

Nanometer MOSFET Variation in Minimum Energy Subthreshold Circuits

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

(Invited Paper)

Abstract—Minimum energy operation for digital circuits typically requires scaling the power supply below the device threshold voltage. Advanced technologies offer improved integration, performance, and active-energy efficiency for minimum energy sub- V_t circuits, but are plagued by increased variation and reduced I_{ON}/I_{OFF} ratios, which degrade the fundamental device characteristics critical to circuit operation by several orders of magnitude. This paper investigates those characteristics and presents design methodologies and circuit topologies to manage their severe degradation. The issues specific to both general logic and dense static random access memories are analyzed, and solutions that address their distinct design metrics are presented.

Index Terms—CMOS digital integrated circuits, leakage currents, logic design, low-power electronics, matching, static random access memory (SRAM), subthreshold, yield estimation.

I. INTRODUCTION

A VARIETY of rich and complex applications are emerging where energy constraints are paramount. Portable battery-powered electronics have staggering capabilities, but their lifetime demands remain stringent. Other applications, like wireless sensor networks and implantable biological systems, preferably power themselves using just the 10–100 μ W [1] available through energy harvesting.

In all cases, voltage scaling is critical to enabling circuits that operate at the required energy levels. Although the resulting speeds might be significantly reduced, proper architectural approaches can be applied, where needed, to meet the performance constraints. Specifically, three classes of energy constrained systems benefit from voltage scaling.

- 1) Low-speed requirement—environment and biological signal monitoring typically requires a circuit speed of 10 to 100s of kilohertz. Voltage scaling can be applied aggressively to operate statically at the minimum energy voltage.

Manuscript received May 29, 2007. This work was supported by the Defense Advanced Research Projects Agency. The work of N. Verma was supported by the Intel Foundation Ph.D. Fellowship Program and NSERC. The review of this paper was arranged by Editor S. Kosonocky.

The authors are with the Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139-4307 USA (e-mail: nverma@mtl.mit.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/TED.2007.911352

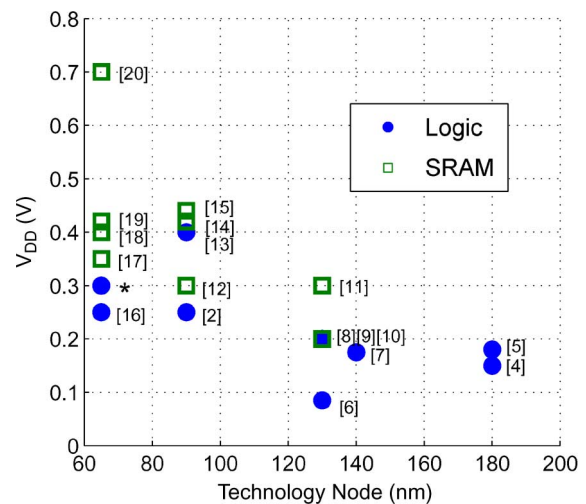


Fig. 1. Minimum V_{DD} for recently reported designs. SRAMs provide the primary barrier to low-voltage operation (*point reported in personal correspondence).

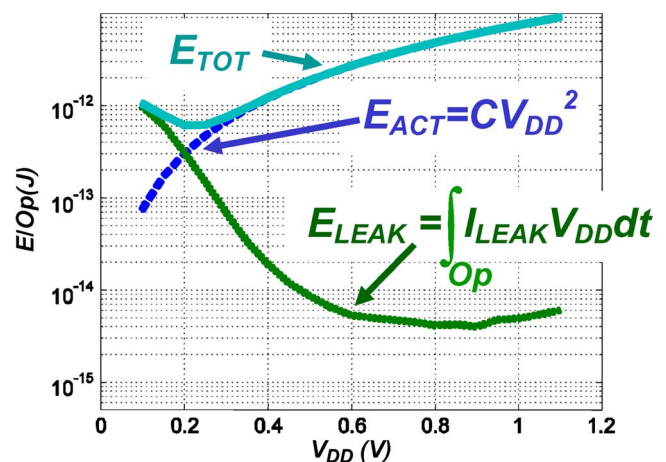


Fig. 2. Energy profiles of the 90-nm carry look-ahead adder with respect to V_{DD} .

