Pipeline Strategy for Improving Optimal Energy Efficiency in Ultra-Low Voltage Design

Mingoo Seok, Dongsuk Jeon, Chaitali Chakrabarti[1], David Blaauw, Dennis Sylvester
University of Michigan, Arizona State University[1]
mgseok@umich.edu

ABSTRACT
This paper investigates pipelining methodologies for the ultra low voltage regime. Based on an analytical model and simulations, we propose a pipelining technique that provides higher energy efficiency and performance than conventional approaches to ultra low voltage design. Two-phase latch based design and sequential circuit optimizations are also proposed to further improve energy efficiency and performance. Silicon results demonstrate a 16b multiplier using the approaches in 65nm CMOS improve energy efficiency by 30% and performance by 60%.

Categories and Subject Descriptors
B.2.1 [**Design Styles**]: Pipeline

General Terms
Algorithms, Design

Keywords
Ultra Low Voltage, Ultra Low Power, Pipeline, Super-pipeline

1. INTRODUCTION
Voltage scaling techniques have been one of the promising methods to minimize energy consumed in integrated circuits. As the supply voltage scales, quadratic to exponential energy savings in switch, subthreshold leakage, and gate leakage energy can be achieved. Although the scaled supply voltage also degrades circuit performance, many applications have relaxed performance requirements, such as implanted medical monitoring or building health monitoring. In these applications, we can reduce the supply voltage to near or below the threshold voltage (V_{th}), referred to as the ultra low voltage regimes, to maximize energy efficiency and prolong battery life. Complementary Metal Oxide Semiconductor (CMOS) gates have been known to be functional in this regime [1] and recently, several stable SRAM designs have been proposed [2][3].

One of the key goals in ultra low voltage operations is to operate at the most energy efficient supply voltage. Zhai [4] and Calhoun [5] showed that energy efficiency actually degrades if we scale the supply voltage too low since the increasingly slow circuits accumulate more and more leakage energy, which eventually offsets the quadratic savings of switch energy. Therefore, the total energy consumption starts to increase once the supply voltage scales down below a certain point, which we refer to as the energy optimal voltage or V_{min}. The optimal energy consumption at V_{min} is referred to as E_{min}, and is illustrated in Figure 1 with silicon measurements of a micro controller [7]. The V_{min} often lies at 0.35-0.45V for circuits in modern sub-micron CMOS technologies.

The energy optimal voltage poses *the* fundamental limit of energy efficiency through voltage scaling. In order to improve energy efficiency beyond this point, it is necessary to minimize leakage energy overhead which causes the saturation of improved energy efficiency. The leakage overhead (i.e., the proportion of total energy consumed by leakage) can be reduced by increasing performance or reducing leakage power since leakage energy is the integration of leakage power over a clock cycle. However, as noted in [4], uniformly reducing leakage of all gates in the design through increasing V_{th} does not actually change V_{min} and E_{min} (to a first order). In ultra-low voltage design, increasing V_{th} reduces leakage power exponentially but also increases circuit delay by the same amount, yielding the same leakage energy consumption per cycle. Hence, to reduce the leakage energy overhead, we need to minimize the number of idling gates during a clock cycle that unnecessarily contribute to leakage energy and/or reduce the cycle time without increasing leakage power. By doing so, the design becomes more switching energy dominated and we can extend the useful voltage scaling, which can establish a new energy efficiency limit beyond what is currently obtained.

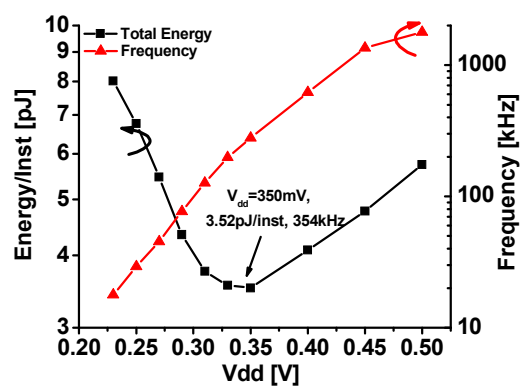

Figure 1. Energy and frequency of a microcontroller measured in silicon, showing to be energy optimal at 0.35V

In this paper, we explore pipeline methodology in ultra low power design space and propose so-called "super pipelining" to create switching dominated designs that have extended voltage scalability by ~25% and energy efficiency limits that are 30-50% reduced beyond traditional ultra-low voltage designs. The scheme also provides a simultaneous performance gain of 30-60%. Pipelining is a commonly employed technique to improve throughput in high performance design at the cost of energy efficiency or conversely to allow for increased Dynamic Voltage Frequency Scaling (DVFS) under a fixed performance constraint [8]. However, both techniques are aimed at the performance constrained super-threshold regimes. To our knowledge, this paper is the first to explore the use of aggressive pipelining in the ultra-low voltage re-

DAC'11, June 5-10, 2011, San Diego, California, USA

gime where there is no pressing performance constraint, and to show its efficacy for improving the fundamental energy efficiency limit by creating designs that are more switching energy dominated.

