A 65 nm Sub-V_t Microcontroller With Integrated SRAM and Switched Capacitor DC-DC Converter

Joyce Kwong, *Student Member, IEEE*, Yogesh K. Ramadass, *Student Member, IEEE*, Naveen Verma, *Student Member, IEEE*, and Anantha P. Chandrakasan, *Fellow, IEEE*

***Abstract*—Aggressive supply voltage scaling to below the device threshold voltage provides significant energy and leakage power reduction in logic and SRAM circuits. Consequently, it is a compelling strategy for energy-constrained systems with relaxed performance requirements. However, effects of process variation become more prominent at low voltages, particularly in deeply scaled technologies. This paper presents a 65 nm system-on-a-chip which demonstrates techniques to mitigate variation, enabling sub-threshold operation down to 300 mV. A 16-bit microcontroller core is designed with a custom sub-threshold cell library and timing methodology to address output voltage failures and propagation delays in logic gates. A 128 kb SRAM employs an 8 T bit-cell to ensure read stability, and peripheral assist circuitry to allow sub-V_t reading and writing. The logic and SRAM function in the range of 300 mV to 600 mV, consume 27.2 pJ/cycle at the optimal $V_{\rm DD}$ of 500 mV, and 1 μW standby power at 300 mV. To supply variable voltages at these low power levels, a switched capacitor DC-DC converter is integrated on-chip and achieves above 75% efficiency while delivering between 10 μW to 250 μW of load power.**

***Index Terms*—CMOS digital integrated circuits, DC-DC conversion, leakage currents, logic design, low-power electronics, SRAM, subthreshold.**

I. INTRODUCTION

VOLTAGE scaling is a compelling approach for energy reduction in digital circuits as it provides quadratic savings in the $CV_{\rm DD}^2$ energy. Although circuits exhibit slower speeds at low supply voltages, the trade-off remains attractive for energy-constrained systems with relaxed throughput constraints. As $V_{\rm DD}$ approaches the sub-threshold region, longer propagation delays eventually lead to a rise in the leakage energy per operation, since the leakage power must be integrated over increasing clock periods. These opposing trends in active and leakage energy give rise to a minimum energy point, which optimizes the energy per operation of a circuit [1], as illustrated in Fig. 1.

The previous argument assumes that the circuit can complete a task at exactly the optimal speed and then shut off, so that it consumes no leakage energy during idle periods. However, certain system components, such as SRAMs, must be powered for arbitrarily long periods unrelated to their own speed. In this case it is essential to also reduce their leakage power. Voltage scaling causes a decrease in leakage current by alleviating drain induced barrier lowering, which, combined with $V_{\rm DD}$ reduction from 1 V to 300 mV, can provide an order of magnitude leakage power savings (Fig. 1).

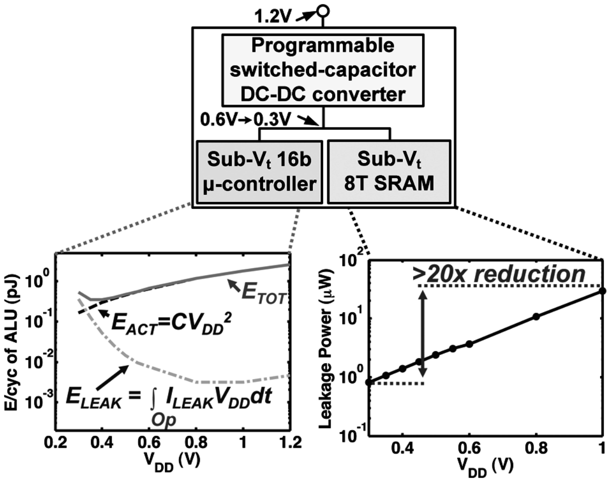


Fig. 1. Voltage scaling enables energy and leakage power reduction in core logic and SRAM of the system.

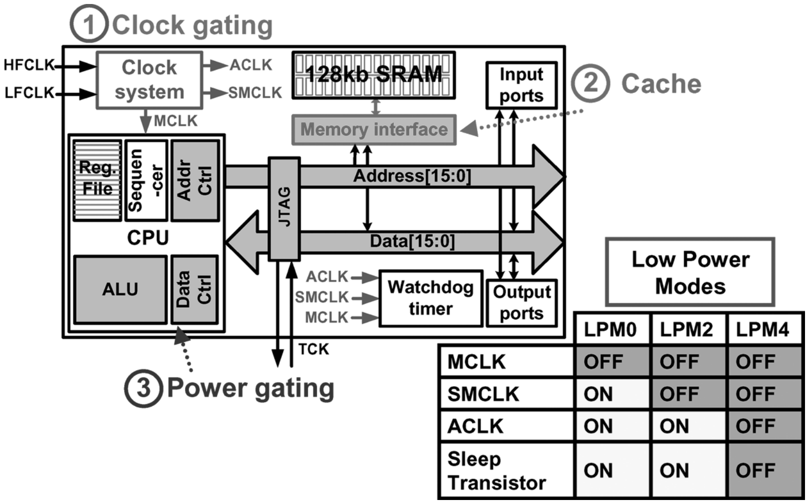
Previous research has demonstrated the energy advantage afforded by ultra-low-voltage operation. For example, a 180 mV, 0.18 μm FFT processor was presented in [2], while a 0.13 μm processor with 8-bit ALU, 32-bit accumulator, and a 2 kb SRAM functional down to 200 mV was implemented in [3]. Body biasing and several gate sizing strategies were examined in a 0.13 μm sub-V_t processor [4].

Looking forward, technology scaling enables reduced $CV_{\rm DD}^2$ energy and increased density, but presents a new challenge in the form of heightened intra-die variation. In [5], a 65 nm 320 mV motion estimation accelerator achieving high throughput employed optimized datapath circuits to address weak I_{ON}/I_{OFF} ratio and threshold voltage variation. For instance, registers contained non-ratioed, upsized keepers, and multiplexers with more than 3 inputs were remapped into 2:1 multiplexers. In [6], a 65 nm SRAM design with a 10 T bit-cell functions down to 400 mV. This paper describes a 65 nm system-on-a-chip with a 16-bit microcontroller and a 128 kb SRAM operating down to 300 mV; both are powered by an integrated DC-DC converter as shown in Fig. 1. Variation-aware design approaches enable the core logic to function in deep sub-threshold. The sub-V_t SRAM

Manuscript received April 15, 2008; revised August 31, 2008. Current version published December 24, 2008. This work was supported by the Defense Advanced Research Projects Agency (DARPA). IC fabrication was provided by Texas Instruments Incorporated. The work of J. Kwong was supported by the Texas Instruments Graduate Women's Fellowship for Leadership in Microelectronics and NSERC. The work of N. Verma was supported by the Intel Foundation Ph.D. Fellowship Program and NSERC.

The authors are with Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: jyskwong@mit.edu; ryogesh@mit.edu; nverma@mit.edu; anantha@mtl.mit.edu).

Digital Object Identifier 10.1109/JSSC.2008.2007160



	LPM0	LPM2	LPM4
MCLK	OFF	OFF	OFF
SMCLK	ON	OFF	OFF
ACLK	ON	ON	OFF
Sleep Transistor	ON	ON	OFF

Fig. 2. Block diagram of microcontroller core.

employs an 8 T bit-cell and peripheral circuit assists to overcome process variation while maintaining density. The DC-DC converter addresses the critical need for efficient power delivery in micro-power systems. Featuring programmable gain settings and optimized control circuitry, the converter can deliver variable load voltage and power levels with high efficiency and low area overhead.

This paper first discusses the challenges in microcontroller logic design and describes approaches to address process variation. Specific circuits and architectures to enable a low-voltage SRAM and a high-efficiency DC-DC converter are then presented. Finally, Section VI provides prototype measurement results.

II. Sub-Threshold Logic Design

A. Microcontroller Overview

Fig. 2 shows a block diagram of the core logic, which is based on the MSP430 microcontroller architecture [7]. The 16-bit RISC CPU supports 27 instructions and 7 addressing modes of the standard MSP430 instruction set. The microcontroller interfaces to 128 kb of unified instruction and data memory, implemented as a custom SRAM, as well as to a watchdog timer and general purpose I/O ports. Programming of the SRAM is performed at startup via a JTAG interface.

Targeting low power applications, the microcontroller provides several power management features as illustrated in Fig. 2. The clock system, which distributes external clocks to the microcontroller logic, supports three low power modes. In the first mode (LPM0), the master clock (MCLK) going to the CPU is gated. At this time, the CPU does not perform any processing, although peripherals remain active. The high frequency clock for the peripherals, or the sub-system master clock (SMCLK), is disabled in the second low power mode (LPM2). However, the auxiliary clock (ACLK), the low frequency clock for peripherals, remains on so that peripherals can function with lower active power. In the standby mode (LPM4), all clocks are shut off. The microcontroller can wake up from any of these modes through an interrupt event generated by the watchdog timer or input port.

