Effects of Global Interconnect Optimizations on Performance Estimation of Deep Submicron Design*

Yu Cao[1], Chenming Hu[1], Xuejue Huang[1], Andrew B. Kahng[2],
Sudhakar Muddu[3], Dirk Stroobandt[4], Dennis Sylvester[5]

[1]EECS Department, UC Berkeley, USA; [2]CS Department, UCLA, USA; [3]Silicon Graphics, Inc., USA;
[4]ELIS Department, Ghent University, Belgium; [5]Synopsys, Inc., USA

ABSTRACT

In this paper, we quantify the impact of global interconnect optimization techniques that address such design objectives as delay, peak noise, delay uncertainty due to noise, power, and cost. In doing so, we develop a new system-performance simulation model as a set of studies within the MARCO GSRC Technology Extrapolation (GTX) system. We model a typical point-to-point global interconnect and focus on accurate assessment of both circuit and design technology with respect to such issues as inductance, signal line shielding, dynamic delay, buffer placement uncertainty and repeater staggering. We demonstrate, for example, that optimal wire sizing models need to consider inductive effects – and that use of more accurate {-1,3} worst-case capacitive coupling noise switch factors substantially increases peak noise estimates compared to traditional {0,2} bounds. We also find that optimal repeater sizes are significantly smaller than conventional models would suggest, especially when considering energy-delay issues.

Keywords

System performance models, interconnect delay, crosstalk noise, inductance, VLSI, technology extrapolation

1. INTRODUCTION

Performance prediction of modern high-performance designs affects the evolution of system architectures, design methodologies, and design tools – as well as broader investment strategy in the semiconductor and electronics sectors. Highly influential performance predictors published in the past decade include [1–5] along with the International Technology Roadmap for Semiconductors [6]. The most critical aspect of performance prediction is the idealized *critical path* in the system of interest. For example, a model on-chip critical path might be described as "12 fanout-4 gates driving average-length local interconnects, plus an optimally buffered corner-to-corner 2μm-wide global wire".[1]

In this paper, we focus on the *optimized global interconnect* portion of on-chip critical paths. Our goal is to assess the impact on critical path models of several potentially important, yet previously unmodeled, *optimization degrees of freedom* and *design constraints*. For example:

- Almost all previous predictions rely on *first-order RC line models*. We assess the impact of adding extracted inductance estimates (and analytic RLC line delay estimates) to the interconnect model.
- Previous works apply simple *"optimal repeater sizing" formulae* (e.g., from [1]). We assess the impact of modern optimizations, such as detailed repeater size and interconnect width optimizations [9].
- Since idealized formulae can result in unrealistic assumptions (extremely large repeater sizes, continuous wire tapering, etc.) we also assess the impact of *engineering considerations* including repeater area/size bounds, deliberate backing off of optimal values to the ''knee of the curve", limiting the number of allowed wire widths, and uncertainty in repeater placement (due to layout constraints).
- Previous works use a heuristic charge-sharing analysis to motivate *switch factor* based bounds on delay uncertainty due to crosstalk from neighboring wires. These previous works set the bounds between {0,2}; we assess the impact of using the correct bounds (for ramp input waveforms) of {-1,3} on optimal design solutions that control delay uncertainty [10].
- Previous works typically do not consider *design constraints* for the critical path. We assess the impact of (upper) bounds on noise margin, delay uncertainty, average wire pitch and device (repeater) area.
- Finally, previous works do not consider *real-world design technology* in their global interconnect models. We assess the impact of repeater staggering, and single- and double-shielding techniques, on global interconnect design [11].[2]

Our work attempts to dispel some of the "vagueness" of current performance predictions that arises from the gaps noted above. We do not make any value judgments with respect to existing models; rather, we simply build a comprehensive modeling environment that allows us to identify the issues that *must* be considered by current and future performance predictions. Our end goal is a reusable, transparent, well-engineered prediction model (or, family of models)

* This research was supported in part by Cadence Design Systems, Inc., Synopsys, Inc., and the MARCO Gigascale Silicon Research Center. Dirk Stroobandt is a Postdoctoral Fellow of the Fund for Scientific Research (F.W.O.)–Flanders; his work on GTX was performed during his stay at UCLA as a visiting researcher. Dennis Sylvester is now Assistant Professor of EECS at the University of Michigan, Ann Arbor. Andrew B. Kahng is now Professor of CSE and ECE at the University of California, San Diego.

[1] In fact, this is basically the critical path model of [7] as well as the Roadmap: according to [8], variation between 12 and 16, or between "corner-to-corner" and "chip-side length", is what distinguishes the Roadmap's cycle time frequency predictions for ASIC, and for cost-performance and high-end microprocessor.

