# Interconnect Effort - A Unification of Repeater Insertion and Logical Effort

Srividya Srinivasaraghavan and Wayne Burleson
*Department of Electrical and Computer Engineering*
*University of Massachusetts, Amherst, MA 01003*
*E-mail: {ssriniva,burleson}@ecs.umass.edu*

## Abstract

*Interconnects are assuming increasing importance in deep submicron design, however there is a significant gap between industrial design practice and classical results of uniform repeater insertion. This work questions two of the standard assumptions of academic interconnect circuits: 1) the uniformity of repeater sizes and spacing and 2) the boundary between logic and interconnects.*

*This research explores the co-design of logic sizing and repeater insertion for improved delay, power and placement. The technique of logical effort is used to develop the sizing scheme for the logic including polarity considerations. Non-uniform repeater insertion is used to combine cascaded sizing and distributed wire buffering. HSPICE simulations carried out for the $0.18\mu$ technology show that combining the sized logic with uniform repeaters is faster than using minimum sized logic circuitry by about 10% while with non-uniform repeaters the gain is about 15%. The average power consumption of non-uniform repeater insertion is less than that of uniform insertion by about 20%. Non-uniform repeater insertion is also less placement sensitive (shifting the position of each repeater in the setup by about 40% results in a delay loss of only about 3%)*

## 1. Introduction

The advancement of integrated circuit technology first predicted by Moore's law [1] in 1965 and projected [2] till 2015 has resulted in exponential reductions of minimum feature size as well as increase in die size and clock rate.While scaling has made devices smaller and faster, it has also increased the importance of the more resistive interconnects due to increasing line lengths and greater impacts of coupling capacitance. This has resulted in interconnect limited designs. It is now a well known fact that below $0.25\mu$ technologies, the performance of global chip communication is primarily limited by interconnect delay. Rapid scaling has introduced many challenges on interconnect performance, modeling and reliability.

Repeater insertion has been the classical and most popular approach to boost the performance of interconnects. Uniform repeater insertion changes the delay dependence on the length from quadratic to linear [3, 6]. There has been a significant work done to determine the optimal placement and size of the repeaters [3, 5, 9, 10], all of which assume uniform sizing and spacing. However the pitfalls of uniform repeater insertion are that the initial cascade of buffers leading to the first repeater (Figure 1) contributes a significant fraction of the delay (about 40% for a linelength of 15mm) as shown in Figure 2. Also this stack of buffers can lead to unfavorable power spikes. Another point of concern as projected in [9] is the placement sensitivity of the uniformly spaced repeaters.The results shown are for the $0.18\mu$ model assuming interconnect parameters of upper metal layers.

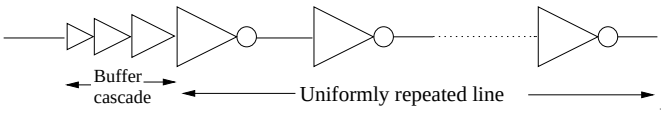


Figure 1: **Schematic of uniformly repeated line with initial cascade**

While improved models for repeater insertion are being developed, scaling trends [19] and discussions with microprocessor designers at Intel and Compaq have indicated that a more pragmatic approach is warranted. In particular, increasing clock rates indicate that very few wires will ever have more than 5-6 repeaters prior to being latched. Also, the capacitance to resistance ratios of modern metal stacks indicate that very large repeaters (200x-300x minimum size inverter) are needed to be within 20% of optimal delay [9]. Thus we argue that it is necessary to consider device sizing in logic and latches along complex interconnects. Logical effort, as developed by Sutherland and Sproull [11] provides an elegant formulation for this problem, which when combined with repeater insertion, including non-uniform

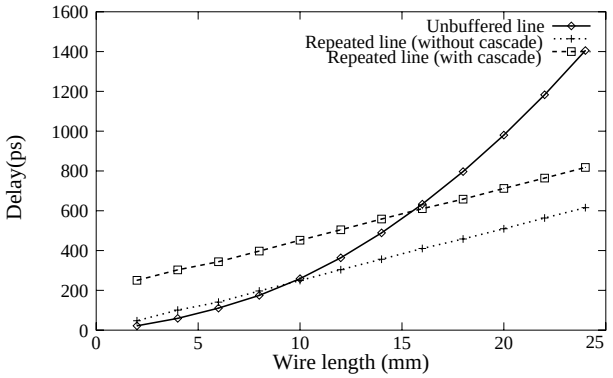


Figure 2: **Delay of buffered line with initial cascade**

insertion (repeaters with progressively increasing sizes driving correspondingly longer lines), provides advantages in terms of delay, power and placement.

