Exploring Asynchronous Design Techniques for Process-Tolerant and Energy-Efficient Subthreshold Operation

Ik Joon Chang, Sang Phill Park, *Student Member, IEEE*, and Kaushik Roy, *Fellow, IEEE*

***Abstract*—Supply voltage scaling is one of the easiest ways to reduce energy dissipation. Therefore, researchers have considered subthreshold logic as a promising option to achieve ultra low energy dissipation. However, circuit propagation delay is extremely sensitive to PVT variations under subthreshold operation. Hence, large delay margin is required for successful operation of conventional synchronous designs. Since leakage energy contributes to a substantial portion of total energy dissipation in subthreshold operation, the leakage energy dissipated for the required delay margin degrades energy efficiency significantly. In addition, even small intra-die variations result in large clock skew and hence, it is difficult to efficiently handle timing issues such as the setup and the hold time violations. In this work, we explore asynchronous design approach to address these challenges in subthreshold operation. We employ critical-path replica to generate completion signals of combinational logic blocks and use classical four-phase handshaking for communication between pipeline flip-flops. Since the proposed design approach uses only local clock buffers, it is easier to handle timing problems compared to synchronous designs. We compared iso-yield minimum energy dissipation of two design approaches (synchronous and asynchronous) in an inverter chain. Despite leakage overhead due to pad delay of critical-path delay line and "return-to-zero" time of four-phase handshaking, the proposed asynchronous design shows 71% energy savings compared to its synchronous counterpart. To demonstrate subthreshold operation of the proposed design approach, we fabricated an 8-tap FIR filter in 90 nm CMOS. Measured oscilloscope plots of handshaking and output bus signals show that the design operates successfully below 300 mV. We also measured energy consumption of the FIR filter from 19 test chips—the average was 4.64 pJ and the standard deviation was 0.3526 pJ.**

***Index Terms*—Asynchronous logic, CMOS digital circuit design, low power VLSI, subthreshold logic design.**

I. Introduction

THE scaling of supply power ($\mathrm{V_{DD}}$) is one of the simplest approaches to reduce energy dissipation of digital circuits. Since dynamic energy dissipation can be expressed as $\mathrm{CV_{DD}^2}$ (C: effective capacitance) [4], supply scaling gives quadratic energy savings. To achieve ultra low energy dissipation, researchers have lowered $\mathrm{V_{DD}}$ below the threshold voltage ($\mathrm{V_t}$) of transistors [1]. In subthreshold operation, circuit speed is slow. However, it can be applied to several applications such as implantable medical devices or wireless sensor networks, where speed is less critical while the battery life time is of utmost importance. In addition, the energy minimum point is in subthreshold region [2], making the subthreshold operation more promising for the machines in which battery longevity is critical. In previous works, subthreshold operation of digital circuits has been demonstrated with real silicon measurements. In [2], A. Wang *et al.* presented an FFT processor working at 180 mV. M. Hwang scaled down $\mathrm{V_{DD}}$ of an FIR filter down to 85 mV using a body-bias technique [3]. The designs were implemented using conventional synchronous design approaches. All pipeline registers are triggered by a common clock signal, whose cycle is determined by the worst propagation delay of combinational logic blocks. It should be noted that circuit propagation delay is extremely sensitive to *process, voltage and temperature* (PVT) variations in subthreshold operation, resulting in several challenging design problems for conventional synchronous designs. For example, large delay margins are required to cope with the extreme sensitivity of logic delay, deteriorating battery life significantly. In addition, even small intra-die variations in the clock buffer tree lead to large clock skew. Hence, it is difficult to handle timing issues such as setup and hold time violations.

In this work, we overcome these challenges using an asynchronous design approach. We present an asynchronous pipeline structure suitable for subthreshold operation employing *critical-path replica* (CPR) delay line to generate a completion signal for a combinational logic block. For the communication of pipeline flip-flops, classical four-phase handshaking [10], [11] is used. We employ parallel delay lines to mitigate the impact of random intra-die variations in CPR delay lines. Long-channel transistors are used for logic gates and flip-flops, improving reliability of subthreshold operation. To demonstrate the feasibility of our approach, we fabricated an 8-tap FIR filter based on our asynchronous design approach in 90 nm CMOS. Hardware measurement results show that the FIR filter operates successfully below 300 mV of $\mathrm{V_{DD}}$.

The rest of this paper is organized as follows. In Section II, we perform a detailed analysis of the challenging problems that conventional synchronous design experiences for subthreshold operation. In Section III, we consider an asynchronous design approach suitable for subthreshold operation. In Section IV, we compare our design approach to its synchronous counterpart, and explain how our design approach addresses the problems

Manuscript received March 26, 2009; revised October 05, 2009. Current version published February 05, 2010. This paper was approved by Associate Editor Stefan Rusu.

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906 USA (e-mail: ichang@purdue.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2009.2036764

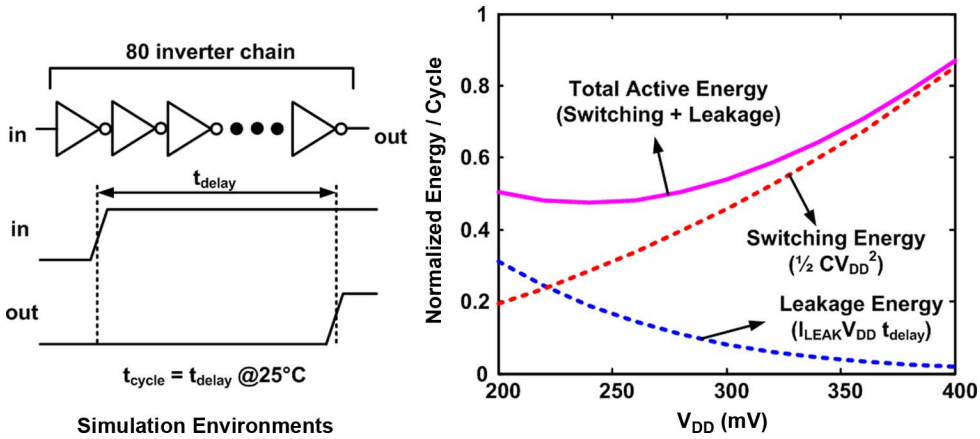


Fig. 1. Energy dissipation curve of an inverter chain is shown when operated at its optimal speed ($\mathrm{T} = 25°\mathrm{C}$).

mentioned in Section II. Then, we discuss fabrication and measurement results of an 8-tap FIR filter fabricated in 90 nm technology. Finally, Section VI concludes the paper.