In contrast to our proposed scheme, traditional ultra-low voltage design has typically avoided pipelining due to variability concerns [13][14]. The drive current of MOSFETs becomes exponentially sensitive to V_{th}, V_{dd}, and temperature variations in ultra-low voltage design. These variations can cause up to 20× delay variability in a single gate delay, compared to nominal voltage operations [6]. Zhai [18] also showed that a large portion of this variation is due to Random Doping Fluctuation (RDF). By using more Fan-out-of-Four (FO4) delays per stage, this random variation can be averaged out, reducing the high sensitivity to variations in ultra low voltage design. In this paper, we therefore explore how to address this issue in highly pipelined designs with different clocking approaches, showing a 6× improvement in sensitivity to process variations over conventionally pipelined design. Finally, pipelining incurs overhead from extra synchronization elements. We therefore present circuit techniques reduce this overhead, resulting in 30% improvement in energy efficiency.

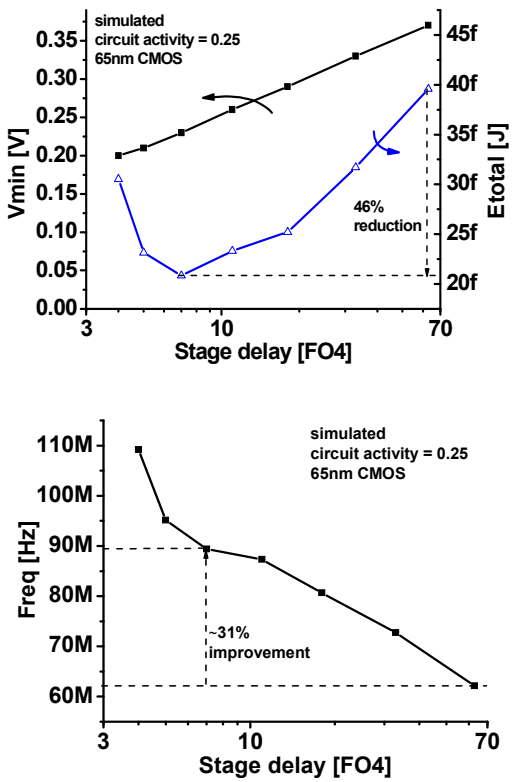

Figure 2. Improvement of energy efficiency and performance as pipelining increases from 1 to 16 in a 60FO4 inverter chain.

We demonstrate the effectiveness of all the proposed methods with silicon implementations as well as simulations of a highly pipelined multiplier in 65nm CMOS technology. The simulation and measurement results show that the proposed methods can reduce energy consumption by 30-50%, improve delay variability by 6× and increase clock frequency by 1.6-5.6×. Finally we also use this super-pipelined multiplier in a Fast Fourier Transform (FFT) core that operates at 30MHz with V_{dd}=0.27V to achieve a new lower limit on energy efficiency per FFT conversion.

In Section 2, we briefly review conventional pipelining both in super-threshold and ultra low voltage regimes. We then introduce the proposed pipelining including an analytical solution of optimal pipelining depth in Section 3. Section 4 discusses a 2-phase latch based design for mitigating variability with aggressively pipelined designs. Finally, we apply our proposed methodologies along with sequential overhead minimization to multipliers to confirm the effectiveness of the methodology in Section 5.

2. CONVENTIONAL PIPELINING

2.1 Pipelining in Nominal Voltage Operations

Pipelining is a well-known scheme to improve circuit performance. Splitting a circuit into multiple stages through register insertions increases the clock frequency linearly to a first order. Additionally, the gained performance from pipelining can be traded off for energy savings: after pipelining, designers can lower supply voltage to a certain points where a target performance is just met, achieving switching energy savings [7].

However, increasing pipeline stages causes energy overhead from both inserted registers and clock distribution. Also a higher number of pipelining stages can cause Cycle per Instruction (CPI) degradation when failed speculative operations need to flush pipeline stages in microprocessors.