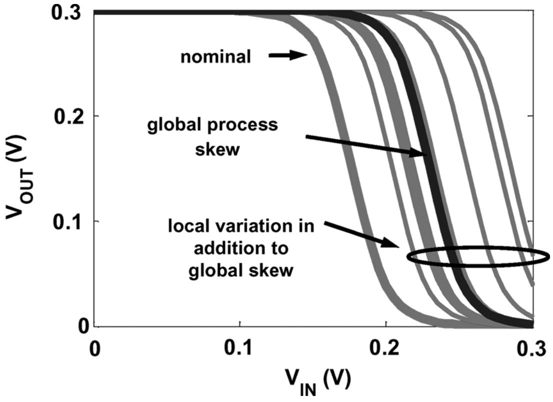


Fig. 3. Effects of variation and reduced I_{ON}/I_{OFF} on sub-V_t inverter voltage transfer curve.

This implementation also contains two features not found in commercial versions of the MSP430 microcontroller. First, the memory interface contains a small cache to reduce the memory access power. One 64-bit row of memory, which contains four 16-bit CPU words, is fetched and stored at a time. Successive 16-bit accesses to the same row require no further memory activity. This provides up to 50% savings in the measured memory access power for applications with a high hit rate. Second, the logic is split into two power domains; the unused blocks shaded in Fig. 2 are power gated during standby mode. Key CPU states are retained such that the microcontroller can continue program execution upon emerging from standby. The on-chip sleep transistor is sized for approximately 5% delay penalty at $V_{\rm DD} =$ 300 mV. Accounting for the energy overhead in turning this transistor on and off, the breakeven time for power gating is less than 100 μs. In other words, the microcontroller only needs to remain in standby for a short period of time in order for power gating to provide a net energy benefit.

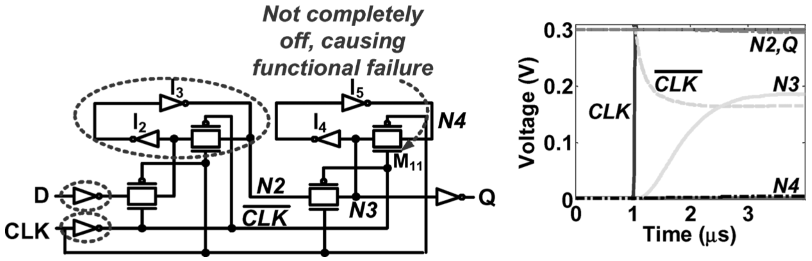
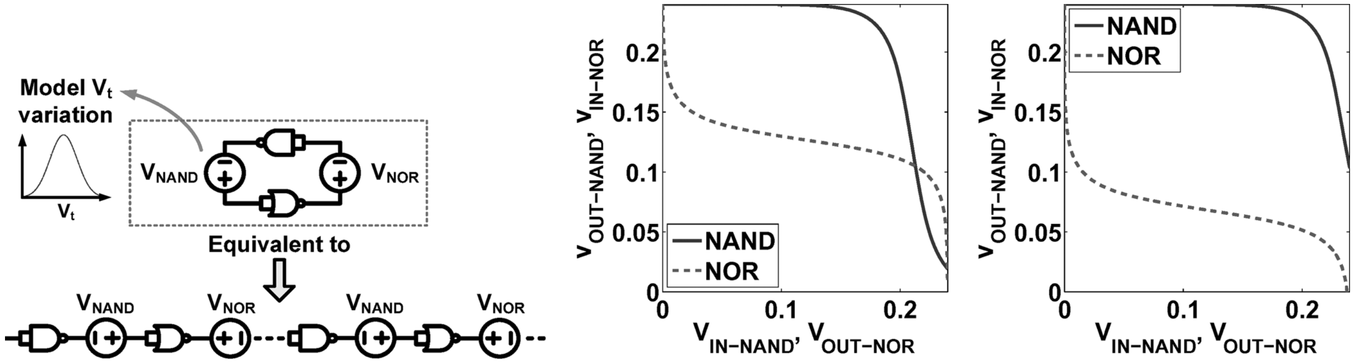


Fig. 4. Reduced voltage swing in sub-V_t can impact hold SNM and signal propagation in registers. The latter issue is shown by transient simulation.



(a) (b) (c)

Fig. 5. (a) Simulation setup for butterfly plot, (b) butterfly plot of functioning logic gates, and (c) butterfly plot of failing gates due to V_t variation.

B. Sub-Threshold Logic Design Challenges

In addition to system-level power management features, voltage scaling is a key strategy in improving the microcontroller energy efficiency. As mentioned in Section I, the energy consumed by a digital circuit can be minimized by operating at the optimal $V_{\rm DD}$, which often lies in the sub-threshold region. However, sub-threshold logic design in a deeply scaled technology node must address two factors which critically impact functionality. In this regime, logic gates exhibit degraded ratios of on to off currents (I_{ON}/I_{OFF}). Moreover, random-dopant-fluctuation is a dominant source of local variation in sub-V_t, causing random, local threshold voltage shifts [8]. The resulting exponential changes in device currents, exacerbating the weak I_{ON}/I_{OFF}, imply that static CMOS logic gates can fail to provide rail-to-rail output swings. The two combined effects are illustrated in Fig. 3 by the voltage transfer curve (VTC) of an inverter at 300 mV. Global variation, which weakens the NMOS relative to PMOS here, skews the VTC towards one side. Additionally, local variation randomly changes the strengths of PMOS and NMOS to cause perturbations in the VTC, in some cases severely degrading the logic levels.

These degraded logic levels can adversely impact functionality, even in typically robust static CMOS circuits. For example, reduced logic swing in inverters I_2, I_3, I_4, I_5 of Fig. 4 decreases the hold static noise margin (SNM) of latches in the classic transmission-gate register. Another failure mechanism is illustrated in the transient simulation of Fig. 4. Here, because the clock buffer has reduced output swing, the transistor M_{11} cannot be completely turned off during the transparent mode of the slave latch. Consequently, a signal cannot propagate successfully from node N2 to N3. Issues such as these motivate the design of a custom library with functionality in the presence of sub-V_t variation as the primary goal.

C. Variation-Aware Logic Design

One approach to mitigate local variation is to upsize transistors, since the standard deviation of V_t varies inversely with the square root of the channel area [9]. However, in the interest of minimizing energy, transistors also should be kept as small as possible, to lower CV^2 energy and leakage currents. To manage this trade-off, the butterfly plot is proposed as a design guideline in building a custom sub-V_t standard cell library.

The butterfly plot is formed by simulating two gates in a back-to-back configuration, as seen in the example of Fig. 5(a). To illustrate the worst case, NAND and NOR are selected here for their inherently skewed VTCs. Because the VTC is input-dependent, all inputs are varied simultaneously to obtain the worst skew. The resulting plot in Fig. 5(b) consists of the VTC of one gate superimposed on the inverse VTC of the other. Intersection points represent stable voltage levels that can be supported by the circuit.

Conceptually, the back-to-back structure, when unrolled, is equivalent to an infinitely long chain of the two gates arranged in an alternating manner [10]. Having two bistable points in the butterfly plot implies that a signal at the input of the logic chain will eventually regenerate to either logic high or logic low.

One way to model local V_t variation in the back-to-back structure is to include it as series noise sources, shown as V_{NAND} and V_{NOR} in Fig. 5(a). Like process variation, these sources cause shifts in the VTCs. Now, when the back-to-back structure is unrolled, these sources affect every other gate in the long logic path in the same manner, shifting their VTCs in the butterfly plot. As shown in Fig. 5(c), the shift due to local variation can

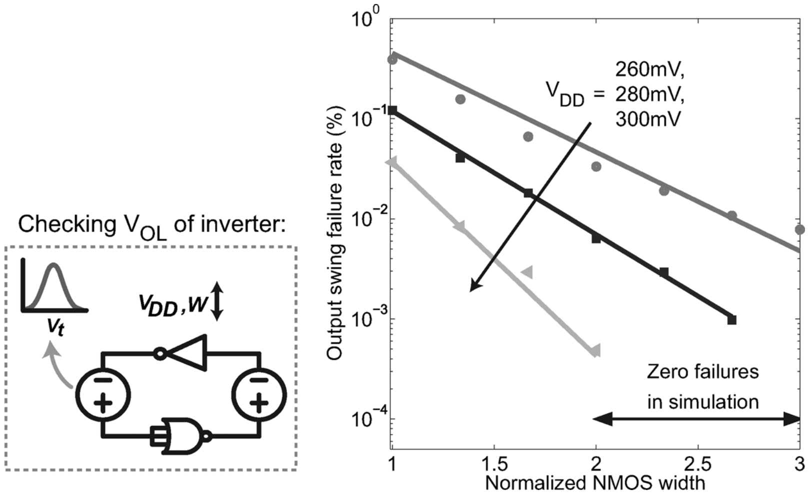


Fig. 6. Failure rate due to degraded V_{OL} in an inverter as $V_{\rm DD}$ and NMOS width (normalized to minimum) are varied. Arrow indicates region where no failures were observed in simulation at 300 mV. Simulation is illustrated conceptually on the left.

be so severe that the VTCs meet at only one monostable point. This implies that any input to the long logic path will ultimately converge to only one logic state, resulting in functional failure.