II. Challenges in Synchronous Subthreshold Logic Design

A. Low Energy Efficiency

Fig. 1 shows the energy dissipation of an inverter chain when a clock cycle ($\mathrm{t_{cycle}}$) is same as the circuit propagation delay ($\mathrm{t_{delay}}$) of 25°C. As $\mathrm{V_{DD}}$ is lowered, the switching energy ($1/2\mathrm{CV_{DD}^2}$) shows quadratic reduction. Consequently, the total active energy dissipation also decreases. On the other hand, the active leakage energy ($\mathrm{I_{LEAK}V_{DD}\,t_{delay}}$, $\mathrm{I_{LEAK}}$: leakage current) increases with the scaling of $\mathrm{V_{DD}}$ due to the increase of $\mathrm{t_{delay}}$. For certain values of $\mathrm{V_{DD}}$, this increase becomes larger than the reduction of switching energy and hence, the overall active energy starts to rise despite $\mathrm{V_{DD}}$ scaling. As shown in the figure, this turning point appears in the subthreshold region. In the standby mode, circuit techniques such as MTCMOS [5] or supply-gating [15], [16] can be used to turn off logic blocks, making their standby leakage extremely small. In such a scenario, the total energy dissipation is mainly dependent on the active energy dissipation. Therefore, we can consider the minimum point of Fig. 1 as the energy optimal point of the inverter chain.

However, for successful operation of conventional synchronous systems, we need to put delay margins to cope with global PVT variations. Hence, the active energy dissipation of conventional synchronous systems can be expressed as $1/2\mathrm{CV_{DD}^2} + \mathrm{I_{LEAK}V_{DD}}(\mathrm{t_{delay}} + \mathrm{t_{margin}})$ ($\mathrm{t_{margin}} = \mathrm{t_{cycle}} - \mathrm{t_{delay}}$). Here, $\mathrm{t_{cycle}}$ is determined by the circuit delay under the worst global PVT situation. For example, to provide successful operation across temperature variations (from $-30°\mathrm{C} \sim 120°\mathrm{C}$), $\mathrm{t_{cycle}}$ should be at least as large as the circuit propagation delay at $-30°\mathrm{C}$; the worst temperature corner for subthreshold operation. Using this $\mathrm{t_{cycle}}$, we estimate the active energy dissipated at 25°C and compare its result (case II) to that of Fig. 1 (case I), shown in Fig. 2. This figure shows that the optimal active energy increases by 73% in case II as compared to case I, resulting from the increase of active leakage energy. As discussed in the introduction, circuit delay shows extremely large sensitivity to PVT variations in subthreshold operation. This makes $\mathrm{t_{cycle}}$ much larger than $\mathrm{t_{delay}}[25°\mathrm{C}]$, causing $\mathrm{t_{margin}}$ to have a larger value as compared to $\mathrm{t_{delay}}[25°\mathrm{C}]$. As a consequence, the leakage energy of $\mathrm{t_{margin}}(\mathrm{I_{LEAK}V_{DD}\,t_{margin}})$ is a significant portion of total active energy dissipation, degrading the energy efficiency of the synchronous design considerably.

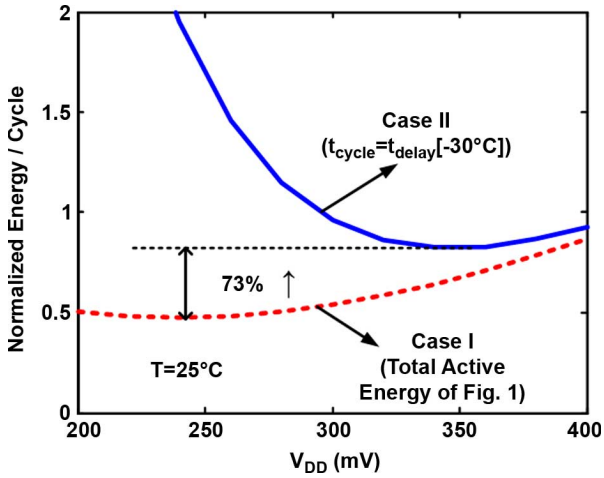


Fig. 2. To provide successful operation across temperature variations ($-30°\mathrm{C} \sim 120°\mathrm{C}$) in conventional synchronous designs, $\mathrm{t_{cycle}}$ should be at least as large as the circuit delay of $-30°\mathrm{C}$. In such a scenario, we estimated the active energy dissipation of an inverter chain at $25°\mathrm{C}$ and compared its result (Case II) to that of Fig. 1(Case I).

In previous literature, an adaptive body-biasing technique [6] has been employed to address this problem. This technique requires adaptive fine-grained body control units to mitigate global PVT variations. Unfortunately, it is extremely challenging to operate such control units successfully in the subthreshold region. For the success of this method, the control units need to be operated in the high $\mathrm{V_{DD}}$ domain, incurring large power penalties. In [7], a reference clock signal is generated using a CPR ring oscillator instead of a crystal oscillator. In such a scenario, the clock period is adjusted by global PVT variations. Since this alleviates the problem of excessive delay margin, we can achieve high energy efficiency. However, in such a scheme, the clock period is influenced by

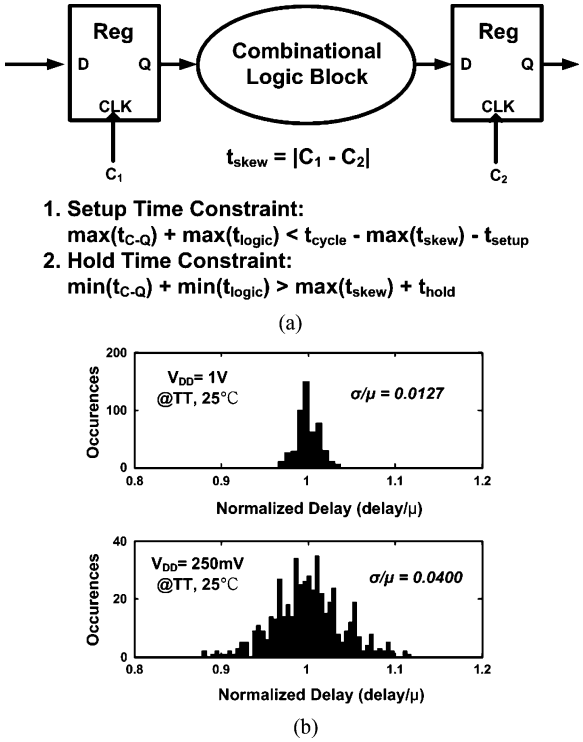


Fig. 3. (a) Timing constraints for successful operation of synchronous pipeline structure. (b) Monte Carlo simulation results of the propagation delay for an inverter chain having 80 inverters at 1 V and 250 mV. The results show that normalized deviation of the timing parameters (t_{skew}, t_{setup} and t_{hold}) increases with the scaling of V_{DD}.

external supply power noise. In [7], [9], in order to achieve high power efficiency, a switching power regulator based on pulse frequency modulation (PFM) method have been employed for subthreshold logic. It should be noted that the switching power regulator produces ripple power noise. Moreover, in the PFM method, the ripple power noise occurs non-periodically [8], where the logics are potentially exposed to considerable clock jitters. This makes it more difficult to handle timing issues; which is another challenge associated with synchronous subthreshold logic design, as discussed in the following subsection.