The benefits and limitations of pipelining have lead to active investigations on performance- and power-constrained pipelining depth (i.e. the delay per a single pipeline stage) [9][10][11]. Hartstein and Hrishikesh [9][10] investigated the optimal pipeline depth for performance improvement with significantly increased power consumption. Srinivasan [11] suggested using less aggressive pipelining to balance power overhead. However, none of work focuses discusses the ultra low voltage operation regime which is energy constrained, and variability mitigation in aggressive pipelining schemes, which this paper primarily discusses.

2.2 Pipelining in Ultra Low Voltage Operations

Contrary to the pipelining practices in nominal voltage operations, low voltage designs have typically employed relaxed pipelining schemes (i.e. more FO4 delays per stage) due to two major benefits. First, the sequential overhead of both registers and clock distributions becomes much smaller with the relaxed pipeline. Given the large energy consumption in clock distribution and latches [12], this can greatly reduce overall energy consumption. Along with the power benefit, long paths per stage also help to mitigate performance variability through averaging of process variations over many gates. In this respects, a much relaxed pipeline in the range of 50-200 FO4 delays per stage, is often the preferred design choice for the recent ultra low voltage designs [13][14].

3. SUPER-PIPELINING STRATEGY

3.1 Concept of Super-Pipeline

Contrary to the conventional pipeline practice in ultra low voltage design, we propose to use significantly shorter pipeline stages and more pipeline registers, which counter-intuitively improves energy efficiency and performance simultaneously. This improvement is achieved by leakage energy reduction with the shorter clock period since leakage energy consumption is the integration of leakage power over a clock cycle. As discussed, this reduced leakage energy consumption extends the useful voltage scaling, which can result in extra switch energy savings. Therefore, as we increase the number of pipeline stages, we can reduce both switch and leakage energy consumption of circuits to obtain an improvement in total energy per operation.

However pipelining also increases the sequential energy overhead. Therefore, the benefit of pipelining on total energy consump-

tion saturates as the number of pipeline stages becomes larger than a certain point. In this respect, it is important to know the energy-optimal pipeline for a given design in ultra-low voltage operation.

3.2 Investigations with Inverter Chains

In order to confirm the validity of the super-pipelining concept, we perform a series of SPICE simulation experiments with 60 FO4 delay inverter chains. The circuit activity ratio is 0.25 which represents typical circuit activity. In these experiments, we increase the number of pipeline stages by inserting registers. Register delay overhead is assumed to be equivalent to a 3 FO4 inverter chain which uses 24 transistors. Note that Master-Slave Flip-Flops (MSFFs) typically use 20-28 transistors depending on topology.

Figure 2 shows that the energy optimal point, or V_{min}, is reduced, from 0.34V to 0.22V when pipeline depth decreases from 63 to 7 FO4 delays per stage (moving from un-pipeline to 16 stage pipeline). This results in an energy savings of ~46% due to the reduction in leakage and switch energy consumption. The experiment also shows that super-pipelining provides a moderate performance improvement of ~31% at the energy optimal supply voltages compared to un-pipeline, even though the supply voltage of the super-pipelined designs is 35% lower. Hence, the increased FO4 delay due to the lower V_{min} is offset by the performance improvement due to the pipelining itself, allowing us to run at lower energy and higher performance simultaneously.

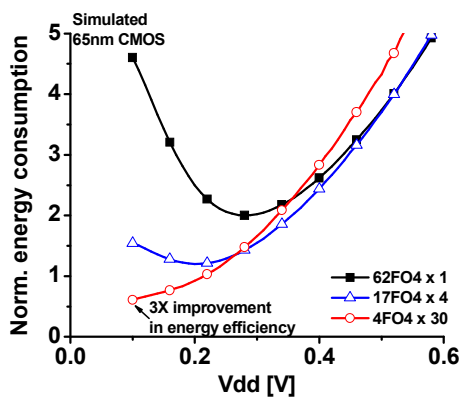

Figure 3. Minimum energy limit improves as sequential element overheads are reduced.

The previous experiments confirm the effectiveness of the proposed pipeline methodology in principle. However, practical circuits can be quite different from simple inverter chains. For example, the inverter chain has a linear increase of register count with increasing number of pipeline stages. However, many circuits have different growth of register count with increased pipeline stages. Srinivasan [11] investigate this issue and find that Latch Growth Factor (LGF) can be 0 to 2 and typically is 1.1, where actual register count increases proportional to the power of LGF or $(\text{constant})^{LGF}$. We can expect that circuits with higher LGF gain less benefit from the super-pipeline scheme since the faster increase of sequential overhead undermines the benefit of the pipeline scheme. Therefore, it is important to minimize the sequential overhead for a given circuit by finding optimal locations for register insertion.