A functional criterion based on the above can be described as follows. Consider selecting two logic gates at random from a circuit, each gate with its associated local V_t variation. The two gates are considered to function properly together if an infinitely long logic path constructed from them can support two logic states, or equivalently, if the butterfly plot contains two bistable points. This is a more stringent requirement than simply cascading the gates and verifying the output voltage after two stages.

D. Sub-V_t Standard Cell Library

To use the above metric in designing a standard cell library, the maximum fan-in of the library is first limited to three. A larger fan-in would require stacking many devices in series, significantly degrading I_{ON}/I_{OFF}. The logic gate to be designed (e.g., inverter, 2-input NAND, 2-input NOR) is put back-to-back with 3-input NAND and 3-input NOR, whose skewed VTCs give the most stringent input-high (V_{IH}) and input-low (V_{IL}) requirements respectively. Sizing of the 3-input gates are fixed to provide a starting point for designing the remaining gates. Then, V_t of transistors in the gate under test and global (inter-die) process conditions are randomized according to models and data provided by the foundry. The Monte Carlo runs are in effect analogous to sampling logic gates across multiple chips. Following the above definition for logic functionality, the failure rate of the gate under test is found from Monte Carlo simulations while varying $V_{\rm DD}$, device sizing, and temperature.

Several trends were observed from the analysis. The failure rate decreases exponentially as either $V_{\rm DD}$ or device width is increased. This is shown in Fig. 6, which plots the failure rate caused by degraded output low voltage (V_{OL}) in an inverter. Starting from an inverter with minimum size devices, the NMOS width is increased at various $V_{\rm DD}$. The arrow marks the region where all samples were functional in a 200 k-point simulation at 300 mV. Other logic primitives, such as two series NMOS in a NAND gate, exhibit similar behavior. Therefore, by increasing the device width or $V_{\rm DD}$, the failure rate can be made sufficiently small.

To examine the effects of temperature, Fig. 7(a) plots the nominal output low (V_{OL}) and output high (V_{OH}) voltages of a sub-V_t inverter from 0 °C to 100 °C. The output voltage levels degrade slightly, but the overall effect is negligible. Fig. 7(b) plots the standard deviation of V_{OH} and V_{OL} with local V_t variation. The spreads in V_{OH} and V_{OL} are seen to increase slightly at high temperature. These observations imply that, in the process technology being used, the high temperature corner is worst case for sub-threshold logic gate functionality.

With these considerations, a 62-cell library was designed which includes various logic functions and drive strengths. Each logic primitive was sized to give the same failure rate. The Monte Carlo simulation effort was reduced by reusing the sizing of logic primitives across several gates. For example, the required sizing for two series NMOS devices was found from the 2-input NAND, where two leaking parallel PMOS oppose the pull-down devices to give the worst case scenario. This sizing can then be reused in other gates with two series NMOS devices (e.g., AND-OR-INVERT).

Although excluding NAND3 and NOR3 from the library allows the remaining cells to be sized smaller, the number of gates needed to synthesize the design would increase. Synthesis results showed that the latter effect dominates in this design; eliminating NAND3 and NOR3 would cause the total transistor area in the logic to increase by approximately 15%.

As discussed in Section II-B, register design also merited special attention. Data retention of the registers under V_t mismatch can be verified by measuring the hold static-noise-margin (SNM) of the master and slave latches while accounting for the voltage drop across transmission gates. As with logic gates, the percentage of latches displaying negative hold SNM, or failure to retain data, decreases exponentially with $V_{\rm DD}$ and device width in the inverters. Additionally, signal propagation issues were addressed by upsizing local clock and data buffers to ensure that their outputs are sufficiently close to $V_{\rm DD}$ and ground. Fortunately, transmission gates were more robust against variation and did not require special upsizing. Compared to a register

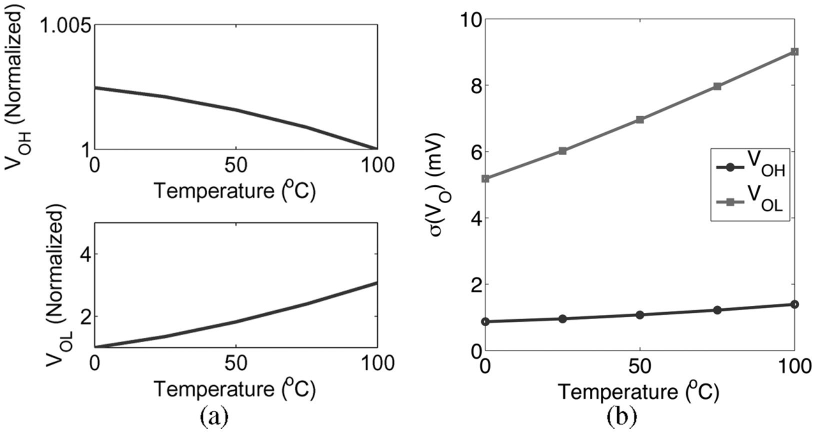


Fig. 7. (a) Inverter output voltage versus temperature at $V_{\rm DD} = 300$ mV and worst case global corner. Each plot is normalized such that the minimum value = 1. (b) Standard deviation of V_{OH} and V_{OL} distributions versus temperature at $V_{\rm DD} = 300$ mV. Both trends indicate that high temperature is the worst case corner for functionality.

optimized for the above-threshold region, the sub-threshold register has 2.4× larger area and 2.1× higher clock loading.

Fig. 8 plots results of the custom library design and motivates the need for variation-aware device sizing. Here, the worst case cells from an unoptimized above-V_t library are compared to the custom sub-V_t library at 300 mV. The left and right columns respectively plot distributions of the V_{OL} of 3-input NAND and the V_{OH} of 3-input NOR, under local variation and at global corners. As expected, the sub-V_t cells exhibit significantly lower output voltage variation at the cost of larger area. Typically, cells with the smallest drive strengths in the sub-V_t library are sized larger than their above-V_t counterparts, but the higher drive strengths can be kept unchanged. For cells such as the inverter, buffer, and 2-input NOR (NOR2), a 10% area increase is sufficient, while NOR3, NAND2, and NAND3 required 190%, 100%, and 270% increases respectively. Nevertheless, the logic synthesis tool was able to reduce the overall area cost by selecting upsized cells less frequently; for example, NAND3 comprised of only 0.31% of the total gate count in this microcontroller.