- 2) Dynamic speed requirement—cellular multimedia handsets have relaxed workloads for the vast majority of time, but can provide bursts of high performance. Dynamic voltage scaling and ultra-dynamic voltage scaling [2] allow higher-voltage operation when required, and

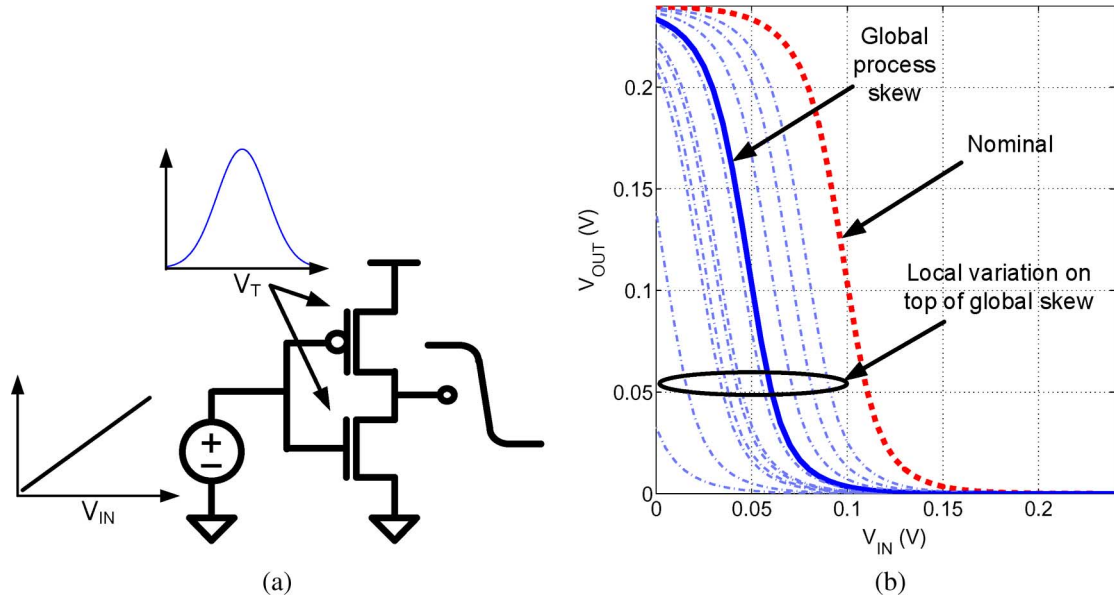


Fig. 3. (a) Simulation setup to measure VTC. (b) VTC of an inverter at 240 mV under process variation.

these systems benefit tremendously from the speed-ups afforded by advanced technologies.

- 3) Constant high-speed requirement—baseband radio processors must meet system specifications for throughput. These systems can leverage the potential for extreme parallelism in scaled technologies to assign unit operations to many separate hardware blocks, each of which operate efficiently at a reduced voltage and rate [3].

All of these systems rely heavily on low-voltage logic circuits and static random access memories (SRAMs). Increasing device variation, however, provides a primary opposition to voltage scaling and limits the energy reductions achievable. SRAMs, in particular, are subject to extreme variation and pose the first failure point in low-voltage designs. Fig. 1 shows the minimum voltage achieved by recently reported designs and highlights the limitation of SRAMs.

This paper starts by describing the energy components of a circuit and the importance of sub- V_t operation in minimizing the total energy. Then, the primary limitations of sub- V_t devices are discussed. Finally, the resulting failure mechanisms, in both logic circuits and SRAMs, are analyzed, and corresponding solutions are presented.