B. Difficulty in Handling Timing Issues

Fig. 3(a) shows two timing constraints for the successful operation of the synchronous design. Here, we need to consider that random intra-die variations result in deviations of timing parameters such as t_{skew}, t_{setup} and t_{hold}. As shown in Fig. 3(b), normalized deviation (σ/μ) of the delay parameters increases considerably as V_{DD} is scaled down. Hence, it is more difficult to satisfy the timing constraints in subthreshold operation compared to standard superthreshold operation. Setup time issue can be easily avoided by increasing t_{cycle}. However, such an approach incurs performance and leakage energy penalties. It should be noted that hold time issues result in functional failures regardless of t_{cycle}, making it extremely challenging to address the hold time problem in subthreshold operation. Large buffering for short paths may not be an efficient solution for the hold time violation due to large switching and leakage energy penalties resulting from the added buffers. To determine the proper length of the buffer, we need to develop accurate models

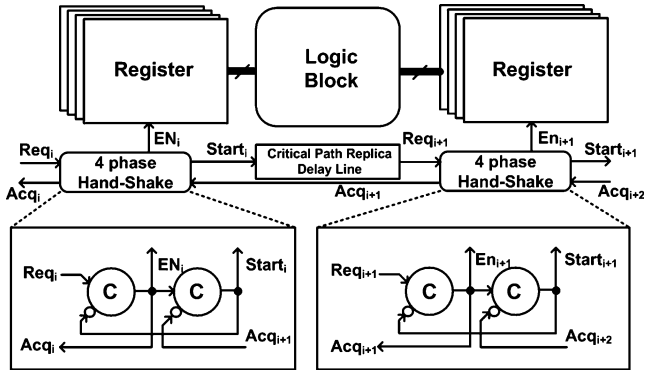


Fig. 4. Basic pipeline architecture of our asynchronous design approach.

of register hold time distribution. Unfortunately, the hold time is not well modeled as closed form expressions in subthreshold operation [9]. Although extensive Monte Carlo simulations give accurate distributions of the hold time, this method requires large simulation time. In [9], J. Kwong *et al.* applied Monte Carlo simulations for selected high-variance paths to reduce the simulation time. However, large number of Monte Carlo simulation is still required, degrading design time efficiency considerably.

III. Asynchronous Design Approach for Subthreshold Operation

A. Basic Pipeline Architecture

In the previous section, we showed that conventional synchronous designs suffer from low energy efficiency and timing problems in subthreshold operation. In this work, we explore an asynchronous design approach as a possible option to overcome these issues. Fig. 4 shows the basic pipeline architecture of our asynchronous design approach. Since this approach does not use a common clock signal, completion signals for each logic stage need to be locally generated. For instance, in dual-rail designs based on dynamic DCVSL logic style [11], a completion signal is enabled when one of two complementary parts completes its evaluation. This method is widely used in superthreshold asynchronous operation. However, the dynamic DCVSL logic style is exposed to leakage noise, and can result in malfunctions at extremely slow speed of subthreshold operation. Moreover, the dual-rail design causes large power dissipation. Hence, such a method may not be suitable for subthreshold operation, requiring strict power constraints. In our proposed work, completion signals are generated using CPR delay lines [4], as shown in the figure. To avoid the problems of dynamic logic style mentioned above, entire circuits are implemented as complementary static CMOS. For communication between pipeline stages, we employ classical four-phase handshaking [11], providing more reliable operation compared to other handshaking techniques [12], [17], [18].

It should be noted that four-phase handshaking has return-to-zero time, incurring performance and leakage energy penalties. Since performance has low priority in subthreshold operation, the performance penalty resulting from the return-to-zero time can be regarded as a minor problem. However,

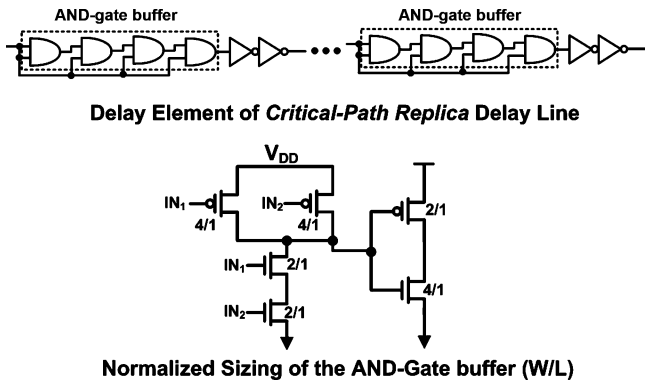


Fig. 5. To reduce the return-to-zero time in four-phase handshaking, the delay elements of CPR delay line are implemented as AND-gate type buffers with some added skew towards falling transition.

active leakage energy dissipated during this time can degrade battery life, which can be one of the most critical issues in subthreshold operation. As shown in Fig. 5, we mitigated this problem using AND-gate type buffers for the CPR delay line. To further reduce the return-to-zero time, some skew is added in the transistor sizing of the AND-gate towards a faster falling transition. Simulation results show that these approaches reduce the return-to-zero time below 1/10 of an operating cycle.

B. Parallel Delay Line

For successful operation of our design approach, the propagation delay of a CPR should be larger than that of a corresponding combinational logic. To cope with delay variations resulting from random intra-die variations, we need to add pad delay to the CPR delay line. As discussed in Section II-B, normalized deviation (σ/μ) of a circuit propagation delay increases as $\mathrm{V_{DD}}$ is scaled down. It implies that much larger pad delay is required for subthreshold operation than that of its superthreshold counterpart. In such a situation, leakage energy dissipated due to the large pad delay deteriorates total energy consumption. To alleviate this problem, we propose parallel delay line (PDL), as shown in Fig. 6. Here, the CPR delay line is divided to several stages. Each stage has two parallel delay elements, the outputs of which are AND-gated. Despite identical circuit structure of two parallel delay elements, random intra-die variations make them have different propagation delay. Since their output is AND-gated, the rising delay of a stage is determined by longer propagation delay of two parallel delay elements. Thus, the rising delay of ith $(\mathrm{i} = 1, \ldots, \mathrm{N_{stage}})$ stage can be expressed as $\mathrm{MAX}\ [\mathrm{t_{delay}(delay\ element\ I_i)},\ \mathrm{t_{delay}(delay\ element\ II_i)}] + \mathrm{t_{delay}(AND\ gate_i)}$. It is clear that the value of $\mathrm{MAX}\ [\mathrm{t_{delay}(delay\ element\ I_i)}, \mathrm{t_{delay}(delay\ element\ II_i)}]$ has smaller normalized deviation σ/μ compared to the propagation delay of single delay element. Hence, the PDL allows us to reduce the pad delay required for reliable operation. We verified our PDL technique using Monte Carlo simulations, where we assume that the CPR delay line has two stages. As shown in Fig. 7, the PDL provides 13.6% less σ/μ compared to the non-parallel counterpart.

It should be noted that the number of stages $(\mathrm{N_{stage}})$ also has impact on σ/μ of the CPR delay line. As we increase $\mathrm{N_{stage}}$, σ/μ of the CPR delay line improves gradually. However, beyond certain point, the delay of AND gates $(\mathrm{A_i}$'s) becomes more dominant than the parallel parts in total propagation delay of the CPR delay line. Then, the increment of $\mathrm{N_{stage}}$ starts to degrade σ/μ of the CPR delay line. Hence, we need to determine optimal $\mathrm{N_{stage}}$ using statistical methods such as Monte Carlo simulation. In our design approach, large combinational logic blocks share one CPR delay line. For example, in our fabricated 8-tap FIR filter, an 8-bit Multiplier and a 16-bit adder use a common CPR delay line. Therefore, the area and energy penalties incurred by the PDL are almost negligible.