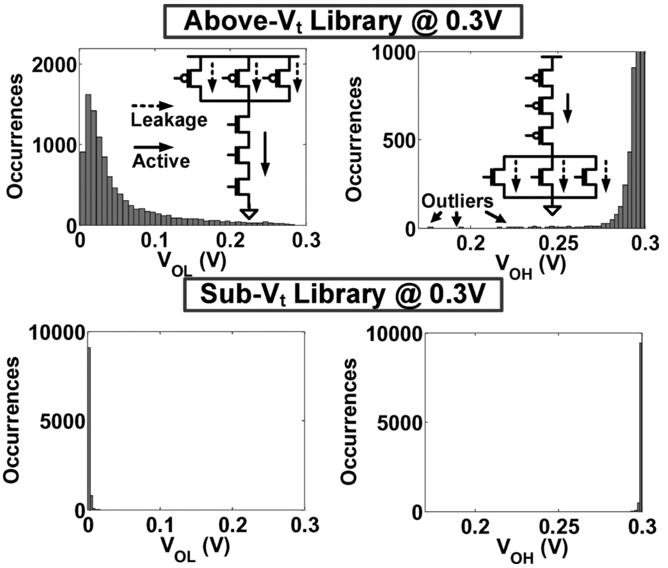


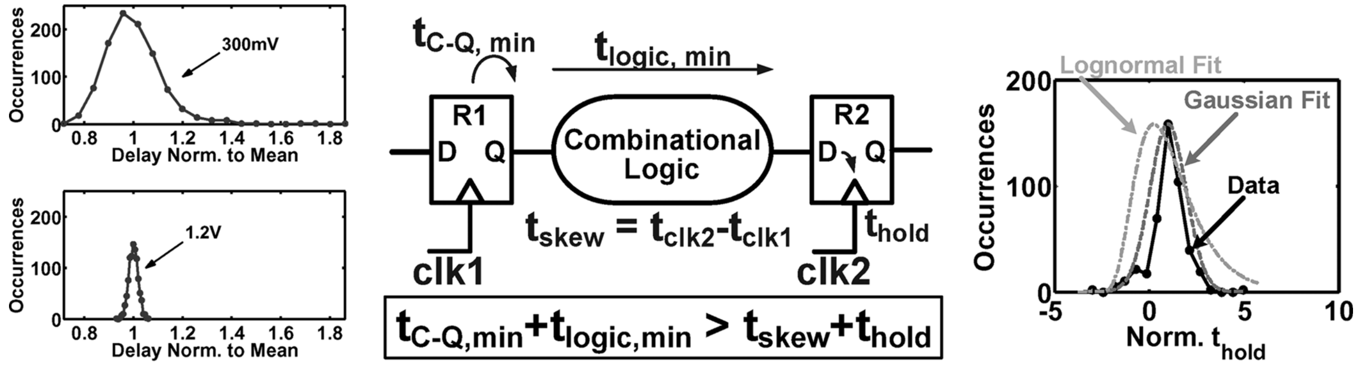
Fig. 8. Variation in V_{OL} and V_{OH} of 3-input NAND and 3-input NOR, from custom sub-V_t and standard above-V_t libraries.

III. SUB-THRESHOLD TIMING ANALYSIS

In addition to affecting functionality, process variation also increases delay uncertainty. In sub-V_t, local variation causes the delay distribution to widen further. Fig. 9(a) plots the normalized delay distributions of a microcontroller logic path, highlighting how variability increases by an order of magnitude at 300 mV compared to 1.2 V. Conventional static timing analysis approaches typically treat logic gate delay as deterministic, taking points at the tails of the distribution to represent the maximum and minimum delay under process variation. However, given the wide distributions in sub-V_t, such approaches would lead to unrealistic results. This motivates statistical timing analysis methodologies [11] which consider the entire delay distribution instead of only the tail points.

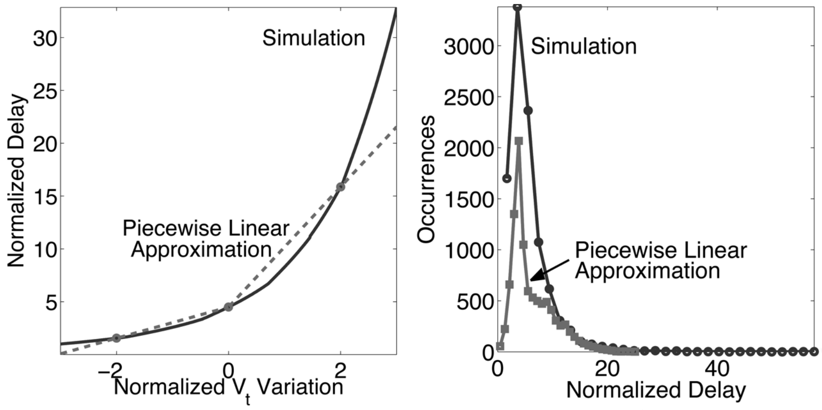
A. Variation-Aware Timing Methodology

Statistical static timing analysis in sub-V_t is complicated by several factors. Unlike in the above-threshold regime, the analysis cannot be easily simplified with linear models due to the exponential dependence of delay on V_t. Fig. 10(a) plots the relationship between delay and V_t shift for a sub-V_t inverter. Characterizing these relationships for all cells in a library, under different input and output conditions, naturally requires substantial effort. Borrowing techniques from the above-V_t regime, one might envision forming a piecewise linear approximation in order to reduce the characterization effort. Following the example of computer-aided design tools in using three points (best, typical, worst), the piecewise linear model plotted by the dashed line in Fig. 10(a) is constructed. An approximate delay distribution can then be derived from the model and the V_t variation statistics. However, as shown in Fig. 10(b), the distribution obtained in this manner does not match well with Monte Carlo SPICE simulation results. Although adding more points to the piecewise linear model can improve accuracy at the expense of longer characterization time, statistical approaches which can capture the nonlinear delay-V_t relationship should also be considered.



(a) (b) (c)

Fig. 9. (a) Delay histograms of microcontroller logic path ($t_{cq,min} + t_{logic,min}$), each normalized to sample mean to highlight the difference in variability. Both histograms contain 1000 samples. (b) Hold time constraint in a generic logic path. (c) Distribution of register hold time (t_{hold}), which does not fit well to Gaussian or lognormal forms.



(a) (b)

Fig. 10. (a) Inverter delay versus V_t variation and a piecewise linear approximation. (b) Inverter delay distribution from Monte Carlo simulation and from linear approximation, both plotted for 10 k samples.

The exponential relationship implies that, when V_t is modeled as having a normal distribution, the resulting delay distribution for a logic gate will be lognormal. However, there is no closed form expression for adding lognormally distributed gate delays to obtain the logic path delay [12]. Instead, this must be done with iterative approaches [13] or analytical models, one example being the expression for the sum of identically distributed sub-V_t gate delays in [8]. Further, register hold time (t_{hold}) is often not well-approximated by standard distributions in sub-V_t. t_{hold} hinges on whether a change in the data input causes a glitch or transition that incorrectly disturbs the output. This, in turn, is influenced by the slew rates of clock and data signals and can be a nonlinear phenomenon. Fig. 9(c) plots the simulated t_{hold} distribution for a register with asynchronous preset and reset. Here, neither the Gaussian nor lognormal models can accurately represent the simulation.

To capture these effects, this design employs an approach based on Monte Carlo simulation while using analytical methods to reduce the total simulation effort. Shown in Fig. 11, the timing analysis flow focuses on hold time violations because they cause functional errors independent of the clock period. An exhaustive timing report listing the data and clock paths is obtained from the placed and routed design. This report is generated under the worst case global conditions—at the fast process corner for verifying hold time, and at low V_{DD} where V_t variation is most prominent. However, the report does not consider local variation. Known paths with very short logic delays (e.g., shift registers) are removed from the timing report and handled separately. The remaining paths are grouped into bins by the nominal hold time margin, and the bins are then analyzed to select paths of interest for further simulation. The

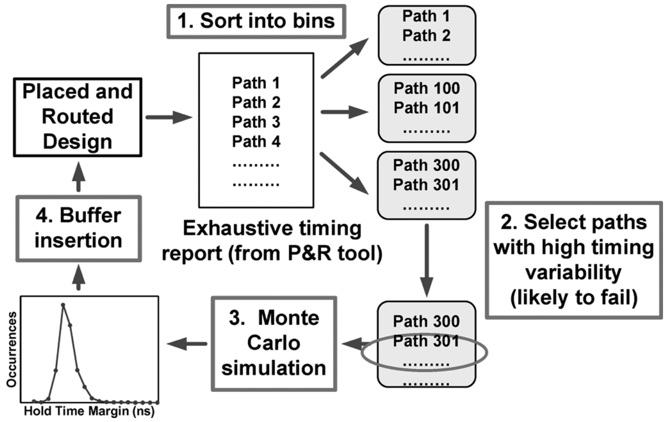


Fig. 11. Overview of timing analysis flow.

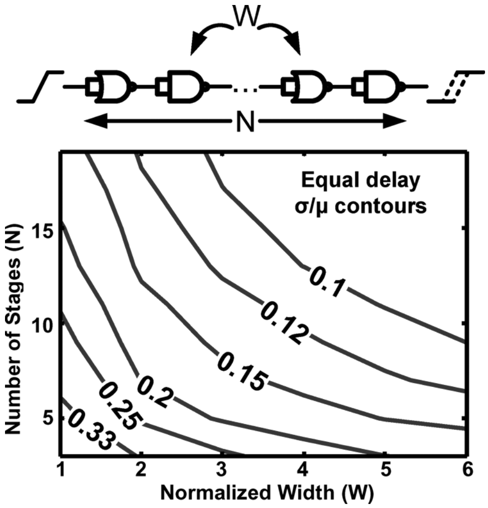


Fig. 12. Equal σ/μ contours for delay of a uniformly sized NAND-NOR chain, showing the decrease in variability as logic depth or device width increases.

hold time margin is derived by rearranging the standard hold time constraint (Fig. 9(b))

$$(t_{clk2} - t_{clk1}) + t_{hold} < t_{c-q,min} + t_{logic,min} \quad (1)$$

and is defined as

$$t_{hold\text{-}margin} = t_{c-q,min} + t_{logic,min} - (t_{clk2} - t_{clk1}) - t_{hold}. \quad (2)$$

It should be greater than zero for proper functionality.

