

# A Robust Edge Encoding Technique for Energy-Efficient Multi-Cycle Interconnect

Jae-sun Seo, *Student Member, IEEE*, Dennis Sylvester, *Senior Member, IEEE*,  
David Blaauw, *Senior Member, IEEE*, Himanshu Kaul, *Member, IEEE*,  
and Ram Krishnamurthy, *Senior Member, IEEE*

**Abstract**—In this paper, we propose a new circuit technique for on-chip communication, the edge encoding technique, to reduce the energy consumption in multi-cycle interconnects. Both average and worst-case energy are reduced by desynchronizing the edges of rising and falling transitions. In a 1.2V 65nm CMOS technology, the proposed approach achieves up to 34% energy reduction with no latency overhead over optimally designed conventional busses due to coupling capacitance reductions. The technique further reduces energy consumption by 39% with iso-throughput at the expense of one-cycle latency. Energy savings are shown to be both larger and more robust to process variations than previous techniques.

## I. INTRODUCTION

In current microprocessors, the number of wires used for intra-module communication has skyrocketed. Furthermore, the increased complexity and high level of integration requires higher wire densities, and coupling capacitance has dominated total wire capacitance for several technologies already. A high coupling capacitance ratio is not favorable in conventional busses due to the possibility of adjacent wires switching in the opposite direction, yielding a worst-case Miller capacitance factor (MCF) of 2. For example, when MCF=2 the coupling capacitance ratio over the total interconnect capacitance is over 80% for a minimum pitch intermediate metal layer in 65nm [6]. It is possible to reduce coupling capacitance by increasing spacing or introducing shielding, but this comes at the cost of significant area penalties [1]. Hence, a key challenge in interconnect design is to reduce the worst-case MCF while maintaining the same physical footprint of the interconnect, thereby reducing the effective wire capacitance and interconnect energy consumption.

There have been several attempts to reduce worst-case MCF to 1 for delay improvement and power reduction. In [4], the authors introduced a delay element on alternating wires, thereby avoiding the MCF=2 switching case through temporal separation. In this approach, however, fine-tuning of the optimal insertion delay is non-trivial, and due to very small inverter delays in sub-90nm technologies, many inverters are needed to sufficiently separate the switching of adjacent wires,

increasing power. Also, this technique is sensitive to process variation since variability in the inserted delay can lead to a lack of sufficient separation for adjacent wires.

Separating the timing of transitions on adjacent wires was also proposed in [5] by assigning different clocks to flip-flops driving adjacent wires. Rather than assigning clocks with different phases, [6] implemented a technique that alternatively used positive-edge triggered and negative-edge triggered flops in every other wire. In this case, however, the wire length associated with the final flop must be short to align to the positive edge at the far end of the wire. In [6], the authors proposed a method to skew alternating wires in the opposite direction using different width, length,  $V_t$  and body bias. In this way the worst-case switching is separated without hurting the best-case switching. However, this technique is also very sensitive to process variations, which can lead to less separation than needed to achieve an MCF of nearly 1. A method using careful staggering of repeater locations is introduced in [7]. This method results in alternating MCF=0 and MCF=2 in neighbor wire segments. However, in terms of physical design, this method incurs significant overhead considering that the repeater location cannot always be arbitrarily selected in industrial designs. Without modifying the repeater locations, techniques to use both inverting repeaters and non-inverting repeaters were proposed in [8], [9], such that one half of the wire segments experience MCF=0 and the other half experience MCF=2.

A number of coding techniques including [10], [11] encode the conventional bus such that adjacent bits never switch in the opposite direction. However, in addition to the special encoder and decoder circuit overhead, these techniques require additional wires for bus encoding which increase routing area. It is not clear whether the conventional bus could better use extra spacing at the same footprint instead of additional wires for better speed and energy consumption.

Pulsed bus techniques [12] also achieve a worst-case MCF of 1. In these pulsed bus techniques, however, the energy dissipation is increased per transition compared to conventional busses due to the pulse encoding. Reference [13] reduced this overhead by selectively using low  $V_{dd}$  with nominal  $V_{dd}$  to drive the interconnect, but this is done at the expense of design complexity since two power supplies are required.

References [4], [5], [7]–[9] reduce the overall worst-case MCF of an interconnect to 1, but also eliminate the best-case MCF of 0 (all adjacent wires switching in the same direction), leading to less advantage in average energy consumption.

Manuscript received May 4, 2009.

Jae-sun Seo, Dennis Sylvester, and David Blaauw are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: jseo@umich.edu, dmcs@umich.edu, blaauw@umich.edu).

Himanshu Kaul and Ram Krishnamurthy are with the Circuit Research Laboratory, Intel Corporation, Hillsboro, OR 97124 USA (e-mail: himanshu.kaul@intel.com, ram.krishnamurthy@intel.com).

Using the technique in [6], best-case switching is maintained, but at the expense of smaller noise margin in the repeaters and more sensitivity to process variations as mentioned above. Furthermore, the amount of skewing required to effectively separate transitions of adjacent wires is heavily dependent on technology.

This paper presents a new encoding technique that achieves a worst-case MCF of 1, while preserving the best-case MCF of 0. This is done by controlling the edges of rising and falling transition in time, namely always performing rising transitions on the negative edge of the clock and falling transition on the positive edge of the clock (or vice versa). Since the worst-case switching is separated by as much as one phase (half clock cycle), this technique remains robust against process variation. Hence, both the average and worst-case energy can be reduced without impacting the sensitivity to process variation. Average energy savings will aid battery life and typical energy costs, but worst-case energy is also a meaningful metric in terms of thermal management and peak demand for power grids and decoupling capacitance [14]. These savings are accomplished at the expense of minimal encoder logic with half cycle latency and additional clocking. However, we find that the logic and clocking overhead is small in long interconnects where interconnect power consumption is dominant, and also show that the potential latency overhead can be eliminated or minimized in multi-cycle interconnects. A preliminary version of this paper appeared in [15].

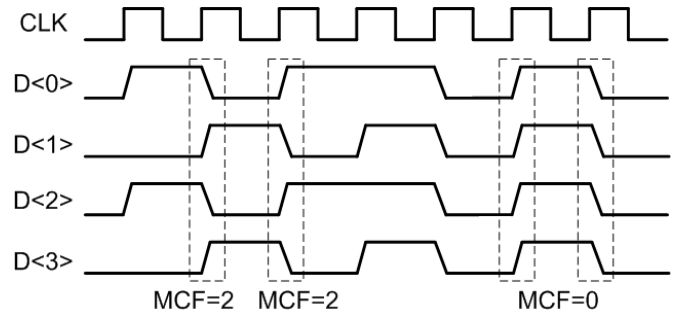
The remainder of this paper is organized as follows. Section II describes the proposed encoding technique and simplified analytical models of savings due to MCF reduction. Section III presents performance, energy, and leakage comparison results. Section IV summarizes the paper.