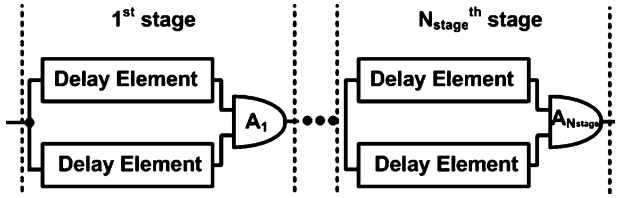


Fig. 6. We propose a parallel delay line technique to mitigate the impact of random intra-die variations in CPR delay lines.

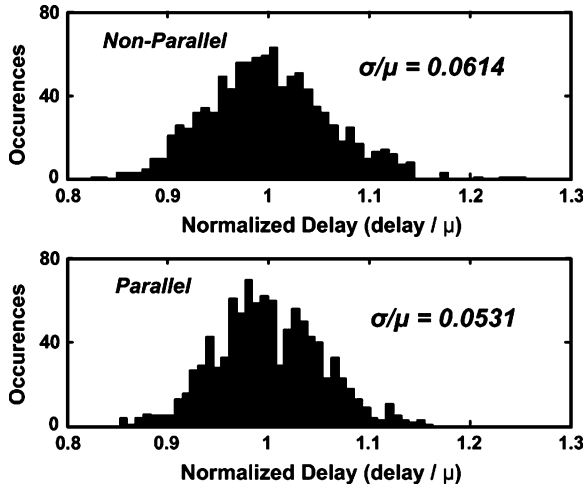


Fig. 7. 1000 Monte Carlo simulation results of a CPR delay ($\mathrm{V_{DD}} = 250$ mV, SS corner, $\mathrm{T} = -30^{\circ}\mathrm{C}$). Our PDL technique provides 13.6% less σ/μ as compared to non-parallel counterpart.

IV. Comparison to Synchronous Design Approach

In the previous section, we presented an asynchronous pipeline structure suitable for subthreshold operation. Since this scheme has only local clock buffers, it is much easier to address timing issues such as setup and hold time violations compared to synchronous pipeline structures. It should be noted that in the proposed asynchronous pipeline structure (Fig. 4), registers cannot accept new data before getting '0' *Acq* signal from the following pipeline stage. Therefore, the hold time constraint of Fig. 3(a) needs to be modified as $\min[\mathrm{t_{C-Q}}] + \min[\mathrm{t_{logic}}] + \min[\mathrm{t_{return-to-zero}}] > \max[\mathrm{t_{skew}}] + \mathrm{t_{hold}}$. Here, $\mathrm{t_{return-to-zero}}$ is as large as two Muller-C gate delay + falling delay of the CPR delay line. Since only local clock buffer is used, $\max[\mathrm{t_{skew}}]$ is much smaller compared to $\mathrm{t_{return-to-zero}}$. Consequently, the hold time constraint can be easily satisfied in our design approach. Hence, extensive Monte Carlo simulations, which are used

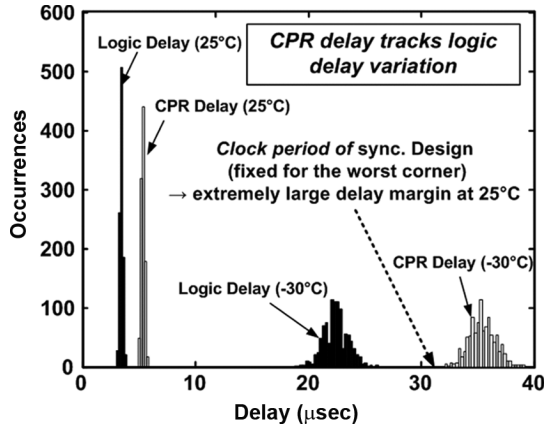


Fig. 8. 1000 Monte Carlo simulation results at 25°C and −30°C (with intra-die variations). While t_{cycle} of synchronous designs is fixed for the worst corner, CPR delay tracks logic delay variations. It alleviates the problem associated with large delay margin at nominal situations.

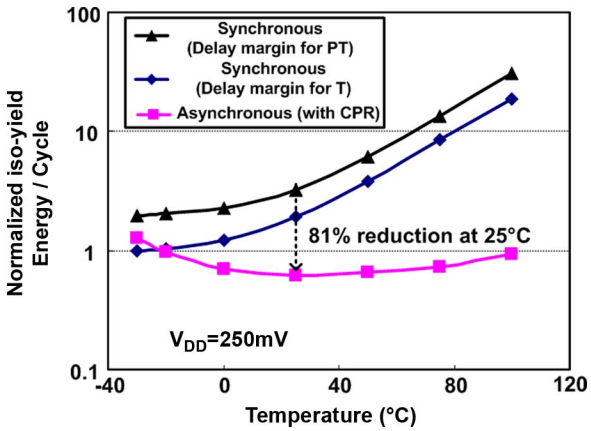


Fig. 9. We estimated the active energy dissipation of an inverter chain. Here, we made some estimation scenarios. In the synchronous design scenarios, we fixed t_{cycle} for the worst T or PT variations. In the asynchronous scenario, a cycle time is determined by the worst CPR delay from 1000 Monte Carlo simulation and its return to zero time.

in [9] to obtain the distribution of register hold time, are not necessary in our design approach. Since entire combinational logic blocks and handshaking circuitries are implemented as complementary static CMOS, it is possible to apply other timing methodologies used in synchronous subthreshold designs [9]. These imply that design time efficiency is improved significantly in our approach compared to its synchronous counterpart.

In addition, our design approach can efficiently handle low energy efficiency of synchronous subthreshold operation. In the proposed scheme, the delay of the CPR line tracks logic delay variations due to global variations, unlike the fixed clock period of synchronous designs for the worst corner, as shown in Fig. 8. This alleviates the problems associated with large delay margin, thereby improving active leakage energy significantly. To demonstrate that the proposed asynchronous design provides better energy efficiency than the synchronous design in subthreshold operation, we compare iso-yield energy dissipation per cycle of two design approaches (synchronous and asynchronous). Here, iso-yield refers to the fact that the two design