Two of the classical results in VLSI theory are closed-form expressions for 1) the step-up ratio of buffer cascades being equal to $e$ [18], and 2) uniform repeater insertion [3]. Although deep sub-micron devices and interconnects have required some modifications to these results, they still provide good intuitive guidelines for designers and initial estimates for CAD tools. This work seeks to combine these results to deal with the overall problem of logic and interconnection.

Existing CAD tools typically synthesize logic blocks and then use some kind of capacitive load model to do sizing [17]. However, it is now well recognized that early floorplanning and global routing estimates are needed to improve these models [8, 16, 4], but interconnect design still remains a largely seperate task from that of block synthesis. This work attempts to reconcile this seperation by combining logic device sizing and repeater insertion (Figure 3).

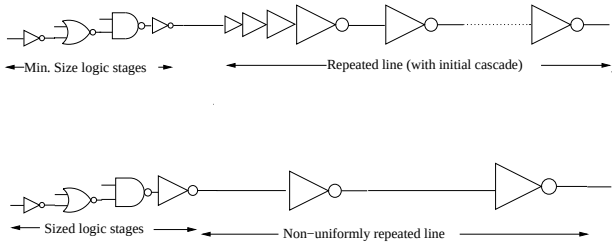


Figure 3: **Elimination of buffer cascade using logic sizing and non-uniform repeater insertion**

Logical effort has been used for device sizing in datapaths[12], SRAMs[13] and also for fan-out optimization[14]. These studies either ignore interconnects or assume that the wires are short so that they can be modeled as an additional load capacitance to carry out the sizing. This work looks at long wires (upto 15mm) and develops a sizing scheme for the overall logic interconnection block.

The methodology employed to size the logic block preceeding a uniformly repeated interconnect and the gains that this offers to interconnect performance is described in Section 2. The basic idea behind non-uniform repeater insertion and the establishment of the optimal values for the number and size of repeaters is expalined in Section 3. The implications of combining the logical block sizing with the non-uniformly repeated line are described in Section 4. Section 5 has a brief overview of the effects of migrating this sizing scheme to lower geometries ($0.13\mu, 0.10\mu, 0.07\mu$). Some conclusions and further development ideas are presented in Section 6.

## 2. Logical Effort in Uniformly Repeated Interconnects

Logical effort refers to the inherent cost of computation in logic functional blocks and is a characterization of the complexity of a gate. It provides a method to estimate delay in a CMOS circuit, develop a scheme for sizing the transistors and also quantify the benefits of various circuit topologies [11].

The total delay, $d$, incurred by a logic gate is composed of a fixed part called the parasitic delay $p$ (which is independent of the size of logic) and a part proportional to the load at the gate output called the effort delay $f$, i.e $d = f + p$. The effort delay in turn depends on the logic properties of the gate (logical effort, $g$) and the load driving capability of the gate (electrical effort, $h$),i.e $f = gh$ The load capacitance at the end of the logic circuit, the input capacitance as well as the topology of the circuit are the primary factors that determine the sizing scheme of the logic.

The traditional process of repeater insertion as well as minimization of delay, power etc. along a repeated line is carried out independent of the nature of logic circuits preceeding the interconnect. The signal from logic devices is assumed to drive the repeated line whose parameters are determined by the Bakoglu optimization procedure [3]. This sometimes necessitates the introduction of the a cascade of buffers leading to the first repeater(so as to enable the signal from the logic capable of driving the repeater, Figure 4). This cascade can be eliminated by taking into consideration the load capacitance provided by the first repeater in the sizing of the logic, and distributing the electrical effort pro-

vided by the cascade along the stages of the logic block (Figure 5). So, the sized logic block not only performs the required logic function but each stage of the also contributes to some fraction of electrical effort required for the signal to be able to drive the first repeater. In other words, while the logical effort (contributed by the logic stages) and the electrical effort (contributed by the cascade of buffers) are separated in Figure 4, they are integrated (as the total path effort) in Figure 5 by appropriate sizing.

