Boosters for Driving Long On-chip Interconnects: Design Issues, Interconnect Synthesis and Comparison with *Repeaters*

Ankireddy Nalamalpu
Intel Corporation
2111 NE 25th Ave
Hillsboro, OR 97124
analamal@ichips.intel.com

Wayne Burleson
ECE Department
University of Massachusetts
Amherst, MA 01002
burleson@ecs.umass.edu

ABSTRACT

Trends in CMOS technology and VLSI architectures are causing interconnect to play an increasing role in overall performance, power consumption and design effort. Traditionally, repeaters are used for driving long on-chip interconnects, however recent studies indicate that repeaters are using increasing area, power, and design resources as well as having an inherent limit in how much they can improve performance [4,10,14]. This paper presents a new circuit called a **booster** which compares favorably with repeaters in terms of area, performance, power and placement sensitivity. Boosters also have the advantage of being bidirectional and providing a low impedance termination to improve signal integrity. Driver edge rates are slower and peak power is drastically reduced compared to repeaters, thus improving signal integrity and mitigating inductive effects. Boosters are shown to be more than 20% faster for driving a variety of interconnect loads over conventional repeaters in 0.16 μm CMOS technology. Boosters are typically inserted three times less frequently than repeaters for optimal performance, resulting in fewer boosters for driving the same interconnect lengths thereby saving on area, power and placement effort.

Unlike differential, dynamic or low-swing techniques which require significantly more sophisticated circuit design and hence are cumbersome for automatic interconnect synthesis tools, boosters can be inserted on lines in a straightforward manner. Based on analytical delay models, we derive rules for insertion and sizing of boosters that can easily be incorporated into a CAD tool. We formulate two design rules that determine 1) the number of boosters needed, 2) their placements and 3) sizes, for driving a given interconnect load, first minimizing delay, and then area and power. Power analysis is slightly more complex than for repeaters so we present a systematic design approach. A placement sensitivity analysis comparing boosters and repeaters is used to study the effects realistic placement constraints that arise in floor-plans. We conclude by discussing various design trade-off's between repeater and booster based interconnect designs. Circuit simulations using a 0.16 μm CMOS technology are used to verify all analytical results.

ISPD'01, April 1-4, 2001, Sonoma, California, USA.

Keywords

Interconnect, Buffering, Timing, Methodology

1. INTRODUCTION

As the die size of CMOS integrated circuits continues to increase and feature sizes decrease, the performance of global communication is primarily limited by the interconnect delay. In general, the propagation delay of the interconnect increases quadratically with the line length[1]. As feature sizes shrink, the delay of the logic continues to decrease and the relative size of the chip tends to grow. Although [10] has shown approaches to a global wiring paradigm down to 50nm technology, future architectures and floorplans require new circuit-level solutions to the interconnect problem.

Repeater insertion is a classical solution which changes the delay dependence on the length from quadratic to linear. Repeater insertion and sizing tools are now commonly used as part of an overall interconnect synthesis approach [14]. Several methods have been proposed in [1-5] to find the optimal number of repeaters and their sizes needed for optimally driving the given interconnect load. In these repeater models, when the interconnect resistance is negligible compared to the driver resistance, increasing the drive strength of the repeater/driver, helps reducing the overall delay. However when the interconnect resistance is comparable to driver resistance, increasing the drive strength of the repeater/driver will no longer be effective in reducing the overall delay. So this approach of using repeaters to improve the performance of interconnect loses some of its advantages in deep sub-micron processes, because of significant interconnect resistance. In addition these repeaters contribute their own propagation delay to the signal delay, which limit the delay reduction that can be obtained. These conventional repeaters can not be used for bi-directional signal lines unless the repeater is replicated and appropriately enabled depending on the direction.

In [9], a self-timed complementary regenerative repeater(CRR) is presented for driving programmable interconnections in

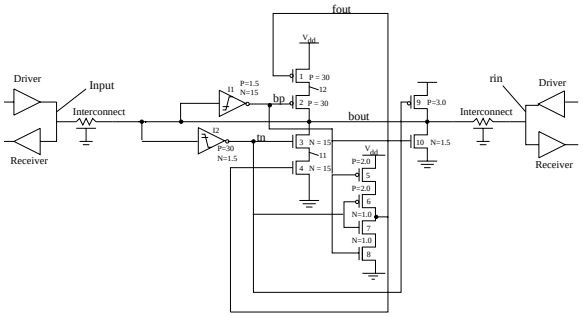

Figure 1: Booster Circuit

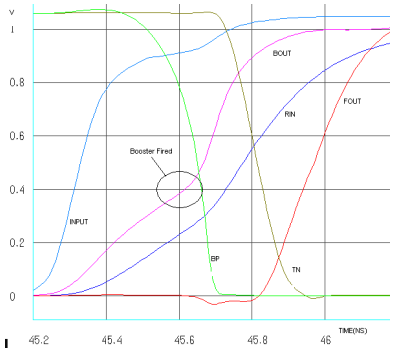

Figure 2: Booster circuit SPICE simulation in 0.16 μm CMOS technology

FPGAs. This design has not been used for driving on-chip interconnects in custom VLSI custom and ASIC circuits because of functional limitations and its inherent metastability. CRR poses strict limits on the minimum signal pulse width and regeneration time to maintain functionality. The design in [9] was designed for the heavily loaded programmable interconnections that exist in FPGAs but does not map well to the long interconnects that arise in advanced microprocessors or SoC ASIC circuits. In US Patents 5920210,5488322,5034623 and [8], different schemes are presented to improve the repeater performance by using dual switching points. Switching points are controlled by using additional pass transistors in the series stack as enabling elements. In all of the above designs, multiple MOS devices in the series stack in the critical path reduce the speed-up that can be obtained.