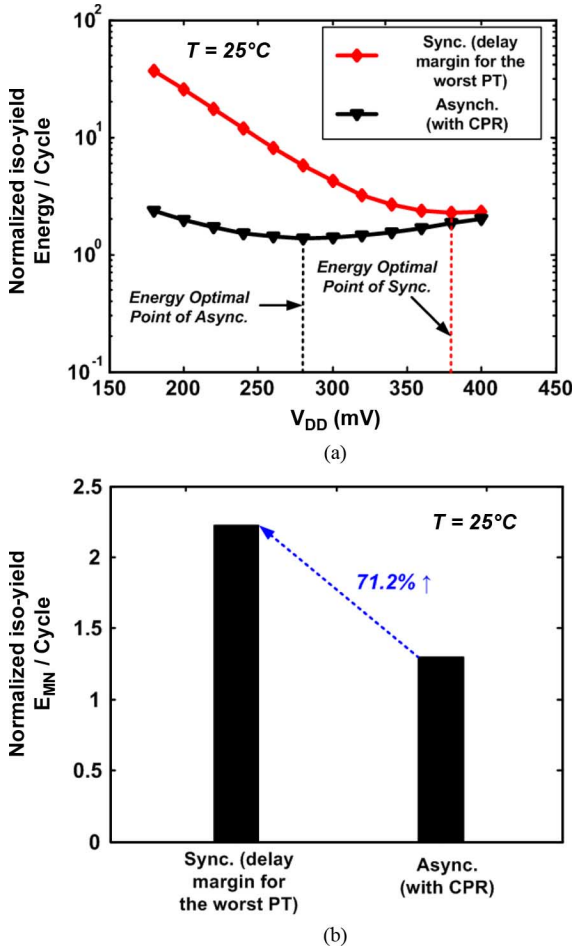


Fig. 10. (a) We simulated energy curves of two design approaches (synchronous and asynchronous). t_{cycle} of the synchronous one is determined by the circuit delay at the worst process and temperature corner. (b) We estimated and compared the minimum energies obtained in (a).

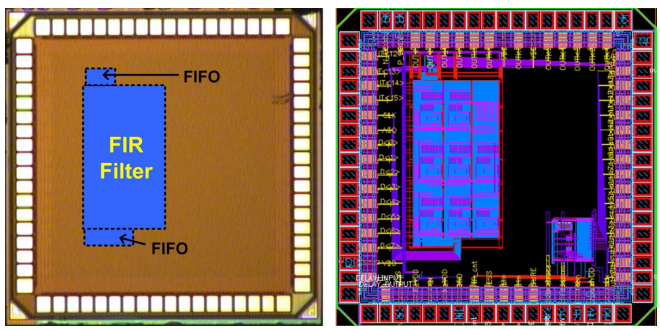


Fig. 11. Die micrograph and layout of the fabricated 8-tap FIR filter.

approaches show same delay margin for 1000 Monte Carlo simulations at the worst global variation corner. For the combinational logic, we employ an inverter chain having 80 inverters. Fig. 9 shows the comparison results simulated at 250 mV of V_{DD}. In these simulations, we considered only process and temperature $(-30°C \sim 120°C)$ variations. When we determine the cycle time of our design approach, we include the return-to-zero time of four-phase handshaking. Moreover, we use the worst CPR delay from 1000 Monte Carlo simulations, providing the worst energy scenario of our design approach. Nonetheless, the

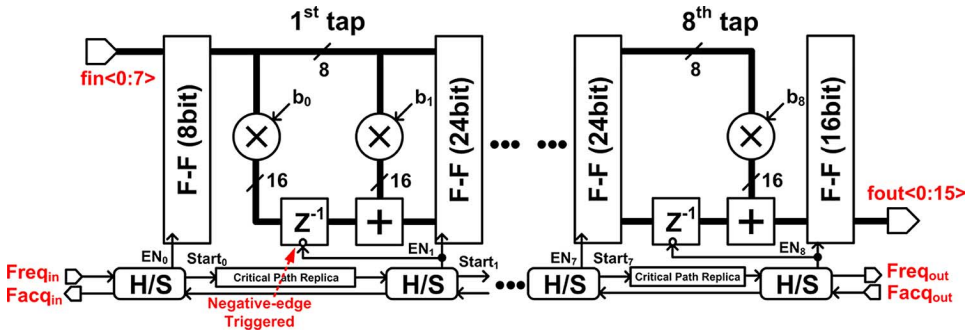


Fig. 12. Architecture of the fabricated FIR filter.

simulation results show that the proposed asynchronous design dissipates lower energy than the synchronous counterpart. We also compared iso-yield optimal energies of the two design approaches at 25°C. As shown in Fig. 10, our design approach shows 71.2% energy savings compared to the corresponding synchronous one at the operating scenario of the optimal energy point. Moreover, the optimal energy point becomes lower by 100 mV in our design approach, reducing standby leakage current considerably. As discussed in Section II-A, supply-gating or MTCMOS technique can improve standby leakage in logic blocks. However, this method cannot be applied for memory arrays in subthreshold operation due to stability issues [9]. Since standby leakage power is substantial in total memory power, the standby leakage reduction significantly contributes to improving energy efficiency.

V. Fabrication and Measurement of an 8-Tap FIR Filter

A. An 8-Tap FIR Filter Design and Test Chip Implementation

As an application of the proposed design approach, we fabricated an 8-Tap FIR filter in 90 nm CMOS (Fig. 11). Fig. 12 shows the architecture of the FIR filter, which is based on the asynchronous pipeline structure of Fig. 4. Here, one tap of the FIR filter is implemented as a pipeline stage. To reduce the overhead of handshaking circuitries such as Muller-C gates [4] and a CPR delay line, they are shared by the entire combinational logic and registers of one pipeline stage. Since the optimal energy point of the FIR filter appears at 250 mV V_{DD} in the simulation (@TT corner, 25 °C), we optimized the design for 250 mV operation. Here, N_{stage} of the CPR delay line is determined to be two since the minimum value of σ/μ is observed at this value, as shown in Fig. 13. To track systemic intra-die variations, we placed the CPR delay line in close proximity to the corresponding combinational logic blocks. Considering $\pm 3\sigma$ variations of the logic delay and the CPR delay, we added pad delay to the CPR delay line. To cope with local clock skew and setup time issues, some delay margins are required. Extensive Monte Carlo simulations show that 5% margin can address these issues successfully. It should be noted that delay units (Z^{-1}) are synchronized to the pipeline flip-flops, where the hold time violation may occur. To prevent hold time violation, we implemented the Z^{-1}'s as negative-edge triggered flip-flops. The Z^{-1}'s are activated after triggering of the pipeline flip-flops and hence, hold time violations can be avoided without any energy or performance penalties.

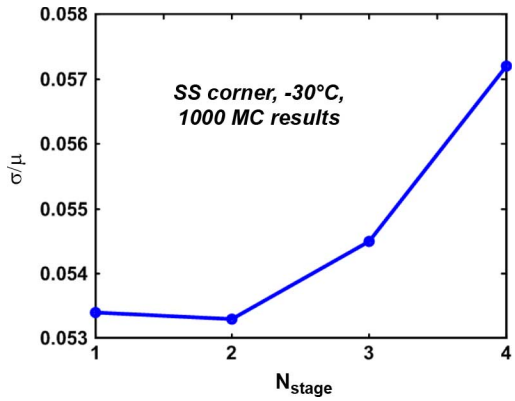
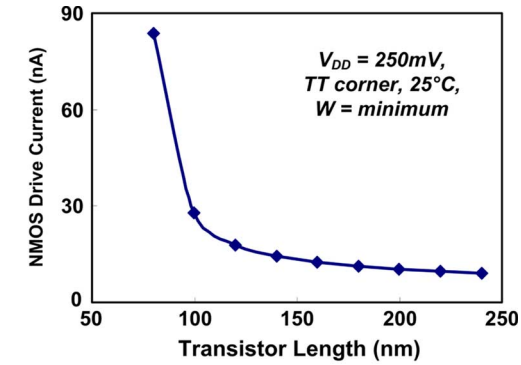
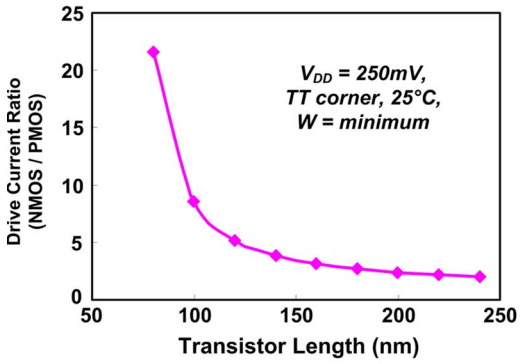


Fig. 13. σ/μ versus N_{stage} (@ 250 mV V_{DD}). This results show that two is the optimal value of N_{stage}.