[2] There are many additional considerations, some of which are discussed in the conclusions. For example, the effect of perturbing the roadmap for material properties (resistivity, dielectric permitivity) is very significant. Capacitance extraction estimates, etc. also affect performance predictions.

for optimized interconnects and on-chip critical paths. To this end, our basic study optimizes global interconnect delay subject to various constraints (noise margin, delay uncertainty, device/wire cost, routing pitch, etc.) – and we then add in the various new considerations from above. Our study is implemented in the public-domain GTX system [12], which is discussed briefly in Section 2.

The remainder of our paper is organized as follows. Section 3 discusses the impact of inductance on critical paths in terms of shielding, driver sizing, and slew rates (and their impact on coupling noise). The cost-performance tradeoffs inherent in signal shielding are also quantified for the first time. In Section 4, we present a comprehensive study on wire sizing and repeater optimization. This analysis attempts to give a realistic depiction of what an optimal repeater topology should look like in terms of repeater sizing, wire widths, pitch allocation, etc.. In addition, the effect on interconnect delay of via parasitics is also discussed. Section 5 presents conclusions and offers directions for future work.

2. STUDY IMPLEMENTATION

We have conducted our studies within the MARCO GSRC Technology Extrapolation (GTX) system [12,13], which provides a robust, portable framework for interactive specification and comparison of modeling choices (e.g., for predicting system cycle time or power dissipation).[3] Unlike previous "hard-coded" systems such as [3,4,5], GTX adopts a paradigm wherein *parameters* and *rules* allow users to flexibly capture an essentially unbounded space of attributes and relationships that are germane to VLSI technology and design. User-defined rules can be composed in numerous ways to define *rule chains*, which are then executed by a *derivation engine* to perform studies. This use model is ideal for testing the sensitivity of performance predictions to various modeling choices: we simply substitute alternatives (for assumed worst-case switch factor, inductance extraction formula, etc.) while keeping the overall study intact, thus saving redundant programming effort. Because alternative rules or sub-chains can be substituted for each other only when their inputs and outputs match up exactly with respect to naming (and therefore semantics), the system automatically helps ensure the validity and meaning of comparisons.

The default technology used in our studies (exceptions will be noted) is a 0.18μm CMOS process with a supply voltage of 1.8V. V_{th} is 0.3V, and the I_{dsat} values for NMOS/PMOS are 700/350 μA/μm. The critical global interconnect we assume is a 1.5cm top-level copper line with thickness of 1.3μm and $\varepsilon_r = 4.0$.

3. INDUCTANCE ANALYSIS

The effect of inductance on the wire delay is well demonstrated in [1]. Interconnects in deep-submicron designs operating at high frequencies, whose inductive impedance cannot be neglected, must be modeled using RLC segment models. Global interconnects have large cross-section and are usually driven by large drivers with small on-resistances, hence inductive impedance is not dominated by resistive impedance (as a consequence of larger widths and lengths in comparison to local interconnects). When the ratio of inductive impedance to resistance exceeds a certain threshold in an interconnect line, a non-monotone voltage response (i.e., oscillation before settling to a steady state value) results. This makes threshold delay calculation much more difficult than in the RC line case. In such regimes, Elmore and other RC line models cannot accurately estimate signal delay.

Inaccuracies in delay estimation are not only harmful to technology projections, but can also damage performance-driven routing methods which try to optimize interconnect segment length, width, spacing, and repeater/buffer sizing, etc. based on analytic delay formulas. Our study quantifies the impact of using analytic threshold delay formulas derived from RLC line models as opposed to RC line models.

3.1 RLC Delay Modeling

Inductance has a larger impact on inductive noise peak and indirectly affects the capacitive coupling noise peak because the slew times at all the nodes of the wire are faster when the line is modeled as RLC. Inductance is calculated based on expressions from [14,15] and the partial inductance concept [16]. We focus on analytical RLC interconnect delay models because their continuous, closed-form nature is well suited to modern iterative-improvement interconnect design methodologies and global optimization techniques. Gate delay is computed separately using a Thevenin model with voltage source and source resistance corresponding to the driver, and the load is modeled with a capacitance.

The two-pole delay model we use in this study was originally presented in [17] and is briefly described here. The transfer function for the two-pole model is given by

$$H(s) = \frac{1}{1 + b_1 s + b_2 s^2} \qquad (1)$$

The coefficients in this transfer function are given by:

$$\begin{aligned} b_1 &= R_S C + R_S C_L + \frac{RC}{2} + RC_L \\ b_2 &= R_S \frac{C^2}{6} + \frac{R_S RCC_L}{2} + \frac{RC^2}{24} + \frac{R^2 CC_L}{6} + \frac{LC}{2} + LC_L \end{aligned} \qquad (2)$$

where R_S is source resistance, C_L is load capacitance at the end of the line, and R, C, L are the total electrical characteristics of the line. When the input at the source is modeled as a step input the output response is computed separately for the underdamped and overdamped cases.

