Wave-Pipelined 2-Slot Time Division Multiplexed (WP/2-TDM) Routing

Ajay Joshi
Georgia Institute of Technology
School of ECE
Atlanta, GA 30332-0250
Tel No. 1-404-894-9362
joshi@ece.gatech.edu

Jeffrey Davis
Georgia Institute of Technology
School of ECE
Atlanta, GA 30332-0250
Tel No. 1-404-894-4770
jeff.davis@ece.gatech.edu

ABSTRACT
The ever-increasing number of transistors on a chip has resulted in very large scale integration (VLSI) systems whose performance and manufacturing costs are driven by on-chip wiring needs. This paper proposes a low overhead wave-pipelined two-slot time division multiplexed (WP/2-TDM) routing technique that harnesses the inherent intra-clock period wire idleness to implement wire sharing in combination with wave-pipelined circuit techniques. It is illustrated in this paper that WP/2-TDM routing can be readily incorporated into future gigascale integration (GSI) systems to reduce the number of interconnect routing channels in an attempt to contain escalating manufacturing costs. Two case studies, one at the circuit level and one at the system level, are presented to illustrate the advantages of WP/2-TDM routing. The circuit level implementation exhibits more than 40% reduction in wire area, a 30% reduction in silicon area with no increase in dynamic power and no loss of throughput performance.

Categories and Subject Descriptors
C.5.4 [**Computer System Implementation**]: VLSI Systems

General Terms
Performance, Design

Keywords
interconnect sharing, wave-pipelining, time division multiplexing, wire area, on-chip interconnects

1. INTRODUCTION
Due to the continuous increase in the number and complexity of global and semi-global interconnects in modern day VLSI systems, ASIC and microprocessor performance is being increasingly restricted by interconnect area, delay, and noise [1], [2]. Consequently, there has been an increase in the number of metal layers for every new technology generation [3] that results in a non-trivial increase in manufacturing cost. It is, therefore, imperative to investigate VLSI interconnect design and implementation methodologies that most efficiently utilize the available wiring tracks in a multilevel wire network. This is especially true in an era when more and more high-speed global wires are flanked on both sides by power and ground lines to control inductive effects.

A variety of techniques have been proposed in an attempt to more efficiently use the wire channels in system-on-chip (SoC) designs. For example, references [4]-[6] discuss the various aspects of the network on chip (NoC) paradigm that controls data exchange between the various intellectual property (IP) cores in an SoC. In particular, authors in [7] and [8] use a multi-slot TDM technique for communication between the different cores of the system. In the cases mentioned above, even where TDM methodologies are used, a significant amount of overhead circuitry and microarchitectural change to the system are necessary. In contrast, [9] explores the question of network complexity, and suggests that a simpler level of network complexity can still provide significant benefits with a small amount of overhead. [9] uses system-level interconnect prediction (SLIP) methods to explore the impact of using simpler 2-slot TDM networks, and concludes that a simpler implementation could still significantly reduce wire area.

This paper proposes a new circuit technique that combines both wave-pipelining and 2-slot time division multiplexing (WP/2-TDM) to produce an interconnect routing technique that can be seamlessly incorporated into existing global and semi-global pipelines. Because of the relative ease of incorporation of this technique into a traditional VLSI design flow, this implementation has the potential to be a ubiquitous routing technique that can be applied to both inter-core and intra-core interconnects in any SoC or microprocessor design. To explore WP/2-TDM routing this paper is organized as follows. Section 2 gives a detailed description of the wave-pipelined 2-TDM routing technique. Section 3 describes and provides preliminary verification of the circuit implementation. Two case studies exhibiting the advantages and ease of application of this wire sharing technique are presented in Section 4.

2. WAVE-PIPELINED 2-SLOT TIME DIVISION MULTIPLEXED (WP/2-TDM) ROUTING
The application of wave-pipelined 2-TDM routing (WP/2-TDM) is primarily driven by the existence of both wire idleness and physical proximity among interconnects.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
GLSVLSI'05, April 17–19, 2005, Chicago, Illinois, USA.
Copyright 2005 ACM 1-59593-057-4/05/0004...$5.00.

This material is based upon work supported by the National Science Foundation under Grant No. 0092450.