In this paper, we present a much simpler new circuit scheme that we coin as a Booster since it is does not repeat the signal but instead provides additional drive to the signal based on an early sensing of the transition. Like the CRR of [9], the Booster attaches along the wire rather than interrupting it and hence can be used for driving bi-directional signals. Thus Boosters do not add any explicit propagation delay except for device parasitics on the interconnect line, which we show to be negligible when compared with the interconnect load. Unlike typically inverting repeaters, Boosters don't impact the polarity of the signal. In section 2, the Booster circuit is presented, both its structure and its functionality verified with SPICE simulations in 0.16 μm CMOS technology. In section 3, Booster-design methodology for driving on-chip interconnects is outlined. As a part of this methodology, two simple rules are presented which will aid in determining the number of boosters needed and their placement to save power, area while maintaining significant speed-advantage over repeaters. In actual floorplans of microprocessor and ASIC designs, numerous constraints exist on the placement of repeaters and boosters on the interconnect lines [7]. Hence, in section 4, Booster placement sensitivity analysis is presented to study the sensitivity of booster performance to its placement, and compare these results with repeater placement sensitivity analysis [7]. In sections 5, various design parameters such as area, power, speed, noise and driving distance are presented for boosters and are compared qualitatively with repeaters. Some conclusions and future work are presented in section 6.

2. BOOSTER CIRCUIT

A transistor level schematic for Booster is shown in Fig.1. A patent has been recently filed for this circuit and several variations. Transistor numbers are indicated underneath each of the respective transistors. Typical transistors widths in 0.16 μm technology are indicated next to each of the transistors in Fig.1 (P=15 indicates that this is PMOS transistor and its width is 15 μm. Similarly N=15, signifies that this a NMOS transistor and its width is 15 μm). Inverter I1 is skewed to lower its switching threshold by sizing it in such a way that the ratio of NMOS/PMOS is equal to 10. Similarly inverter I2 shown in Fig.1. is skewed to raise its switching threshold by sizing it such a way that the ratio of PMOS/NMOS is far greater than 2.5 (the mobility ratio between holes and electronics). Transistors (5),(6),(7) and (8) will turn-on one of transistors (1), (4) in the steady state. If the interconnect steady-state voltage is V_{ss}, then the next future transition on the interconnect can only be rising edge. Transistors (5),(6),(7) and (8) will turn-on transistor (1) if the next future transition is rising edge. This ensures that the node (12) is charged in the steady state and transistor (1) is not on critical-path. Similarly if the next transition is falling edge, transistor (4) is turned-on in the steady state, there by ensuring the node (11) is discharged. Hence the the firing time of booster depends only on how fast one of the inverters I1, I2 respond depending on the next transition. In the steady-state both pull-up and pull-down paths of the driver are off(only one of the transistors in both pull-up and pull-down paths are active). This ensures that for the next transition on interconnect, there never exists a dc path between Vdd and Vss at any point of time , even though there exists skewed devices namely I1, I2 which respond to two different ends of the voltage transition. Transistors (9), (10) form a full-keeper and drive the interconnect in the steady-state. Neither of transistors (5), (6), (7), (8), (9) and (10) are on speed critical path and these transistors can be of small size.

Notice that transistors (5), (6), (7) and (8) will never turn-off booster prematurely as they are driven by inverters (I_1, I_2) which respond to opposite ends of voltage transition. Hence there is no minimum delay requirement on feed-back configuration formed through transistors (5), (6), (7) and (8). However, there is a maximum delay requirement and is determined by minimum pulse width on the line. In a glitch environment, skewed inverters I_1, I_2 can accidently be turned-on for short time. However, since transistors (5),

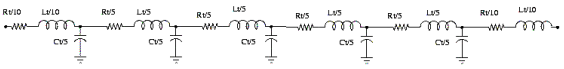

Figure 3: T-Model of the Interconnect

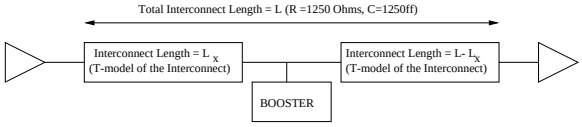

Figure 4: SPICE simulation model of booster

(6), (7) and (8) are driven by two oppositely skewed inverters I_1, I_2 $fout$ will never flip its state due to glitches. Hence feed-back configuration formed through transistors (5), (6), (7) and (8) reject glitches, there by making booster more immune to glitch and meta-stability environment. Subsequently any charge lost due to accidentally turning-on one of inverters I_1, I_2 for short-time will be recovered by full-keeper formed through transistors (9) and (10).

Transistors(4) and (1) will help minimizing short-circuit power. Unlike in repeaters, there will be virtually no short-circuit power between V_{dd} and V_{ss} path, formed through transistors (1), (2), (3), and (4). Immediately one notices that the booster input and output are tied together thus providing a high-gain positive feed-back loop. Booster in action is shown in Fig.2 and can be observed that once the booster is fired, the edge rate of the signal becomes much faster. As shown in Fig.2, $fout$ is not on the critical path and ensures that in the steady-state, appropriate nodes are pre discharged/charged to detect the next signal edge on interconnect faster.