(a)



(b)

Fig. 14. Since nMOS shows extreme Vt-roll-off [14] in this 90 nm CMOS, pull-down of nMOS is much stronger than pull-up of pMOS in short-channel length region. (a) Minimum width nMOS drive current in 90 nm CMOS. (b) Drive current ratio between minimum width nMOS and pMOS.

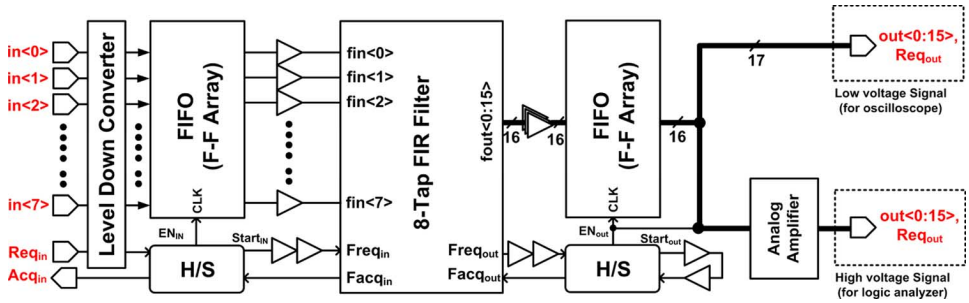
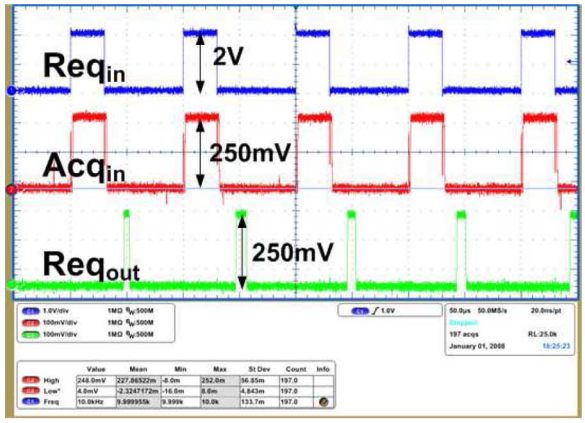


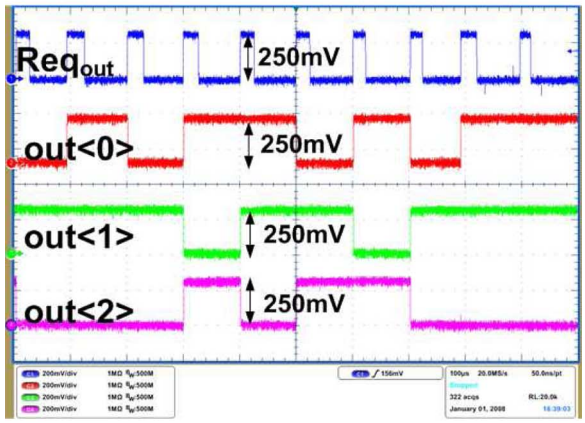
Fig. 15. Test-chip organization.

In sub-100 nm technologies, V_t is highly correlated to length and width of transistors [13]. Since subthreshold current of a MOSFET is an exponential function of V_t, it has profound impact on the reliability of subthreshold operation. For example, in this 90 nm technology, nMOS transistors experience large V_t roll-off [13] in short-channel length region, as shown in Fig. 14(a). It makes pull-down transistor much stronger than pull-up transistor at short-channel lengths [Fig. 14(b)]. Under such circumstances, the stability of logic gates and flip-flops degrades significantly. In addition, the short-channel V_t roll-off makes their stability more sensitive to other parameter variations such as line-edge roughness [14]. To alleviate these problems, we increased the basic length of transistors in this work. Since the drive current ratio between nMOS and pMOS starts to stabilize around 120 nm gate length, we utilize this length as the basic size of logic gates and flip-flops in this design.

Fig. 15 illustrates the organization of the test-chip. At the input and output stages of the FIR filter, FIFO arrays are placed to communicate with test equipments which are synchronous. For the generation of input signals and the analysis of the output signals, we employ a pattern generator and a logic analyzer. Since these test equipments require at least 2 V swing for their input and output signals, we inserted level-down converters (for input) or analog amplifiers (for output) at I/O pads. In the test chip, we separated the body of pMOS and nMOS transistors from power lines in order to control their body voltage. Due to the small number of available test-chips and the limitation of our test environment, we could not obtain real worst case process or temperature corner. However, the control of the body voltage allows us to emulate different process corners. In our simulation, ± 150 mV body-biasing at typical process corner provides drive current comparable to Fast nMOS-Fast pMOS (FF) and Slow nMOS-Slow pMOS (SS) process corners, respectively.



(a)

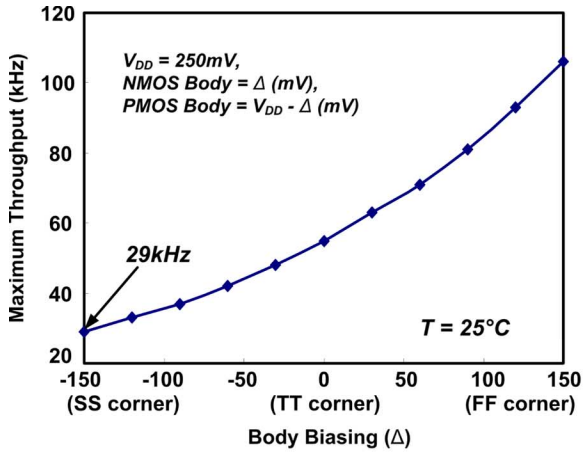


(b)

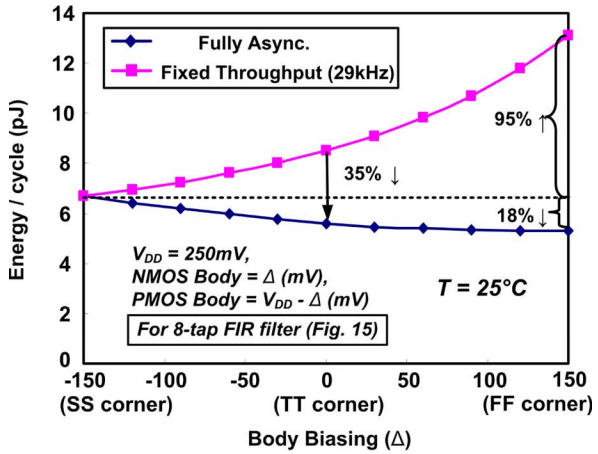
Fig. 16. Measured oscilloscope plots. (a) Handshaking signals. (b) Output bus signals.