## II. EDGE ENCODING APPROACH

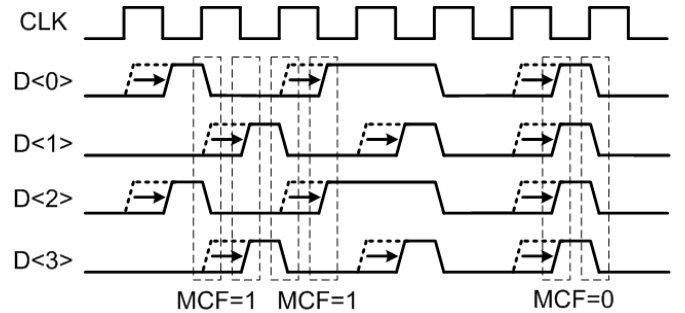
### A. Basic Concept

In a multi-cycle bus structure, the transitions between neighboring wires are synchronized at every flip-flop as the signal propagates down the bus. This often generates simultaneous switching of adjacent wires in the opposite or same direction. In Figure 1(a), the worst-case (MCF=2) and best-case (MCF=0) switching of the conventional bus are shown. The MCF=2 case, where every other wire switches in the opposite direction, generates the worst-case delay, which defines the clock frequency and also consumes the worst-case energy due to maximum coupling capacitance. To avoid this, we propose to selectively shift rising and falling edges and separate them by as much as half cycle. For example, as seen in Figure 1(b), if we selectively delay only the rising transitions by a half cycle and keep the falling transitions unaltered, the worst-case MCF is reduced from 2 to 1. We refer to this selective edge shifting as *edge encoding*. Since edge encoding shifts the same transitions together, the advantage of best-case switching (MCF=0) is still maintained, which is unachievable in most other approaches [4], [5], [7]–[9].

Since the edge-encoded signal transitions at both positive and negative edges of the clock, we use dual-edge triggered flip-flops to propagate the signal along long multi-cycle interconnects. The methodology for the placement of dual-edge



(a) Conventional wire switching.



(b) Proposed wire switching.

Fig. 1. Conventional and proposed wire switching scenario in adjacent wires.

flip-flops in the edge encoding technique to maximize energy-efficiency will be described in Section II-C.

### B. Theoretical Energy Savings

The total interconnect capacitance is the sum of ground capacitance ( $C_g$ ) and coupling capacitance ( $C_c$ ). The effective interwire coupling capacitance depends on the switching behavior of adjacent wires, which is characterized by MCF in Equation 1 below. MCF is 0 when all adjacent wires switch in the same direction where the total wire capacitance is only  $C_g$ , and MCF is 2 when every alternating wire switches in the opposite direction resulting in total capacitance of  $C_g + 4C_c$ . Note that MCF is an approximate value since transitions in adjacent wires can occur arbitrarily. Actually, [16] reports the true worst case MCF of 3 if the slew rate of the aggressor is twice as fast as that of the victim, but MCF of 2 is used as a rule of thumb for worst-case switching here to compute theoretical energy savings. In reporting results later in Section III, we use SPICE to reflect the actual interwire coupling in multi-bit busses.

If we can control the transitions as shown in Figure 1(b), the worst-case MCF is reduced to 1, and reduction of wire energy consumption is achievable, as expressed in Equation 2. The maximum wire energy savings we can ideally achieve is dependent on the ratio of ground capacitance and the coupling capacitance in the interconnect.

$$C_{wire} = C_g + 2 \times MCF \times C_c \quad (1)$$

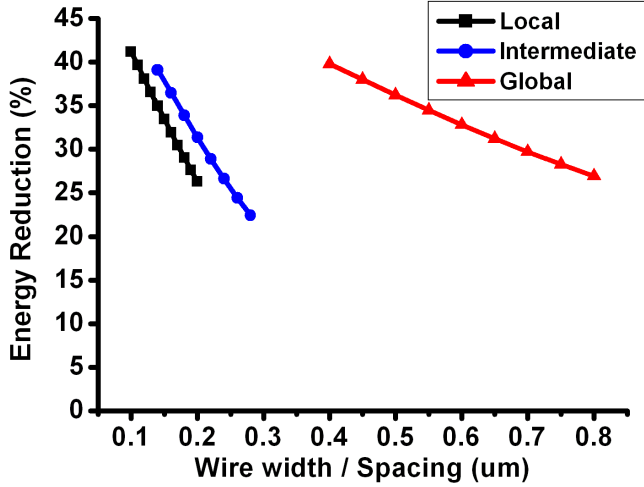


Fig. 2. Ideal wire energy savings due to MCF reduction based on 65nm interconnect dimensions [3].

$$E_{wire,saving} = 1 - \frac{C_{wire}(MCF=1)V_{dd}^2}{C_{wire}(MCF=2)V_{dd}^2} = 1 - \frac{C_g + 2C_c}{C_g + 4C_c} \quad (2)$$

Closed-form equations from [2] compute capacitance values for a given wire geometry, namely wire width, spacing, thickness and dielectric thickness. With typical wire dimensions given in [3] for local, intermediate and global wires in the 65nm technology node, the expected energy savings are calculated using Equation 2. A range of wire pitches are shown, with  $W=S$  being swept from minimum to double pitch. The ideal energy savings are shown in Figure 2.

Besides the wire energy, the energy dissipation due to the capacitance of repeaters should be also included in the total energy consumption of optimally repeated interconnects, as shown in Equation 3.  $C_{tr}$  is the sum of gate and drain capacitance of a unit-sized repeater. The total capacitance of repeaters is proportional to the number of repeaters inserted ( $N_R$ ) and the size of the repeaters ( $H_R$ ), where these two parameters heavily depend on the resistance and capacitance of the interconnect.

As the pitch increases, the achievable energy savings due to MCF reduction decreases as expected since the interwire coupling capacitance diminishes. In general, even for less favorable non-minimum pitches, Figure 2 shows that manipulation of MCF can lead to appreciable (25-40%) energy savings.

$$E_{total} = (C_{wire} + N_R H_R C_{tr}) V_{dd}^2 \quad (3)$$

Equations 4 and 5 from [17] show the interaction between the repeater parameters and wire parasitics for energy-delay optimal repeater insertion.  $R_{tr}$  is the average transistor resistance unit-sized repeater,  $R_{wire}$  is the total interconnect resistance, and  $C_{wire}$  is the total interconnect capacitance from Equation 1.

$$N_R = \frac{1}{\sqrt{3}} \sqrt{\frac{0.4 R_{wire} C_{wire}}{R_{tr} C_{tr}}} \quad (4)$$

$$H_R = \frac{0.6}{\sqrt{3(0.4)(0.7)}} \sqrt{\frac{R_{tr} C_{wire}}{0.7 R_{wire} C_{tr}}} \quad (5)$$

Given a 40% reduction in  $C_{wire}$  (minimum pitch wires in Figure 2), both  $N_R$  and  $H_R$  are reduced by 23% ( $1 - \sqrt{1 - 0.4} = 0.23$ ). In this simple energy model, peak MCF reduction decreases  $C_{wire}$  and  $N_R \times H_R$  by the same ratio. Therefore, the total energy reduction (Equation 3) will be identical to the wire energy reduction from Equation 2, regardless of the absolute values of wire capacitance and repeater capacitance. The total energy savings including repeaters for local, intermediate, and global interconnects is equivalent to Figure 2.