2.1 Interconnect idleness opportunities

It is assumed that all interconnects on a tier have approximately the same wiring pitch, and this pitch is proportional to the length of the longest interconnect on that tier [10]. A tier in this paper is defined as a pair of orthogonal routing levels with the same pitch. A result of this assumption is that the shorter interconnects on a particular tier require less than the allotted time period for signal transmission. Hence, shorter interconnects on the semi-global or global tiers, which are not in a critical path, remain idle during part of the clock period. The WP/2-TDM technique takes advantage of this wire idleness and sends one additional data signal during this idle period.

To illustrate the amount of wire idleness that is present in a current system, a system level simulator similar to [10] is used to simulate a 40M-transistor logic core that is implemented in 0.1µ technology with a 1.3 Ghz clock and a 1.2cm^2 core area. Figure 1 shows interconnect delay normalized to the clock period for all wire lengths on different wire tiers of this simulated logic core.

It can be observed from Figure 1 that the multilevel interconnect network has been designed such that the longest interconnect on each tier requires a maximum of 80% of the clock period for data transfer from source to sink. The extra 20% of the clock period accounts for clock skew and provides necessary guardband to ensure a robust transfer of data from source to sink. It can be calculated that 67% of wires with length greater than 0.1mm require less than 60% of the available clock period for data transmission. WP/2-TDM routing takes advantage of the resulting idle time and sends a second signal during the idle portion of the clock cycle in a wave-pipelined fashion.

A modified wave-pipelining technique similar to [11] is adopted for sending multiple signals. [11] gives an expression for calculating the minimum sustainable pulse width (t_{pulse}) that can travel along a repeater interconnect circuit without any loss of signal integrity. In WP/2-TDM routing, two signals are transmitted during one clock period. The first signal is scheduled at the beginning of the clock period and the second signal is scheduled after t_{pulse} seconds. Both signals will arrive at the respective sinks within a single clock period as long as:

$$\frac{t_{latency} + t_{pulse}}{t_{clk}} \leq 0.8 \qquad (1)$$

where $t_{latency}$ is the 50% latency of the wire channel and t_{clk} is the clock period. The condition in (1) ensures that the second signal reaches the appropriate sink before the end of the current clock period. Figure 2 shows the plot of the left hand side of (1) for different wire lengths in the 40M transistor logic core described above. In addition, the corresponding stochastic interconnect demand function [12] for this system is also plotted as a function of wire length. The shaded regions in Figure 2 illustrate the range of interconnects to which the WP/2-TDM routing technique can be applied without any loss of throughput or latency performance. In case of the longer interconnects that do not satisfy the delay constraints given by (1), the WP/2-TDM technique is further modified. Even in this case, the first and second signals are sampled and transmitted at the beginning of the clock cycle ($t = 0$) and at $t = t_{pulse}$ respectively. However, both the signals do not reach the appropriate sinks within one clock cycle and hence, they are available to the receiver side circuitry after $t = t_{clk}$ (i.e. during second clock cycle). Since, we have assumed that all the circuits

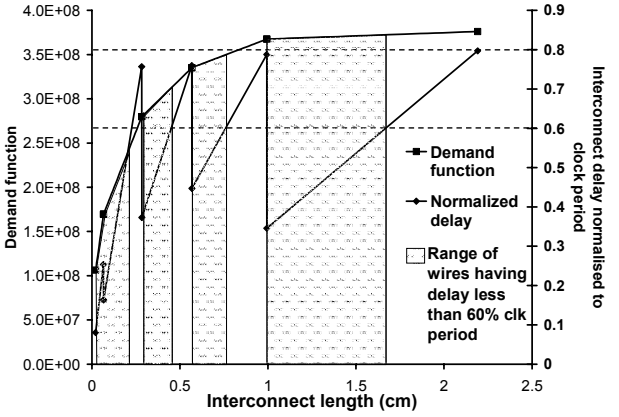

Figure 1. Interconnect Delay normalized to clock period for different interconnect lengths and stochastic wire distribution.

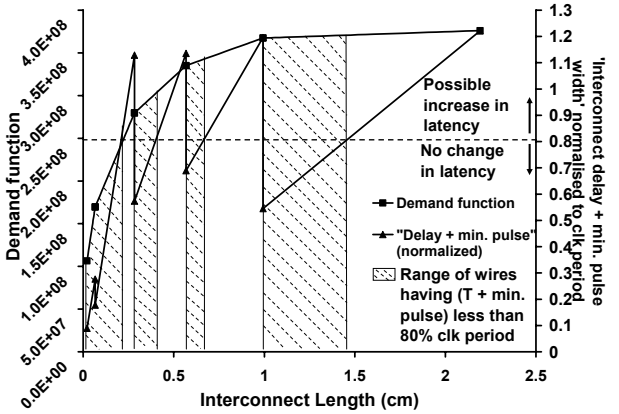

Figure 2. 'Interconnect delay + minimum sustainable pulse width' normalized to clock period for different interconnect lengths and stochastic wire distribution.

of our system sample data at the beginning of the clock period, the data sent at $t = 0$ and $t = t_{pulse}$ can be used only at $t = 2 * t_{clk}$. As a result there is an increase in the signal latency.