Within a bin, an algorithm selects paths with high variance, whose long distribution tails result in a higher probability of timing violation. To estimate the variance of path delays, the algorithm employs the standard deviation over mean (σ/μ), which decreases with larger device sizes and higher logic depth [8], as shown in Fig. 12 for a generic logic path with stacked devices. It is important to note, however, that the trends eventually reach diminishing returns, which must be considered in the timing methodology.

For logic gates, the σ/μ of delay is first characterized at the device sizes used in the standard cell library. Input slew and load capacitance also have a slight effect on σ/μ. These dependencies are summarized in lookup tables used by the algorithm. For a logic path, the relative variation becomes smaller as the logic depth increases, since variation tends to average out across stages. To account for this effect, the algorithm assigns a weighting factor to every path according to its logic depth. The factor is found empirically by simulating the delay variability in paths of different lengths and varied transistor sizes. The analysis did not consider spatial correlation since several studies [14]–[17] reported small spatial correlation coefficients, which showed weak or no discernible dependence on separation distance over the ranges of interest in this design. Further, it would be impractical to model the position dependence of spatial variation during the design phase, since this is very difficult to predict without the final layout [16].

The high variance paths are then selected to undergo Monte Carlo simulation with local variation and at the global fast corner. This gives an accurate hold time margin distribution, accounting for the local clock skew and the hold time requirement (t_{hold}) of the destination register. Probability of a hold time violation $(t_{hold\text{-}margin} < 0)$ can then be estimated. The distribution of $t_{hold\text{-}margin}$ is generally not Gaussian nor lognormal according to the Anderson–Darling test [18]. Nevertheless, the data was fitted to a Gaussian curve, since this gave on average a more pessimistic probability of violation compared to finding the percentage of violating samples in the raw data.

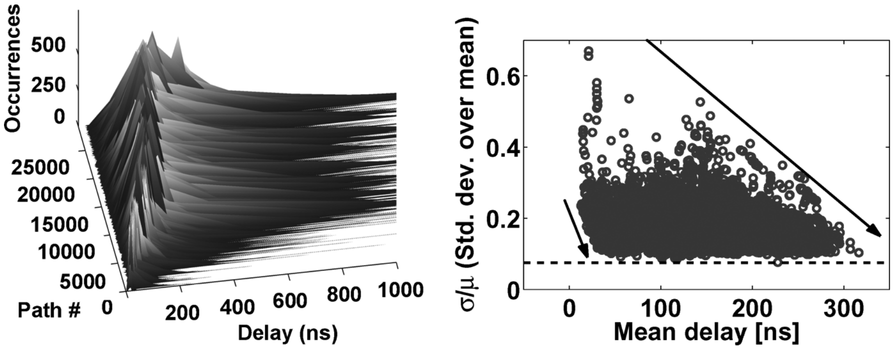
If the probability is above a set threshold, as determined by the number of paths in the design and the desired timing yield, then extra delay buffers are applied to increase the hold time margin. To be conservative, buffers are also applied to unsimulated paths belonging to the same bin. Paths requiring extra buffering were concentrated in small bins with low average hold time margin. It should be noted that a variation-aware approach typically results in fewer delay buffers inserted compared to worst case timing analysis. For instance, a common worst case methodology uses two deterministic values to model fast and slow delay in a cell under local variation. One such example would be to use the $\pm 1\sigma$ points as the slow and fast delays. Hold time constraint is verified by assuming that all cells in the data path have fast delays, while those in the capture clock path have slow delays, in order to obtain the worst case scenario. However, in reality, it is unlikely that all cells in the data path uniformly exhibit fast delay due to local variation. Because of this pessimism, the worst case methodology identified 929 timing paths for hold time fixing, several times more than the 151 paths selected by the variation-aware approach.

B. Comprehensive Delay Variation Data

Apart from the analysis described above, comprehensive Monte Carlo SPICE simulations were performed for 30000 timing paths in the microcontroller over several months. The results serve to illustrate trends in sub-V_t delay variability. In Fig. 13(a), each horizontal cross section is the delay distribution of one timing path under local variation, at 300 mV and global fast corner. The rightward skew is typical of a lognormal distribution. Fig. 13(b) shows a scatter plot of the corresponding timing path statistics. Each point represents one path, with mean delay plotted on the x-axis and σ/μ shown on the y-axis. Initially, the lower range of σ/μ decreases with the mean delay, which reflects how variation tends to average out in longer paths. However, this quickly reaches diminishing returns, and σ/μ does not decrease far below 0.1, even for very long paths. The same trend is observed when logic depth, instead of mean delay, is plotted on the x-axis. Since σ/μ depends on both device sizes and logic depth, the lower bound observed reflects the inherent variability given the device sizes used in the standard cell library. Additionally, the upper range indicates that outliers with large amounts of variation occur less frequently in very long paths. However, when examining critical paths for hold time, it is important to consider both the shortest paths and slightly longer paths that may exhibit higher variability.

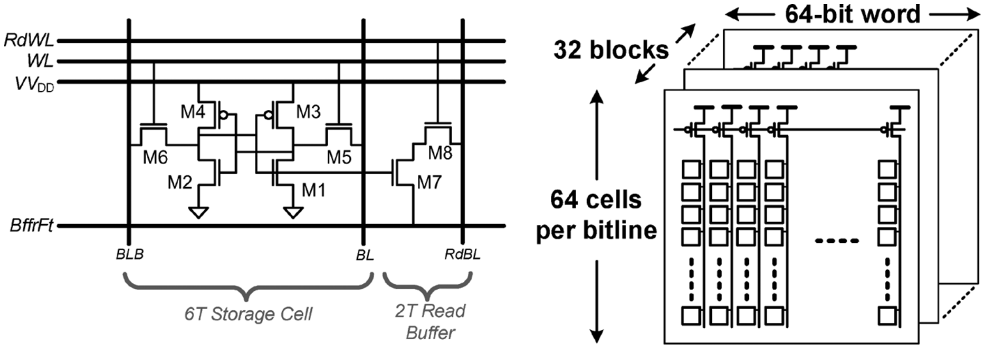
IV. ULTRA-LOW-VOLTAGE SRAM

Although the 6 T SRAM bit-cell provides a good balance between density, stability, and performance for conventional



(a) (b)

Fig. 13. (a) Delay distributions of 30 k microcontroller timing paths at 300 mV, fast corner. Each horizontal cross section represents distribution of one path. (b) Scatter plot of microcontroller timing path statistics corresponding to data in (a).



(a) (b)

Fig. 14. (a) Sub-V_t 8 T bit-cell with peripheral assists for write-ability and bit-line leakage management. (b) SRAM array architecture.

applications, its high sensitivity to variation leads to very unfavorable trade-offs for ultra-low-voltage applications (i.e., below 500 mV). Most critically, its read static noise margin (SNM) [19] is severely degraded by the greatly amplified effect of random dopant fluctuations (RDF) [20], and up-sizing, to manage variation and ensure sufficient margin, leads to an impractically large bit-cell layout. Similarly, correct write operation requires that stored data be overwritten by the access devices; however, the relative device strengths necessary to ensure this cannot practically be guaranteed. Further, the increased sensitivity to variation also results in extremely low worst case read-current. The resulting effect on performance is drastic, but, even more importantly, the effect on functionality can be fatal, where the read-current can be exceeded by the aggregate bit-line leakage-current [21].

A. Sub-V_t SRAM Design

In this ultra-low-voltage design, an SRAM based on the 8 T bit-cell shown in Fig. 14(a) is used to provide full operation down to 300 mV. Though the cell area is increased by the read-buffer, it obviates the stringent read SNM, which is less than 80 mV (with sigma of approximately 40 mV); the remaining hold SNM is over 130 mV (with sigma of approximately 30 mV). Meanwhile, write-margin is ensured by control of VV_{DD}, which selectively weakens the PMOS loads, $M3/4$, and bit-line leakage, to enable a high-level of column integration, is managed by control of $BffrFt$, which gates the sub-V_t leakage from unaccessed read-buffers.

