor, ground and right neighbor per 1000 μm are respectively 60fF, 80fF and 60fF. In Technology II, capacitive couplings to left neighbor, ground and right neighbor per 1000 μm are respectively 80fF, 160fF and 80fF.

- We assume a *period* between inverters (repeaters) of 4000 μm. So that HSPICE cannot introduce any error in its RC analysis, we manually distributed the line and coupling parasitics into 40 μm segments, *i.e.*, repeaters occurred every 100 segments, and line lengths were 250 or 350 segments. Each segment is modeled as a double-pi model. This segmenting is chosen such that any finer-grain representation does not change the HSPICE-computed delays.
- We always place the inverters on the middle line with "phase = 0", *i.e.*, at positions 4000, 8000, ... microns along the line. Inverters on the left and right neighbors are placed according to all combinations of phase = $0, 0.1, 0.2, \ldots, 0.9$ (again with respect to the period of 4000 μm). There are 100 different phase combinations. Figure 3 shows the three-line configurations with left/right neighbor phase combinations of $(0, 0)$ and $(0.5, 0.5)$.
- We stimulate the three lines with the periodic waveform, with the first transition either rising (R) or falling (F). There are eight combinations of directions for the first transisions, *i.e.*, RRR, RRF, ..., FFF.
- Finally, we may offset the input waveforms of the left and right neighbors by -100 ps, 0 ps or $+100$ ps with respect to the input waveform of the middle line. There are nine combinations of these input offsets.

Table VII shows HSPICE delays for systems of three lines of length 10000μm, using Technology I, for all combinations of rising (R) and falling (F) initial transition on the input waveform. The Table

TABLE VII HSPICE delays (ns) for three lines of length 10000 μm, using Technology I, for all combinations of rising (R) and falling (F) initial transition on the input waveform. We show delays for inverter phases $(0, 0)$ and $(0.5, 0.5)$ on the left and right neighbors of the middle line (phase 0)

Input waveforms (Left neighbor, victim, right neighbor)	Interconnect delay (ns)					
	Left, right neighbor buffer phases: 0, 0			Left, right neighbor buffer phases: 0.5, 0.5		
	Left neighbor delay	Victim delay	Right neighbor delay	Left neighbor delay	Victim delay	Right neighbor delay
R, R, R	0.361	0.361	0.361	0.510	0.630	0.510
R, R, F	0.428	0.584	0.676	0.533	0.697	0.499
R, F, R	0.546	0.994	0.546	0.483	0.689	0.483
R, F, F	0.676	0.584	0.428	0.499	0.697	0.533
F, R, R	0.676	0.584	0.428	0.499	0.697	0.533
F, R, F	0.546	0.994	0.546	0.483	0.689	0.483
F, F, R	0.428	0.584	0.676	0.533	0.697	0.499
F, F, F	0.361	0.361	0.361	0.510	0.630	0.510

shows delays for inverter phases (0, 0) and (0.5, 0.5) on the left and right neighbors of the middle line (phase 0). The effect of Miller coupling is clearly shown.

Table VIII shows the worst-case delays (with respect to all eight possible combinations of rising and falling inputs) for the middle line, for each combination of phases for the inverter locations on the left and right neighbor lines. Input offsets are all 0, *i.e.*, the waveforms start at the same time. All four combinations of Technology and line length are shown. In every case, the optimum phase

TABLE VIII Worst-case middle line delays over all input rise/fall combinations, for each phase combination on left and right neighbors. Input offsets are all 0 ps

A. Line length 10000 μm, Technology I

		Right neighbor phase									
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0	**0.994**	0.988	0.971	0.954	0.929	0.910	0.874	0.900	0.930	0.962
	0.1	0.988	0.974	0.960	0.938	0.911	0.885	0.854	0.881	0.917	0.952
Left	0.2	0.971	0.960	0.941	0.917	0.887	0.848	0.829	0.863	0.897	0.932
	0.3	0.954	0.938	0.917	0.890	0.855	0.806	0.801	0.834	0.872	0.912
Neighbor	0.4	0.929	0.911	0.887	0.855	0.818	0.753	0.766	0.805	0.841	0.885
	0.5	0.910	0.885	0.848	0.806	0.753	**0.697**	0.735	0.778	0.822	0.867
Phase	0.6	0.874	0.854	0.829	0.801	0.766	0.735	0.739	0.768	0.799	0.832
	0.7	0.900	0.881	0.863	0.834	0.805	0.778	0.768	0.796	0.827	0.859
	0.8	0.930	0.917	0.897	0.872	0.841	0.822	0.799	0.827	0.860	0.894
	0.9	0.962	0.952	0.932	0.912	0.885	0.867	0.832	0.859	0.894	0.924