We have implemented three different interconnect models and compared with SPICE results. **Figure 1** shows the results with varying line lengths. Line width is fixed at 1μm. The driver size and receiver size are fixed at (Wp, Wn)=(54,18)μm. For long wire lengths or for narrow line widths, the line tends to be more resistive (RC dominant), and Bakoglu's RC model produces results closely matching with SPICE. However, when delay is more LC dominated (i.e., large inductance value), the RC model underestimates delay by more than 10%. Friedman's model [18] matches well with SPICE for LC-dominated cases but overestimates delay by up to 30% in RC-dominated cases. Finally, the two-pole model of Kahng and Muddu [17] described above matches SPICE for both RC and RLC cases within 10% error. (Given its very acceptable accuracy, we use the two-pole model for subsequent studies below.) Note that with increasing line length, the 2-pole model changes from the complex pole case (overdamped or LC-dominated) to the real pole case (underdamped or RC-dominated). The condition to determine the case is from $b_1^2 - 4b_2 > 0$ (real poles) or ≤ 0 (complex or double poles).

We also study the reduction of threshold delay by controlling overshoot/undershoot of the voltage response. Typically, circuit design guidelines will define the amount of overshoot and undershoot allowed in a response. These can be translated into a condition between the first and second moments of the interconnect transfer function, which are in turn functions of driver and interconnect parameters. As shown in **Table 1**, undershoot conditions in 0.18 μm

[3] We obtain the GTX program itself – as well as documentation of supporting grammars, parameter naming conventions, extension mechanisms, etc. – from the website [12]. We have contributed all the studies that we have described to the collection maintained at the website.

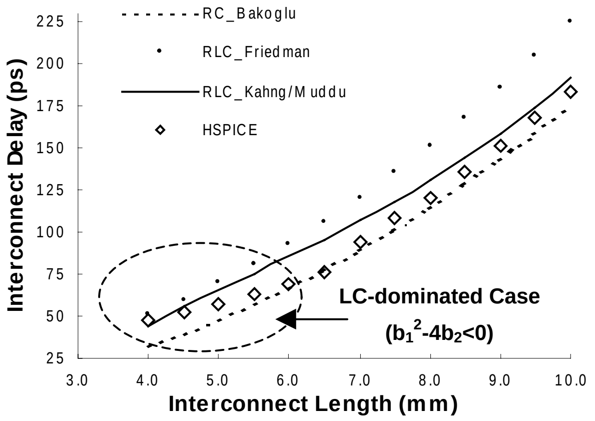

Figure 1. Comparison of RC/RLC delay models

technology can be easily avoided with proper repeater sizing and by providing reasonable signal return paths.

Table 1. Undershoot voltage normalized to V_{dd} with varying drive strengths and return path distances; width = 2 μm

Repeater Size (Wn/Ln)	Return path distance (μm)			
	25	50	100	150
200	0.0004	0.008	0.021	0.029
300	0.0098	0.035	0.061	0.074
400	0.0262	0.062	0.093	0.108
500	0.042	0.083	0.116	0.132

3.2 Shielding Topologies

Shielding is an important technique that designers can leverage to maximize interconnect performance at the cost of increased routing area [19]. By inserting ground and V_{dd} shield wires, current return paths can be clearly defined and loop inductance can be reduced compared to cases without explicit shielding. The extreme case of shielding is described in [20] where every signal wire has a ground and V_{dd} wire as its two nearest neighbors. In this study, we seek to minimize the cost of a design while achieving good performance. The width of the shield wires (W_{shield}) and signal wires (W_{sig}), the spacing between signal wires (S_{sig}), and the spacing from signal to neighboring shield wires (S_{shield}) are all parameters in this study. We examine the following three scenarios:

- No shielding (NS) – all current returns through a regular power grid. Wiring pitch is equal to (W_{sig} + S_{sig}).
- Single shielding (1S) – each signal wire has one shield wire as a nearest neighbor, while the other neighbor is another signal wire. If signal wires are denoted by S and shield (ground) wires by G, the order is G-S-S-G-S-S-G-S-S-G. Wiring pitch is ($2S_{shield}$ + S_{sig} + $2W_{sig}$ + W_{shield})/2.
- Double shielding (2S) – signal and shield wires alternate. This case is identical to the dense wiring fabric in [22]. Wiring pitch is (W_{sig} + W_{shield} + $2S_{shield}$).