3. BOOSTER DESIGN METHODOLOGY

In this section firstly we present the SPICE simulation results of driving the interconnect load using boosters, motivating the need for design methodology which can be integrated into an interconnect synthesis tool. We then formulate design rules that determine 1)number of boosters needed and their spacings, first minimizing delay, and then area and power.

A Simulation model for the booster is presented in Fig.4. Unlike the repeater, booster does not repeat the signal but attaches along the wire rather than interrupting it. Let L be the length of interconnect line being simulated. The total resistance and capacitance of the interconnect line of length L are $1250fF$ and 1250Ω respectively (on the order of .5 -1.1 cm depending on metal layer and geometry). Each of the virtually separated interconnect segments formed by attaching boosters are represented by 5 stage pi RLC section as shown in Fig.3.

SPICE simulation results in 0.16 μm CMOS technology of driving the interconnect load of 1250Ω and $1250fF$ by varying the number of boosters are presented below in Table.1. Boosters are placed uniformly along the line i.e each booster drives identical interconnect load.

From the above results, it can be observed that placing too many boosters (> 3) didn't hurt or improved the performance. This is in-contrast to the repeater based interconnect design, where performance degrades first slowly

C_L (pF)	R_L (kΩ)	Number of Boosters	Booster placements	Delay (ps)
1.25	1.25	1	$\frac{L}{2}$	546
1.25	1.25	2	$\frac{L}{3}, \frac{2L}{3}$	505
1.25	1.25	3	$\frac{L}{4}, \frac{2L}{4}, \frac{3L}{4}$	521
1.25	1.25	4	$\frac{L}{5}, \frac{2L}{5}, \frac{3L}{5}, \frac{4L}{5}$	532
1.25	1.25	5	$\frac{L}{6}, \frac{2L}{6}, \frac{3L}{6}, \frac{4L}{6}, \frac{5L}{6}$	536

Table 1: Delay results varying number of boosters in .16μm CMOS

but then significantly when too many/few repeaters (non-optimal) are used for driving a given interconnect load.

Placing too many boosters does increase the silicon area and power, even though it doesn't hurt the performance as much. In-order to integrate booster into an interconnect synthesis tool, we need a design methodology. Later in this section, two simple analytical rules are presented which can be used to find the number of boosters needed for driving the given interconnect line, first minimizing delay, then area and power. These analytical rules can easily be integrated into a CAD tool.

A simplified model of the interconnect with booster placed on the interconnect is shown in Fig.5. Only the relevant transistors in booster circuit are shown in Fig.5. Assume R_1, C_1 are resistance and capacitance of interconnect line of length L_1 and R_2, C_2 are resistance and capacitance of interconnect line of length L_2. Assume booster is placed at location L_1(node a) measured from driver-end and $L_2 + L_3$ distance measured from receiver-end as shown in Fig.5. Interconnect load is represented by pi-model, which is shown in [3] to be accurate enough for first-order approximation of the interconnect delay. Assume r_0 is resistance of the driver and C_{in} is the gate capacitance of the receiver as shown in Fig.5. V_a is interconnect voltage at node (a), location L_1.V_b is interconnect voltage at node (b), location $L_1 + L_2$.

3.1 Rule on Booster Placement

The following analysis is applicable for rising transition at node (a), and this could be extended for falling transition. For rising transition, pull-up stack formed by connecting PMOS transistors (1) and (2) in series should be turned-on so that rising voltage transition at node (a) can be improved. Assume sV_{dd} is switching threshold of inverter I_1 ($sV_{dd} << \frac{V_{dd}}{2}$). To place a booster at location L_1, inequality (1) needs to be met.

$t_{(V_a = sV_{dd})}$ is the time it takes for the voltage at node (a) to reach sV_{dd} from V_{ss}. $t_{(V_{bp} = \frac{V_{dd}}{2})}$ is the time it takes for the voltage at node (bp) to reach $\frac{V_{dd}}{2}$ after node (a) reached sV_{dd} and this will be equal to the delay of the inverter I_1 (Here delay of the inverter is measured from 50% output to $100s$% input, since the inverter I_1 is skewed to move its switching threshold from $\frac{V_{dd}}{2}$ to sV_{dd}). $t_{(V_a = \frac{2}{3}V_{dd})}$ is the time it takes for the voltage at node (a) to reach $\frac{2}{3}V_{dd}$ from V_{ss}, assuming booster is not fired (This delay is also equal to dominant time constant at node (a)). The left hand side expression in inequality (1), is the absolute time measured from signal transition at input, after which booster will be fired. If the above inequality (1) is not met, indicating that the absolute time it takes for the booster to fire is more than the dominant time constant, booster should not be placed

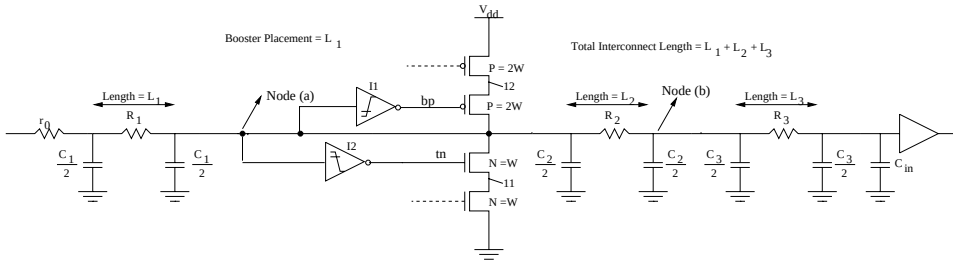

Figure 5: Analytical model of booster

at that particular location.