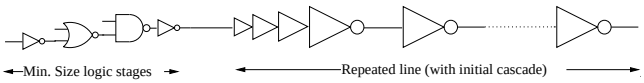


Figure 4: **Minimum sized logic leading to the cascaded first repeater**

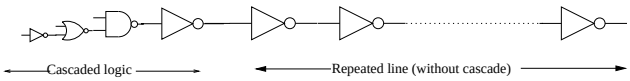


Figure 5: **Cascaded logic leading to repeated line**

A comparison of the delays (determined from HSPICE simulations) for the cases represented in Figures 4 and 5 for various wirelengths is shown in the Figure 6. The results are for the $0.18\mu$ technology with $V_{dd} = 1.8V$. The device models are provided by the Berkeley Predictive Technology Method [15] and wire models are based on their typical global interconnect parameters . As can be seen from the plots, there is about a 10% gain in the delay when logic sizing is carried out while the average power consumption almost remains the same in both cases. An obvious disadvantage of this sizing procedure is the greater area requirement for the sized logic blocks because of their bigger sizes.

## 3. Non-Uniform Repeater Insertion

Non-Uniform Repeater Insertion is an alternate design technique to eliminate the initial buffer cascades of uniformly repeated lines. In this method of repeater insertion, as shown in Figure 7, buffers of increasing sizes drive progressively longer segments of the wire.

The schematic shows a minimum sized inverter driving a line length of 'a'. Successive repeaters are 'f' times bigger than the previous one driving a line that is 'r' times longer than the previous one. There are 'n' such stages along the line. The 50% propagation delay for such a setup can be analytically derived as a function

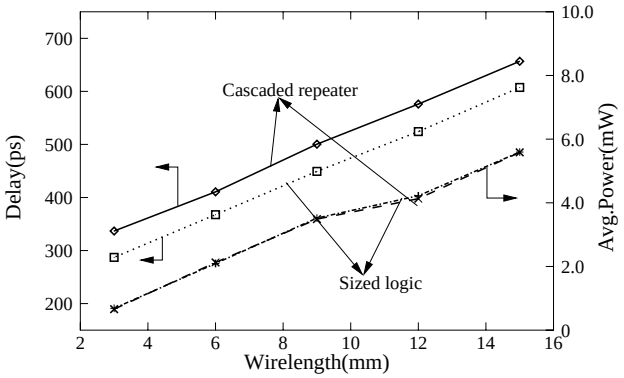


Figure 6: **Comparison of delay and average power**

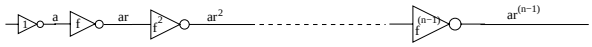


Figure 7: **Schematic for non-uniform repeater insertion**

of a,r,n and f.

$$T_{50\%} = 0.7R_oC_o[(n-1) + \frac{1}{f^{(n-1)}}] + \frac{0.7R_oC_{int}}{L}[a(\frac{1-(\frac{r}{f})^{n-1}}{1-(\frac{r}{f})}) + \frac{z}{f^{n-1}}] + \frac{0.7R_{int}C_o}{L}[af(\frac{1-(rf)^{n-1}}{1-(rf)}) + z] + \frac{0.4R_{int}C_{int}}{L^2}[a^2(\frac{1-r^{2(n-1)}}{1-r}) + z^2]$$

where $R_o$ and $C_o$ are the output resistance and input capacitance of a minimum sized inverter respectively and $R_{int}$ and $C_{int}$ are the resistance and capacitance of the interconnect.

The complexity of the delay expression does not allow for analytical determination of optimal solutions for 'a','r' and 'f'. However exhaustive HSPICE simulations are carried out to determine the combination of the design parameters to obtain the minimum delay. The delay of the non-uniformly repeated line for various line lengths is shown in Figure 8. The design parameters leading to the plot are: a=0.01mm, r=f=3.25.

Even though the delay of the non-uniformly repeated line is greater than that of the uniformly repeated line, this method is attractive because it possible to have longer unrepeated lines. Table 1 compares some design features of uniform and non-uniform repeater insertion techniques.