In our proposed scheme, the analytical energy reduction will be degraded by additional clock and encoder energy, however for long intermediate and global interconnects, this additional energy will be small compared to the total wire energy consumption. Detailed results will be shown in Section III.

### C. Edge Encoding Technique

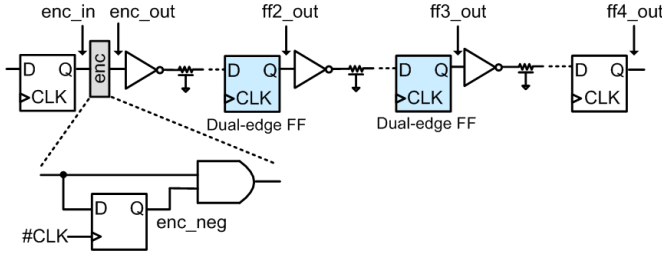
As described in Section II-A, the objective of the edge encoder is to selectively shift the rising and falling transition by different amounts. This encoding is done simply by performing an AND operation between the original signal and the half-cycle delayed version of itself. In this way, only the rising edge is delayed by a half cycle, separating simultaneous rising and falling transition by a half cycle. Since the encoder logic is very simple, the encoding overhead in terms of power and area is very small. This makes the edge encoding technique a highly practical approach.

We propose two schemes to effectively use the edge-encoding technique in multi-cycle interconnect. The two methods differ in the procedure to cope with the initial half cycle latency required for edge encoding and to address the issue of aligning back to the positive-edge triggered signal at the far end of the wire.

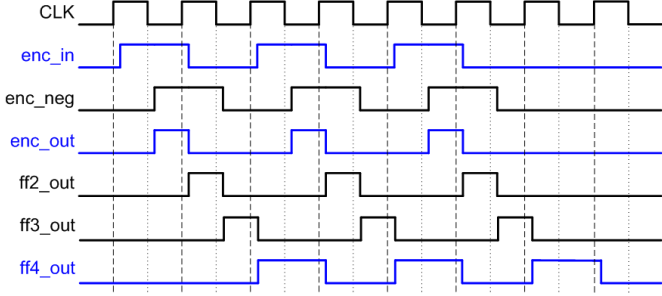
1) *Zero Latency (ZL) Scheme*: The zero latency (ZL) scheme reduces energy consumption in multi-cycle interconnects without any latency overhead although encoding requires a half-cycle delay at the near end of the wire. This scheme exploits the fact the signal propagation in the edge-encoded bus is faster than that in the conventional bus due to reduced MCF.

The block diagram of a multi-cycle interconnect with simple encoder logic is shown in Figure 3(a). The encoding procedure and the propagation of the encoded signal are shown in Figure 3(b). When data toggles every cycle, the encoder generates a half-cycle pulse (enc\_out). As this half-cycle pulse propagates through an even number of dual-edge flip-flops, it automatically aligns back to a positive edge triggered signal (ff4\_out) at the far end. Therefore, there is no need for any decoder circuit.

To achieve overall zero-latency, the interconnect system is set up as shown in Figure 4. If the conventional scheme requires  $n$  cycles to propagate through the entire interconnect, the edge-encoded bus must propagate through in  $(2n - 1)$



(a) Encoder logic and block diagram of ZL scheme.



(b) Timing diagram of ZL scheme.

Fig. 3. Block and timing diagrams of ZL scheme.

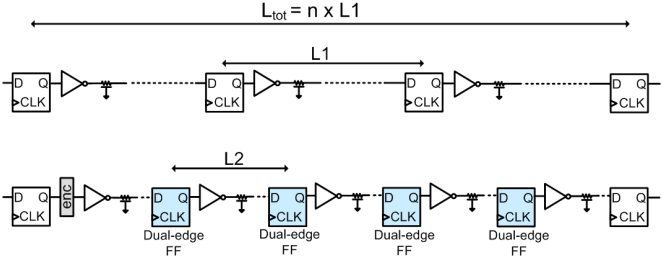
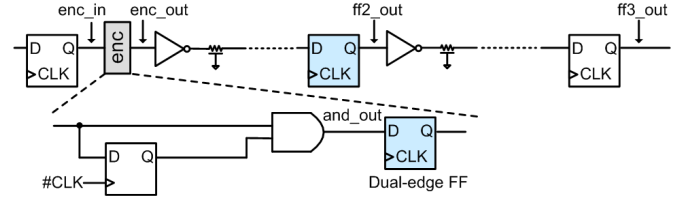


Fig. 4. Flip-flop placement in conventional and ZL edge encoding scheme.

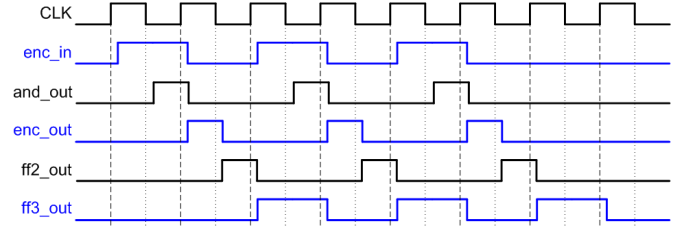
half cycles, considering that the encoding takes one half cycle to synchronize at the far end of wire. In Figure 4,  $L_1$  is the distance between positive-edge triggered flip-flops in the conventional bus, and  $L_2$  is the distance between dual-edge triggered flip-flops in the edge-encoded bus. If  $L_2$  is defined by  $L_1$  and  $n$  in Equation 6, overall zero latency is achievable. For example, in a 9mm interconnect, when  $n=3$  and  $L_1=3$ , the edge-encoded signal will propagate 1.8mm every half cycle while the conventional signal will propagate 3mm every cycle. Effectively, the edge-encoded signal is traveling 20% longer (1.8mm vs. 1.5mm) during the same time period, which is possible when at least a 17% ( $1 - 1/1.2$ ) speedup is achieved in the edge encoded bus due to coupling capacitance reduction.

$$L_2 = L_1 \times \frac{n}{2n - 1} \quad (6)$$

2) *One Cycle Latency (OCL) Scheme*: In multi-cycle interconnects, multiple cycles are required to propagate across the entire wire. In these cases, one additional cycle latency may be acceptable if a clock frequency increase or aggressive energy



(a) Encoder logic and block diagram of OCL scheme.