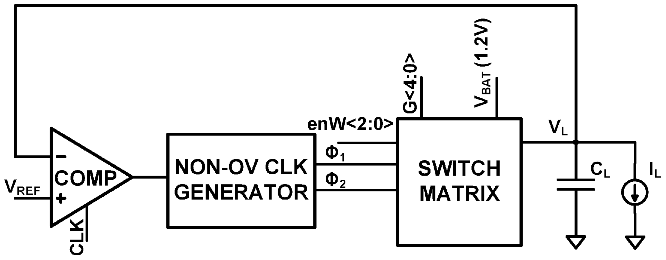


Fig. 15. Architecture of the switched capacitor DC-DC converter.

Though based on the techniques used in [22], the SRAM requires several design changes for this application: a new bit-cell provides larger read-current by taking advantage of the reverse-short-channel effect; shorter column configurations reduce the bit-line leakage and loading, ensuring reliable sensing with 10× lower access-time at 500 mV, and an interface buffer allows independent optimization of SRAM word-length and CPU word-length.

Since the read-buffer devices of an 8 T bit-cell have no impact on stability, they can be sized primarily for optimal cell read-current. In above-V_t designs, this typically leads to nearly minimum length devices, even though the cell layout height, which is limited by the other devices, permits longer lengths. However, longer read-buffer devices have the advantage of lower effective threshold voltage, through the reverse-short-channel effect [23], and less RDF variation, which is particularly critical in sub-V_t where its impact is greatly amplified. As a result,

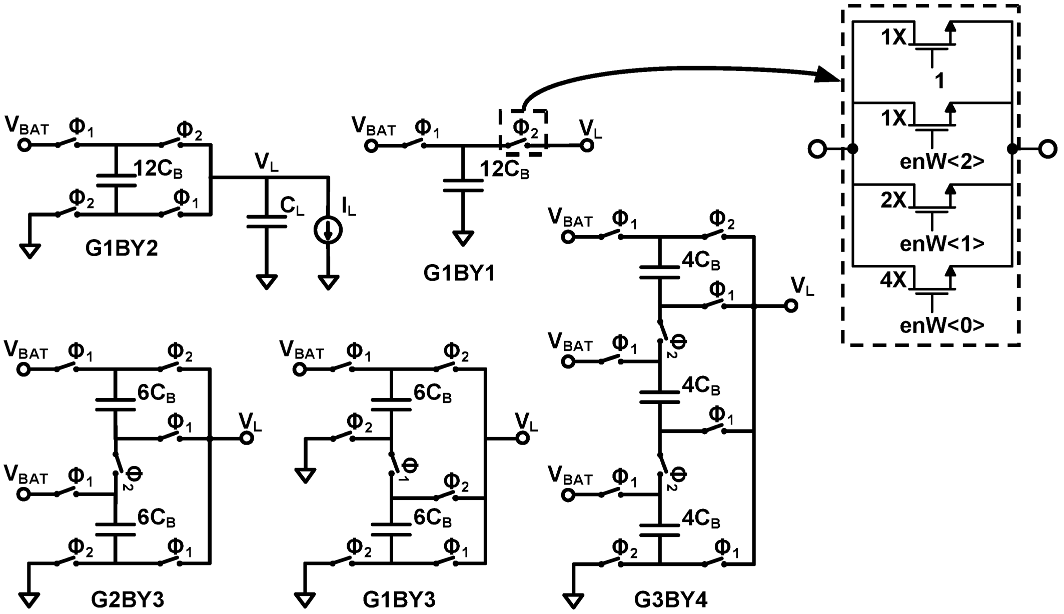


Fig. 16. The different gain settings used within the switch matrix. A simplified representation of the switch size control is shown in the inset.

at ultra-low-voltages, both the mean and weak-cell read-current can be much higher [24]. This has been exploited to improve the cell's write-ability in sub-V_t SRAMs [25]; but in this design, it is applied by aggressively lengthening the read-buffer devices to increase read-current. It is worth noting that, despite the read-current improvement, there is no significant change in the sub-V_t bit-line leakage current, since raising $BffrFt$, as described in [22], eliminates this source of leakage. Lastly, the increase in dynamic power to drive $RdWL$ and $BffrFt$ is negligible, as their wire capacitance greatly dominates over gate and diffusion loading.

Additionally, reducing the number of cells per bit-line, from 256 (in [22]) to 64, as shown in Fig. 14(b), mitigates both secondary bit-line leakage sources and bit-line loading. Consequently, the higher read-current and lower bit-line capacitance enable a performance increase of 10× at 500 mV, which is critical in this application since system clocking requires the SRAM to operate at a higher clock rate than the logic.

Lastly, a local buffer is used to provide an interface between the CPU and the SRAM. Peripheral read and write assists in the SRAM are critical to ensure robust low-voltage operation. However, amortizing their overhead leads to large SRAM access words. Accordingly, the local buffer abstracts these constraints from the CPU, providing optimal data alignment.

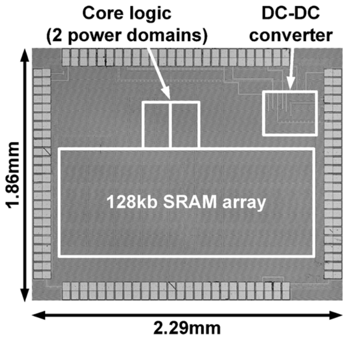
V. DC-DC CONVERTER

The previous sections have described the energy savings that can be achieved by reducing the $V_{\rm DD}$ of logic and memory circuits. To realize the full energy savings of sub-V_t operation, a DC-DC converter supplying ultra-low voltages at high efficiencies is essential. Since the power consumption of the logic and SRAM load circuits drops exponentially at sub-V_t voltages, the DC-DC converter was designed to deliver a maximum of 500 μW of load power. This reduced load power demand makes switched capacitor DC-DC conversion an ideal choice for this application. The switched capacitor (SC) DC-DC converter is based on [26], and makes use of 600 pF of total on-chip charge transfer (flying) capacitance to provide scalable load voltages from 300 mV to 1.1 V. The logic and SRAM circuits in this system utilize voltages up to 600 mV.

Fig. 15 shows the architecture of the DC-DC converter. The converter uses an all-digital pulse frequency modulation (PFM) mode of control to regulate the output voltage. In this method of control, the converter stays idle until the load voltage V_L falls below the reference voltage ($V_{\rm REF}$), at which point a clocked comparator enables the switch matrix to transfer one charge packet to the load. A PFM mode control is crucial to achieving high efficiency for the extremely low power system being built. The switch matrix block contains the charge transfer switches and the charge transfer capacitors.

One of the main efficiency limiting mechanisms in a switched capacitor DC-DC converter is the linear conduction loss [26]. To maintain efficiency over the wide load voltage range of 300 mV to 1.1 V, this converter employs five different gain settings ($\mathrm{G}\langle 4{:}0\rangle$). Fig. 16 shows how the different gain settings are achieved from a total charge transfer capacitance of $12C_B$ (600 pF). The external voltage input to the system is 1.2 V. Each gain setting at no-load provides a voltage ratioed output of the input voltage. A suitable gain setting ($\mathrm{G}\langle 4{:}0\rangle$) is chosen off-chip, depending on the proximity of its no-load voltage to the load voltage being delivered, and its ability to provide the load power demand [26]. Since the logic and SRAM load circuits utilize voltages up to 600 mV, in the actual testing of the chip, only gain modes G2BY3, G1BY2 and G1BY3 were used.

The switching losses in the converter are dominated by the energy expended in turning the charge transfer switches ON and OFF. The switch widths are designed such that the charge transfer capacitors just settle at the end of a charge transfer cycle. In order to scale switching losses with load power, the charge transfer switches have adjustable widths which are enabled by the signal $\mathrm{enW}\langle 2{:}0\rangle$ as shown in the inset of Fig. 16.

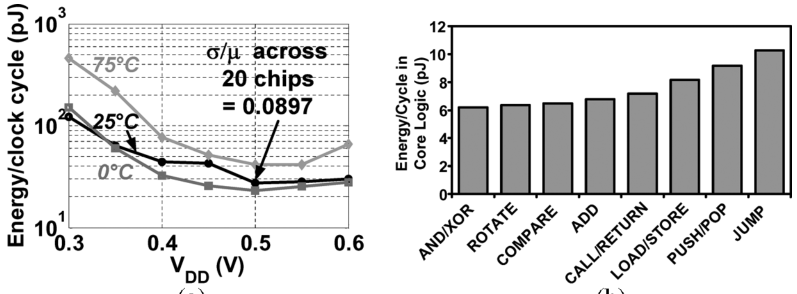


(a)

Process	65nm CMOS
Area	
DC-DC Converter	0.12mm²
SRAM	1.36mm²
Logic	0.14mm²
Performance	μ-controller (Logic and SRAM)
Minimum Energy Point	500mV
Minimum Functional V_{DD}	300mV

(b)

Fig. 17. (a) Die micrograph and (b) summary of microcontroller test chip at room temperature.