B. Line length 10000 μm, Technology II

	0	**1.437**	1.422	1.400	1.370	1.332	1.299	1.259	1.300	1.343	1.388
	0.1	1.422	1.405	1.379	1.347	1.306	1.258	1.234	1.278	1.324	1.372
Left	0.2	1.400	1.379	1.352	1.315	1.270	1.206	1.199	1.247	1.296	1.347
	0.3	1.370	1.347	1.315	1.274	1.223	1.144	1.158	1.208	1.261	1.314
Neighbor	0.4	1.332	1.306	1.270	1.223	1.167	1.075	1.109	1.161	1.216	1.273
	0.5	1.299	1.258	1.206	1.144	1.075	**1.015**	1.069	1.124	1.180	1.239
Phase	0.6	1.259	1.234	1.199	1.158	1.109	1.069	1.079	1.120	1.163	1.209
	0.7	1.300	1.278	1.247	1.208	1.161	1.124	1.120	1.160	1.203	1.250
	0.8	1.343	1.324	1.296	1.261	1.216	1.180	1.163	1.203	1.246	1.293
	0.9	1.388	1.372	1.347	1.314	1.273	1.239	1.209	1.250	1.293	1.339

C. Line length 14000 μm, Technology I

	0	**1.474**	1.467	1.448	1.429	1.401	1.383	1.341	1.340	1.382	1.427
	0.1	1.467	1.454	1.439	1.414	1.385	1.356	1.308	1.324	1.370	1.417
Left	0.2	1.448	1.439	1.418	1.393	1.359	1.320	1.267	1.299	1.345	1.395
	0.3	1.429	1.414	1.393	1.362	1.328	1.276	1.217	1.267	1.319	1.375
Neighbor	0.4	1.401	1.385	1.359	1.328	1.287	1.223	1.174	1.229	1.285	1.342
	0.5	1.383	1.356	1.320	1.276	1.223	**1.105**	1.146	1.203	1.263	1.323
Phase	0.6	1.341	1.308	1.267	1.217	1.174	1.146	1.110	1.162	1.220	1.281
	0.7	1.340	1.324	1.299	1.267	1.229	1.203	1.162	1.192	1.240	1.287
	0.8	1.382	1.370	1.345	1.319	1.285	1.263	1.220	1.240	1.283	1.330
	0.9	1.427	1.417	1.395	1.375	1.342	1.323	1.281	1.287	1.330	1.377

D. Line length 14000 μm, Technology II

	0	2.123	2.108	2.085	2.052	2.011	1.983	1.925	1.938	1.995	2.056
	0.1	2.108	2.092	2.064	2.029	1.985	1.943	1.876	1.913	1.974	2.039
Left	0.2	2.085	2.064	2.036	1.996	1.947	1.889	1.816	1.878	1.944	2.012
	0.3	2.052	2.029	1.996	1.952	1.898	1.823	1.765	1.833	1.903	1.977
Neighbor	0.4	2.011	1.985	1.947	1.898	1.837	1.743	1.703	1.778	1.854	1.932
	0.5	1.983	1.943	1.889	1.823	1.743	**1.590**	1.664	1.744	1.823	1.903
Phase	0.6	1.925	1.876	1.816	1.765	1.703	1.664	1.630	1.686	1.763	1.843
	0.7	1.938	1.913	1.878	1.833	1.778	1.744	1.686	1.741	1.801	1.867
	0.8	1.995	1.974	1.944	1.903	1.854	1.823	1.763	1.801	1.860	1.925
	0.9	2.056	2.039	2.012	1.977	1.932	1.903	1.843	1.867	1.925	1.989

combination is (0.5, 0.5), while the traditional phase combination of (0.0, 0.0) is actually the *worst* possible. The worst-case delay is reduced by anywhere from 25% to 30% when the repeaters are placed with optimum phase. Finally, Table IX shows the same worst-case delays for the middle line, this time taken over all eight rise/fall combinations and all nine combinations of input waveform offsets. Again, even when the inputs do not switch perfectly simultaneously, the best phase combination is (0.5, 0.5) and the worst phase combination is the traditional (0.0, 0.0) methodology.