In addition, we assume in the inverter experiment that the pipeline register is equivalent to 3 FO4 delays. However, this depends on the choice of register - e.g., Srinivasan [11] uses 2 FO4 delays for registers. Figure 3 shows how a smaller register overhead results in a higher optimal number of pipeline stages and lower overall energy limit. It is therefore critical to reduce the sequential overhead both in register count and register circuit overhead since it can enable effective pipelining. This minimization can be achieved through circuit as well as micro-architecture optimizations, which we discuss in Section 5 in the context of applying the scheme to a 16b multiplier and an FFT core.

3.3 Analytical Solution for Optimal Pipelining

The proposed pipeline scheme raises the need for finding the optimal number of pipeline stages for a given design without actually designing and simulating every different pipeline configuration. We approach this problem by modeling the total energy consumption as a function of total width of pipeline registers. This is motivated by the fact that leakage and switch energy consumptions are typically proportional to transistor widths. Once we find the optimal total width of the sequential elements for minimum energy, we can easily estimate the number of stages based on a given circuit topology.

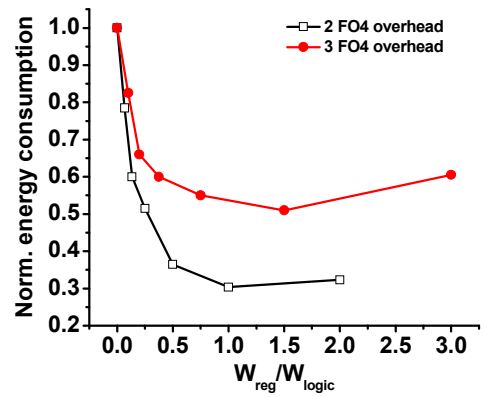

Figure 4. Energy efficiency improvement as a function of the total width of pipeline registers.

Figure 4 shows the SPICE simulation results of energy efficiency similar to Figure 2 except the x-axis is the width ratio of all pipeline registers to the other non-register circuitry. Small W_{reg}/W_{logic} represents less number of pipeline stages are used in the design while large W_{reg}/W_{logic} shows the design is heavily pipelined. Similar to Figure 2, too many register insertions, i.e. higher W_{reg}/W_{logic} results in energy efficiency degradation.

We derive the optimal W_{reg}/W_{logic} ratio, α_w, for an N-stage inverter chain. The switch energy consumption of the inverter chain is shown in EQ1 while EQ2 represents the leakage energy. α_w and α_d represents the extra switching capacitance and delay from added pipeline registers. In this N-stage inverter chain case, these two values are linearly related. The η is a fitting coefficient from Zhai [4] and $Coeff_{eff\_width}$ is a scaling constant for register width since all added capacitance from pipeline registers does not contribute switching and leakage energy consumption. For a simple MSFF, we use 0.67 for $Coeff_{eff\_width}$.

The total energy consumption is a sum of these two components (EQ1 and EQ2). To find the value of α_w resulting in the minimum energy consumption, we can differentiate the total energy consumption with respect to α_w, which is shown in EQ3. The α_w is a strong function of supply voltage: if voltage is high, smaller α_w is preferred since adding more pipeline stage always increase switching energy while leakage energy consumption is less important in this voltage regime. On the other hand, in the low voltage regime, optimal α_w becomes larger and approaches 1/k or α_w/α_d. Smaller k results in higher α_w: using more pipeline stages for energy optimal-

ity. In other words, if pipeline register induces less delay overhead for stages, we can use more stages for achieving an energy optimal design.

With the optimal α_w, we can calculate the total energy consumption at several supply voltages using EQ1 and EQ2, and then find the optimal supply voltage that gives the minimal energy consumption. Finally we can partition the given circuits and calculate the optimal number of pipeline stage with the guidance of the found α_w.