$$t_{(V_a=sV_{dd})} + t_{(V_{bp}=\frac{V_{dd}}{2})} < t_{(V_a=\frac{2}{3}V_{dd})} \quad (1)$$

$$t_{(V_a=sV_{dd})} = r_0[(C_1+C_r)+R_1(\frac{C_1}{2}+C_r)]ln(\frac{1}{1-s})$$

$$t_{(V_{bp}=\frac{V_{dd}}{2})} = \frac{(2w)k}{k_n V_{dd}}$$

$$t_{(V_a=\frac{2}{3}V_{dd})} = \tau = r_0[(C_1+C_r)+R_1(\frac{C_1}{2}+C_r)] \quad (2)$$

where $2w$ is the the width of PMOS transistor in the pull-up stack that is connected to the output of the inverter I_1. k is constant, represents gate cap per micron width. k_n represents the drive strength of NMOS transistor in inverter I_1. C_r is the extra capacitance added to the interconnect load due to booster and will be negligible when compared with interconnect load.

$$\frac{(2w)k}{k_n V_{dd}} < r_0[(C_1+C_r)+R_1(\frac{C_1}{2}+C_r)]ln(e(1-s)) \quad (3)$$

The above delay expressions doesn't take into consideration the shape of the input signal, since the delay is modeled by discrete RC elements. However, since the inequality (1) does not require absolute time measurement, both the left and right hand side of inequality (1) will be effected in the same fashion. Inequality (3) is accurate enough, and determines if the booster placement on the interconnect is warranted.

$\frac{\tau}{t_{lexp}}$	Delay w/o booster(ps)	Delay with booster(ps)	speed-up
≤ 1	98	96	2%
1.5	162	153	5.6%
2.0	205	179	12%
2.5	243	197	19%
3	302	224	27%

Table 2: Delay results varying booster firing time in .16μm CMOS

Let's assume, the left hand expression of inequality (1) is t_{lexp}. SPICE simulation results for varying values of $\frac{\tau}{t_{lexp}}$ in 0.16μm CMOS technology are shown in Table.2. $\frac{\tau}{t_{lexp}}$ values are varied to study the sensitivity of booster to input signal transition. Interconnect loads and sizes of inverters I1, I2 in booster circuit (Fig.1) are chosen such a way that $\frac{\tau}{t_{lexp}}$ can be varied over the desired range. Hence, in Table.2 absolute delay values are of no significance but only the speed-up. In Table.2, various values of $\frac{\tau}{t_{lexp}}$ are chosen to test sensitivity of booster with incoming signal transition time.

We can observe that when, for all values of $\frac{\tau}{t_{lexp}} < 2.0$, total speed-up due to booster is marginal (Interconnect model chosen is shown in Fig.4. Speed-up is measured as the delay ratio at the output of receiver with and with-out using booster). Hence, to save on power and area at the expense of marginal speed-up, it is recommended that boosters are placed at a location on the interconnect where $\frac{\tau}{t_{lexp}} > 2.5$. This also means fewer boosters for driving the interconnect loads, which saves significant power (section.5). From Table.2, it seems that speed-up that could be obtained keep increasing with increasing values of $\frac{\tau}{t_{lexp}}$. However, this does mean that this is not the optimal performance(absolute delay) that can be achieved with boosters and may be even the repeater based design is also not optimal. So, this should not be confused with the actual speed-up numbers, that will be later presented in section.5 which are obtained by comparing with optimal repeater interconnect design. Significantly higher values of $\frac{\tau}{t_{lexp}}$, indicates that more boosters needs to be placed along the way on the interconnect and more than one booster is required for optimally driving the interconnect, hence in the next section we present another rule, that guides on how frequently these boosters could be placed.

3.2 Rule on Number of Boosters

Let's assume a booster is placed at location L_x (assume at location L_x, $\frac{t_{lexp}}{T} = 3$ hence a booster is placed at L_x). Let's assume we are contemplating placing another booster at location L_y ($L_y > L_x$). After the booster is fired at L_x, to ensure that any booster that could be placed downstream i.e at L_y benefit from the booster that is placed up-stream i.e at L_x, the fast signal transition at L_x resulted due to booster at L_x has to propagate to L_y before booster at L_y could be fired. Mathematically the above can be represented as shown below (Assume L_x = location L_1, node a and L_y = location $L_1 + L_2$, node b as shown in Fig.5)

$$t_{(V_a=sV_{dd})} + [t_{(V_{bp}=\frac{V_{dd}}{2})}]_a + [t_{(V_{bp}=\frac{V_{dd}}{2})}]_b + t_{ext} \leq t_{(V_b=sV_{dd})} + [t_{(V_{bp}=\frac{V_{dd}}{2})}]_b \quad (4)$$

t_{ext} is the propagation delay between point L_a and L_b when booster at L_a is turned-on. $t_{(V_b=sV_{dd})}$ is the time it takes for the voltage at node (b) to reach sV_{dd} from V_{ss} when the booster at L_a is turned-off. $[t_{(V_{bp}=\frac{V_{dd}}{2})}]_a$, delay of inverter I_1 in booster placed at node a, can be approximated to be equal to $[t_{(V_{bp1}=\frac{V_{dd}}{2})}]_b$ (delay of inverter I_1 in booster

placed at node b).