(b) Timing diagram of OCL scheme.

Fig. 5. Block and timing diagrams of OCL scheme.

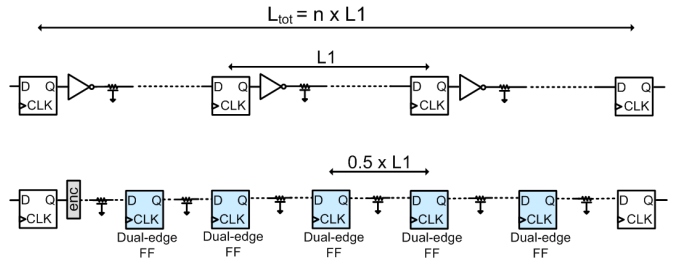


Fig. 6. Flip-flop placement in conventional and OCL edge encoding scheme.

reduction is a higher design priority. The one cycle latency (OCL) scheme is intended to achieve further performance improvement and energy reduction for a fixed throughput at the expense of one-cycle latency. After the encoding, the data must eventually align to the positive edge of the clock at the far end of the wire. To achieve this, we can align the transition at the near end to the positive edge of the clock by encoding with a full one cycle delay, and then allow for normal signal propagation along the wire. The one-cycle latency is therefore introduced once at the beginning of the wire and the throughput is not hampered.

The block and timing diagrams of the OCL edge encoding scheme are shown in Figure 5. The difference in the encoder in Figure 5(a) compared to ZL is that a dual-edge flip-flop is added at the output to intentionally delay  $enc\_in$  by one cycle and align the rising edge of  $enc\_out$  at the positive edge of the clock as shown in Figure 5(b). The corresponding flip-flop placement in the OCL edge encoding scheme is shown in Figure 6. Dual-edge flip-flops are placed at intervals equal to half the flop distance of the conventional bus. In the OCL edge-encoded bus, since the worst-case wire delay is reduced due to MCF reduction, we can either increase the clock frequency for high-performance busses or downsize

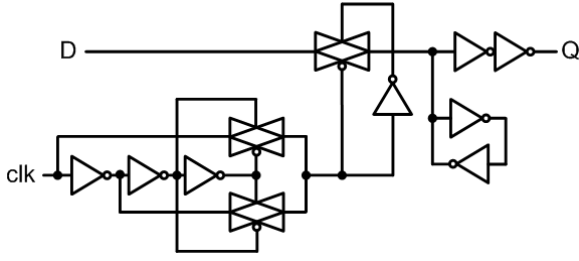


Fig. 7. Schematic of time-borrowing pulsed dual-edge flip-flop.

the repeaters for iso-performance to the conventional bus for aggressive energy reduction.

#### D. Flip-flops and setup time in edge encoding

Since the edge encoding technique requires dual-edge flip-flops, the number of flip-flops placed in multi-cycle interconnects is inevitably increased compared to conventional multi-cycle interconnects with single-edge flip-flops. Therefore, along the critical path of multi-cycle interconnects, both the total setup time and CLK-Q delay in flip-flops increase as well. Hold-time is not a concern even when using dual-edge flip-flops because the interconnect paths are well-defined with several repeaters and large wire load, and thus short paths between flip-flops do not exist.

It is well known that time-borrowing flip-flops have zero or negative setup time, providing performance benefits over master-slave flip-flops [18], [19]. Also, particularly for multi-cycle interconnects, time-borrowing single-edge flip-flops were suggested [20] for better tolerance against within-die variation and higher maximum frequency. Therefore, in the proposed edge encoding techniques, time-borrowing flip-flops are considered for the inserted flip-flops to mitigate the increase in total setup time and variation. However, the benefits of time-borrowing flip-flops come at the expense of additional transistors and higher energy consumption. In the proposed edge encoding techniques, since additional dual-edge flip-flops are placed in the multi-cycle interconnect path, the energy-delay tradeoff of the time-borrowing dual-edge flip-flops has to be carefully investigated.

In our experiments, we performed each edge encoding technique (ZL and OCL) with both conventional master-slave dual-edge flip-flops and time-borrowing dual-edge flip-flops. The master-slave dual-edge flip-flops have positive setup time, and the additional setup time and D-Q delay will be considered in the overhead involved with the proposed techniques. For the time-borrowing dual-edge flip-flops, several types of flip-flops from [19] were considered to absorb the setup time and mismatch between interconnect paths. We selected the pulsed triggered dual-edge flip-flop in Figure 7 for the proposed edge encoding because it is more area-efficient and the D-Q path is shorter. Throughout this paper, we designed the transparent window to be  $\sim 1.6$  FO4 delay. Since the transparent window will be applied on both the rising and falling edge of a high frequency clock in our schemes, the size of the transparent window cannot be arbitrarily large. The results of using both

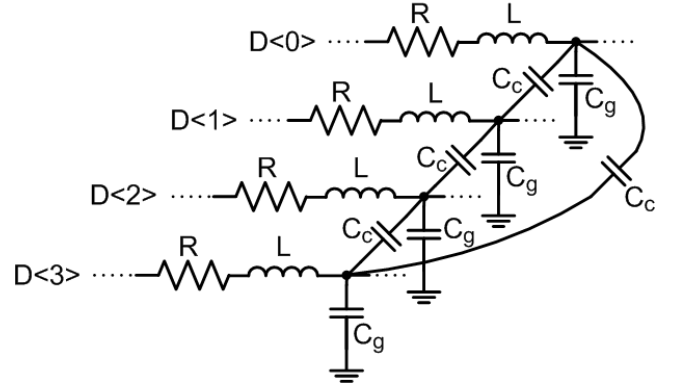


Fig. 8. 4-bit cyclic bus model.

master-slave flip-flops and time-borrowing pulsed flip-flops for the proposed techniques are shown in the following section.

### III. EXPERIMENTAL RESULTS

To accurately capture the effect of coupling capacitance in adjacent wires, we use the 4-bit RLC cyclic model [6] for the interconnect shown in Figure 8. Interconnect parasitic values are extracted for a minimum pitch intermediate layer metal 4 in 65nm technology and all results are obtained from SPICE simulations with a 1.2V supply.

For various flop distances, the conventional repeater bus is optimized by sweeping both the number and sizes of repeaters. Energy, delay, clock frequency, and leakage power are measured for the optimally designed conventional busses, with this serving as the baseline for comparison with edge encoded busses. Unless mentioned otherwise, activity factor of 50% (data switches on every positive edge of clock) is assumed. We now show results for the two edge encoding schemes as proposed in Section II-C.