TABLE IX Worst-case delays with all combinations of input offsets

A. Line length 10000 μm, Technology I

		Right neighbor phase									
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0	**1.090**	1.071	1.051	1.021	0.988	0.942	0.948	0.984	1.018	1.051
	0.1	1.071	1.054	1.026	0.995	0.957	0.905	0.920	0.958	0.997	1.035
Left	0.2	1.051	1.026	0.998	0.964	0.921	0.865	0.890	0.930	0.970	1.008
	0.3	1.021	0.995	0.964	0.924	0.876	0.825	0.854	0.894	0.936	0.980
Neighbor	0.4	0.988	0.957	0.921	0.876	0.825	0.782	0.813	0.856	0.900	0.944
	0.5	0.942	0.905	0.865	0.825	0.782	**0.760**	0.791	0.824	0.860	0.900
Phase	0.6	0.948	0.920	0.890	0.854	0.813	0.791	0.816	0.849	0.879	0.911
	0.7	0.984	0.958	0.930	0.894	0.856	0.824	0.849	0.880	0.911	0.945
	0.8	1.018	0.997	0.970	0.936	0.900	0.860	0.879	0.911	0.944	0.982
	0.9	1.051	1.035	1.008	0.980	0.944	0.900	0.911	0.945	0.982	1.016

B. Line length 10000 μm, Technology II

	0	**1.526**	1.502	1.471	1.430	1.382	1.335	1.329	1.379	1.427	1.476
	0.1	1.502	1.475	1.440	1.396	1.343	1.284	1.292	1.345	1.398	1.449
Left	0.2	1.471	1.440	1.400	1.350	1.291	1.229	1.249	1.305	1.361	1.416
	0.3	1.430	1.396	1.350	1.295	1.231	1.171	1.200	1.258	1.315	1.373
Neighbor	0.4	1.382	1.343	1.291	1.231	1.167	1.114	1.148	1.205	1.262	1.321
	0.5	1.335	1.284	1.229	1.171	1.114	**1.074**	1.124	1.175	1.226	1.279
Phase	0.6	1.329	1.292	1.249	1.200	1.148	1.124	1.148	1.190	1.234	1.281
	0.7	1.379	1.345	1.305	1.258	1.205	1.175	1.190	1.234	1.280	1.328
	0.8	1.427	1.398	1.361	1.315	1.262	1.226	1.234	1.280	1.327	1.376
	0.9	1.476	1.449	1.416	1.373	1.321	1.279	1.281	1.328	1.376	1.425

C. Line length 14000 μm, Technology I

	0	**1.572**	1.551	1.530	1.502	1.465	1.419	1.391	1.429	1.474	1.521
	0.1	1.551	1.534	1.507	1.472	1.438	1.388	1.362	1.406	1.451	1.499
Left	0.2	1.530	1.507	1.474	1.442	1.400	1.345	1.323	1.373	1.423	1.475
	0.3	1.502	1.472	1.442	1.401	1.353	1.293	1.279	1.334	1.388	1.443
Neighbor	0.4	1.465	1.438	1.400	1.353	1.297	1.241	1.231	1.288	1.348	1.406
	0.5	1.419	1.388	1.345	1.293	1.241	**1.171**	1.203	1.256	1.310	1.365
Phase	0.6	1.391	1.362	1.323	1.279	1.231	1.203	1.206	1.247	1.291	1.339
	0.7	1.429	1.406	1.373	1.334	1.288	1.256	1.247	1.288	1.332	1.377
	0.8	1.474	1.451	1.423	1.388	1.348	1.310	1.291	1.332	1.374	1.424
	0.9	1.521	1.499	1.475	1.443	1.406	1.365	1.339	1.377	1.424	1.471