$$E_{switch} = N \cdot \frac{1}{2} \cdot C_{inv} V_{dd}^2 (1+\alpha_w) = \frac{1}{2} \cdot C_{tot,inv} V_{dd}^2 (1+\alpha_w) \quad \text{[EQ1]}$$

$$E_{leak} = t_{stage,logic} \cdot P_{tot,inv,leak} \cdot (1+\alpha_d)(1+\alpha_w) \quad \text{[EQ2]}$$

$$= \frac{N}{P} \cdot N \cdot \frac{1}{2} \cdot \eta \cdot C_{inv} \cdot V_{dd}^2 \cdot \exp(-\frac{V_{dd}}{mV_T}) \cdot (1+\alpha_d)(1+\alpha_w) \quad (\because [2])$$

$$= \frac{W_{ff}}{\alpha_w \cdot W_{inv}} \cdot \frac{1}{2} \cdot \eta \cdot C_{tot,inv} \cdot V_{dd}^2 \cdot \exp(-\frac{V_{dd}}{mV_T}) \cdot (1+k \cdot \alpha_w)(1+\alpha_w)$$

$$\frac{\partial E_{total}}{\partial \alpha_w} = \frac{1}{2} \cdot C_{tot,inv} V_{dd}^2 + \frac{1}{2} \frac{W_{ff}}{W_{inv}} \cdot \eta \cdot C_{tot,inv} \cdot V_{dd}^2 \cdot \exp(-\frac{V_{dd}}{mV_T}) \cdot (k - \frac{1}{\alpha_w^2}) = 0 \quad \text{[EQ3]}$$

$$\alpha_w = \sqrt{\frac{W_{ff}/W_{inv} \cdot \eta \cdot \exp(-\frac{V_{dd}}{mV_T})}{1 + W_{ff}/W_{inv} \cdot \eta \cdot \exp(-\frac{V_{dd}}{mV_T}) \cdot k}}$$

where N = number of inverters, C_{inv} = single inverter capacitance,

$C_{tot,inv} = N \cdot C_{inv}$, P = *number of pipeline stage*

W_{ff} = effective width of a flip-flop $= Coeff_{eff\_width} \cdot W_{ff,real} = 0.67 \cdot W_{ff,real}$

$$\alpha_w = \frac{W_{reg}}{W_{logic}} = \frac{P \cdot W_{ff}}{N \cdot W_{inv}}, \quad \alpha_d = \frac{t_{ff}}{N/P \, t_{inv}} = k \cdot \alpha_w$$

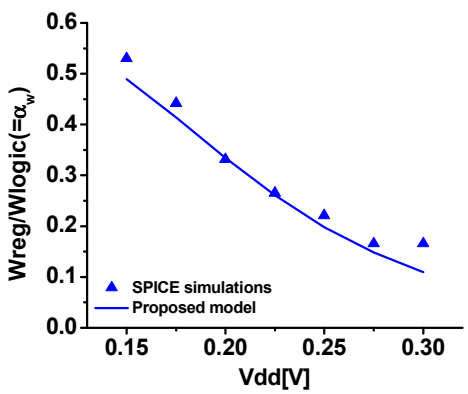

Figure 5. Optimal α_w for inverter chains via the proposed model and SPICE simulations

For confirming the validity of this model, we compare our model to the SPICE simulation results with inverter chains pipelined different stages. Figure 5 confirms that the model is well matched with SPICE simulation results. As expected, lower V_{dd} results in higher α_w since leakage energy consumption takes a dominant portion in total energy consumption.

4. LATCH-BASED DESIGNS

4.1 Mitigating Delay Variability in Ultra Low Voltage Operations

Delay variability is a critical issue for voltage scaling techniques. The source-drain current of MOSFETs exhibit large variability since subthreshold leakage current, which dominates the driving current in ultra low voltage regimes, is exponentially modulated by the changes in V_{th}, V_{dd} and temperature. Figure 6 shows that the delay variability from random process variations at scaled supply voltages is heightened by up to 4~7×, compared to nominal supply voltage operations.

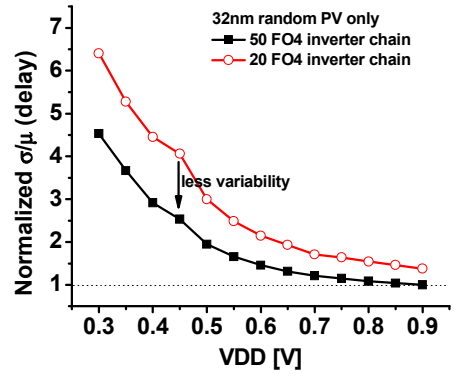

Figure 6. Delay variability from random process variations in the inverter chains of different length

These variations can be categorized as global and local variations. If a variation affects all the transistors in the same direction and magnitude, it is defined as global. Conversely, if every transistor experiences its own direction and magnitude, it is considered a local or random variation.

While delay variability from both variations are important and must be addressed, they require different methods. For global variations, knobs such as body bias and supply voltage have been shown to be effective [7]. However, these methods are ineffective for local variations since they impact all devices in the same way. A well-known method to address local variation is to use long pipeline stages since long paths average random variations through a series of gates. Figure 6 confirms the effectiveness of this method, showing ~30% reduction in delay variability using from 20 FO4 to 50 FO4 delay inverter chains at 0.3V.