Another attractive feature of this design is its low sensitivity to placement. For an interconnect (with 6 repeater stages) having non-uniformly sized and spaced repeaters, reducing the length of wire driven by each repeater (except the last one) by 40% and having the last repeater drive the remaining line length increases

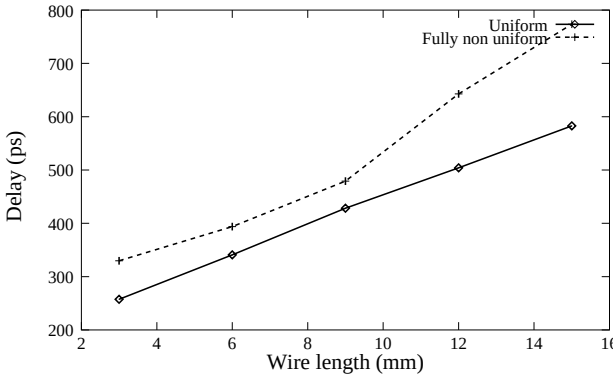


Figure 8: **Comparison of delays between uniform and non-uniform repeater insertion**

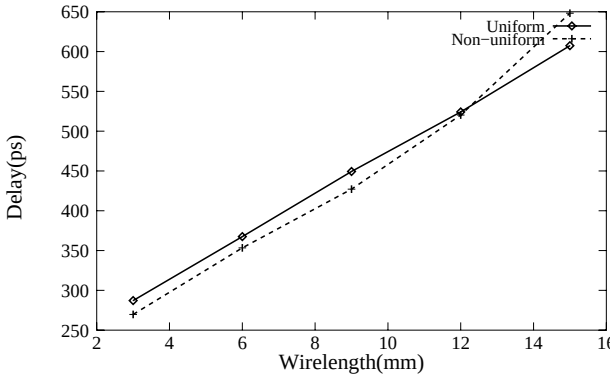


Figure 9: **Comparison of delays between uniform and non-uniform repeater insertion when logic is sized**

| Wire-length (mm) | Uniform | | Non-Uniform | |
|---|---|---|---|---|
| | # of buffers (incl.cascade stages) | Segment length(mm) | # of buffers | Longest seg. length(mm) |
| 3 | 6 | 3 | 6 | 1.39 |
| 6 | 7 | 3 | 6 | 4.39 |
| 9 | 9 | 2.25 | 7 | 3.77 |
| 12 | 10 | 2.4 | 7 | 6.77 |
| 15 | 11 | 2.5 | 7 | 9.77 |

Table 1: **Comparison of design features of Uniform and Non-uniform repeaters**

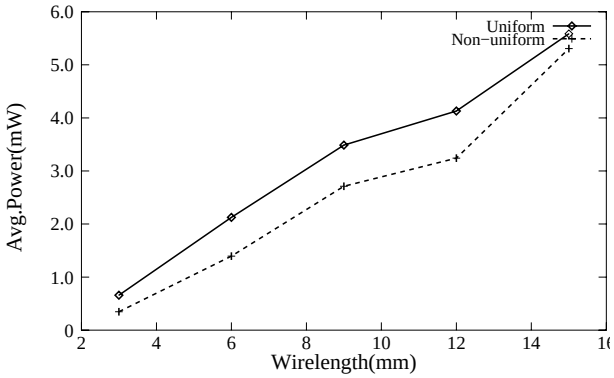


Figure 10: **Comparison of average power between uniform and non-uniform repeater insertion when logic is sized**

the delay by about 20ps whereas an identical placement done for a uniformly repeated line increases the delay by about 300ps [9].

## 4. Combining Logical Effort & Non-uniform Repeater Insertion

The design flexibility offered by non-uniform repeater insertion procedure can be combined with the sizing of logic blocks. The sizes of the logic stages can be reduced and the ouput of the logic can drive a small repeater which then drives a bigger repeater and a longer wire segment. A comparison of the interconnect delays for various line lengths when the sized logic block drives a uniform and non-uniform repeated interconnect is shown in Figure 9

The sized block followed by a non-uniformly repeated line provides delay gains for smaller wirelengths while for lines longer than 12mm, it loses by about 6%. However the average power consumption gain by using the non-uniformly repeated line is very significant (about 22% for 12mm) as can been seen in Figure 10

The gains in power consumption can be attributed to the fact that fewer of repeater stages (each driving a longer segment) are required in the case of non-uniform repeater insertion.

