

A Practical Repeater Insertion Method in High Speed VLSI Circuits

Julian Culetu
Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto, CA 94303

julian.culetu@Eng.Sun.com

Chaim Amir
Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto, CA 94303

chaim.amir@Eng.Sun.com

John MacDonald
Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto, CA 94303

john.macdonald@Eng.Sun.com

Abstract

In today's design of VLSI high speed circuits, frequency has a major impact on the number of repeaters that needs to be inserted. A microprocessor operating at less than 200Mhz might require several hundred repeaters, while one operating at greater than 500Mhz may require a number in the thousands. The following paper describes an efficient and simple way to automatically determine buffer placement based on maintaining equal transition time for all gate input signals across the net. A maximum allowable transition time is determined (limited by the frequency of the circuit), and correlated with the interconnect Elmore Delay. A Spice RC model having nodes with physical locations (X, Y coordinates) can be obtained by extraction tools providing standard parasitic format (SPF). This can then be used with the results of the algorithm for repeater placement to determine the exact physical location desired for each repeater.

1. Introduction

In a high performance VLSI design, signal propagation over on chip metal interconnect is significant due to RC delay. In general, the interconnect delay increases with the square of the length of the line. Delay can be reduced by inserting inverting or non-inverting buffers. A large scale integration design has certain objectives when it comes to inserting repeaters. These design objectives are (a) to minimize interconnect delay and (b) limit transition times while limiting impact on (c) area and (d) power.

Algorithms for post-layout buffer insertion in interconnect wiring were studied by van Ginneken [4] with further contributions by Kannan, Suaris and Fang [3] and by Lillis, Cheng and Lin [4, 6]. Given required arrival times at cell inputs and analyzing Elmore delay on a globally routed net, van Ginneken presented a dynamic programming algorithm that positioned and inserted buffers such that the latest possible arrival time at the interconnect driving point could be achieved. Buffers were then connected during a detail route stage. Work by Lillis, et. al. [1, 2] employed a similar approach in providing polynomial-time algorithms for simultaneous buffer insertion and wire sizing and, later, algorithms for power minimization and multiple-source nets.

In practical application there are additional considerations that must be applied in a large, high performance design. One such consideration is the necessity for signals propagated through wiring with RC degradation to achieve acceptable voltage levels for recognition of logic low and high. This translates into a limitation on signal transition or rise and fall times at gate inputs. In this paper

we present a simple and efficient method using transition time limitation for post-layout buffer insertion and show how the solution achieved is near optimal, providing a practical trade-off among timing, power and area.

Our repeater insertion method starts with a Spice analysis to find the optimum delay and to limit the transition times while not losing track of area and power. The goal here is to find the wire sizing of each metal and buffer size in the design that best fit the above mentioned requirements (delay, area, and power). Once the buffer and the wire sizes have been selected, a correlation between transition time (measured in Spice) and the Elmore delay is performed (Section 2). The Elmore delay corresponding to the transition time (limited by the frequency of the design) can be determined and used as a limiting value in the repeater insertion algorithm. When inserting repeaters, equal transition times are maintained at the input of each gate in the net. The placement of repeaters can be easily obtained given that the extraction tools provide standard parasitic format (SPF) with X and Y coordinates for each node in the net (Section 3). The implementation of this method is described in Section 4 followed by the conclusions in Section 5.

2. Transition Time and Elmore Delay Correlation

2.1 Background

The Elmore delay, as described in [6], is an efficient way of calculating the delay for an RC tree. A first order approximation of RC delay at any node i on an RC tree is given by the Elmore time constant

$$T_i = \sum_{k=1}^n R_{ki} C_k \quad (1)$$

where R_{ki} is the resistance of the portion of the (unique) path between the input and node i , that is common with the (unique) path between the input and node k , and C_k is the capacitance at node k .

The first order approximation of the waveform at node i is

$$V_i(t) = V_{DD} \left(1 - e^{-t/T_i} \right) \quad (2)$$

where T_i is the time constant Elmore delay given by (1). The time at which the voltage at node i reaches any value V_x is determined by

$$T_x = T_i \ln \frac{V_{DD}}{V_{DD} - V_x} \quad (3)$$

A quick approximation of 20-80% (or 10-90%) transition time can be calculated by substituting V_x with voltages at 20 and 80% of

VDD.

$$T_i \cong 0.72 \cdot T_{20-80} \quad (4)$$

The ratio coefficient of the transition time (20-80%) based on first order approximation is about 1.39.