(a) (b)

Fig. 18. (a) Energy versus V_{DD} of logic and memory over temperature. The σ/μ of measurements across 20 chips at 500 mV is shown. (b) Energy of microcontroller core logic while it executes different instructions at 500 mV, room temperature.

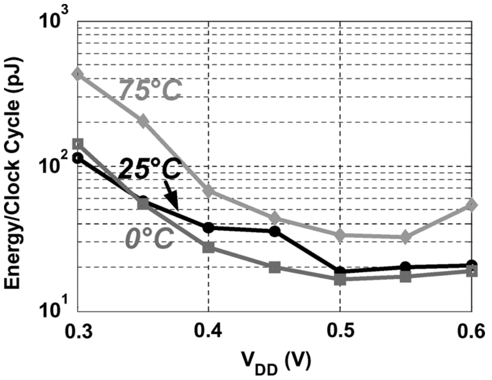


Fig. 19. Energy versus V_{DD} of the SRAM array per system clock cycle.

For any decrease (increase) in the load power by a factor of 2, the clock frequency (CLK) of the comparator is halved (doubled) and correspondingly, the width of the charge transfer switches is also halved (doubled). This helps to decrease the switching power by 4× when the load power decreases by 2×, leading to an increase in efficiency at lower load power levels. While the signal $\mathrm{enW}\langle 2:0\rangle$ was set externally in this implementation, [26] describes a method to automatically determine the signal $\mathrm{enW}\langle 2:0\rangle$ as the load power varies. In Fig. 20, the gain in efficiency as the load power decreases close to 380 μW and 200 μW is due to the scalable switch width design. However, at very low load power levels (sub-5 μW), leakage and other fixed losses in the control circuitry reduce the efficiency of the switched capacitor DC-DC converter.

VI. PROTOTYPE MEASUREMENTS

A summary and die micrograph of the test chip, fabricated in 65 nm CMOS, is shown in Fig. 17. The DC-DC converter, including charge transfer capacitors, occupies just 0.12 mm^2. The minimum energy point of the microcontroller occurs at 500 mV, and functionality was verified down to 300 mV.

A. *Active Energy and Performance*

Fig. 18(a) plots the measured energy per cycle versus supply voltage for the microcontroller logic and SRAM at 0 °C, 25 °C, and 75 °C. The energy is measured while the system executes test code which cycles through the available instructions and addressing modes. Since the I/O pads, logic, and memory array are operated at the same voltage, level shifters are not required on-chip. Level converters are used on the test board to interface the low-voltage I/Os to the logic analyzer. Memory and logic together consume 27.2 pJ per clock cycle at 500 mV and 25 °C. The optimum energy does not vary much across 20 chips; the measurements have a σ/μ of 0.0897.

Shown in Fig. 18(b) is the energy consumption of the microcontroller core logic while it executes specific instructions. Generally, instructions for arithmetic or boolean operations (e.g., add, and, compare), executed on operands stored in CPU registers, require roughly the same amount of energy per cycle. Instructions that involve memory accesses for data (e.g., load/store, push/pop) exhibit higher energy consumption as expected. The jump instruction, which generates high switching activity on the address bus, requires the most energy.

The energy consumed by the SRAM array per system clock cycle is shown in Fig. 19. The memory greatly influences the minimum energy point of the system since it consumes a major portion of the total system energy, highlighting the importance

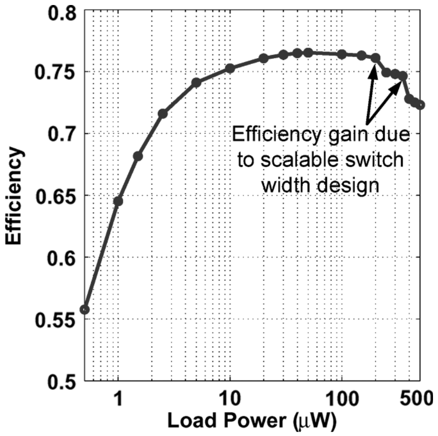


Fig. 20. DC-DC converter efficiency while delivering 500 mV. The DC-DC converter is powered by a 1.2 V supply. Arrows mark efficiency gain from scalable switch width design as discussed in Section V.

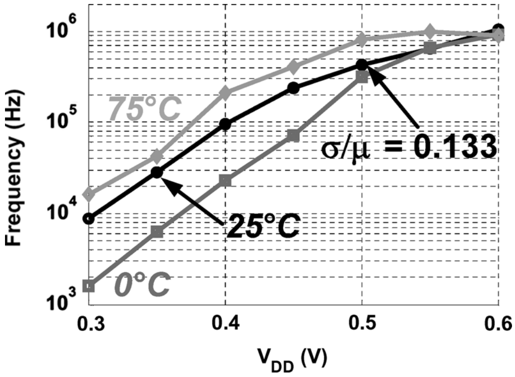


Fig. 21. Frequency versus V_{DD} across temperature. The σ/μ of measurements across 20 chips at 500 mV is shown.

of reducing memory energy through voltage scaling and other circuit techniques.

The efficiency of the DC-DC converter delivering 500 mV is shown in Fig. 20. The converter achieves more than 75% efficiency with an order of magnitude change in load power, between 10 μW to 250 μW. With the microcontroller as a load, the converter provides 75% efficiency at 12 μW. When measured standalone, the converter reaches a peak efficiency of 78%.

Fig. 21 plots the microcontroller performance versus supply voltage at 0 °C, 25 °C, and 75 °C. The measured frequency, accounting for logic and memory delays, is 434 kHz at 25 °C and 500 mV. The frequency ranges from 8.7 kHz to 1 MHz across the operating range of 0.3 V to 0.6 V. The σ/μ of measurements across 20 chips at 500 mV is 0.133.

B. Standby Power

The inclusion of a DC-DC converter enables the system to dynamically scale V_{DD} to 300 mV during standby mode, where memory and logic together consume less than 1 μW, as shown in Fig. 22. Accounting for the DC-DC converter efficiency loss at such low power levels, this represents a 2.1× reduction in leakage power compared to keeping V_{DD} constant at 500 mV during standby.

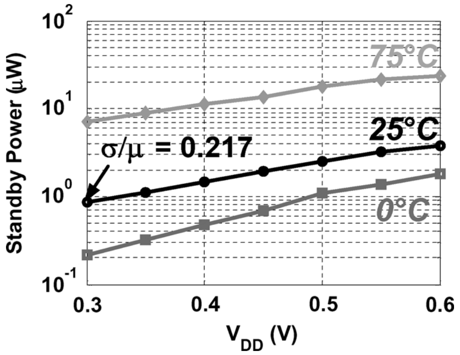


Fig. 22. Standby power versus V_{DD} across temperature. The σ/μ of measurements across 20 chips at 300 mV is shown.

VII. CONCLUSIONS AND SUMMARY

Voltage scaling enables energy minimization and leakage power reduction in micro-power systems. However, design techniques and circuit assists are necessary to overcome process variation in the ultra-low-voltage regime. The 65 nm sub-V_t microcontroller presented here demonstrates several approaches to enable operation down to 300 mV. A standard cell library design methodology addresses the degraded V_{OH} and V_{OL} levels in sub-V_t, which, at deeply scaled process nodes, can render logic gates non-functional. Circuit delays are similarly affected by variation, exhibiting an order of magnitude higher variability at low voltages. Conventional timing analysis approaches that treat delays as deterministic are insufficient. Instead, a variation-aware methodology combining Monte Carlo simulation and analysis was developed to verify hold time constraints. The SRAM represents a dominant portion of area and power in this system. Therefore, energy and leakage reduction through voltage scaling is highly desirable. In conventional 6 T SRAMs, V_t variation causes severely degraded read-current and increased cell instability, limiting the minimum functional voltage. The SRAM in this system employs an 8 T bit-cell to address these limitations. Further, peripheral circuit assists enforce the relative device strengths needed for read and write functionality, despite significant variation. The fully integrated, switched capacitor DC-DC converter provides highly efficient power delivery at the low voltage and power levels required by energy-constrained systems. Employing multiple gain settings and efficient control circuitry, the DC-DC converter achieves above 75% efficiency while supplying 500 mV across an order of magnitude change in load power.

ACKNOWLEDGMENT

The authors are grateful to D. Buss, K. Huber, M. Koesler, H. Moormann, and A. Wang for their support and feedback.