Most practical circuits have to handle both fan-in and fan-out. The method of logical effort accounts for the fan-in/fan-out by including a 'branching effort' in the sizing scheme. The branching effort depends on the load capacitance (along the branch where the delay has to be minimized) and the total capacitance as seen at the branch point. Here we consider a 16-input AND gate structure (composed of 2-input NAND and NOR gates) driving two interconnect branches. The schematic is shown in Figure 11. The logic in this case has to drive a greater load due to the presence of a big device (first repeater of the interconnect branch) on each leg and so the sizes of constituent logic gates also becomes larger.

The delay and power consumption results for both uniform and non-uniform interconnect are shown in

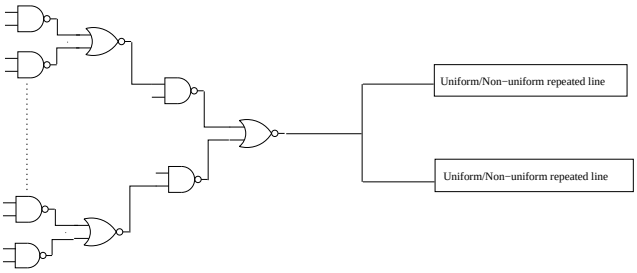


Figure 11: **Schematic for logic block driving branched interconnect**

the Table 2. Here again the logic, driving branched

| Wire-length (mm) | Delay(ps) | | Avg.Power(mW) | |
|---|---|---|---|---|
| | Uniform | Non-uniform | Uniform | Non-uniform |
| 3 | 305.2 | 298 | 0.45 | 0.3 |
| 6 | 382.8 | 383.6 | 1.3 | 0.95 |
| 9 | 471.1 | 459.8 | 2.18 | 1.63 |
| 12 | 553.6 | 552.2 | 2.6 | 1.92 |
| 15 | 618.2 | 656.5 | 3.3 | 2.39 |

Table 2: **Comparison of delay and avearge power when logic drives branched interconnects**

non-uniformly repeated interconnects offers significant power gains over the case when the logic is followed by branched uniformly repeated interconnects.

## 5. Technology Scaling Impacts

The technology scaling trends as projected in the ITRS Roadmap[2] point toward growing wire delays with shrinking geometries. Table 3 lists the relevant interconnect parameters over the technology generations from 180nm to 70nm. The wirelength of the global in-

| Technology (nm) | **180** | **130** | **100** | **70** |
|---|---|---|---|---|
| Clock (GHz) | 1.2 | 1.6 | 2.0 | 2.5 |
| $V_{dd}$(V) | 1.8 | 1.5 | 1.1 | 0.9 |
| Chip area($mm^2$) | 340 | 372 | 408 | 468 |
| $R_{int}$ $per$ $mm$ ($\Omega$) | 20.0 | 27.78 | 33.33 | 37.04 |
| $C_{int}$ $per$ $mm$ ($fF$) | 243.77 | 258.01 | 259.67 | 228.48 |

Table 3: **Relevant interconnect parameters (based on the 1999 ITRS roadmap)**

terconnects does not get scaled but grows proportional to the die sizes. It has also been shown that the number of repeaters present in a die increases. Figure 12 shows the trends pertaining to the number and size if repeaters across different technology generations.

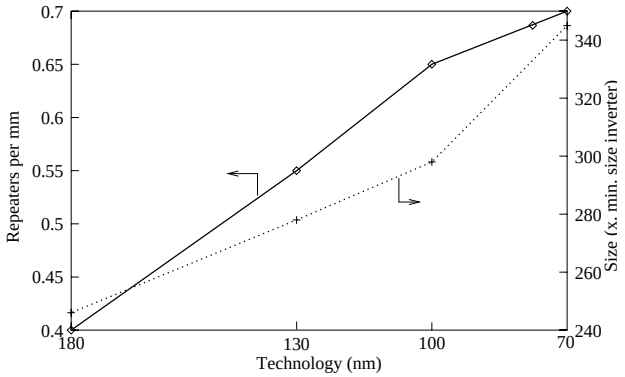


Figure 12: **No.of repeaters per mm and size across technology generations**

Since the sizing of the logic device preceeding the interconnect is dependent on the repeater size, an increase in their sizes as the technology shrinks will consequently increase the sizes of the logic stages. The increased electrical effort (reflected by the increased repeater sizes) is distributed uniformly over each of the 4 logic stages so that the sizes of the devices in each stage (w.r.t to the dimensions of the minimum size inverter in that technology) does not increase appreciably. The gain in delay due to logic sizing across technology generations for a specific wirelength is shown in Figure 13.