However, even if the first set of signals do not reach their respective sinks at $t = t_{clk}$, the second set of signals can be scheduled at $t = t_{clk}$ without losing signal integrity. The second set of signals will reach the respective sinks in the third clock cycle and by that time the first set of signals would have already been used by the receiver side circuitry. Thus, the second set of signals can be used at $t = 3 * t_{clk.}$ Therefore, signals can be transmitted at the source side in every clock cycle and be sampled at the sink side in every clock cycle, and the overall throughput performance of the system is maintained. Since the latency is two clock cycles for this case, the shared interconnect can have total delay of 1.8 * t_{clk} (remaining 0.2 * t_{clk} for clock skew and guardband). Hence, the timing constraint in (1) can be relaxed and both the signals would safely reach the appropriate sinks as long as:

$$\frac{t_{latency} + t_{pulse}}{t_{clk}} \leq 1.8 \qquad (2)$$

Initially, the interconnects were designed such that they would have a maximum delay of 0.8 * t_{clk}. However, under the new constraint in (2), t_{pulse} and $t_{latency}$ can be larger. This provides an opportunity whereby it might be possible to reduce both silicon *and* wire area.

2.2 Source-sink and run length proximity

The possibility of using WP/2-TDM routing instead of dedicated routing is also determined by the physical placement of a given source-sink pair, or the existence of shared run length between two interconnects. A wide variety of routing configurations of WP/2-TDM technique could be used in both regular and irregular routing. Consider two wires that have both sources and sinks close to each other. For application of WP/2-TDM routing, the two sources should be at a distance less than 'r' from each other and so should be the sinks. Here the distance 'r' is proportional to the average of the two wire lengths and is chosen such that deviation in routing will have minimal impact on delay.

On the other hand, two interconnects of unequal lengths, could have shared run length and be at a distance less than 'd' from each other. Here, the distance 'd' is proportional to the length of the longer interconnect. In this case, one can replace the shorter interconnect and part of the longer interconnect by a shared wire. The source-sink pair of the shorter wire will transfer data over the shared wire, while the data that was transmitted over the longer dedicated wire earlier will now be transmitted partially over the shared wire and partially over the dedicated wire. Here, the interconnects can be of equal lengths too. As long as they have some shared run length, one can replace whole or part of the two interconnects by a single shared interconnect.

3. CIRCUIT DESIGN AND TIMING ISSUES

Figure 3a and 3b show the schematic diagram of the circuitry required for conventional routing and WP/2-TDM routing respectively. Pipeline registers are used at the source and sink side in both the routing techniques for data storage. For conventional routing, a driver, receiver and suboptimal number [10] and suboptimal size [13] of repeaters are used. Each repeater consists of an inverter pair. For WP/2-TDM routing a 2:1 multiplexer and a 1:2 demultiplexer are placed at the input and output, respectively, of the shared wire. Buffers are placed at the receiver side to ensure that the integrity of the first data signal is maintained while the second data signal is being sampled.

The signals from the two different sources are given as input to the two input lines P0 and P1 of the multiplexer, respectively. A signal (φ_{min}) having cycle period equal to global clock cycle and which remains at logic 1 only for $t = t_{pulse}$, calculated using [11], is given as input to the select line of the multiplexer. When φ_{min} is high, (beginning of the clock cycle; $t = 0$) input at P0 is sampled by transmission gate A and transmitted over the shared interconnect while on low φ_{min} ($t = t_{pulse}$), the input at P1 is sampled by transmission gate B and transmitted.