The cost function is defined as the product of wiring pitch, repeater sizing factor, and the number of repeaters inserted in the path. We attempt to minimize this cost function based on the following constraints:

1. Maximum delay is set at 1 ns and calculated according to each of the three delay models we have incorporated.
2. Peak noise is fixed at 20% of V_{dd} and calculated based on the exponential model in [21].
3. Delay uncertainty is constrained and defined to be the difference between the RC (2-pole) and RLC delays. This constraint helps minimize potential modeling errors in neglecting line inductance.
4. We set the maximum allowable slew time at the input of any repeater to be 0.5 ns.

Using these constraints, we can examine the impact of shielding topology on circuit performance via coupling capacitance (constraints 2,4) and inductance (constraints 1–3). Since inductance will yield faster slew rates, the peak noise due to capacitive coupling will be indirectly impacted by inductance. We use switch factors of 1, 2, and 3 in this study.

We sweep repeater size, number of repeaters, W_{sig}, and S_{sig} to find the minimal layout cost while meeting the above constraints. We also set W_{shield} to $2W_{sig}$ and S_{shield} equal to S_{sig} to reduce the total number of variables. Results are presented in **Table 2**, which shows the achievable cost (in arbitrary units) with varying switch factors and delay models. Perhaps the most interesting result is that the 2S case can yield the minimal cost when a high switching factor is used. This is true in both RC models – these two models show very similar results from the optimization runs. The RLC model gives the overall best-cost results. Also, the slew time constraint can be more easily met if inductive effects are accounted for. The third constraint described above turns out to be a limiting factor for many input combinations – we find that RLC delay uncertainty is within bounds for smaller repeater sizes and for the 1S and 2S cases where inductance is small due to nearby current return paths.

Table 2. Cost function comparison for varying switch factors, delay models, and shielding scenarios.

Model	Shielding	SF = 1	SF = 2	SF = 3
RC, 1 pole	NS	3.45	5.75	8.75
	1S	5.55	7.4	9.25
	2S	7.65	7.65	7.65
RC, 2 pole	NS	3.45	6.25	9.0
	1S	5.55	7.4	9.25
	2S	7.65	7.65	7.65
RLC	NS	2.85	4.6	6.75
	1S	5.1	6	7.4
	2S	7.05	7.05	7.05

4. DESIGN OPTIMIZATION STUDIES

4.1 Wire Sizing

We next turn to the impact of wire sizing on important design metrics such as delay, noise, and cost. We begin with an expression for optimal wire width as a function of line length, l, from [9]:

$$W_{opt}(l) = \sqrt{\frac{R_{int}(C_f l + 2C_L)}{2R_D C_a}} \qquad (3)$$

Here C_a and C_f denote the area and fringing capacitances per unit length, R_{int} is the per unit length line resistance, and C_L is the load capacitance at the end of the line.[4] We first examine the impact of line spacing on optimal wire width by changing spacing from 0.5 to 2 μm – **Figure 2** plots the optimal line width + spacing for a 1.5 mm line, versus spacing alone on the x-axis. This plot shows an inflection point for switch factors 0 and –1, which corresponds to the optimal *pitch*, not just the optimal line width. Nominal and pessimistic

[4] Fringing capacitance is taken as the difference between the total line capacitance and the parallel-plate capacitance from [20].

switch factors may have such an inflection point, but they do not fall in the design space of the process technology. **Figure 2** uses Equation (3) to calculate optimal line width.

We compare line widths obtained using Equation (3) to the optimal line widths as found by sweeping W in GTX, for a range of driver and interconnect topologies. In addition, we incorporate inductance into the delay expressions and again perform exhaustive sweeping to find optimal line widths based on minimizing RLC as well as RC stage delay. As shown in **Figure 3**, our results demonstrate that (3) matches the GTX results within 10% and often less than 5% error. However, the presence of inductance causes the optimal line width to shrink substantially and (3) therefore overestimates W_{opt} for RLC lines. Also, increasing repeater size leads to a rise in W_{opt} for all models studied – expression (3) shows slightly more error for larger drivers.

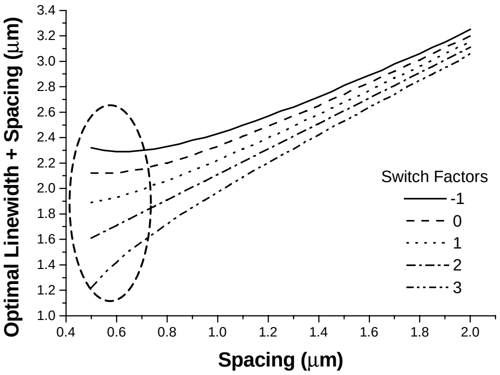

Figure 2. Translation from optimal line width to optimal pitch demonstrates inflection points for certain switch factors.