B. Measurement Results

For the first measurement, we fixed V_{DD} at 250 mV since the design is optimized for 250 mV operation. Throughout the measurements, the ambient temperature was maintained at $25 \sim 27^{\circ}$C. Fig. 16 shows the measured oscilloscope plots of handshaking and output signals, which shows that the FIR filter operates successfully at 250 mV. The performance measurement results of the FIR filter is shown in Fig. 17(a). Since our design does not have a clock signal unlike synchronous designs, we measured maximum throughput for performance instead of clock frequency. Based on this performance result, we can consider two scenarios for the energy measurement: fully asynchronous and synchronous. In the fully asynchronous scenario, the throughput can be adjusted by global variations. Hence, we measured the active energy dissipation at the maximum throughput corresponding to each process corner. However, the throughput of the synchronous scenario was fixed for SS corner since the clock cycle is determined by the operating speed of the worst corner in conventional synchronous designs. Fig. 17(b) shows the energy measurement results of the two scenarios, where the energy dissipation of the synchronous scenario increases gradually as the process corner becomes

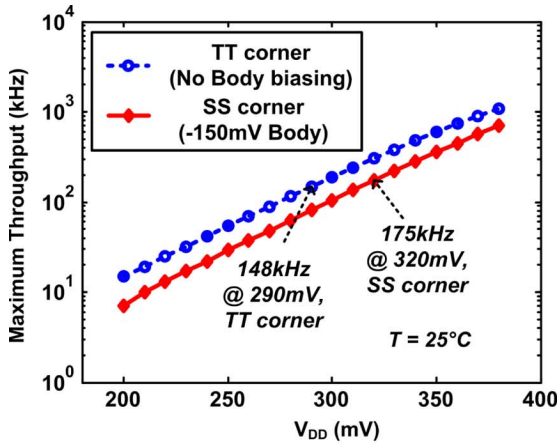


(a)

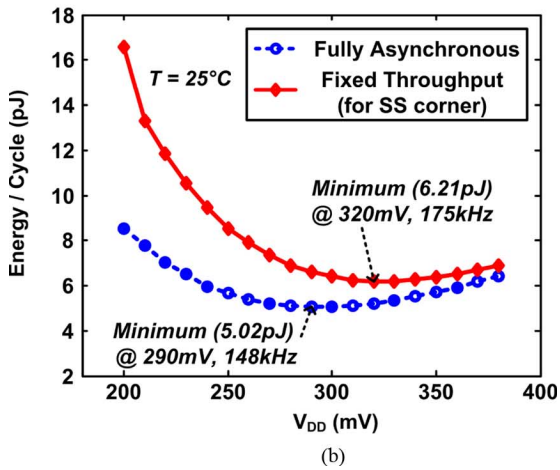


(b)

Fig. 17. We measured (a) performance and (b) active energy dissipation of the FIR filter at 250 mV V_{DD}. For the energy measurement, we consider two scenarios, which are fully asynchronous and synchronous. In the fully asynchronous scenario, we measured the energy at the throughput corresponding to each process corner. However, the throughput was always fixed for SS corner (29 kHz) in the synchronous one.



(a)



(b)

Fig. 18. (a) Performance measurement results at typical process corner and SS corner. (b) Based on the measured performance in (a), we measured energy curves of two measurement scenarios (full asynchronous and synchronous with a fixed throughput).

faster. As a result, the energy dissipation of FF corner is 95% larger compared to that of the SS corner. On the other hand, the fully asynchronous scenario shows almost constant energy dissipation (below 18% variation) despite extreme process corner variation. These observations are due to the improvement of active leakage energy in the asynchronous design as discussed in Section IV.

We also compared the minimum active energy dissipation for the two measurement scenarios. We measured the performance for typical process corner (no body biasing, for asynchronous scenario) and SS corner (−150 mV body biasing, for synchronous scenario). The results of the performance measurements are shown in Fig. 18(a). For the energy measurement, the body voltage of pMOS and nMOS transistors are connected to V_{DD} and ground lines, respectively. Then, we varied the throughput corresponding to the measurement scenario and V_{DD}, and measured active energy dissipation [Fig. 18(b)]. In these curves, we observe the minimum energy points of the two measurement scenarios at 290 mV and 320 mV, respectively. Here, the fully asynchronous design (dissipation of 5.02 pJ @ 290 mV, 148 kHz) shows 23.7% energy savings compared to the synchronous scenario (6.21 pJ @ 320 mV, 175 kHz). It should be noted that in these measurements, we considered only inter-die variation. It is clear that the energy saving will further increase when the impact of temperature and voltage variations are included in the measurements. Since we do not use real synchronous design, the measurement results above cannot show real energy comparison between an asynchronous design and its synchronous counterpart. However, they successfully demonstrate that the timing margin of synchronous designs has profound impact on active energy dissipation and our asynchronous design approach can handle this problem efficiently. Fig. 19 shows minimum active energy measurement results from 19 chips at room temperature, where the average energy dissipation (μ) was 4.64 pJ and the standard deviation (σ) was 0.3526 pJ.

VI. Conclusion

We present an asynchronous design approach suitable for subthreshold operation. Previous implementation of subthreshold logic mainly used conventional synchronous design methodology. Hence, the clock period has to be fixed for the

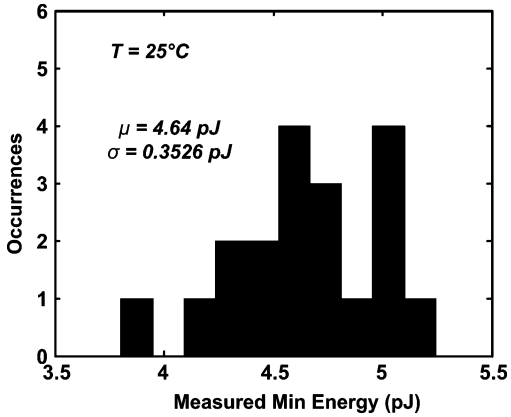


Fig. 19. Statistical data of the optimal energies measured from 19 chips.

worst PVT corner. Since the circuit propagation delay of subthreshold operation is extremely sensitive to PVT variations, the design approach results in excessively large delay margin. In such a situation, the leakage energy (corresponding to the delay margin) deteriorates total energy dissipation considerably. Moreover, even small intra-die variations in the clock buffer trees lead to large clock skews, making it difficult to handle timing issues, such as setup and hold time violations. In the proposed design approach, a critical-path replica delay line is used to generate the completion signal of the corresponding logic block. The delay of the critical-path replica delay line is adjusted by global PVT variations. This alleviates the delay margin problem associated with subthreshold operation, providing high energy efficiency. In addition, timing issues are regarded as minor problems in the asynchronous design since it uses only local clock buffers. Hence, our design approach allows us to handle the setup and hold time violations more easily as compared to its synchronous counterpart. For an inverter chain, we simulated optimal energy dissipation of the two design approaches, where the clock period of the synchronous design is fixed for the worst process and temperature variations. The results showed that our design approach dissipates 71.2% lower energy than its synchronous counterpart at the nominal condition (typical process corner and room temperature). To demonstrate the feasibility of the proposed design approach, we fabricated an 8-tap FIR filter in 90 nm CMOS. The measured oscilloscope plots show that the FIR filter operates successfully below 300 mV V_{DD}. We measured minimum energy dissipation of the FIR filter from 19 chips—the average energy dissipation was 4.64 pJ and the standard deviation was 0.3526 pJ.