4.2 Two-Phase Latch-Based Designs

Using long pipeline stages is contrary to the aggressive pipelining which can potentially improve the performance and energy efficiency as discussed in Section 3. Therefore, it is critical to mitigate the delay variability without resorting to long paths. It is not preferred to add a delay margin since the larger amount of delay variability compared to nominal voltage operations significantly hurts performance and energy efficiency.

We propose two-phase latch based pipeline instead of hard-edge flip-flop due to its well-known cycle borrowing ability [15]. The simple comparison of flip-flop and two-phase latch approaches is shown in Figure 7. The cycle borrowing can re-establish the averaging of random process variations through long paths that used to be present in less-pipeline design while still increasing clock frequency of circuits. It provides a cycle borrowing window that is slightly shorter than half the clock period. This large window is well-suited for the high variability in ultra low voltage circuits, compared to other soft-edge clocking approaches like soft-edge flip-flops and pulsed latches [16][17].

A hold time violation is one of the critical challenges in latch based design. Two phase latches have a hold time constraint when there is an overlap between clock and complementary clock signals while flip-flop often has negative hold time constraints and use a single clock. Non-overlapping clock generation is one way of eliminating hold time violations but can causes overhead in energy

consumption and design complexity. In Section 5, we discuss circuit techniques to eliminating hold time violations with less overhead.

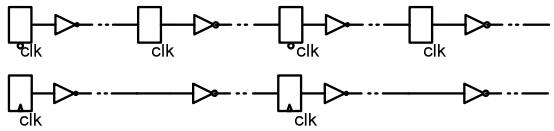

Figure 7. Sequential element choice: 2-phase latches, and flip-flops

5. APPLICATION TO MULTIPLIERS

In Sections 3 and 4, we propose super-pipeline and 2-phase latch schemes for improving energy efficiency, performance, and delay variability with simple inverter chains. In this section, we apply our methods to a more practical circuit, a 16b multiplier to confirm the applicability of the schemes. We also show the results of an FFT engine that uses the multiplier.

5.1 Super-pipelining

We first estimate the energy-optimal register counts or total width of pipeline registers for a 16b carry save multiplier. The multipliers in these experiments eliminate 14 Least Significant Bits (LSB), generating 18b outputs for two 16b inputs. The un-pipelined multiplier uses a Ripple Carry Adder (RCA), since it minimized total transistor width, thereby minimizing energy consumption in an un-pipelined design.

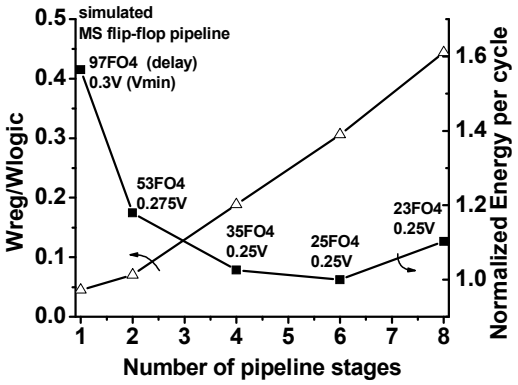

Figure 8. Total register width and energy consumption with different pipeline stages in multipliers

Figure 8 shows the register width and energy efficiency of the multipliers pipelined with MSFFs. More pipeline stages initially improve energy efficiency and then saturate due to sequential overheads when W_{reg}/W_{logic} becomes 0.2~0.3 (i.e. 4~6 stage). The energy optimal voltage reduces from 0.3V to 0.25V or 16% with more pipeline stages. The optimal width ratio matches with the FO4 inverter chain experiments in Figure 4 as well as the modeling in Figure 5 at 0.2-0.3V ranges.

5.2 Circuit Techniques and 2-Phase Latch Design

As briefly mentioned in Section 3.2, circuit techniques can reduce the overhead of registers and improve energy efficiency and performance further. In Figure 9, we apply several circuit optimizations from the 6 flip-flop pipelined multiplier (FF-6), which is the optimal design with basic MSFF pipelining. We first embed two registers in a full adder cell, saving 2 transistors per register (FF-6-EM). Sharing local clock buffers is another optimization (FF-6-SH) in addition to the embedding. The local clock buffer is implemented with multiple minimum-width fingers, which improve drivability at iso-switching power due to narrow width effect. We can replace the pipelined RCA with a Variable Carry Skip Adder (CSK) for final accumulations (FF-5-CSK), which works faster and thus consumes less energy despite containing more gates than RCA.