2.2 Transition Time vs Elmore Delay

The resulting ratio of the transition time vs the Elmore delay of the above equation (4) would not be accurate enough for this repeater insertion method. The reason being is that the optimum distance between repeaters, and the size of the repeater itself that will satisfy the above design requirements have been determined by running Spice simulations. Therefore using the above method for correlating Elmore delay and transition time is inaccurate. The idea is to determine a limiting value for the Elmore delay, that corresponds to the maximum transition time obtained by running Spice simulations.

Once the optimum wire lengths/widths, buffer sizes and maximum transition time allowed for a design, have been determined, a comparison of the Spice waveform vs the above first order waveform approximation (2) can be performed. This can be accomplished by finding the ratio coefficient between transition time (measured in Spice for optimum wire length/width and buffer size) and the Elmore delay.

To find the optimum delay in our design, that will not affect the area and power, a set of Spice simulations were run. This was done for different buffer and wire sizes. Since area and power for our processor were strict limitations, the buffer and wire sizes have been chosen accordingly. As a limit for metal 1-4 wire width/space, a double the minimum width (2w) and a triple the minimum space (3s) was enough to satisfy our area and power goals. A limitation to buffer sizes was also necessary due to the fact that a number of these were placed in the channels (32 x minimum size). Placing large buffers in the channels and making wires unnecessarily wide would present significant blockage to the routing. This would have a significant impact on the area, due to wire congestion. Figure 1 shows the result of running Spice on different metal 1-4 wire and buffer sizes.

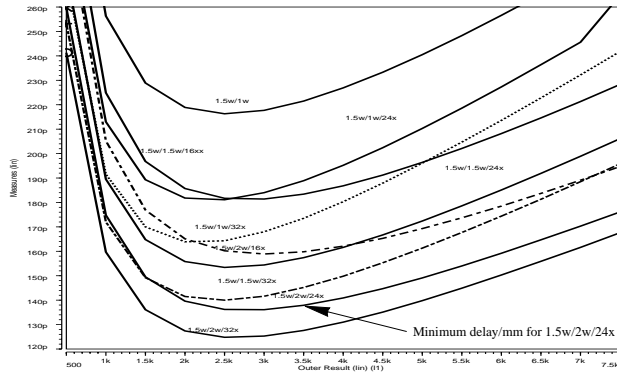


Figure 1. Delay per mm for 1.5w wire width, 1/1.5/2 w space, and 32/24/16 x minimum buffer size

As an observation, when the width or space between wires decreases the delay/mm increases. Different wire widths and spaces were chosen for our design based on needs. For instance, for critical nets that required optimum delays, wide widths and spaces were selected. In all the other nets, more conservative solutions were chosen such that area and power would not be affected. Table 1 shows some measurements of delay/mm and the

optimum length between repeaters of 24 and 32 x minimum size.

As an example, it was found that, for a metal 1-4 wire with 1.5 w width, 2w space and 24 x minimum size buffer, the minimum delay per mm (including repeater delay) occurs at 2.5 - 3.0mm between the chosen sized repeaters (see figure 1). Looking at the curve (figure 1), the delay per mm does not change much between 2.5mm and 3.5mm. Since our main goal is to insert as less repeaters as possible while still be close to the optimum delay, 3.5 mm looks like a better option than 2.5 and 3.0mm.

Table 1: M1 - M4 Repeater data

Repeater Size			24x		32x	
width (um) w - min	space (um) w - min	pitch (um) w - min	l (mm)	d/l (ps/mm) rise/fall	l (mm)	d/l (ps/mm) rise/fall
1.0w	1.2w	1.1w	2.68	187/191	2.85	176/184
1.5w	1.0w	1.2w	2.75	183/183	3.02	168/172
	1.5w	1.5w	3.24	156/157	3.51	145/150
	2.0w	1.7w	3.65	139/141	3.90	130/136

Using this method, we have reduced the number of repeaters by ~29%. As a result area and power are reduced considerably. But there is a limit to how long a wire can be in between two repeaters at a certain frequency. For our example, as shown in figure 2, the waveform dies out due to RC degradation as the length is increased. Making the wire longer than 3.5 mm will considerably affect the transition time to a point where the waveform does not reach the voltage ramps (VDD and VSS). The length that we have chosen initially (3.5mm) looks safe.