#### A. Zero Latency (ZL) Scheme

As described in Section II-C1, both the conventional and ZL edge encoding schemes operate at the same clock frequency, however the flop to flop distance in the ZL scheme is effectively larger. From Figure 4,  $L_2$  in the ZL scheme depends on  $L_1$  in the conventional scheme as defined by Equation 6. The optimized set of flop distances and interconnect lengths using the ZL scheme is summarized in Table I. The number of cycles is set to 3 for all cases for simplicity. A flop distance of 1mm in a conventional bus was found to be too short for the edge encoding technique to gain enough speedup for the ZL scheme to be applicable, thereby 2-5mm are selected for  $L_1$ . This gives a range of applicability for the proposed technique in this particular technology - note that more advanced processes should allow for benefits at even shorter wire lengths.

For each configuration in Table I, we found the maximum clock frequency at which we compared the total energy consumption in the conventional and ZL edge encoding schemes. The resulting energy reduction obtained in the ZL scheme and the clock frequency achievable at each flop distance ( $L_1$ ) are shown in Figure 9 for both master-slave dual-edge flip-flops

TABLE I  
FLOP DISTANCE AND TOTAL WIRE LENGTH SETTINGS FOR  
CONVENTIONAL AND ZL EDGE ENCODING SCHEME.

$n$ (number of cycles)	L1	L2	Total wire length ( $n \times L1$ )
3	2mm	1.2mm	6mm
3	3mm	1.8mm	9mm
3	4mm	2.4mm	12mm
3	5mm	3mm	15mm

(MS-FF) and time-borrowing pulsed dual-edge flip-flops (TB-FF). Both peak energy and average energy are shown. For average energy, we generated random data over 100 cycles with activity factor of 25% for each of the 4-bit input. As L1 increases, more energy reduction can be achieved using edge encoding, while the maximum clock frequency degrades. Using time-borrowing flip-flops allows negative setup time in the dual-edge flip-flops and improves performance. In the case using time-borrowing flip-flops, the repeaters and flip-flops can be sized down for iso-performance at each flop distance (L1), which leads to additional energy reduction as shown in Figure 9.

A detailed comparison for a flop distance (L1) of 3mm is shown in Table II. Both schemes operate at 2GHz, and we can see that considerable energy savings are achieved for various activity factors. The amount of energy saving decreases at lower activity factors, because the edge encoding scheme consumes additional clock energy and the portion of clock energy increases as the data activity rate is lowered (this may be ameliorated by clock gating or other similar techniques). Using time borrowing flip-flops accelerates this trend since time-borrowing flip-flops have additional transistors and higher clock energy consumption than normal master-slave flip-flops. Due to the reduction of effective capacitance on the wire fewer repeaters are required in the ZL scheme than the conventional scheme for optimal performance and energy, yielding less leakage power and total transistor width, as seen in Table II.

### B. One Cycle Latency (OCL) Scheme

As we saw in Section II-C2, the OCL edge encoding scheme can either reduce energy at iso-performance or improve the performance at iso-energy at the expense of one-cycle latency. To quantify the performance gain or energy reduction, we optimized both the conventional and OCL edge encoding schemes (MS-FF and TB-FF) for a minimum pitch 3mm wire. Figure 10 shows the energy per cycle versus clock frequency for each scheme. Edge encoding shows a potential 29% performance improvement at iso-energy or a 34% energy reduction at iso-performance (2GHz).

Figure 11 shows the breakdown of energy for a 5mm wire in both an optimally-designed conventional repeater bus and OCL edge encoded bus at iso-throughput. The wire energy, which is the dominant source of total energy consumption, is reduced considerably using edge encoding at the expense of minimal encoder logic and additional clocking energy. Overall,

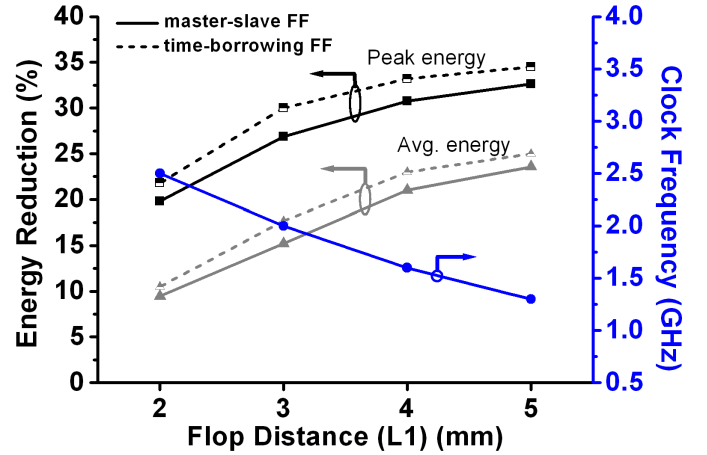


Fig. 9. Comparison of the ZL edge encoding scheme to conventional buses in worst-case/average energy and clock frequency for flop distances (L1) of 2-5mm.

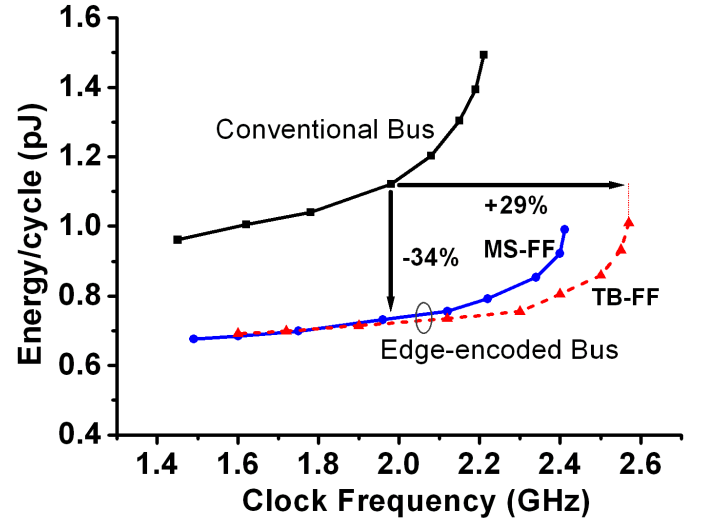


Fig. 10. Energy-clock frequency comparison for a 3mm OCL edge encoded bus.

the OCL approach can achieve 36% energy reduction with MS-FF and 39% energy reduction with TB-FF. Similar to the ZL scheme, larger energy reductions are achieved with TB-FF due to smaller repeaters and flip-flops at a fixed performance.

Also, in the OCL scheme the placement and number of repeaters can be unaltered, allowing the designer to simply drop-in the encoder and additional flip-flop to enable edge encoding. Results of this approach using identical repeater placement and sizes to the conventional repeater scheme are summarized in Table III. Total wire length of 10-12mm is assumed for flop distances of 1-5mm, and the latency overhead is calculated as the relative overhead of encoding (1 cycle) to the number of cycles needed to propagate through the entire interconnect for each flop distance. As flop distance increases, the wire energy sufficiently dominates, allowing OCL edge encoding to achieve larger performance improvements and energy reductions, at the expense of larger relative latency overhead.