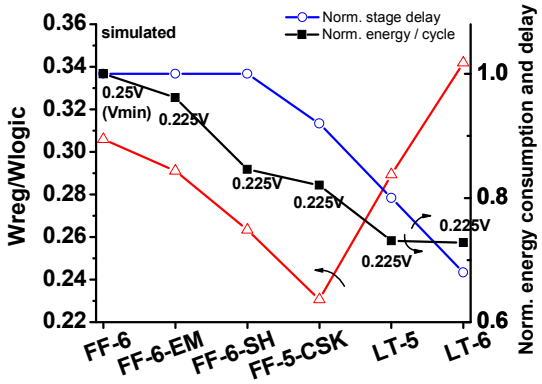

Figure 9. Improvement of energy efficiency and performance with circuit techniques and latch design

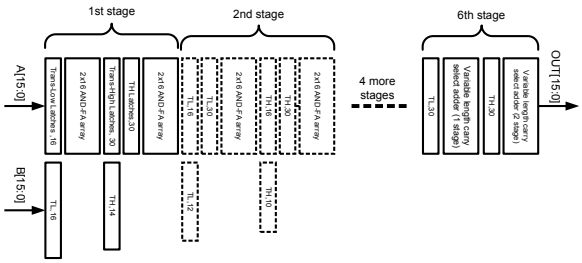

Figure 10. Diagram of a 6 stage multiplier

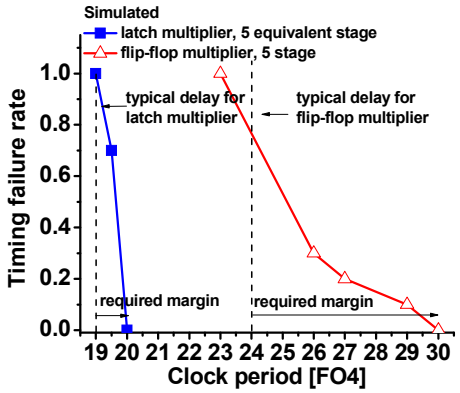

Figure 11. Required delay margin for process variations

We also investigate the use of 2 phase latches for pipeline registers. We replace MSFFs with 2-phase latches (LT-5) for utilizing cycle borrowing ability. The use of latch based design exhibits the increase of optimal W_{reg}/W_{logic} since two latches use more transistors than a single MSFF. However, k in EQ3 is smaller for latch based design. In other words, latches allow shorter stage delay due to cycle borrowing ability, which causes

higher optimal W_{reg}/W_{logic} ,as dictated by EQ3. Finally we add one more pipeline stage (LT-6). The proposed schematics of the LT-6 multiplier are shown in Figure 10, where 12 banks of latch banks are used for pipelining. The single stage takes 17FO4 delays.

The proposed multiplier (LT-6) with super-pipelining, circuit optimizations and latch-base design improves energy efficiency and performance by roughly 2×, compared to one stage multipliers at their own energy optimal voltage. At iso-V_{dd} the clock cycle can increase by 5.7×.

We also investigate if the latch-based design can mitigate delay variability. We run Monte-Carlo SPICE simulations to find the required delay margins against random process variations. As shown in Figure 11, the latch multiplier needs 6× smaller delay margin, compared to the flip-flop based multiplier of the same number of stages. The typical delay is also noticeably reduced with the latch design due to the cycle borrowing ability, which matches the results in Figure 9.

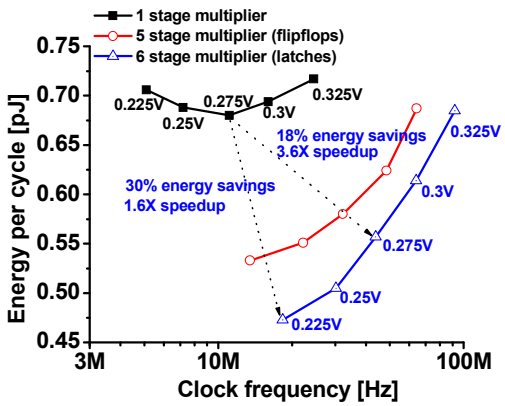

Figure 12. Measurement results of three multipliers

5.3 Fixing Hold Time Violations

As mentioned in Section 4.2, hold time violations must be eliminated in latch based design when non-overlapping clocks are not employed. In this respect, we identify the potential short paths and pad them with delay elements. The paths are verified with 150k random process Monte-Carlo and corner simulations at 0.2V, to guarantee ~99% functional yield for 2k path instances in the multipliers. This added delay elements causes 2.4% of energy overhead for the multiplier.