4.2 Repeater Optimization

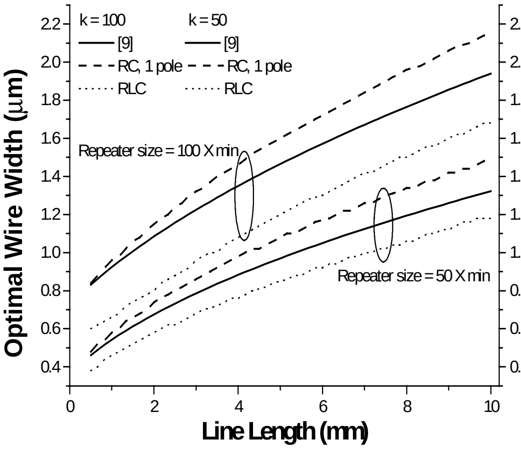

Figure 3. Optimal wire width expression [9] has up to 30% error with respect to RLC model, less error with respect to RC.

In this subsection, we introduce a number of techniques to optimize the use of repeaters in critical paths. Models are developed and used to account for many effects that are currently dealt with in an *ad hoc* manner.

4.2.1 Repeater sizing

The most commonly cited optimal buffer sizing expression is that of Bakoglu [1]:

$$S = \sqrt{\frac{R_D C_{int}}{R_{int} C_{in}}} \qquad (4)$$

R_D reflects the minimum-sized driver resistance, C_{in} is the input gate capacitance of a minimum-sized inverter, and R_{int} and C_{int} are respectively the line resistance and capacitance per unit length. Although this expression can give accurate results in some cases when optimizing for delay only, the delay vs. device size relationship lends itself to further optimization due to its insensitivity near the optimal point. Results obtained from Equation (4) are often unrealistically large – typical standard cell libraries may include inverters or buffers up to 54-96X the minimum size (W_n=L_{drawn}) whereas (1) can give results in the range of 400-700X minimum. To compensate for this, an expression was derived in [5] to optimize a weighted delay-area product rather than purely delay – it gave results on the order of 50-60% smaller than (4). Even with this modification, however, so-called "optimal repeater sizes" seem impractical in the face of power and area constraints.

Here and in the remainder of the subsection, we present a more experimental approach to finding optimal repeater size. For various wire geometries, noise conditions, area and placement constraints and delay models, we develop a more complete picture of the optimal repeater topology solution. We begin with a simple sweep of the repeater size for a single stage of a chain, and examine both delay and energy-delay product vs. repeater size in **Figure 4**.

As **Figure 4** shows, the optimal buffer sizing as calculated from (4) is 480 times the minimum-sized inverter.[5] From pure delay analysis, GTX optimization results indicate that the ideal buffer size for our standard critical path is ~140-150 times the minimum size. When optimizing the energy-delay product, that value drops all the way to 50-60 times minimum. Any range of weighting functions can be easily incorporated into the rule chains – for instance, (energy-delay)^2 or (energy-delay)^3. Results from such functions are not included here, but will push the optimal size towards the delay-only size of 140-150 times minimum. It is also important to note from **Figure 4** that the path delay function around the delay-optimal repeater size is very flat: a buffer which is 43% smaller than optimal yields only a 6.8% delay penalty. Since the energy-delay optimal size is found in the steep part of the delay curve, a truly ideal choice would more closely reflect the knee of the delay curve. In the case of **Figure 4**, our choice of "optimal repeater size" is in the range of 80-100 times the minimum inverter size.

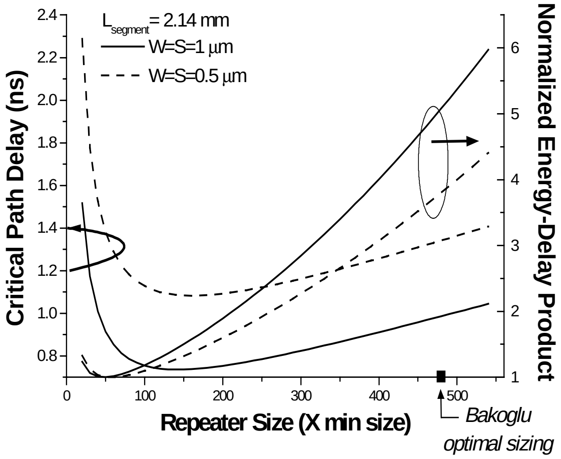

Figure 4. This plot clearly demonstrates the severe oversizing resulting from simple expressions such as Eq. (4).