$$t_{ext} \leq r_0 C_2 + R_1 C_2 + \frac{R_2 C_2}{2} - \frac{(2w)k}{k_n V_{dd}} \quad (5)$$

The above inequality (5) guides the number of boosters that could be placed on interconnect saving area, power. From the above inequality, we can observe that delay between two consecutive boosters should at-least be delay of inverters I_1, I_2 in booster circuit shown in Fig.1. It has been shown earlier in section.3 that placing too many boosters didn't hurt/improved the performance significantly. This now can be explained by the fact that if the boosters are closely placed, with out meeting the inequality (5), indicates that booster placed up-stream has no influence on the booster that is placed down-stream. Hence up-stream booster is redundant, as booster at down-stream is fired before the faster voltage transition resulted due to booster up-stream propagated to down-stream. Placing too many boosters didn't hurt the performance since the extra capacitance added to interconnect due to boosters is negligible when compared with interconnect capacitance. The above fact that placing extra boosters didn't hurt/improved the performance is very good characteristic of boosters unlike in repeater based interconnect design where placing non-optimal number of repeaters degrades the performance significantly. However, inequality (3) and (5) can guide to find minimum number of boosters that needs to be placed on given interconnect line saving area, power at the expense of losing only marginal speed.

It has been shown earlier in section.3 that placing too-many boosters didn't hurt the performance as much even though it costs area and power. In next section, we show that boosters have the very nice characteristic that they are less sensitive to placement variation. Hence, for all practical purposes, pi-model representation of interconnect load used for the analytical booster design methodology will suffice. However, if it warrants more accuracy, the in-equality's (3) and (5) can be derived by using alpha-power MOSFET model [6] to model short-channel effects in deep sub-micron CMOS. Unlike in elmore delay estimation, where transistor is represented by discrete RC elements, I-V equations of alpha-power model are used to find more accurate delay of an inverter driving interconnect load [5]. From [5], the in-equality (3),(5) can be re-written for a falling transition on the interconnect as follows

$$\frac{(w)k}{k_p V_{dd}} \leq \frac{C_1(bV_{dd} - V_{d0})}{I_{d0}} - ln(\frac{V_{dd}}{3V_{d0}})(R_{d0} + R_1)(C_1 + C_r) \quad (6)$$

$$t_{ext} \leq C_2[\frac{0.8V_{dd}}{I_{d0}} + \frac{V_{d0}}{I_{d0}} ln(\frac{5V_{d0}}{eV_{dd}}) + (R_1 + R_2) ln(\frac{5V_{d0}}{V_{dd}})] + \frac{(1-b)V_{dd}C_2}{I_{d0}} + C_1 R_2 ln(\frac{5V_{d0}}{V_{dd}}) - \frac{wk}{k_p V_{dd}}, R_{d0} = \frac{V_{d0}}{I_{d0}} \quad (7)$$

The above expressions are for falling voltage transition on interconnect and similar expressions can be obtained for rising voltage transition [5]. Assume switching threshold of inverter I_2 in Fig.1 is $bV_{dd}(> \frac{V_{dd}}{2})$. V_{do} is the drain saturation voltage at $V_{gs} = V_{dd}$ which is process dependent constant. I_{d0} represents the drivability of the MOSFET and is proportional to W/L, in this case it represents the drive strength of the driver. k_p represents the drive strength of PMOS transistor in inverter I_2.

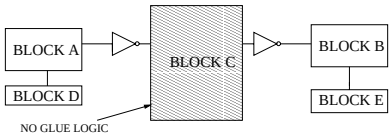

Figure 6: Inter-block Routing

4. BOOSTER PLACEMENT SENSITIVITY ANALYSIS

All of the design methodologies presented in [1-5] for optimally driving interconnects assume that there are no constraints on repeater/driver placement. However, in the practice of microprocessor and ASIC design, due to placement constraints, it is not always possible to place repeaters/boosters on the interconnect at the exact optimal locations. Let y represent the percentage deviation of the actual repeater/booster placement from its original optimal location. Here are some of the possible scenario's where placement constraints might exist on interconnect.

As shown in Fig.6, when a signal to be routed from block A to block B through block C, repeaters/boosters may not be allowed to place on block C. This might result in higher values of y. For signal integrity reasons to reduce the capacitive and inductive coupling, repeater placement on a bus are often intentionally staggered, which results in very high values of y (up to 50%). Staggering ensures that half of the interconnect length will have current ejection and the remaining half will have current injection so that total effective inductive coupling current will be close to zero. Similarly, in the case of Cache design, dedicated channels will be provided in each of the memory bank's for placing random logic such as latches, boosters, repeaters etc..This preserves the regularity of the design and ensures that the design manageable(e.g. verifiable, reusable).

It has been shown in [7] that, for repeater-based interconnect designs, performance will degrade by as much as 35%, for realistic floor plans(which have placement constraints, y can be as high 30%) arising out of VLSI microprocessor designs. Repeater-based interconnect design performance degradation is shown to have a square dependence on placement variation [7]. Hence, in this section, a placement sensitivity analysis of boosters is presented.

Booster simulation model presented in Fig.4 is used. Let L be the length of interconnect line being simulated. The total resistance and capacitance of the interconnect line of length L are $1250fF$ and 1250Ω respectively. Only one booster is assumed to drive the entire interconnect load. Let L_x be the booster location on the interconnect line measured from driver end. For example, L_x equals to $0.3L$, indicates that booster is placed at $0.3L$ measured from driver end and $0.7L$ measured from receiver end. Hence in this case, booster is placed more close to the driver than the receiver. Each of the two virtually separated segments of the interconnect resulted due to booster placement, are represented using RLC T-Model(Fig.3) in SPICE. The simulation results varying booster placement are shown in Table.3.