5.4 Measurement Results

Along with simulation results, we also fabricate three different multipliers in 65nm CMOS technology. The fabricated multipliers are the proposed 6 stage (LT-6), 1 stage baseline, and 5 stage flip-flop based multipliers (FF-5-CSK). The measurements for energy efficiency and performance are shown in Figure 11. The proposed multiplier outperforms the baseline by 30% in energy efficiency and 1.6× in performance when each operates at its energy optimal voltage. At the iso-voltage of 0.275V, the proposed multiplier still improves energy efficiency by 18% and performance by 3.6×. It is also shown that the latch based design has better energy efficiency and performance than the 5 stage flip-flop pipelined design. We also implement an FFT core with the proposed multipliers and achieve 17.7nJ per 16b 1024-pt complex FFT at 0.27V along with the remarkable performance of 30MHz.

The measurement results confirm the merits of using super-pipeline and 2-phase latches in the practical circuits. These techniques are also enhanced by circuit level techniques such as latch optimizations.

6. CONCLUSION

In this paper we investigate pipeline methodology for ultra low power design. We propose the use of an aggressively pipelined architecture for higher energy efficiency and performance, which is radically different from the existing practices in low voltage designs. Analytical modeling of simple inverter chains is presented. We also propose 2-phase latch design for mitigating delay variability. The effectiveness of these techniques is successfully demonstrated in the multiplier test chip in 65nm CMOS for improving energy efficiency by 18-30%, performance by 1.6-3.6×, and delay variability by 6×.

Acknowledgement

IC fabrication support of STMicroelectronics is gratefully acknowledged. Authors also acknowledge Multiscale Systems Center and Army Research Laboratory for their support.

References

[1] R. Swanson, et al., "Ion-implanted complementary MOS transistors in low-voltage circuits," *Journal of Solid-State Circuits*, Vol. 7, No. 2, pp. 146-153, 1972.

[2] B. Calhoun, et al., "A 256kb Sub-threshold SRAM in 65nm CMOS," *International Solid-State Circuits Conference*, pp.2592-2601, 2006

[3] I.-J. Chang, et al., "A 32kb 10T Sub-Threshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90nm CMOS," *Journal of Solid-State Circuits*, pp.650-658, 2009

[4] B. Zhai, et al, "Theoretical and Practical Limits on Dynamic Voltage Scaling", *Design Automation Conference*, 2004

[5] B. Calhoun et al., "Characterizing and modeling minimum energy operation for subthreshold circuits, " *International Symposium on Low Power Electronics and Design*, 2004

[6] M. Seok, et al., "CAS-FEST 2010: Mitigating Variability in Near-Threshold Computing," *Journal of Emerging Technology in Circuits and Systems*, 2011

[7] S. Hanson et al., "Performance and variability optimization strategies in a sub-200mV, 3.5pJ/inst, 11nW subthreshold processor," *Symposium on VLSI Circuits*, 2007.

[8] A. Chandrakasan, et al., "Low-Power CMOS Digital Design," Journal of Solid-State Circuits, vol.27, pp.473-484, 1992

[9] A. Hartstein et al., "The optimum pipeline depth for a microprocessor," *International Symposium on Computer Architecture*, May 2002.

[10] M. Hrishikesh, et al., "The optimal logic depth per pipeline stage is 6 to 8 FO4 inverter delays," *International Symposium on Computer Architecture*, pages 14–24, May 2002.

[11] V. Srinivasan, et al., "Optimal Pipelines for Power and Performance," *International Symposium on Microarchitecture* , 2002

[12] N. Magen, et al, "Interconnect power dissipation in a Microprocessor," *International Workshop on SLIP*, 2004

[13] A. Wang, et al., "A 180mV FFT Processor using Subthreshold Circuit Techniques," *International Solid-State Circuits Conference*, 2004

[14] M. Seok et al, "The Phoenix Processor: A 30pW Platform for Sensor Applications," *Symposium on VLSI Circuits*, 2008.

[15] D. Harris, "Skew-Tolerant Circuit Design," Morgan Kaufmann, 2000

[16] M. Wieckowski, et al., "Timing Yield Enhancement Through Soft Edge Flip-Flop Based Design," *Custom Integrated Circuits Conference*, Sep., 2008

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

[18] B. Zhai et al., "Analysis and Mitigation of Variability in Subthreshold Design," *International Symposium on Low Power Electronics and Design*, 2005