TABLE II  
MULTI-CYCLE INTERCONNECT RESULTS (PEAK ENERGY, LEAKAGE AND AREA) FOR THE ZL EDGE ENCODING SCHEME. RESULTS FOR BOTH MASTER-SLAVE FLIP-FLOPS (MS-FF) AND TIME-BORROWING PULSED FLIP-FLOPS (TB-FF) ARE SHOWN.

Scheme	Frequency	Energy/cycle @25% activity	Energy/cycle @15% activity	Energy/cycle @10% activity	Leakage power	Total transistor width
Conventional	2GHz	1.83pJ	1.26pJ	0.77pJ	16.9 $\mu$ W	492.4 $\mu$ m
Proposed (ZL) MS-FF	2GHz	1.39pJ (-24.2%)	0.96pJ (-23.6%)	0.64pJ (-16.5%)	14.2 $\mu$ W (-16.2%)	424.7 $\mu$ m (-13.7%)
Proposed (ZL) TB-FF	2GHz	1.34pJ (-27.3%)	0.94pJ (-25.4%)	0.67pJ (-13.0%)	11.5 $\mu$ W (-26.0%)	375.1 $\mu$ m (-23.8%)

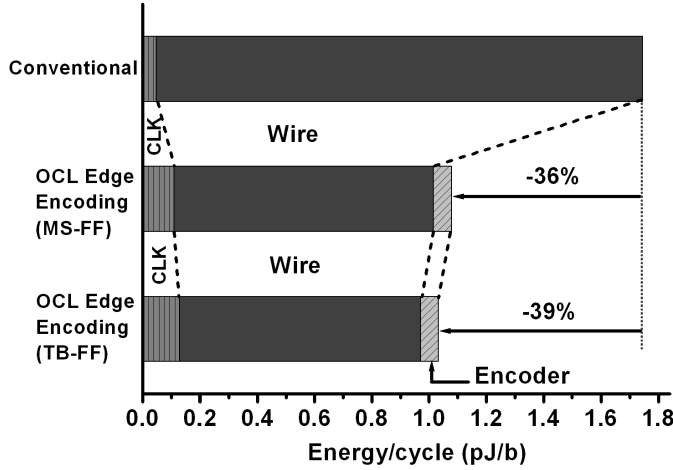


Fig. 11. Energy breakdown of a 5mm wire for conventional and OCL edge encoding scheme.

TABLE III  
PERFORMANCE, ENERGY, AND LATENCY COMPARISON FOR IDENTICAL REPEATER SIZES IN OCL EDGE ENCODING SCHEME.

Total Length	Flop Distance	Performance Gain	Energy Reduction	Latency Overhead
10mm	1mm	-8.2%	-4.9%	10%
10mm	2mm	4.6%	10.0%	20%
12mm	3mm	11.7%	15.3%	25%
12mm	4mm	17.1%	18.1%	33%
10mm	5mm	22.2%	21.1%	50%

### C. Leakage Power Comparison

In sub-90nm technologies, leakage power in repeaters has become problematic [21]. Given that the proposed schemes includes more flip-flops than the conventional case, it is worthwhile to investigate the impact on static power. Table II shows a 26% reduction in leakage power for the TB-FF ZL edge encoding scheme over the conventional approach. This is achieved due to the use of smaller repeaters in the ZL edge encoding scheme.

Since the signal propagation in the ZL scheme needs to be completed a half cycle earlier than the OCL scheme, the repeaters in the OCL edge-encoded bus can be smaller than

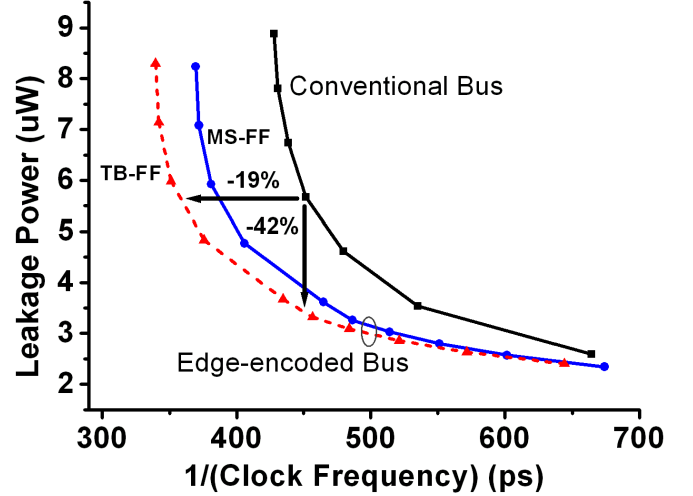


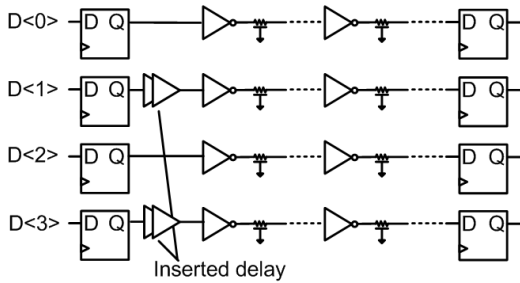
Fig. 12. Leakage power comparison between a conventional and OCL edge-encoded bus. Ten repeaters are used in both cases and wirelength is 3mm.

those in the ZL edge-encoded bus. Therefore, the leakage power, which is proportional to the size of repeaters, is expected to be less in the OCL scheme.

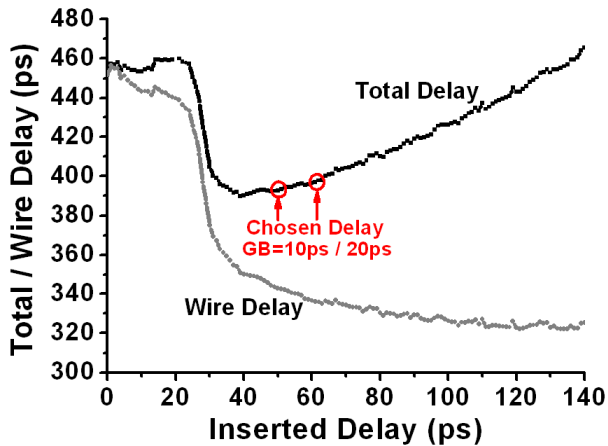
Given an identical number of repeaters in both conventional and OCL edge encoding schemes, Figure 12 shows the leakage power and delay characteristics for a 3mm wire. Due to the performance benefit of the edge-encoded bus, the repeaters in the conventional bus must be upsized to achieve the same performance as the edge-encoded bus, leading to higher leakage power. On the other hand, for identical repeater sizes and hence similar leakage power, the edge-encoded bus can operate at a higher clock frequency than the conventional bus. The OCL edge encoding performance is further improved by using TB-FF rather than MS-FF as expected. Therefore, for iso-performance, we can see that the reduction of coupling capacitance allows fewer and smaller repeaters, resulting in leakage power reduction of 42%.