[5] With the 2 different pitches in the figure, the optimal sizing from (4) actually varies slightly from 485 to 500.

4.2.2 Repeater placement uncertainty

The placement of repeaters in a deep submicron design is non-trivial – many thousands of repeaters must be used to meet timing and noise objectives, and this number will increase with process scaling. As a result, the area consumed by these buffers is substantial and may no longer be ignored during the floorplanning design phase. Particularly in a hierarchical design methodology, such as that proposed in [23], it may not be possible to place repeaters at any given location either inside a pre-designed block or at the top-level of the hierarchy. A potential solution to this problem involves the formation of repeater block regions located around the chip at the floorplanning stage which provide specified areas for repeaters to be placed [24]. However, with such an approach the feasible distances between repeaters are discrete, not continuous.

Here, we study the impact on critical path delay of this inability to place repeaters at arbitrary locations. As before, we examine a top-level metal 1.5 cm route in the default technology. We define an uncertainty parameter, ε, which can range from 0 (no uncertainty) to 1 (maximum uncertainty). We express the location uncertainty as $(1\pm\varepsilon)*L_{seg}$ where L_{seg} is the nominal distance between repeaters when there are no placement restrictions. Given these bounds on segment length between consecutive buffers, we examine the worst-case scenario when half of the segments in the critical path have length $(1-\varepsilon)*L_{seg}$ and the other half are of length $(1+\varepsilon)*L_{seg}$ while total path length is fixed. Given uniform buffer sizing, half of these segments will be overdriven while the other half are underdriven.

While sweeping ε, we vary the switch factor and plot the path delay and peak noise normalized to the $\varepsilon = 0$ case. (Recall that switch factor accounts for the capacitive Miller effect – the impact of neighboring wires switching in the same (opposite) directions can be modeled by lumping their coupling capacitances to ground and multiplying by some switch factor.) Results shown in **Figure 5** indicate that the impact of repeater placement uncertainty is small for total path delay but large for peak noise. This can be understood by realizing that the path delay effectively averages out the resulting fast and slow stages while peak noise is a function of the segment length $(1\pm\varepsilon)*L_{seg}$ and not the total path length. Since the peak noise results are normalized to the $\varepsilon = 0$ case, the switching factor does not play a major role. With a conservative ε of 0.3, the worst-case peak noise increases by approximately 30%.

4.2.3 Staggered repeaters

The use of staggered repeaters for global buses was first described in [11]. The layout structure is shown in **Figure 6**. This approach uses offset buffers in a bus-like structure to minimize the impact of coupling capacitance on delay and crosstalk noise. If repeaters are offset so that each gate is placed in the middle of its neighboring gates' interconnect loads, the effective switching factor is limited to one. This is because potential worst-case simultaneous switching on adjacent wires can be present for only half the victim line's length, and in such conditions the other half of the victim line will consequently experience best-case neighboring switching activity.

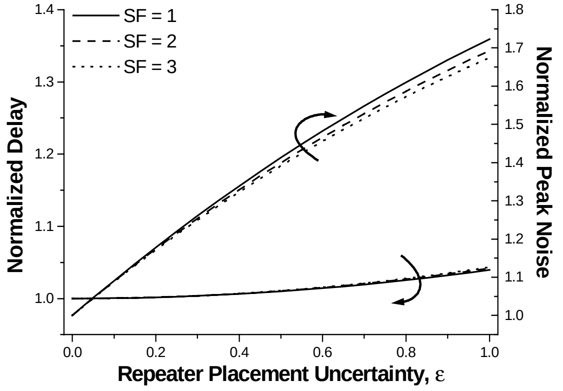

Figure 5. Repeater placement uncertainty has a large impact on noise but little on delay.

In our analysis, we examine the potential reduction in delay uncertainty, as well as in peak crosstalk noise, due to staggered repeaters. **Figure 7** shows that the noise reductions can easily be greater than 10% of V_{dd} for realistic spacing and switch factors. The delay uncertainty when using non-staggered repeaters can exceed 50% of the nominal delay – but staggering almost completely eliminates this uncertainty that stems from capacitive coupling.

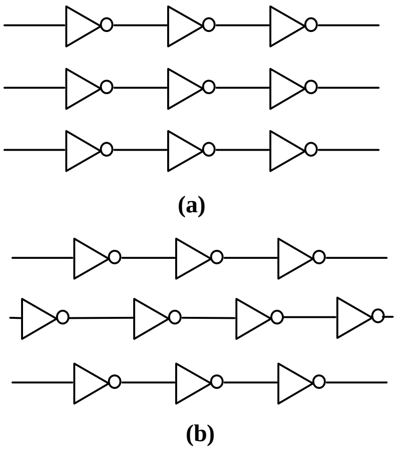

Figure 6. Reduction of worst-case Miller coupling by staggered repeaters. In usual layouts (a), inverters on left and right neighbors are at phase=0 with respect to the inverters on middle line. Staggering (b) places inverters on left and right neighbors at phase=0.5.