From Table.3, it can be observed that booster performance is relatively insensitive to its placement. The variation is less than 20% even for worst-case placement scenarios, when booster is placed close to the driver at 0.1L(placement variation is 40%) and farther away from driver. This is in-

C_L (pF)	R_L (kΩ)	Booster placement (L_x)	Delay (ps)
1.25	1.25	0.5L	546
1.25	1.25	0.4L	497
1.25	1.25	0.3L	494
1.25	1.25	0.2L	567
1.25	1.25	0.1L	635

Table 3: Delay results varying booster placement in .16μm CMOS

contrast with repeater based interconnect design, where repeaters shown to be highly sensitive to its placement [7]. Worst-case repeater placement scenarios, resulted in performance lose by as much as 40% for similar values of y, and performance degradation shown to have square dependence on placement variation [7]. From Table.3, it can also be seen that booster performance will be optimum when it is placed little closer to driver (0.3L - 0.4L) than placing at exact midpoint (0.5L). This can be explained as, due to finite speed of electromagentic waves, there will be finite lag in response along interconnect line due to voltage excitation at one-end of interconnect line and this lag will be maximum at the receiver. From elmore delay constants, we can find one of the inverters (I1, I2) in the critical path of booster circuit shown in Fig.1 can be fired earlier by placing the booster close to the driver end. This ensures that booster responds earlier to the voltage excitation at the driver end, improving its speed. When the booster is placed too close to the driver, this advantage will be offset by the the excessive interconnect distance that the resulted electro magnetic waves after firing booster, has to propagate before reaching the receiver.

5. BOOSTER VERSUS REPEATER FOR DRIVING INTERCONNECT

5.1 performance/Area

Silicon area and performance comparison between repeater based interconnect design and booster based interconnect design are presented below in Table.4. RLC T-model of the interconnect shown in Fig.3 is used to represent each segment of interconnect length.

Let n be the total number of repeaters each of size w (w is NMOS transistor width and PMOS transistor width in the repeater is assumed to be twice that of NMOS) are inserted to drive interconnect in a repeater based interconnect design. n_{opt}, w_{opt} be the optimal values for n and w respectively, for which the total delay of driving the given interconnect load is minimum. C_L, R_L are the total capacitance and total resistance of the interconnect load respectively. $n_{booster}$ is the number of boosters inserted for driving the given interconnect load. $D_{booster}, D_{repeater}$ are the delays of driving the interconnect load using boosters and repeaters respectively. $W_{booster}, W_{repeater}$ are total device widths of the transistors in boosters and repeaters respectively($W_{repeater} = n_{opt} * W_{opt} * 3$). Closed-form expressions for n_{opt} and w_{opt} are presented in [5] are shown below

$$n_{opt} = \sqrt{\frac{R_L C_L K_i K2}{K_g K1}} \quad (8)$$

$$W_{opt} = \sqrt{\frac{C_L K1}{R_L K_i K_g K2}} \quad (9)$$

K_i,K_g,$K1$ and $K2$ are process dependent constants. Both the analytical rules outlined in the section.3. are used to determine the number of boosters required and their placements to drive the interconnect loads in Table.4., saving area, power with out losing significant performance. $D_{booster}$ is the simulated delay value of driving the interconnect load using the number of boosters and their placements obtained from two analytical rules presented in section.3. Interconnect model in Fig.3, where 5 pi RLC representation of interconnect is used to simulate boosters in o.16μm CMOS technology.

Table.4, compares design trade-off's between booster and repeater based interconnect designs for driving variety of interconnect loads. For repeaters, optimal repeater insertion is chosen, where performance of driving the interconnect is maximum. Boosters are optimized first for delay, then area/power. Speed-up numbers in Table.4 show that repeaters can't be used to improve performance beyond some point. It has been shown in [10] that optimal repeaters (maximum performance) are often very expensive in area/power and that non-optimal repeaters save silicon area/power. In Table.4, optimal repeater insertion is chosen to find worst-possible speed-up that could be obtained by using boosters. Hence, in case of non-optimal repeater insertion, boosters speed-up numbers in Table.4 will improve greatly furthur, and area/power of repeaters will decrease. To obtain similar speed using both repeaters and boosters, booster-based interconnect design still requires less silicon area as fewer boosters are required.

From the above results, we can observe that boosters has longest driving distance and can drive approximately thrice as much as the driving distance of repeaters. Hence, number of boosters required to drive the same driving distance is one-third of repeaters that are needed to drive the same distance. As shown above a significant reduction in area is obtained with boosters while doing better in performance than repeaters. Speed-up of up to 28% is obtained with boosters when compared with the optimal performance that could be obtained with repeaters. From the Table.4, we can also observe that better speed-up using boosters is obtained for highly resistive interconnects, where repeaters tend to under-perform because of the shielding offered by resistance.