At the receiver end, φ_{min} is delayed and this delayed signal is used for sampling the data received on the shared wire. $\varphi_1$and φ_2 are the signals given to the nFETs of transmission gates C and D respectively, while, Line_out gives the signal transmitted over the shared interconnect and given as input to the demultiplexer on the receiver side. Figure 3b shows two delay circuitries at the receiver side. The delay circuitry 1 delays the signal φ_{min} to give φ_1 such that signal P0 gets sampled by transmission gate C as soon as it reaches the input (Line_out) of the demultiplexer. The second signal P1 follows P0 on the shared wire with a time difference of t_{pulse}. Hence, the delay circuitry 2 further delays φ_1 to give φ_2 such that transmission gate D samples signal P1 at the appropriate time. It should be noted that only one of the two transmission gates C and D is on during sampling of signals received on the shared interconnect. These delay circuitries will be shared among multiple shared interconnects to distribute the resulting overhead. Buffers are used at both the outputs of the demultiplexer to maintain signal integrity, and to hold the received value dynamically. It is assumed that the necessary shielding mechanism has been used for the delay circuitry, in order to prevent any crosstalk noise. In addition, a study of the leakage current of receiver side circuitry confirms that the large transistors and the high data transmission rate prevent any loss of data on the dynamic nodes due to leakage.

Figure 4 shows the timing waveforms, generated using HSPICE, for the two data signals sent over a 0.7cm long shared interconnect. A pitch of 1.05e-4 cm is used for this interconnect. The pitch value is selected based on the interconnect network design obtained for the 40M-transistor logic core described in Section 2. A bit stream of 0110010 and 0110110 are given as input to P0 and P1 respectively. When φ_{min} goes high, the transmission gate A samples and transmits the signal at P0 over the shared interconnect. When φ_{min} goes low, the input signal at P1 is sampled and transmitted by the transmission gate B. At the receiver side, whenever, φ_1 is high, transmission gate C samples the data at the input of the demultiplexer (Line_out) and gives it as output at OP0. At this time transmission gate D is cutoff. When φ_2 goes high, transmission gate D samples and transmits data on the shared wire. This corresponds to signal at OP1. It can be observed from Figure 4 that both input signals at P0 and P1 reach the appropriate sinks within one clock cycle.

For interconnects that do not satisfy the delay constraint in (1) but exhibit source-sink or run length proximity the same circuit in Figure 3b is used. As explained in section 2, the latency of the signals is two clock cycles and the constraint in (2) is used.

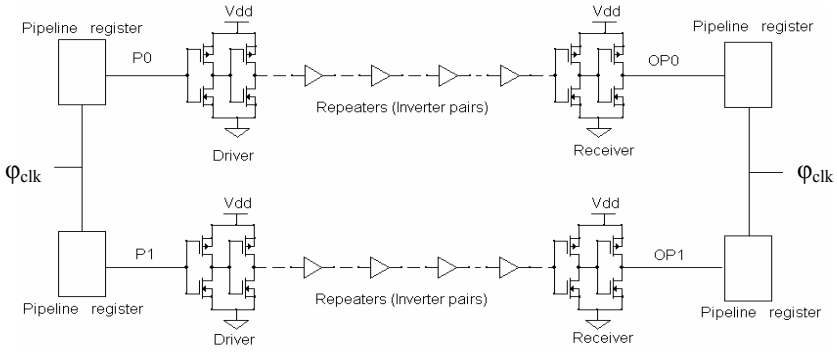

Figure 3a. Schematic diagram of conventional routing.

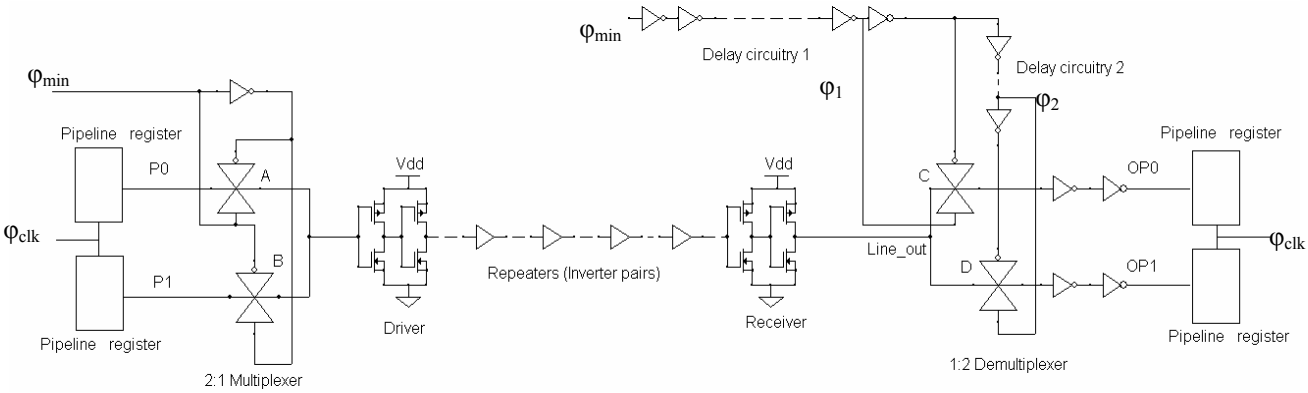

Figure 3b. Schematic diagram of normal WP/2-TDM routing.