4.3 Via Parasitics

The impact of via resistance is commonly ignored in calculating delays for on-chip critical paths. With the steady increase in the number of metallization levels and the shrinking size of vias (and hence, increase in via resistance), this assumption needs to be reexamined. In our analysis, we incorporate the via resistance associated with moving from the silicon level up to the top metal level (where many critical signals are routed) and back down again, into the RC delay expressions in GTX. Via resistance values are taken from an industrial 0.15 μm technology. A somewhat pessimistic derivation upper-bounds the total via resistance from one repeater to the next by 92 Ω; this value is equal to the resistance of a 4.2 mm global line with cross-sectional area of 1 μm^2. Using the 1-pole RC delay expression, the error incurred by ignoring vias is shown in

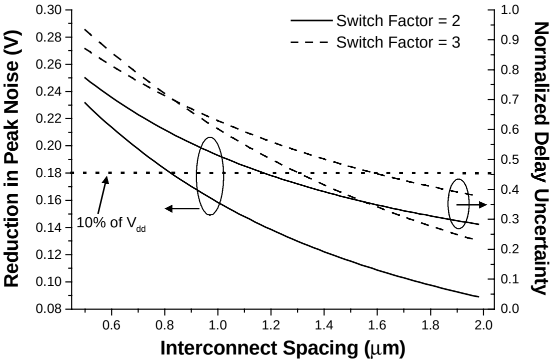

Figure 7. Staggered repeater topology substantially reduces peak noise and delay uncertainty compared to traditional case.

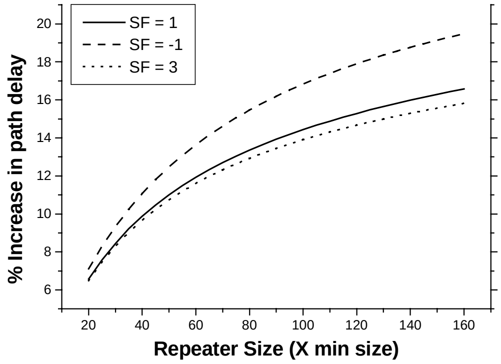

Figure 8. Ignoring parasitic via resistance can lead to significant (10-20%) underestimation of delay, even with modest buffer sizes.

Figure 8. For nominal switching factor of 1 and optimal buffer size (from Section 4.1) of 90X minimum, the via resistance contributes an additional 14% to the total critical path delay.

5. CONCLUSIONS

In this paper, we have attempted to quantify the effects of a number of important deep submicron design issues in the framework of global interconnect optimization. Using a flexible system performance modeling engine implemented as a set of studies within the MARCO GSRC Technology Extrapolation (GTX) system, we examined the topics of RLC delay modeling, optimal repeater and wire sizing, repeater staggering, repeater placement uncertainty effects and via parasitics. We demonstrated that when including inductance, errors in estimates of optimal line delay could increase up to 30%, implying that an RLC-based model could be necessary. A closed-form wire sizing expression was evaluated and found to yield good results compared to a 1-pole RC delay model, but more substantial error compared with an RLC model. We also found that conventional models for optimal repeater sizing [1] are insufficient – our examples show significant overestimation of repeater size up to 500%. A more effective sizing criterion would weight energy and delay so that the size closely approximates the knee of **Figure 4**. We have also modeled the impact of a number of design issues, including repeater staggering (a layout technique which limits delay uncertainty and peak noise due to capacitive Miller effect) and repeater placement uncertainty (which may result in underestimated noise peaks if left unmodeled).

There are a number of open issues left unaddressed in our paper. For example, critical-path modeling can be strongly affected by the calculation of effective capacitance (to account for resistive shielding) or assumptions regarding gate fan-out. Notable materials and process technology issues include grain boundary and cladding layer effects on material resistivity (especially for copper), the effects of manufacturing variability on the predicted performance of VLSI interconnects, and the possible optimism of current roadmaps for dielectric permitivity.[6] Our ongoing research aims to incorporate these and other considerations into our unified test bench for performance prediction of leading-edge designs.

6. REFERENCES

[1] H.B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley, 1990.

[2] G.A. Sai-Halasz, "Performance Trends in High-Performance Processors," *Proc. IEEE*, Jan. 1995, pp. 20-36.