D. Line length 14000 μm, Technology II

	0	**2.213**	2.190	2.157	2.116	2.069	2.031	1.974	2.027	2.084	2.147
	0.1	2.190	2.161	2.125	2.081	2.029	1.982	1.930	1.991	2.053	2.119
Left	0.2	2.157	2.125	2.085	2.035	1.977	1.920	1.879	1.946	2.013	2.084
	0.3	2.116	2.081	2.035	1.980	1.913	1.846	1.818	1.893	1.965	2.041
Neighbor	0.4	2.069	2.029	1.977	1.913	1.837	1.775	1.750	1.831	1.909	1.989
	0.5	2.031	1.982	1.920	1.846	1.775	**1.666**	1.730	1.804	1.879	1.955
Phase	0.6	1.974	1.930	1.879	1.818	1.750	1.730	1.713	1.773	1.835	1.901
	0.7	2.027	1.991	1.946	1.893	1.831	1.804	1.773	1.830	1.892	1.957
	0.8	2.084	2.053	2.013	1.965	1.909	1.879	1.835	1.892	1.951	2.015
	0.9	2.147	2.119	2.084	2.041	1.989	1.955	1.901	1.957	2.015	2.079

6. CONCLUSIONS

To our knowledge, this work has provided the first technology-specific studies of interconnect tuning in the literature. We have described experimental approaches to interconnect tuning issues related to bus routing, repeater insertion, and choice of shielding/spacing rules for signal integrity and performance. In particular, four questions have been addressed: allocation of width and spacing to maximize performance for a given pitch, finding the optimal interval for repeater insertion, assessing the potential benefits of shield wiring, and optimizing the insertion of repeaters in global buses. Our answers to these questions are at times surprising: in answering (3), we demonstrate that current shielding methodologies may be suboptimal when compared with alternate width/spacing rules, and in answering (4), we propose a new repeater offset technique that can reduce worst-case cross-chip delays by over 30% in current technologies. Ongoing efforts extend our interconnect tuning research to encompass layer thicknesses, more detailed analyses of noise coupling and tuning to meet noise margins, and the delay/noise behavior in emerging technology regimes (Cu interconnect and low-K dielectrics or air-gaps). Finally, we seek to develop more complete full-chip interconnect tuning approaches based on analyses of the interconnect structure, speed target, and power dissipation target for a given design.

ALPHA NUMERIC CHARACTERS

Zero	0
Capital letter O	O
Lowercase L	l
Number one	1
Mu (micro)	μ
Omega	Ω
Pico Henry	pH
Pico Farad	pF
Alpha	α
Beta	β
Rho	ρ
Femto Farad	fF
Theta	θ
Tau	τ
Pi	π
Laplace Variable	s

References

[1] Alpert, C. J. and Devgan, A., "Wire Segmenting for Improved Buffer Insertion", *Proc. Design Automation Conf.*, June 1997, pp. 588–593.

[2] Cong, J., He, L., Kahng, A. B., Noice, D., Shirali, N. and Yen, S. H.-C., "Analysis and Justification of a Simple, Practical 2 1/2-D Capacitance Extraction Methodology", *Proc. Design Automation Conference*, June 1997.

[3] Dartu, F. and Pileggi, L. (1997). "Calculating Worst-Case Gate Delays Due to Dominant Capacitance Coupling", *Proc. ACM/IEEE Design Automation Conf.*, pp. 46–51.

[4] Deutsch, A., Kopcsay, G. V., Surovic, C. W., Rubin, B. J., Terman, L. M., Dunne, R. P. and Gallo, T., "Modeling and Characterization of Long On-chip Interconnections for High-Performance Microprocessors", final report, ARPA HSCD Contract C-556003, September 1995. Also appeared In: *IBM Journal of Research and Development*, **39**(5), Sept. 1995, 547–567.

[5] Fisher, P. D., "Clock Cycle Estimations for Future Microprocessor Generations", *Proc. IEEE Innovative Systems in Silicon*, Austin, October 1997.

[6] Gwennap, L., "IC Vendors Prepare for 0.25-Micron Leap", *Microprocessor Report*, September 16, 1996, pp. 11–15.