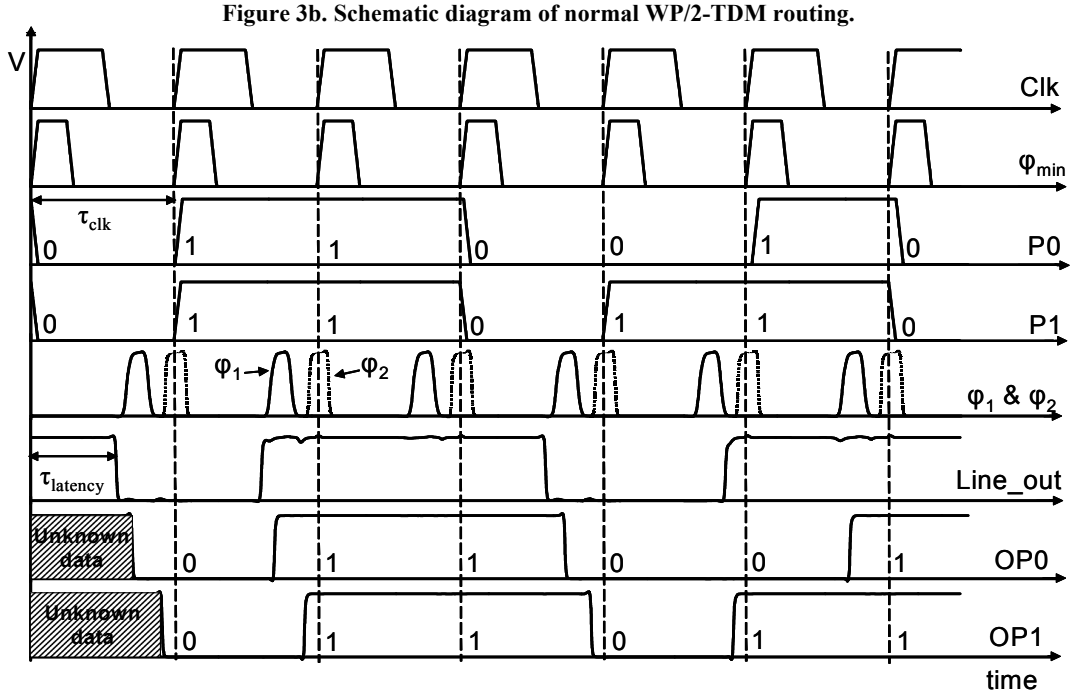

Figure 4. Timing waveforms of a WP/2-TDM circuit using HSPICE.

For example, if the first signal at P0 is sampled at $t = 0$, then signal at P1 will be sampled at $t = t_{pulse}$ by the multiplexer. Assuming the first signal reaches Line_out at $t = 1.5 * t_{clk}$ (accounting for any clock skew and guardband), the second signal at P1 will reach Line_out at $t = 1.5 * t_{clk} + t_{pulse}$. These signals will be used by the appropriate circuits at $t = 2 * t_{clk}$. Meanwhile, the second set of signals input at P0 and P1 will be sampled and transmitted at $t = t_{clk}$ and $t = t_{clk} + t_{pulse}$ respectively and will be used by the receiver side circuits at $t = 3 * t_{clk}$. The delay circuitries will have to suitably designed so that φ_1 will go high at $t = 1.5 * t_{clk}$ to sample first signal and φ_2 will go high at $t = 1.5 * t_{clk} + t_{pulse}$ to sample second signal.

Thus, signals can be transmitted at both the sources, and will also be available at both the sinks, at the beginning of every clock period resulting in maintenance of total communication throughput of the system. Depending on the physical layout of the macrocells, there are various opportunities for incorporating the WP/2-TDM wire sharing technique. The maximum advantage of the WP/2-TDM wire sharing technique can be obtained by incorporating its design approach in the CAD layout algorithms.