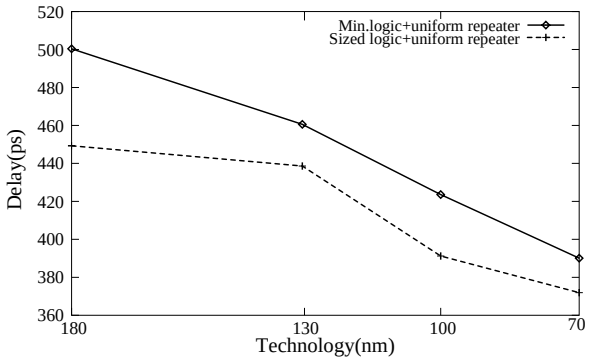


Figure 13: **Delay gain due to logic sizing across technologies**

The formulation of non-uniform repeater insertion is such that, the number and size of repeaters is independent of the line resistance and capacitance but depends only on the line length. This means that the

design specifications (including the logic stage sizing) can remain unaltered for all technology generations. The comparison of delays between using uniform and non-uniform repeaters following the sized logic for a specific wirelength is shown in Figure 14.

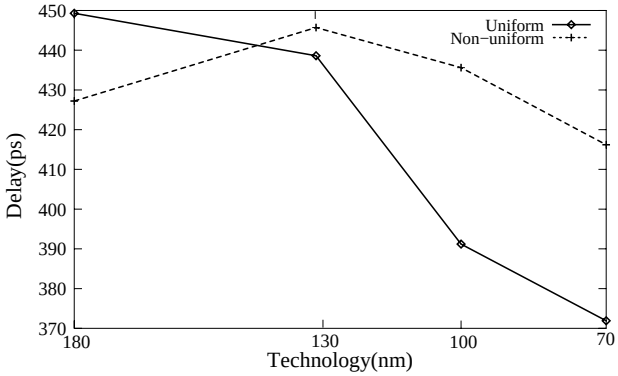


Figure 14: **Comparison of delay across technologies when sized logic is followed by uniform and non-uniform repeated line**

As can be seen from the Figure 14, using non-uniform repeater insertion technique is lower technologies loses over uniform repeater insertion. This can be attributed to the fact while the interconnects become more resistive and capacitive in these smaller geometries, the insertion procedure does not account for this in its optimal sizing and spacing of the repeaters. By carrying out the optimization of parameters explicitly in each technology generation, the non-uniform insertion procedure can be made to be provide gains over uniform repeaters. However the average power consumption is lesser when the sized logic is followed by non-uniformly repeated line as seen in Figure 15.

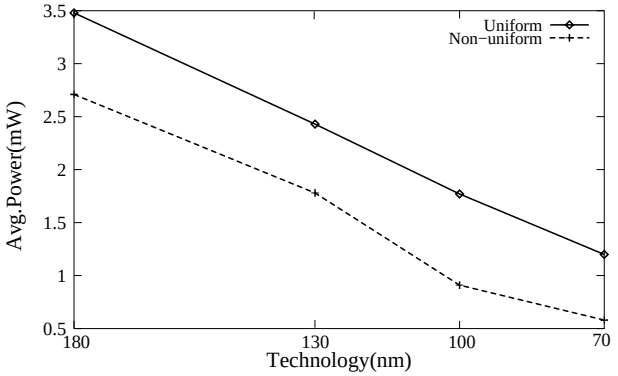


Figure 15: **Comparison of avg. power across technologies when sized logic is followed by uniform and non-uniform repeated line**

## 6. Conclusions and Future Work

Sizing of the logic preceeding an interconnect is shown to provide gains in interconnect performance. Even though the gains with respect to delay are marginal, they do offer benefits in terms of power consumption down to 70nm technology. A greater flexibility in terms of placement was observed when using a non-uniform repeated interconnect as opposed to having a uniformly repeated wire. The method of logical effort for sizing the logic gives a compact means of establishing optimal device sizes and helps avoid extensive iterations that are typical of other sizing schemes. This work is motivated by our experience with advanced microprocessor designs [20, 21, 22] which typically use manual non-uniform repeater insertion to deal with the inherent non-uniformities in routing layers, loads, and coupling. However they also use a step-up in the repeaters to account for insufficient sizing in the preceding logic blocks and latches. Designing a latch with 200x-300x devices is not reasonable from an area or power standpoint. In this context, appropriate sizing of the logic stages will prove to be beneficial.