ACKNOWLEDGMENT

The authors thank Hyungmin Lee of Georgia Tech for helping with the measurements.

REFERENCES

[1] H. Soeleman and K. Roy, "Ultra-low power digital subthreshold logic circuits," in *Proc. ISLPED*, 1999, pp. 94–96.
[2] A. Wang and A. P. Chandrakasan, "A 180-mV subthreshold FFR processor using a minimum energy design methodology," *IEEE J. Solid-State Circuits*, vol. 40, pp. 310–319, Jan. 2005.
[3] M. Hwang, A. Raychowdhury, K. Kim, and K. Roy, "A 85 mV 40 nW process-tolerant subthreshold 8 × 8 FIR filter in 130 nm technology," in *Symp. VLSI Circuits Dig.*, Jun. 2007, pp. 154–155.
[4] J. Rabaey, A. Chandrakasan, and B. Nicolic, *Digital Integrated Circuits*. Englewood Cliffs, NJ: Prentice Hall, 2003.
[5] K. Roy and S. C. Prasad, *Low-Power CMOS VLSI Circuit Design*. New York: Wiley-Interscience, 2000.
[6] S. Hanson *et al.*, "Exploring variability and performance in sub-200-mV processor," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 881–891, Apr. 2008.
[7] Y. K. Ramadass and A. P. Chandrakasan, "Minimum energy tracking loop with embedded DC-DC converter delivering voltages down to 250 mV in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 256–265, Jan. 2008.
[8] A. J. Stratakos, "High-efficiency low-voltage DC-DC conversion for portable applications," Ph.D. dissertation, Univ. of California, Berkeley, 1998.
[9] J. Kwong *et al.*, "A 65 nm sub-Vt microcontroller with integrated SRAM and switched capacitor DC-DC converter," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 115–126, Jan. 2009.
[10] I. E. Sutherland, "Micropipelines," *1988 Turing Award Lecture, ACM*, vol. 32, no. 6, pp. 720–738, Jun. 1989.
[11] G. M. Jacobs and R. W. Brodersen, "A fully asynchronous digital signal processor using self-timed circuits," *IEEE J. Solid-State Circuits*, pp. 1526–1537, Dec. 1990.
[12] M. Ferretti and P. A. Beerel, "High performance asynchronous design using single-track full-buffer standard cells," *IEEE J. Solid-State Circuits*, pp. 1444–1454, Jun. 2006.
[13] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. New York: Cambridge Univ. Press, 2002.
[14] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness," *IEEE Trans. Electron Devices*, vol. 50, pp. 1254–1260, May 2003.
[15] M. Johnson, D. Somasekhar, and K. Roy, "Models and algorithms for bounds on leakage in CMOS circuits," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, pp. 714–725, Jun. 1999.
[16] Y. Ye, S. Borkar, and V. De, "A new technique for standby leakage reduction in high performance circuits," in *Symp. VLSI Circuits Dig.*, Jun. 1998, pp. 40–41.
[17] I. E. Sutherland and S. Fairbanks, "GasP: A minimal FIFO control," *Proc. ASYNC*, pp. 46–53, 2001.
[18] M. Singh and S. M. Nowick, "High performance asynchronous pipelines for fine-grain dynamic datapaths," *Proc. ASYNC*, pp. 198–209, 2000.

Ik Joon Chang received the B.S degree in electrical engineering from Seoul National University, Seoul, Korea (*summa cum laude*) and the M.S. degree from the School of Electrical and Computer Engineering, Purdue University, IN, in 2002 and 2005, respectively. Since 2005, he is pursuing the Ph.D. at Purdue University.

He worked with LG Electronics as a software engineer in 2003. During summer and fall of 2007, he had an internship in Qualcomm, CA, where he performed research related to low power circuit technique. He is awarded from Samsung Scholarship Foundation in 2005. His current research interest is robust and low power circuit design in nano-scaled CMOS technologies.

Sang Phill Park (S'07) received the B.S. degree in computer engineering from University of Arizona, Tucson, AZ, in 2004. He is currently working toward the Ph.D. degree in electrical engineering at Purdue University, West Lafayette, IN.

In summer of 2008, he was with Exploratory VLSI design group in IBM Austin Research Laboratory as an intern. His research interests include variation tolerant circuit design and ultra low power circuit design.

Kaushik Roy (F'02) received B.Tech. degree in electronics and electrical communications engineering from the Indian Institute of Technology, Kharagpur, India, and Ph.D. degree from the electrical and computer engineering department of the University of Illinois at Urbana-Champaign in 1990. He was with the Semiconductor Process and Design Center of Texas Instruments, Dallas, where he worked on FPGA architecture development and low-power circuit design. He joined the electrical and computer engineering faculty at Purdue University, West Lafayette, IN, in 1993, where he is currently a Professor and holds the Roscoe H. George Chair of Electrical & Computer Engineering. His research interests include VLSI design/CAD for nano-scale Silicon and non-Silicon technologies, low-power electronics for portable computing and wireless communications, VLSI testing and verification, and reconfigurable computing. Dr. Roy has published more than 450 papers in refereed journals and conferences, holds 8 patents, and is coauthor of two books on Low Power CMOS VLSI Design (John Wiley & McGraw Hill).

Dr. Roy received the National Science Foundation Career Development Award in 1995, IBM faculty partnership award, ATT/Lucent Foundation award, 2005 SRC Technical Excellence Award, SRC Inventors Award, Purdue College of Engineering Research Excellence Award, and best paper awards at 1997 International Test Conference, IEEE 2000 International Symposium on Quality of IC Design, 2003 IEEE Latin American Test Workshop, 2003 IEEE Nano, 2004 IEEE International Conference on Computer Design, 2006 IEEE/ACM International Symposium on Low Power Electronics & Design, and 2005 IEEE Circuits and system society Outstanding Young Author Award (Chris Kim), 2006 IEEE Transactions on VLSI Systems best paper award. Dr. Roy is Purdue University Faculty Scholar. Dr. Roy was a Research Visionary Board Member of Motorola Labs (2002). He has been in the editorial board of IEEE Design and Test, IEEE Transactions on Circuits and Systems, and IEEE Transactions on VLSI Systems. He was Guest Editor for Special Issue on Low-Power VLSI in the IEEE Design and Test (1994) and IEEE Transactions on VLSI Systems (June 2000), IEE Proceedings—Computers and Digital Techniques (July 2002).