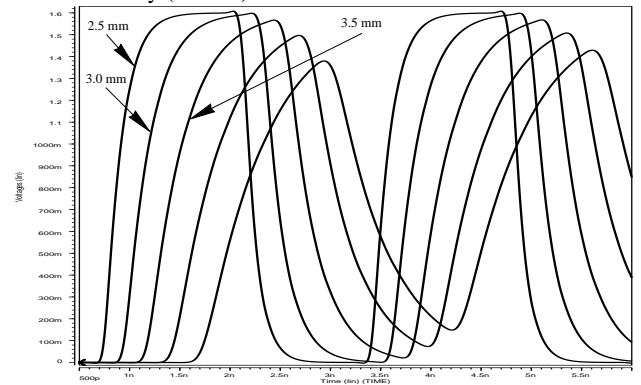


Figure 2. Transition time analysis

The voltage waveform corresponding to 3.5 mm is captured and shown in figure 3. Also the corresponding first order approximation (see equation (2)) of the same wire is overlapped on the same plot for comparison purposes. The Elmore delay constant T_i in (2) was calculated for the 3.5mm wire and used in the first order approximation. Thus Equation (2) becomes a simple exponential function of time. As can be observed, the waveforms almost match at the 50% of VDD point, but not at 20 and 80%

points.

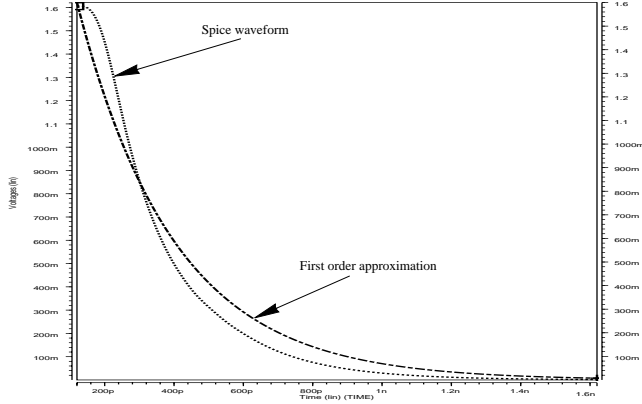


Figure 3. Spice waveform and first order approximation

To get the Elmore time constant that corresponds to the transition time at the load, a ratio of the Elmore delay and transition time was plotted with respect to the wire length (see figure 4). The ratio corresponding to 3.0 mm of wire is 0.98 and for 3.5mm is 1.01.

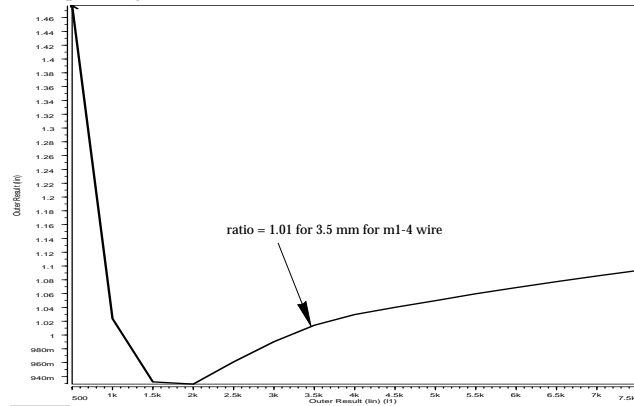


Figure 4. Ratio of Elmore delay and transition time vs length

It is interesting to notice that the ratio does not change much as the wire length between repeaters increases. The ratio seems to be in the range of ~0.9 - 1.1 for lengths greater than 1mm.

While we have only described the Elmore delay and transition time correlation only for point to point wires, this procedure can also be applied for more complex branching nets. Experiments showed that, for branching nets with total wire lengths similar to the ones described above, the ratio does not differ significantly from the above mentioned range (0.9 - 1.1).

3. Insertion Method

After the maximum transition time for a particular circuit performance has been determined, and the correlation with the interconnect Elmore delay has been performed, a maximum allowable RC delay is set as input to our algorithm. In addition an optimum repeater is determined by Spice simulations and a timing model is created. To obtain RC data from physical design, an RC extraction tool can be used, that can provide an SPF (Standard Parasitic Format) output to be used by the insertion tool. Standard Parasitic Format is a way to represent RC data and it is compatible with the Spice format. Block timing models are also needed to provide driver and load information.

In an SPF (Standard Parasitic Format) data, the RC nets are

represented as lumped RC models (see figure 5). To be able to insert repeaters, there is a need for transforming the lumped RC model into a distributed model.