[3] J.C. Eble, V.K. De, D.S. Wills and J.D. Meindl, "A Generic System Simulator (GENESYS) for ASIC Technology and Architecture Beyond 2001," *Proc. ASIC*, 1996, pp. 193-196.

[4] Rensselaer Interconnect Performance Estimator (RIPE), http://latte.cie.rpi.edu/ripe.html

[5] D. Sylvester and K. Keutzer, "System-Level Performance Modeling with BACPAC – Berkeley Advanced Chip Performance Calculator," *Proc.SLIP*,1999,pp.109-114, http://www.eecs.berkeley.edu/~dennis/bacpac/

[6] International Technology Roadmap for Semiconductors," December 1999, http://www.itrs.net/

[7] P. D. Fisher and R. Nesbitt, "The Test of Time: Clock-Cycle Estimation and Test Challenges for Future Microprocessors,*" IEEE Circuits and Devices Magazine* 14(2) 1998, pp. 37-44.

[8] A. E. Dunlop and P. D. Fisher, *personal communication*, 1999.

[9] J. Cong and D.Z. Pan, "Interconnect Estimation and Planning for Deep Submicron Designs," *Proc. DAC*, 1999, pp. 507-510.

[10] A.B. Kahng, S. Muddu and E. Sarto, "On Switch Factor Based Analysis of Coupled RC Interconnects," *Proc. DAC*, 2000, pp. 79-84.

[11] A.B. Kahng, S. Muddu and E. Sarto, "Tuning Strategies for Global Interconnects in High-Performance Deep Submicron IC's," *VLSI Design* 10(1), 1999, pp. 21-34.

[12] http://vlsicad.cs.ucla.edu/GSRC/GTX/

[13] A E. Caldwell, Y. Cao, A.B. Kahng, F. Koushanfar, H. Lu, I. Markov, M. Oliver, D. Stroobandt and D. Sylvester, "GTX: The MARCO GSRC Technology Extrapolation System," *Proc. DAC*, 2000, pp. 693-698.

[14] L. He, N. Chang, S. Lin, and O.S. Nakagawa, "An Efficient Inductance Modeling for On-Chip Interconnects," *Proc. CICC*, 1999, pp. 457-460.

[15] X. Qi, G. Wang, Z. Yu, R.W. Dutton, T. Young and N. Chang, "On-Chip Inductance Modeling and RLC Extraction of VLSI Interconnects for Circuit Simulation," *Proc. CICC*, 2000.

[16] A. E. Ruehli, "Inductance calculations in a complex integrated circuit environment,*" IBM J. Res. Dev.*, September 1972, pp. 470-480.

[17] A.B. Kahng and S. Muddu, "An analytical delay model for RLC interconnects," *IEEE Trans. CAD* 16(12) (1997), pp. 1507-1514.

[18] Y.I. Ismail, E. G. Friedman, J.L. Neves, "Equivalent Elmore delay for RLC trees", *IEEE Trans. CAD* 19(1) (2000), pp. 83-97.

[19] Y. Massoud, S. Majors, T. Bustami and J. White, "Layout Techniques for Minimizing On-Chip Interconnect Self-Inductance," *Proc. DAC*, 1998, pp. 566-571.

[20] S. P. Khatri, A. Mehrotra, R. K. Brayton, A. Sangiovanni-Vincentelli, and R.H.J.M. Otten, "A Novel VLSI Layout Fabric for Deep Submicron Applications," *Proc. DAC*, 1999, pp. 491-496.

[21] K. L. Shepard *et al*., "Design Methodology for the S/390 Parallel Enterprise Server G4 Microprocessors,*" IBM J. Res. Dev.*, July-Sept. 1997, pp. 515-554.

[22] T. Sakurai, "Closed-Form Expressions for Interconnect Delay, Crosstalk, and Coupling in VLSI's,*" IEEE Trans. Electron Devices*, Jan. 1993, pp. 118-124.

[23] D. Sylvester and K. Keutzer, "Getting to the Bottom of Deep Submicron," *Proc. ICCAD*, 1998, pp. 203-211.

[24] J. Cong, T. Kong and D. Z. Pan, "Buffer Block Planning for Interconnect-Driven Floorplanning," *Proc. ICCAD*, 1999, pp. 358-363.

[6] Many experts believe that the low-κ roadmap is far too optimistic due to the barrier layers that are needed between inter-layer dielectrics and interconnects. Corrected values may have very large implications on various critical-path projections. Of course, the overall roadmap, along with critical-path projections for MPU and ASIC architectures, shows great sensitivity to other issues as well: SRAM and logic layout densities, die cost models (which impact die size assumptions), etc.