[7] Gwennap, L., "IC Makers Confront RC Limitations", *Microdesign Resources Microprocessor Report*, August 4, 1997, pp. 14–18.

[8] Kahng, A. B. and Muddu, S., "Effcient gate Delay Modeling for Large Interconnect Loads", *Proc. IEEE Multi-Chip Module Conf.*, Feb. 1996, pp. 202–207.

[9] LaPotin, D. P., Ghoshal, U., Chiprout, E. and Nassif, S. R., "Physical Design Challenges for Performance", *International Symposium on Physical Design*, April 1997, pp. 225–226.

[10] Meindl, J., "GigaScale Integration: 'Is the Sky the Limit'?", keynote presentation slides, Hot Chips IX, Stanford, CA, August 25–26, 1997.

[11] Oh, S.-Y., Chang, K.-J., Chang, N. and Lee, K., "Interconnect modeling and design in high-speed VLSI/ULSI systems", *Proc. International Conference on Computer Design: VLSI in Computers and Processors*, October 1992, pp. 184–189.

[12] Rahmat, K., Nakagawa, O. S., Oh, S.-Y., and Moll, J. (1995). "A Scaling Scheme for Interconnect in Deep-Submicron Processes", *Intl. Electron Devices Meeting. Technical Digest*, pp. 245–248.

[13] Scheffer, L., "A Roadmap of CAD Tool Changes for Submicron Interconnect Problems", *International Symposium on Physical Design*, April 1997, pp. 104–109.

[14] Sechler, R. F., "Interconnect design with VLSI CMOS", *IBM Journal of Research and Development*, Jan.–March 1995, pp. 23–31.
[15] Semiconductor Industry Association, National Technology Roadmap for Semiconductors, December 1997.
[16] Sylvester, D. and Keutzer, K. (1998). "Getting to the Bottom of Deep-Submicron", *Proc. IEEE Intl. Conference on Computer-Aided Design.*

Authors' Biographies

Sudhakar Muddu received his B. Tech degree in Electronics and Communications Engineering from Indian Institute of Technology at Madras in 1990, and the M.S. and Ph.D. degrees in computer science from the University of California at Los Angeles in 1994 and 1996. His Ph.D. thesis was in the area of analysis and modeling of VLSI interconnects. He has published over 25 papers in the area of VLSI interconnects. He has also served on technical program committees for many leading conferences. He has had summer internships at Intel Corporation in Santa Clara, IBM T. J. Watson Research Center in Yorktown Heights, and AT and T Bell Laboratories in Holmdel. He has also worked for MIPS Technologies in the VLSI CAD group and currently working for Silicon Graphics, Inc. in the high performance RISC microprocessor development team. Dr. Muddu has four U.S. patents pending in the areas of interconnect analysis and ATM congestion control. His research interests include computer aided-design of VLSI circuits; analysis, modeling and simulation of VLSI interconnects; VLSI architecture; CMOS circuit design; computer network modeling and analysis; and discrete and combinatorial algorithms.

Andrew B. Kahng received the A.B. degree in applied mathematics (physics) from Harvard College, and the M.S. and Ph.D. degrees in computer science from the University of California at San Diego. He joined the computer science faculty at UCLA in July 1989, and it currently professor and vice-chair for graduate studies. From April 1996 through September 1997, he was on sabbatical leave and leave of absence from ULCA, as a Visiting Scientist at Cadence Design Systems, Inc. Professor Kahng has received NSF Research Initiation and Young Investigator awards, and a DAC Best Paper award. His research interests include VLSI physical layout design and performance analysis, combinatorial and graph algorithms, and stochastic global optimization.

Egino Sarto received the B.S. degree in Electrical Engineering from Universidade de Brasilia, Brasilia, Brazil, and the M.S. degree in Electrical Engineering from Stanford University, Stanford, CA in 1979 and 1981 respectively. Prior to joining MIPS/Silicon Graphics in 1987, he was a circuit design engineer at Intel, working on fast SRAMs and microprocessor development. At Silicon Graphics, where he is a Principal Engineer, he has worked on the design of high performance RISC microprocessors and recently engaged on CAD flow development for high density ASICs. His interest include general circuit design, timing analysis, physical layout design, IC interconnect, silicon processing technologies.