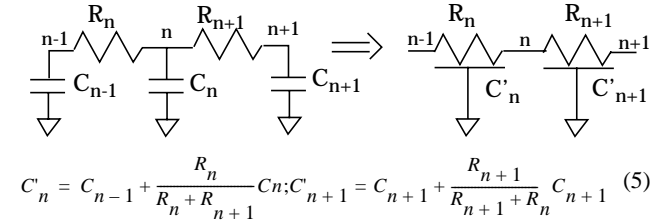


Figure 5. Lumped to distributed RC model transformation

Once the distributed capacitance has been calculated, the Elmore delay to each of the loads in the net from driver is calculated (figure 6). The repeaters are being inserted starting from the load with the highest Elmore delay towards the driver. While parsing through the RC net segment by segment, Elmore delay is calculated considering a repeater as driver of the segment. If the delay exceeds the maximum, a repeater is inserted in the segment, forming a new subnet. The process starts again from the highest Elmore delay load. In the case of branching, the assumed branching node is moved from an assumed repeater input to a repeater output. This continues until the maximum delay is exceeded, in which case a repeater is inserted with its output at the branching node (figure 7). The process continues until all the segments in the net have been explored.

<p>INPUT: SPF RC net and MAX_DELAY OUTPUT: SPF RC net with repeaters inserted</p> <ol style="list-style-type: none"> 1: Read net 2: Convert lumped into distributed RC model 3: Compute Elmore delay to each load (in subnet) 4: Identify load with largest Elmore delay 5: Identify unprocessed segment from the load 6: Calculate Elmore delay assuming repeater as driver 7: IF delay > MAX_DELAY { <ol style="list-style-type: none"> a: Interpolate to determine point of repeater placement in the segment b: Insert repeater, creating subnet with driver, connected segments and loads, counting repeater as load c: Reset data structures d: Go back to (3) 8: IF branching node present { <ol style="list-style-type: none"> a: Move an exiting branch from (assumed) repeater input to repeater output and find Elmore delay b: IF delay > MAX_DELAY { <ol style="list-style-type: none"> Move last branch back to repeater input Call (7b), (7c) and (7d) c: IF more exiting segments { <ol style="list-style-type: none"> Go back to (8a) 9: IF more segments on path { <ol style="list-style-type: none"> a: Go back to (5) 10: IF delay < f * MAX_DELAY { <ol style="list-style-type: none"> a: new_max_delay = MAX_DELAY + delay / (max # of repeaters in the path) b: IF new_max_delay > MAX_DELAY + p * MAX_DELAY { <ol style="list-style-type: none"> new_max_delay = Average Elmore delays c: MAX_DELAY = new_max_delay d: Go back to (7c) 11: Output repeater inserted nets

Figure 6. Repeater insertion algorithm

To optimize the total delay of the net with the repeaters inserted, there are some special cases where a repeater could be removed if

it is too close to the driver. This is presented as step (10) in figure 6. The process of elimination is being done by resetting the new maximum delay to a new value lower than the initial one plus some set margin error.

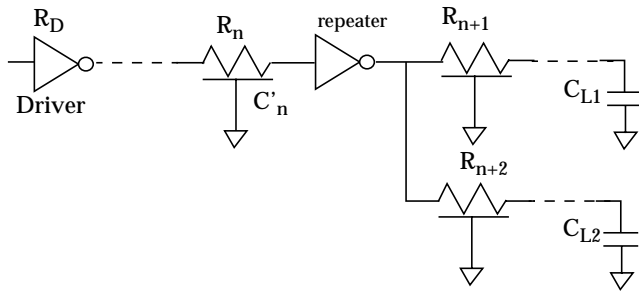


Figure 7. Net with repeater inserted

A Spice RC model having nodes with physical locations (X, Y coordinates) can be obtained by extraction tools providing standard parasitic format (SPF) data. This can then be used with the results of the above algorithm for repeater placement to determine the exact physical location desired for each repeater. This can be done by calculating the X, Y coordinates for the point at which the repeater is to be inserted along a segment extending between two nodes with coordinates given.