### D. Sensitivity to Variation

Variability has become a major concern in modern CMOS technologies. Previous techniques [4], [6] rely on inserted delays and p/n skewing to separate the worst-case switching scenario, which are sensitive to process, voltage and temperature variation. We implemented the techniques proposed in [4] and [6] in the 65nm technology used in this paper to compare



(a) Interconnect system of staggered firing bus.



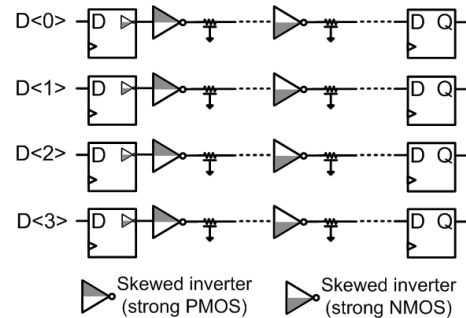
(b) Total/wire delay versus inserted delay.

Fig. 13. Delay selection in staggered firing bus [4].

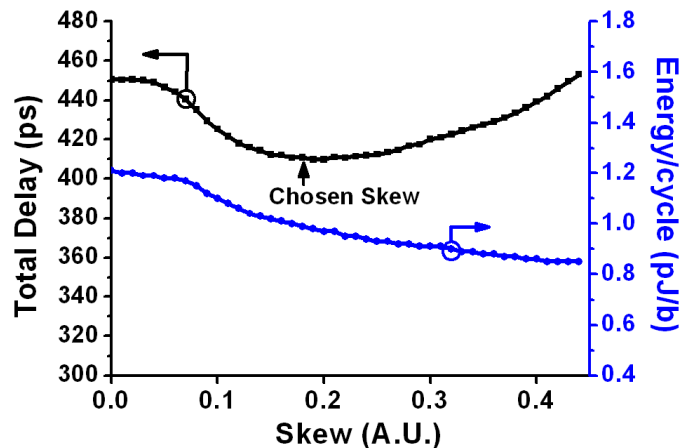
the overall robustness of the achievable gains in the presence of different sources of variation. For on-chip communication circuits with long interconnects, local variation is less of a concern than global variation, because the inserted repeaters are large and the wire delay or energy is not that sensitive due to device mismatch. For 1,000 Monte Carlo runs, the spread in delay for the conventional bus was only 2%. Since global variation has a more dominant effect on the area of concern, we focused on global variation of process, supply, and temperature to analyze the robustness of each technique.

The technique in [4] inserts additional delay elements at the beginning of alternating wires. As more delay is added to adjacent wires, the worst-case switching is further separated, however this additional delay is included in the total delay leading to an optimal inserted delay as shown in Figure 13. Based on the total delay curve in Figure 13 we selected an inserted delay of 50ps and 60ps, which is guard-banded by 10ps and 20ps, respectively, to avoid the steep slope of the total delay curve for inserted delay of less than  $\sim 40$ ps.

The technique in [6] propagates a specific transition (either 0-1 or 1-0) faster down the wire, such that the worst-case MCF pattern naturally separates as the signal propagates. However, this benefit is achieved at the expense of a greatly slowed transition in the non-preferred direction. In this experiment, alternating repeaters are skewed by modifying the width and



(a) Interconnect system of skewed repeater bus.



(b) Total delay and energy versus skew.

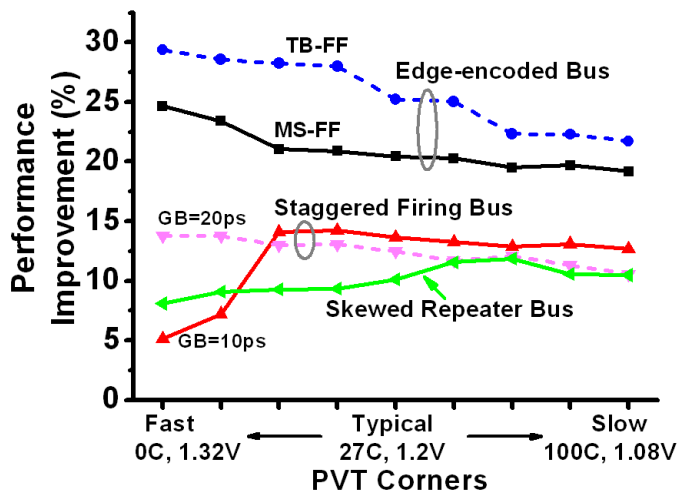
Fig. 14. Skew selection in skewed repeater bus [6].

length of repeaters and flip-flops.

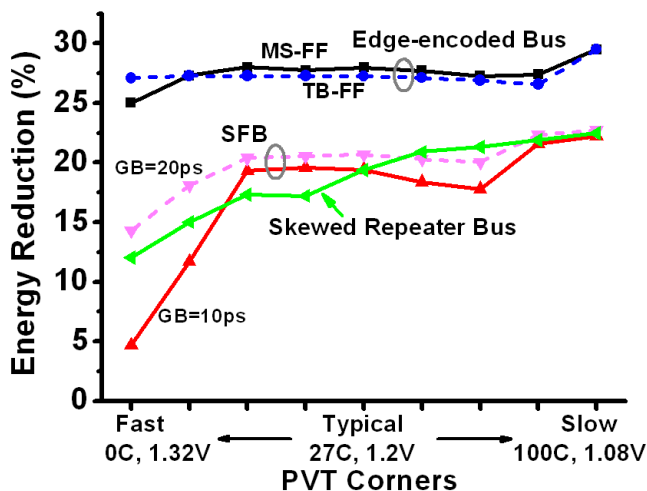
Simulation results for the OCL edge-encoded bus with both MS-FF and TB-FF, staggered firing bus with 10ps and 20ps guard-banding, and skewed repeater bus at different process, voltage, and temperature (PVT) corners are shown in Figure 15. We consider 10% supply voltage variation and 0-100C temperature variation. We first see that edge encoding provides the best overall performance improvement and energy savings. Edge-encoded bus with TB-FF provides additional performance improvement at the cost of small increase in energy consumption compared to edge-encoded bus with MS-FF across PVT corners. The added delay on alternate wires (staggered firing bus) or the intentional skew added to the repeater bus do not retain the delay and energy characteristics of the MCF=1 case when they separate worst-case switching events. This can be seen in Figure 13(b), where the wire delay at the chosen delay points is still larger than the minimum wire delay, and in Figure 14(b), where the energy at the chosen skew is still larger than the energy consumption with further separation of adjacent switching.