5.2 Power

Power consumption has become one of the premier issues in VLSI design. There are two primary components of the total power dissipation: dynamic power and short-circuit power. Dynamic power dissipation can be quantified by CV^2f,its estimation is straight forward. On the other-hand, short-circuit power is often neglected since dynamic power is assumed to be dominant. In modern processes, especially in the interconnect design short-circuit power is a significant portion of total power due to slow transitions time resulting due to larger interconnect loads. For given supply voltage and frequency, dynamic power dissipation depends only in the load capacitance and does not depend on the input

C_L (pF)	R_L (kΩ)	$n_{booster}$	n_{opt}	W_{opt} (μm)	$D_{booster}$ (ps)	$D_{repeater}$ (ps)	speed-up	$W_{booster}$ (μm)	$W_{repeater}$ (μm)
1.0	1.0	2	5	24	349	434	19.2%	300	360
1.25	1.25	2	6	24	451	567	20.6%	300	432
4.0	1.0	4	11	46	688	829	17%	750	1518
2.0	5.0	5	17	16	933	1287	27.6%	750	816
20	1.0	9	22	102	1380	1792	23.1%	1350	6732

Table 4: Performance and area comparisons for varying interconnect loads in .16μm CMOS

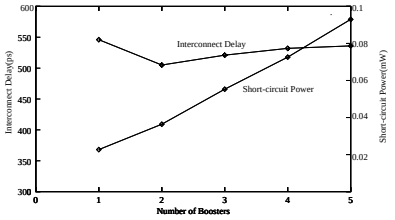

Figure 7: Booster Delay and Short-circuit Power for Interconnect load of $1250ff$ and 1250Ω in 0.16μm CMOS technology

waveform shape and output resistance. In contrast, short-circuit power is both load and signal dependent [11-13] and its measurement is not straight forward.

In the booster circuit shown in Fig.1, transistors (1) and (4) present to ensure that short-circuit power is minimized. There never is a direct path between V_{dd} and V_{ss} through transistors (1), (2), (3) and (4) even though there exists skewed inverters responding to the opposite ends of voltage transition which are feeding directly transistors (2) and (4). In the steady-state, both pull-up and pull-down paths formed through transistors (1), (2), (3), (4) are turned-off (only one device in each of pull-up and pull-down stacks are turned-on). If the voltage transition on interconnect is a rising edge, short-circuit power is minimized by turning-off transistor (4) in steady-state even though transistor (3) is fully-on, driven by inverter I_2. If the voltage transition on interconnect is a falling edge, short-circuit power is minimized by turning-off transistor (1) in steady-state even though transistor (2) is fully-on, driven by inverter I_1.

Both inverters I_1 and I_2 output capacitance is very small, driving only transistors (2) and (3) respectively. Hence, inverter I_2 output rise time is substantially smaller than input fall time. So, there will be maximum short-circuit current for most of the transition period, equal to saturation current of PMOS in inverter I_2 during falling voltage transition on interconnect, while short circuit current in I_1 is minimum for same transition. Similarly, inverter I_1 output fall time is substantially smaller than input rise time. There will be maximum short-circuit current for most of the transition period, equal to saturation current of NMOS in inverter I_1 during rising voltage transition on interconnect, while short circuit current in I_2 is minimum for same transition. Short-circuit power dissipation for the paths between V_{dd} and V_{ss} through transistors (5,6,7,8) and (9, 10) is negligible, since the transistors are of smaller size and for both of these paths, input moved through the transient region before output starts to change.

Overall, single booster short-circuit current is high when compared with repeater, where the repeater is just an inverter consisting of only one PMOS and one NMOS transistor. However as shown earlier, the driving distance of a booster is approximately thrice that of repeaters. Hence, for driving a given interconnect load, the number of repeaters required will be at least thrice that of boosters. Hence, over-all for driving a given interconnect load, short-circuit current of booster-based interconnect design is comparable to repeater-based interconnect design. Mathematical estimation of short-circuit power dissipation of the booster is very complicated unlike that of repeater short-circuit power [13].

As shown in Fig.7, interconnect delay is relatively flat when too-many (> 3) boosters are placed, short-circuit power consumption grows linearly with number of boosters placed. By applying rules (3) and (5) which are formulated to minimize the over-all power consumption while minimizing the impact on performance, we can find from Table.4 that driving an interconnect load of 1250Ω and $1250ff$ requires just 2 boosters. SPICE simulation in 0.16 μm CMOS technology shows that the total short-circuit current I_{SC} is $4.8(mA)(nS)$ for optimally driving 1250Ω and $1250ff$ interconnect load using repeaters(From Table.4, for driving optimally the above interconnect load requires 6 repeaters, with each repeater NMOS transistor size is 24 μm). Only two boosters are required for driving an identical interconnect resulting in a total short-circuit current of $2.8(mA)(nS)$. Short-circuit power dissipation is computed by $I_{SC}V_{dd}f$. If the switching frequency(f) of the above interconnect is 10MHz and $V_{dd} = 1.3$V, short-circuit power dissipation using a booster-based interconnect design is 36.4μW as opposed to 62.4μW for the optimal repeater-based interconnect design. However, power dissipation in repeater-based design can also be reduced, by moving away from optimal performance [10]. If non-optimal repeater insertion is used, booster speed-up over repeaters will greatly be improvised than in Table.4. and this will also result in less area/power of repeaters. As pointed out in section 5, the point of comparison in Table.4 is primarily to point out what is the worst speed-up that boosters can be obtained there by showing how boosters can be used to drive faster than optimal repeaters and to compare the over-head cost to obtain similar speed using both schemes. It seems that by inserting more repeaters, each repeater drives less interconnect load, thereby improving signal transition time, hence short-circuit current for each individual repeater will decrease but might increase the over-all power consumption of the system, since now there are more repeaters for driving the same interconnect load. If each repeater drives more interconnect load, there by decreasing total number of repeaters for the system, each repeater short-circuit power will increase but might decrease the over-all power of the system. In general, dynamic power consumption increases linearly with increasing device width

and short-circuit power dissipation grows non-linearly with increasing device width. Rules (3) and (5) are formulated to minimize the over-all power consumption while minimizing the impact on performance.