4. Implementation

The above method has been implemented in the design of Ultra Sparc III microprocessor at Sun Microsystems. A repeater map that shows the distribution of repeaters throughout the chip is shown in figure 9 (a similar map can be obtained which indicates the number of repeaters per square mm) as a result of running the repeater insertion program on SPF data with node coordinates. This is not to be considered as the final picture of where the repeaters are physically located in the chip, but rather the optimum locations from the timing point of view. Since placing the repeaters at those exact locations was not physically possible, a more realistic approach was to place blocks of repeaters in the very heavy areas, as shown in the map by the darker color indicating a large number of repeaters. The rest of the repeaters can be placed in the vicinity blocks.

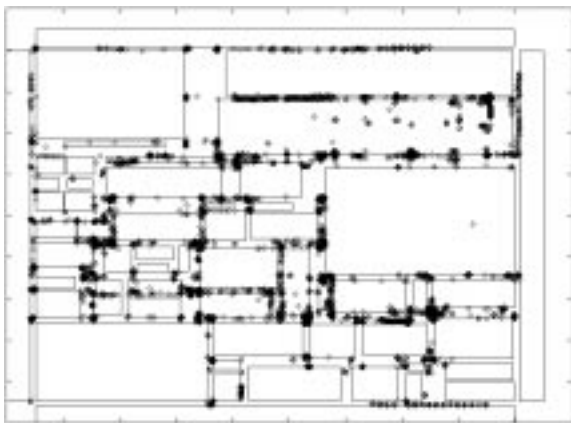


Figure 8. Repeater Map Distribution

After blocks of repeaters are placed in the floorplan in the designated locations (indicated by the dark color areas or by the number of repeaters per square mm) and the remaining ones in the logic blocks, nets are rerouted with a routing tool. Timing and distance constraints can be put on certain nets to enforce optimum routing, but even that will not give the needed results after the first shot. Several iterations are required through: Fullchip RTL, Routing, RC extraction, Fullchip Timing Analysis (mainly identifying the critical paths and long wires) and back, until a satisfying result is obtained in terms of area, power and timing.

5. Conclusion

This paper presents a practical repeater insertion method that satisfies the three main goals in designing a high speed microprocessor: delay, area and power. We have shown, step by step, how the method has been developed by first (a) identify the maximum wire length between repeaters and transition time by running spice simulations, (b) correlate the transition time with the Elmore delay based on the chosen wire lengths and repeaters resulting in a ratio coefficient, (c) use this coefficient to find the maximum Elmore delay to be used as an input to our repeater insertion program, (d) develop a simple and applicable algorithm that ensures equal transition time (limited by the frequency of the design), for all the gate input signals across the net and (e) a map of repeater distribution is to be obtained given a standard parasitic format (SPF) with node coordinates specified. The repeater insertion method has been implemented in the design of Ultra Sparc III microprocessor at Sun Microsystems.

6. Acknowledgments

As a closing statement, we would like to thank Fabian Klass, Farzad Chehrazai, and Tony Todesco from our physical design group for their help and support in the areas of circuit simulation and power analysis.

7. References

- [1] J. Lillis, C.-K. Cheng, T.-T. Lin, "Optimal Wire Sizing and Buffer Insertion and Low Power and a generalized Delay Model," *IEEE Journal of Solid State Circuits*, vol. 31, no. 3, March 1996, pp. 437-447.
- [2] J. Lillis and C. -K. Cheng, "Timing Optimization for Multi-Source Nets: Characterization and Optimal Repeater Insertion," *Proc. 34th Design Automation Conf.*, 1997, pp. 214-219.
- [3] L. N. Kannan, P.R. Suaris, H. -G Fang, "A Methodology and Algorithms for Post-Placement Delay Optimization," *Proc. ACM/IEEE Design Automation Conf.*, 1994, pp 327-332.
- [4] L. P. P. van Ginneken, "Buffer Placement in Distributed RC-tree Networks for Minimal Elmore Delay," *Proc. International Symposium on Circuits and systems*, 1990, pp 865-868.
- [5] M. A. Horowitz, "Timing Models for MOS circuits," Ph.D. thesis, Stanford Univ., Stanford, CA, Jan. 1984.
- [6] P. Penfield and J. Rubinstein, "Signal Delay in RC tree networks," in *Proc. 18th Design Automation Conf.*, 1981, pp. 613-617.
- [7] T. H. Cormen, C.E. Leiserson, R.A. Rivest, *Introduction to Algorithms*, 1990, MIT Press and McGraw-Hill Book Co., pp. 558-562.
- [8] W. C. Elmore, "The transient analysis of damped linear networks with particular regard to wideband amplifiers," *J. Appl. Phys.*, vol. 19, no. 1, pp. 55-63, 1948.