Furthermore, edge encoding is more robust across all PVT corners, achieving consistent energy savings. In both [4] and [6] improvements in delay and energy are achieved at the





(a) Performance improvement across PVT corners



(b) Energy savings across PVT corners

Fig. 15. Sensitivity of improvements against process, supply, and temperature (PVT) variation for OCL edge-encoded bus (with MS-FF and TB-FF), staggered firing bus (with 10ps and 20ps guard-banding) [4], and skewed repeater bus [6] (flip distance: 3mm).

expense of susceptibility to PVT variation. Additional guard-banding to improve robustness to variation will result in less delay savings, as shown in the staggered firing bus curve with 20ps guard-banding in Figure 15(a). In Figure 15(b), the energy consumption of 20ps guard-banding is slightly better than that of 10ps guard-banding due to further separation of opposite switching, but the amount of energy reduction will saturate as we increase the guard-banding while the performance will continue to degrade.

#### IV. CONCLUSION

In this paper, we proposed two new edge encoding techniques to improve energy efficiency and performance for multi-cycle on-chip interconnects. Both master-slave flip-flops

and time-borrowing flip-flops were studied for optimal delay and energy. Compared to previously proposed techniques, edge encoding reduces both peak and average energy with improved robustness to process, supply, and temperature variation. For typical flip-flop distances of 2-5mm (corresponding to clock speeds of 1.3-2.5GHz in 65nm CMOS), the new techniques achieve 20-34% energy reduction without any overall latency, and 26-39% at the same throughput when one-cycle latency is introduced, comparing to conventional static bus in a multi-cycle interconnect.

#### ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation and the Samsung Scholarship Foundation.

#### REFERENCES

- [1] R. Arunachalam, E. Acar, and S. Nassif, "Optimal shielding/spacing metrics for low power design," *Proc. of IEEE Computer Society Annual Symposium on VLSI*, pp. 167-172, 2003.
- [2] S.-C. Wong, T. G.-Y. Lee, D.-J. Ma, and C.-J. Chao, "An empirical three-dimensional crossover capacitance model for multilevel interconnect VLSI circuits," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 13, pp. 219-227, 2000.
- [3] <http://www.eas.asu.edu/~ptm/interconnect.html>
- [4] K. Nose and T. Sakurai, "Two schemes to reduce interconnect delays in bi-directional and uni-directional buses," *Symposium on VLSI Circuits Dig. Tech. Papers*, pp. 193-194, 2001.
- [5] K. Hirose and H. Yasuura, "A bus delay reduction technique considering crosstalk," *Proc. of Design, Automation and Test in Europe (DATE)*, pp.441-445, 2000.
- [6] M. Khellah, M. Ghoneima, J. Tschanz, Y. Ye, N. Kurd, J. Barkatullah, S. Nimmagadda, and Y. Ismail, "A Skewed repeater bus architecture for on-chip energy reduction in microprocessors," *Proc. of Int. Conf. on Computer Design (ICCD)*, pp. 253-257, 2005.
- [7] A. B. Kahng, S. Muddu, and E. Sarto, "Interconnect optimization strategies for high-performance VLSI designs," *Proc. of Int. Conf. on VLSI Design*, pp. 464-469, 1999.
- [8] H. Kaul, J. Seo, M. Anders, D. Sylvester, and Ram Krishnamurthy, "A robust alternate repeater technique for high performance busses in the multi-core era," *Proc. of Int. Symp. on Circuits and Systems*, pp. 372-375, 2008.
- [9] C. J. Akl and M. A. Bayoumi, "Reducing interconnect delay uncertainty via hybrid polarity repeater insertion," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 16, pp. 1230-1239, 2008.
- [10] B. Victor and K. Keutzer, "Bus encoding to prevent crosstalk delay," *Proc. of Int. Conf. on Computer-Aided Design (ICCAD)*, pp. 57-63, 2001.
- [11] P. P. Sotiriadis, A. Wang, and A. Chandrakasan, "Transition pattern coding: an approach to reduce energy in interconnect," *Proc. of European Solid-State Circuits Conference (ESSCIRC)*, pp. 348-351, 2000.
- [12] M. Khellah, J. Tschanz, Y. Ye, S. Narendra, and V. De, "Static pulsed bus for on-chip interconnects," *Symposium on VLSI Circuits Dig. Tech. Papers*, pp. 78-79, 2002.
- [13] H. Deogun, R. Senger, D. Sylvester, R. Brown, and K. Nowka, "A dual-Vdd boosted pulsed bus technique for low power and low leakage operation," *Proc. of Int. Symp. on Low Power Electronics and Design (ISLPED)*, pp. 73-78, 2006.
- [14] H. Kaul, D. Sylvester, M. Anders, and R. Krishnamurthy, "Design and analysis of spatial encoding circuits for peak power reduction in on-chip buses," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 13, pp. 1225-1238, 2005.
- [15] J. Seo, D. Sylvester, D. Blaauw, H. Kaul, and R. Krishnamurthy, "A robust edge encoding technique for energy-efficient multi-cycle interconnect," *Proc. of Int. Symp. on Low Power Electronics and Design (ISLPED)*, pp. 68-73, 2007.
- [16] A. B. Kahng, S. Muddu, and E. Sarto, "On Switch Factor Based Analysis of Coupled RC Interconnects," *Proc. of Design Automation Conference (DAC)*, pp. 79-84, 2000.
- [17] J. Eble, V. De, D. Wills, and J. Meindl, "Minimum repeater count, size, and energy dissipation for gigascale integration (GSI) interconnects," *Proc. of Int. Interconnect Technology Conference*, pp. 56-58, 1998.

- [18] H. Partovi, R. Burd, U. Salim, F. Weber, L. DiGregorio, and D. Draper, "Flow-through latch and edge-triggered flip-flop hybrid elements," *Dig. Tech. Papers on IEEE Int. Solid-State Circuits Conference (ISSCC)*, pp. 138-139, 1996.
- [19] J. Tschanz, S. Narendra, Z. Chen, S. Borkar, M. Sachdev, and V. De, "Comparative delay and energy of single edge-triggered and dual edge-triggered pulsed flip-flops for high-performance microprocessors," *Proc. of Int. Symp. on Low Power Electronics and Design (ISLPED)*, pp. 147-152, 2001.
- [20] K. Bowman, J. Tschanz, M. Khellah, M. Ghoneima, Y. Ismail, and V. De, "Time-borrowing multi-cycle on-chip interconnects for delay variation tolerance," *Proc. of Int. Symp. on Low Power Electronics and Design (ISLPED)*, pp. 79-84, 2006.
- [21] K. Bernstein, C.-T. Chuang, R. Joshi, and R. Puri, "Design and CAD challenges in sub-90 nm CMOS Technologies," *Proc. of Int. Conf. on Computer-Aided Design (ICCAD)*, pp. 129-136, 2003.