6. CONCLUSIONS

A novel circuit technique called Booster is presented for driving long on-chip interconnect lines in VLSI circuits. Boosters have an advantage that, they don't contribute their own propagation delay to the signal delay and they can be used for driving bi-directional signals. SPICE simulation in 0.16 μm CMOS technology show that boosters are more than 20% faster for driving variety of interconnect loads over repeaters. Unlike in repeater-based interconnect design, booster placement sensitivity analysis shows that booster possesses a nice characterstic that they are relatively in-sensitive to placement variation and placing non-optimal number of boosters doesn't hurt the over-all performance. This makes them extremely attractive for worst-case placement scenarios resulting in realistic VLSI floor-plans. We presented as a part booster-design methodology two rules that will help determining number of booster needed and their placements for driving given interconnect distance, saving area/power while obtaining significant speed-up over repeaters.

They are 1) Rule guiding the placement of boosters, which compares the firing time of booster with dominant RC time constant. Results show that booster firing time should be at-least 2.5 times less than dominant RC time constant for placing a booster at a point on interconnect. 2) Rule guiding the number of boosters that should be placed to obtain significant speed-advantage over repeaters, with-out wasting area/power. This rule indicates that the delay between two consecutive boosters should at least be the absolute time it takes for booster to fire. Circuit simulations in 0.16 μm CMOS technology shows that boosters interconnect driving distance is approximately thrice as much that of repeaters, resulting in requiring fewer boosters for driving identical interconnect length there by saving on area and power.

Repeaters still have a place and the limitations of boosters need to be recognized. Sometimes it is desirable to actually break the line for buffering reasons. This can be to limit long lines for inductive reasons, however it appears that boosters also mitigate inductive effects, by providing a low-impedance termination along the line. However reflections could also be at booster so a thorough transmission line analysis would be required. Repeaters can also be combined with latches, multiplexors and even logic, unlike boosters. Ultimately we expect that repeaters and boosters will both be used to generate the sophisticated interconnect systems required for gigascale integration.

7. FUTURE WORK

Noise analysis, Design guidelines, Electrical rule checking (i.e MOSCritic), Timing analysis (how will a static timing analyzer account for boosters), Integration into interconnect synthesis tool with repeaters and Impact on bidirectional multi-source lines could dramatically impact VLIW, FPGA, routers, multi-processor, memory and other highly connected architectures

8. REFERENCES

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

[2] H.B. Bakoglu and J.D.Meindl., "Optimal Interconnection Circuits for VLSI," *IEEE Trans. Electron Devices*, vol. ED-32, pp 903-909, May 1985.

[3] M.Nekili and Y.Savaria., "Parallel regeneration of interconnections in VLSI & ULSI circuits,"*Proc.IEEE Int. Symp. Circuits and Systems*, May 1993, pp2023-2026

[4] V.Adler and E.G.Friedman.,"Repeater Design to Reduce Delay and Power in Resistive Interconnect," *IEEE Trans. on Circuits and Systems II*, Vol.45, pp 607-616, May 1998

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

[6] T.Sakurai and A.R.Newton., "Alpha-Power Law MOSFET model and its applications to CMOS inverter delay and other formulas,"*IEEE Journal of Solid-State Circuits*, vol. 25 No. 2, pp.584-595,Apr.1990

[7] A.Nalamalpu and W.Burleson., "Quantifying and Mitigating the effects of Repeater Placem ent Constraints on Interconnect Performance," submitted to *IEEE Trans. on VLSI Design*, 2000

[8] Iima et al, "Capacitance Coupling Immune, Transient Sensitive Accelerator for Resistive Interconnect Signals of Subquarter Micron ULSI, " *Proc.IEEE Trans. on Electron Devices*, July 1995, pp 940-945

[9] Dobbelaere et al, "Regenerative Feedback Repeaters for Programmable Interconnections," *IEEE Journal of Solid-State Circuits*,Nov 1995, pp.1246-1253

[10] D.Sylvester and K.Keutzer., "Getting to the bottom of deep sub-micron II: A global paradigm," *Proc. of IEEE Int. Symp. Physical Design*,1999, pp.193-200

[11] A.M.Hill and S.M.Kang, "Statistical Estimation of Short-circuit Power in VLSI Systems, " *Proc. of IEEE Int. Symp. Circuits and Systems*, pp.4.469-4.471, May 1996

[12] Bisdounis et al, "Modeling the CMOS Short-circuit Power Dissipation," *Proc. of IEEE Int. Symp. Circuits and Systems*, pp.4.489-4.491, May 1996

[13] H.J.M. Vendrick , "Short-circuit Dissipation of Static CMOS Circuitry and its impact o n the Design of Buffer Circuits," *IEEE Journal of Solid State Circuits*, Vol.SC-19, pp.468-473, Aug.1984

[14] J.Cong and D.Z.Pan, "Interconnect Estimation and Planning for Deep Submicron Designs" *Design Automation Conf.*, New Orleans, LA., June, 1999, pp. 507-510