REFERENCES

[1] A. Wang, A. Chandrakasan, and S. Kosonocky, “Optimal supply and threshold scaling for sub-threshold CMOS circuits,” in *IEEE Computer Society Annual Symp. VLSI*, Apr. 2002, pp. 5–9.

[2] A. Wang and A. Chandrakasan, “A 180-mV subthreshold FFT processor using a minimum energy design methodology,” *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, Jan. 2005.

[3] B. Zhai, L. Nazhandali, J. Olson, A. Reeves, M. Minuth, R. Helfand, S. Pant, D. Blaauw, and T. Austin, "A 2.60 pJ/Inst subthreshold sensor processor for optimal energy efficiency," in *Symp. VLSI Circuits Dig.*, Jun. 2006, pp. 154–155.

[4] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, and D. S. Blaauw, "Performance and variability optimization strategies in a sub-200 mV, 3.5 pJ/inst, 11 nW subthreshold processor," in *Symp. VLSI Circuits Dig.*, Jun. 2007, pp. 152–153.

[5] H. Kaul, M. Anders, S. Mathew, S. Hsu, A. Agarwal, R. Krishnamurthy, and S. Borkar, "A 320 mV 56 μW 411 GOPS/Watt ultra-low voltage motion estimation accelerator in 65 nm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2008, pp. 316–317.

[6] B. H. Calhoun and A. P. Chandrakasan, "A 256-kb 65-nm sub-threshold SRAM design for ultra-low-voltage operation," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 680–688, Mar. 2007.

[7] MSP430. Texas Instruments. [Online]. Available: http://focus.ti.com/lit/ug/slau056g/slau056g.pdf

[8] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "Analysis and mitigation of variability in subthreshold design," in *Proc. Int. Symp. Low-Power Electronics and Design (ISLPED)*, Aug. 2005, pp. 20–25.

[9] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[10] J. Lohstroh, E. Seevinck, and J. D. Groot, "Worst-case static noise margin criteria for logic circuits and their mathematical equivalence," *IEEE J. Solid-State Circuits*, vol. SC-18, no. 6, pp. 803–807, Dec. 1983.

[11] A. Srivastava, D. Sylvester, and D. Blaauw*, Statistical Analysis and Optimization for VLSI: Timing and Power*. New York: Springer, 2005.

[12] N. C. Beaulieu and F. Rajwani, "Highly accurate simple closed-form approximations to lognormal sum distributions and densities," *IEEE Commun. Lett.*, vol. 8, no. 12, pp. 709–711, Dec. 2004.

[13] S. Schwartz and Y. Yeh, "On the distribution function and moments of power sums with log-normal components," *Bell Syst. Tech. J.*, vol. 61, no. 7, pp. 1441–1462, Sep. 1982.

[14] K. Agarwal, F. Liu, C. McDowell, S. Nassif, K. Nowka, M. Palmer, D. Acharyya, and J. Plusquellic, "A test structure for characterizing local device mismatches," in *Symp. VLSI Circuits Dig.*, Jun. 2006, pp. 67–68.

[15] L.-T. Pang and B. Nikolic, "Impact of layout on 90 nm CMOS process parameter fluctuations," in *Symp. VLSI Circuits Dig.*, 2006, pp. 69–70.

[16] N. Drego, A. Chandrakasan, and D. Boning, "An all-digital, highly scalable architecture for measurement of spatial variation in digital circuits," in *Proc. IEEE Asian Solid-State Circuits Conf.*, Nov. 2008, pp. 393–396.

[17] T. Mizuno, J. Okumtura, and A. Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 2216–2221, Nov. 1994.

[18] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes," *Ann. Mathemat. Statist.*, vol. 23, no. 2, pp. 193–212, 1952.

[19] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 5, pp. 748–754, Oct. 1987.

[20] A. Bhavnagarwala, X. Tang, and J. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 658–665, Apr. 2001.

[21] N. Verma and A. P. Chandrakasan, "A 256-kb 65 nm 8 T subthreshold SRAM employing sense-amplifier redundancy," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 141–149, Jan. 2008.

[22] N. Verma and A. Chandrakasan, "A 65 nm 8 T sub- V_t SRAM employing sense-amplifier redundancy," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2007, pp. 328–329.

[23] C.-Y. Lu and J. M. Sung, "Reverse short-channel effects on threshold voltage in submicrometer salicide devices," *IEEE Electron Device Lett.*, vol. 10, no. 10, pp. 446–448, Oct. 1989.

[24] N. Verma, J. Kwong, and A. P. Chandrakasan, "Nanometer MOSFET variation in minimum energy subthreshold circuits," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 163–174, Jan. 2008.

[25] T.-H. Kim, J. Liu, J. Keane, and C. H. Kim, "A 0.2 V, 480 kb subthreshold SRAM with 1 k cells per bitline for ultra-low-voltage computing," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 518–529, Feb. 2008.

[26] Y. K. Ramadass and A. P. Chandrakasan, "Voltage scalable switched capacitor DC-DC converter for ultra-low-power on-chip applications," in *Proc. Power Electronics Specialists Conf.*, 2007, pp. 2353–2359.

Joyce Kwong (S'02) received the Bachelor of Applied Science degree from the University of Waterloo, Canada, in 2004, and the Master's degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2006, where she is currently pursuing the Ph.D. degree. Ms. Kwong received the 2007 Texas Instruments Graduate Woman's Fellowship for Leadership in Microelectronics and the NSERC Postgraduate Fellowship. Her research interests include sub-threshold design methodology and system implementation.

Yogesh K. Ramadass (S'03) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, in 2004, and the S.M. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 2006, where he is currently a doctoral candidate. From May 2007 to August 2007, he worked in the Wireless Analog Technology Center at Texas Instruments, Dallas, TX, designing power converters. His research interests include low-power circuit design, DC-DC converters and energy harvesting/processing circuits. Mr. Ramadass received the President of India Gold Medal in 2004, the Beatrice Winner award for editorial excellence at ISSCC 2007, the 7th International Low Power Design Contest award at ISLPED 2007, and the 2008–2009 Intel Foundation Ph.D. Fellowship.

Naveen Verma (S'04) received the B.A.Sc. degree in electrical and computer engineering from the University of British Columbia, Vancouver, Canada, in 2003, and the M.S. degree from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 2005. He is currently pursuing the Ph.D. degree at MIT, where his research interests include low-power mixed-signal circuits in the areas of analog-to-digital converters, SRAMs, and implantable biological systems. Mr. Verma received the Intel Foundation Ph.D. Fellowship and the NSERC Postgraduate Fellowship.

Anantha P. Chandrakasan (M'95–SM'01–F'04) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer sciences from the University of California at Berkeley in 1989, 1990, and 1994, respectively. Since September 1994, he has been with the Massachusetts Institute of Technology (MIT), Cambridge, MA, where he is currently the Joseph F. and Nancy P. Keithley Professor of Electrical Engineering. He is the Director of the MIT Microsystems Technology Laboratories. His research interests include low-power digital integrated circuit design, wireless microsensors, ultra-wideband radios, and emerging technologies. He is a co-author of *Low Power Digital CMOS Design* (Kluwer Academic, 1995), *Digital Integrated Circuits* (Pearson Prentice-Hall, 2003, 2nd edition), and *Sub-threshold Design for Ultra-Low Power Systems* (Springer, 2006). He is also a co-editor of *Low Power CMOS Design* (IEEE Press, 1998), *Design of High-Performance Microprocessor Circuits* (IEEE Press, 2000), and *Leakage in Nanometer CMOS Technologies* (Springer, 2005).

Dr. Chandrakasan has been a co-recipient of several awards, including the 1993 IEEE Communications Society's Best Tutorial Paper Award, the IEEE Electron Devices Society's 1997 Paul Rappaport Award for the Best Paper in an EDS publication in 1997, the 1999 DAC Design Contest Award, the 2004 DAC/ISSCC Student Design Contest Award, the 2007 ISSCC Beatrice Winner Award for Editorial Excellence, and the 2007 ISSCC Jack Kilby Award for Outstanding Student Paper. He served as a technical program co-chair for the 1997 International Symposium on Low Power Electronics and Design (ISLPED), VLSI Design '98, and the 1998 IEEE Workshop on Signal Processing Systems. He was the Signal Processing Subcommittee Chair for ISSCC 1999–2001, the Program Vice-Chair for ISSCC 2002, the Program Chair for ISSCC 2003, and the Technology Directions Subcommittee Chair for ISSCC 2004–2008. He was an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS from 1998 to 2001. He served on the SSCS AdCom from 2000 to 2007 and was the meetings committee chair from 2004 to 2007. He is the Technology Directions Chair for ISSCC 2009.