Only RC interconnects have been considered in this study. Considering the fact that inductance effects are becoming more important in lower technologies, it is important to extend this sizing scheme to include RLC interconnects also. The methodology for logic sizing would be the same for RLC interconnects but the physical sizes might vary due to the different repeater sizes in RLC interconnects. This idea of sizing based on logical effort can be extended to latches and other dynamic blocks. Also, the requirements when the signal from a logic and traversing through a long wire has to drive another logic circuitry should also be studied. Such analysis would further highlight the practical concerns to be accounted for while developing an optimal sizing scheme.

**Acknowledgement:** This work has been partially supported by Semiconductor Research Corporation under contract 2000-HJ-766 and Intel Corporation.

## 7. References

[1] G.E. Moore, "Cramming More Components onto Integrated Cicruits,"*Electronics Magazine, Vol. 38*, 1965, pp. 114-117.

[2] International Technology Roadmap for Semiconductors (ITRS), *http://public.itrs.net*, 2001.

[3] H. B. Bakoglu, “Circuits, Interconnections and Packaging for VLSI,” *Addison Wesley,* 1990.

[4] R.H.J.M. Otten and R.K. Brayton,“Performance Planning,”*Integration, The VLSI Journal*, 2000, pp. 1-24.

[5] H. B. Bakoglu and J. D. Meindl, “Optimal Interconnection Circuits for VLSI," *IEEE Transactions on Electron Devices,* 1985, pp. 903-909.

[6] J. Rabaey, “Digital Integrated Circuits, A Design Perspective,” *Prentice-Hall,* 1996.

[7] D. Sylvester and K. Keutzer, "Getting to the Bottom of Deep Submicron II : A Global Paradigm," *Proceedings of IEEE International Symposium on Physical Design,* 1999, pp. 193-200.

[8] Z. Pan, L. He, C. K. Koh, K. Y. Khoo and J. Cong, "Interconnect Design for Deep Submicron ICs," *Proceedings of IEEE Conference on Computer Aided Design,* 1997, pp. 478-485.

[9] A. Nalamalpu and W. Burleson, "Repeater Insertion in deep sub-micron CMOS : Ramp-based Analytical Model and Placement Sensitivity Analysis," *IEEE International Symposium on Circuits and Systems,* 2000, pp. 766-769.

[10] V. Adler and E. G. Friedman, "Repeater Design to Reduce Delay and Power in Resistive Interconnect," *IEEE Transactions on Circuits and Systems - II,* 1998, pp. 607-616.

[11] I. Sutherland, B. Sproull and D. Harris,*Logical Effort - Design of fast CMOS circuits*, Morgan Kaufmann, 1999.

[12] X. Yu, V. Oklobdzija and W. Walker, “Application of Logical Effort on Design of Arithmetic Blocks,” *ASILOMAR 2001,* pp.872-874.

[13] B. Amrutur and M. Horowitz, “Speed and Power Scaling of SRAMs,”*Journal of Solid State Circuits*,2000.

[14] P. Rezvani, A. Ajami, M. Pedram and H. Savoj, LEOPARD - Logical Effort Based Fanout Optimizer for Area and Delay, *Proc of ICCAD*, 1999.

[15] Berkeley Predictive Technology Model (BPTM), *http://www-device.eecs.berkeley.edu/ptm*

[16] J. Cong and C. Koh,“Simultaneous Driver and Wire Sizing for Performance and Power Optimization,”*Proceedings of ICCAD*, 1994.

[17] R. Rudell,”Logic Synthesis in VLSI Design,” *Ph.D Thesis, U. Berkeley*, 1989.

[18] C.A. Mead and L. Conway, *An Introduction to VLSI Systems*. Addison Wesley, 1980.

[19] A. Maheshwari, S. Srinivasaraghavan and W. Burleson, "Quantifying the Impact of Current Sensing on Interconnect Delay Trends", *Proc. of IEEE International ASIC/SoC Conference*, Sept 2002.

[20] A. Alvandpour, Private Communication, *Intel Corp.*

[21] C. Chun, Private Communication, *Motorola Inc.*

[22] M. Gowan, Private Communication, *Intel Corp.*