4. CASE STUDIES

In order to illustrate the potential advantages of WP/2-TDM routing, two case studies are presented here. The first case study is applicable to the optimal design of a global wire tier that will support semi-global and global wires that extensively utilize WP/2-TDM techniques. The second example will elucidate the effectiveness of incorporating WP/2-TDM into an existing global routing tier whose wire dimensions and pitch already fixed.

4.1 Global wire tier design

Consider two dedicated global wires, each 1 cm long. Here, it is assumed that these two wires are the longest on a tier and are initially designed such that their delay is 80% of a 1.3 Ghz clock. HSPICE and RAPHAEL are used to accurately model the wire transients. Assuming that they satisfy the aforementioned proximity constraints, we replace these two wire channels with a single WP/2-TDM routing channel. This new routing channel is redesigned so that it will transfer both data bits within 80% of the clock period. Hence, it will have slightly larger wire dimensions and transistor sizing to avoid any loss of performance.

Figure 5 shows the variation in pitch values for the two designs. The pitch value is larger for wires having a smaller number of repeaters to ensure timely data transfer. As the number of repeaters increases, the interconnect pitch decreases initially; however, as the number of repeaters increases beyond the optimal point, repeaters contribute significantly to the overall interconnect delay which makes it necessary to again have fatter interconnects to satisfy delay constraints.

Even though this new WP/2-TDM global wire tier results in wider wire routing channels, the total global wire area still decreases because of extensive wire sharing. Figure 6 illustrates the variation in wire area with number of repeaters, for the conventional design and the WP/2-TDM design. The wire area follows the same trend as that of the wire pitch. Overall, more than 40% reduction in wire area is obtained by application of the WP/2-TDM routing. A significant amount of silicon area can be vacated due the elimination of the repeaters on the eliminated wire. Figure 7 shows this decrease in the total transistor area for different number of repeaters. At the optimal wire area design point, one can obtain close to 30% decrease in the transistor area.

As a result of this decrease in the transistor area, one would expect a decrease in the total power of the system. Here, the static power of the system decreases; due to the smaller silicon area, however, there is a slight increase in the dynamic power. The elimination of interconnects and the repeaters on those interconnects does not decrease the dynamic power due to the proportional increase in the activity factor of the shared interconnects and logic circuits. In addition, the use of overhead circuitry i.e. multiplexer, demultiplexer and delay elements, contributes to the power equation resulting in an increase in the total dynamic power of the system. Figure 8 shows the increase in the dynamic power of the system for the different number of repeaters. On an average a 6% increase in the dynamic power is observed.

One could reduce the dynamic power by increasing the spacing between metal wires to reduce coupling capacitance. In addition, the transistor area would decrease as smaller drivers would be required due to decrease in the coupling capacitance. The increase in wire spacing will of course increase wire area; however, if the power budget is extremely tight this tradeoff between wire area and power might be advantageous. Figures 5-7 show the change in wire pitch, wire area and transistor area for a design exhibiting no dynamic power change or loss of performance. One can still observe more than 40% reduction in wire area and close to 30% reduction in transistor area for this design.

4.2 Custom routing example

[14] gives a description of the 1.3 GHz fifth generation SPARC64 microprocessor design. Using the die micrograph in [14], approximate length of the interconnects between the floating point (FP) macrocell and the Load/Store (LS) macrocell, and the fixed point (FX) macrocell and the LS macrocell are estimated to be 1.023cm and 0.75cm respectively. It is assumed that the interconnects travel from the center of one macrocell to the center of the other macrocell. Given that it is a 64 bit microprocessor and it has 2 FP units, one can assume that there will be 4 read ports (therefore 4 x 64 interconnects) and 2 write ports (therefore 2 x64 interconnects) on the FP macrocell that sends/receives data from the LS unit. In addition to these data lines, there will be additional control lines to send and receive various handshaking signals between the two macrocells; however, these control lines have been ignored for this case study. Thus there will be a total of 384 interconnects (set A) between the two macrocells. Similarly one can assume that there will be 384 interconnects (set B) between the FX macrocell and the LS macrocell.

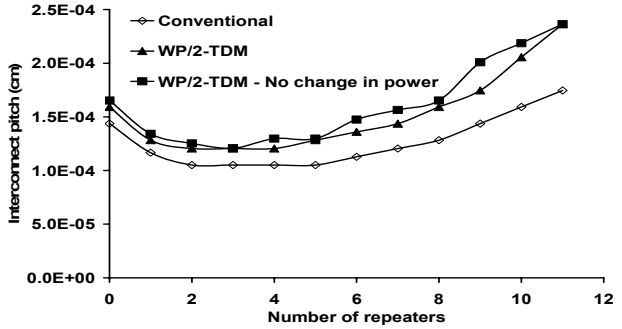

Figure 5. Interconnect pitch vs number of repeaters.

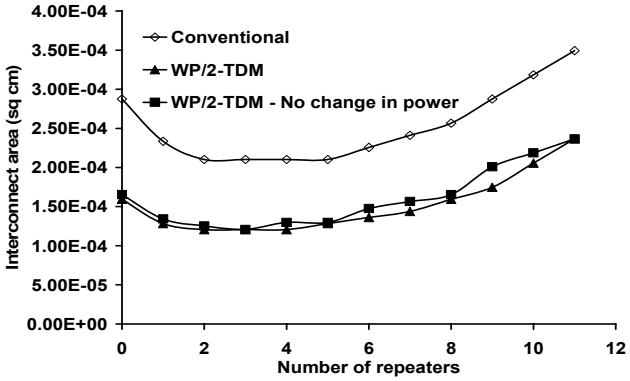

Figure 6. Wire area vs number of repeaters.

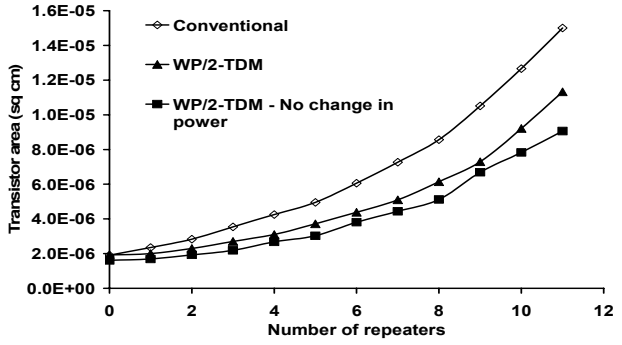

Figure 7. Transistor area vs number of repeaters.

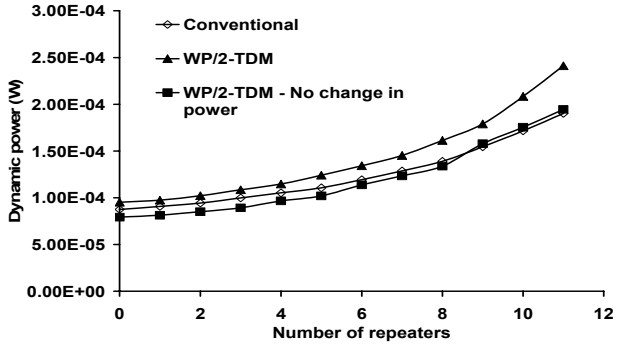

Figure 8. Dynamic power vs number of repeaters

Table 1. Delay for different interconnect lengths.

Interconnect length (cm)	Interconnect delay (ns)	Normalized delay
0.75	0.427	0.55
1.023	0.585	0.76

In order to determine any existence of wire idleness, one interconnect from set A and one from set B are modeled using Level 49 HSPICE models for 130nm technology [15]. The wire pitch and thickness values for the processor design are obtained from [16]. The processor design in [14] has a die size of 1.81cm x 1.599cm and hence, the interconnects of length 1.023cm and 0.75cm are assumed to be global interconnects that are routed in metal 7 and 8. Hence, the interconnect width is considered to be 900nm [16]. A sub-optimal number of repeaters [10], having sub-optimal size [13], are inserted on the interconnects.

Table 1 shows wire delay, calculated using HSPICE, for the two wire lengths. The delay for 0.75cm long wire is just 0.427ns i.e. 0.55 times the clock period and from [11] the minimum pulse width evaluates to 0.184ns. The sum of wire delay and minimum pulse width is 0.611ns which is less than 0.8 times the clock period. Thus, delay constraint (1) is satisfied. On the other hand, the wire of length 1.023cm has a delay of 0.585ns which is 0.76 times the clock period. The minimum sustainable pulse width evaluates to 0.228ns using [11] for this case. Hence, it does not satisfy the WP/2-TDM delay constraint (1).

Thus, the normal WP/2-TDM routing can be applied to all interconnects in set A if they satisfy the proximity constraints. One can then reduce the number of routing channels by 50% without any loss of throughput performance and the latency of t_{clk} will be maintained. For interconnects in set B, though the single clock period latency constraint is not satisfied, a slightly modified WP/2- TDM routing can still be applied and the routing channel count can be reduced by 50%, given that the proximity constraints are satisfied. Here, though the latency would increase to twice the clock period, the throughput performance would be maintained. Interconnects of set B could require a more extensive re-design at the RTL stage to account for this data latency change. Once the system is appropriately redesigned, the WP/2-TDM could be seamlessly incorporated at the logic and circuit levels of design.

5. CONCLUSION

This paper proposes a new circuit that combines both wave-pipelining and 2-slot time division multiplexing (WP/2-TDM) to produce an interconnect routing technique that can be seamlessly incorporated into existing global and semi-global pipelines. Because of the relative ease of incorporation of this technique into a traditional VLSI design flow, this implementation has the potential to be a ubiquitous routing technique that can be applied to both inter-core and intra-core interconnects in any SoC or microprocessor design.

Two case studies are presented to demonstrate the advantages of the application of the WP/2-TDM technique. More than a 40% reduction in the wire area and close to 30% reduction in silicon area can be observed for a simple two interconnect system with no increase in dynamic power and no loss in performance. The custom routing example illustrates opportunities whereby the WP/2-TDM technique can be incorporated into the system design and the number of the required routing channels can be reduced by up to 50% with no loss in throughput performance. Requirements for deepening interconnect pipelines for the longest wires are discussed.

6. REFERENCES

[1] J. Meindl, "Low-power microelectronics: Retrospect and prospect," *Proc. IEEE*, vol. 83, pp. 619–635, Apr. 1995.

[2] J. Davis, et. al., "Interconnect limits on gigascale integration (GSI) in the 21st century," *Proc. IEEE*, vol. 89, pp. 305–324, Mar. 2001.

[3] International Technology Roadmap for Semiconductors (http://public.itrs.net/)

[4] S. Kumar, et. al., "A Network on Chip Architecture and Design Methodology," *Proc. IEEE Comp Soc*, pp. 105-112, April 2002.

[5] J. Liu, et. al., "A Global Wire Planning Scheme for Network-on-Chip," *Proc. ISCAS 2003*, vol.4, pp IV-892 – IV-895, May 2003.

[6] P. Bhojwani et. al., "Interfacing cores with on-chip packet switched networks," *Proc. VLSI design*, pp. 382-387, Jan. 2003.

[7] J. Liu, et. al., "System Level Interconnect Design for Network-on-Chip using Interconnect IPs," *Proc. IEEE/ACM International Workshop on System Level Interconnect Prediction (SLIP), 2003*, pp. 117-124, Apr 2003.

[8] K. Lahiri, et .al., "LOTTERYBUS: A New High-Performance Communication Architecture for System-on-Chip Designs," *Proc. DAC, 2001*, pp. 15-20, June 2001.

[9] A. Joshi**,** et. al., "A 2-slot time-division multiplexing (TDM) interconnnect network for gigascale integration (GSI)," *Proc. IEEE/ACM International Workshop on System Level Interconnect Prediction (SLIP), 2004*, pp. 64-68, Feb 2004.

[10] R. Venkatesan, et. al., "Optimal n-tier Multilevel Interconnect Architectures for Gigascale Integration (GSI)," *IEEE Trans. VLSI systems,* vol. 9, pp. 899-912, Dec. 2001.

[11] V. Deodhar et. al., "Optimization for throughput performance for low power VLSI interconnects", to be published in *IEEE Trans. VLSI systems,* March 2005.

[12] J. Davis, et.al., "A stochastic wire-length distribution for gigascale integration (GSI)—Parts I and II," *IEEE Trans. Electron Dev.,* vol.45, pp. 580-597, Mar. 1998

[13] Y. Cao, et. al., "Effects of global interconnect optimizations on performance estimation of deep submicron design," *Proc ICCAD 2000. IEEE/ACM International Conference,* pp. 56–61, Nov. 2000.

[14] H.Ando et. al., "A 1.3GHz Fifth Generation SPARC64 Microprocessor," *Proc. ISSCC, 2003,* pp. 246-255, Feb. 2003.

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

[16] H. Ando, et al., "A 1.3-GHz Fifth-Generation SPARC64 Microprocessor," *IEEE Journal of Solid State Circuits,* vol. 38, pp. 1896-1905, Nov. 2003.