II. SUB- V_t OPERATION FOR MINIMUM CIRCUIT ENERGY

In modern digital systems, active switching and leakage are the dominant sources of energy. The total energy is given by (1)

$$E_{TOT} = E_{ACT} + E_{LEAK} = CV_{DD}^2 + \int_{Op} I_{LEAK} V_{DD} dt. \quad (1)$$

A circuit's leakage energy is the integral of its leakage power over the time it requires to complete an operation. Once the operation is complete, the circuit can be power-gated, using high-

threshold devices, to suppress the idle leakage currents [21]. It should be noted that power-gating itself increases E_{ACT} due to the overhead of controlling the gating device and restoring the power/ground node voltage before subsequent circuit operation.

Although voltage scaling reduces the active CV_{DD}^2 energy, it also reduces circuit speed and results in a longer leakage-power integration time, during which the circuit cannot be power-gated. Hence, leakage energy increases [22]. These opposing trends are shown for the case of a 32-b carry look-ahead adder in 90-nm CMOS in Fig. 2, where the minimum total energy occurs at approximately 250 mV. This result is typical as most practical digital circuits have a low minimum energy voltage that occurs below the threshold voltage of the devices [4], [5].

The argument, however, is slightly modified for SRAMs, where data buffering requirements imply a need for long-term retention capability. In this scenario, SRAMs must stay on for an arbitrary length of time unrelated to their own access delay. Here, leakage power is more important than leakage energy, and the two critical metrics (leakage power and active energy) both benefit from supply voltage scaling. In fact, the reduction in leakage power can be quite large since, in short-channel technologies, a reduced V_{DS} significantly alleviates drain induced barrier lowering (DIBL). For example, a supply voltage reduction from 1 to 0.3 V can reduce the leakage current by a factor of nearly five, resulting in a leakage power savings of over $15\times$ [17]. Low-voltage SRAMs are thus essential, particularly since, in modern systems, SRAMs occupy a dominating portion of the total power, energy, and area.

A. Sub- V_t Device Variation and Current Degradation

Random dopant fluctuation (RDF) and processing variation are dominating effects in modern nanometer technologies. Both prominently change the resulting threshold voltage of devices. In sub- V_t , where V_t has an exponential effect on the drain current, the resulting impact is overpowering; for instance, in

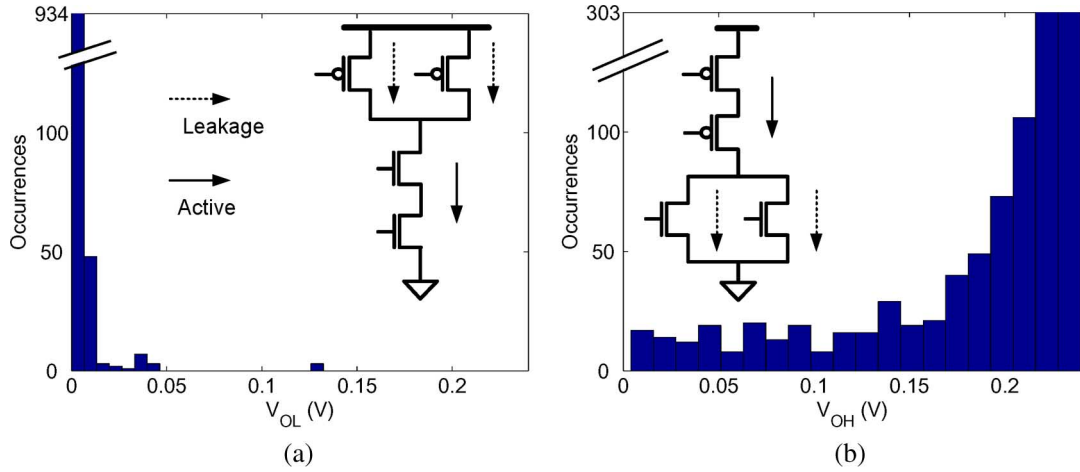


Fig. 4. (a) V_{OL} distribution for two-input NAND gate, and (b) V_{OH} distribution for two-input NOR gate, simulated at the respective skewed global process corners.

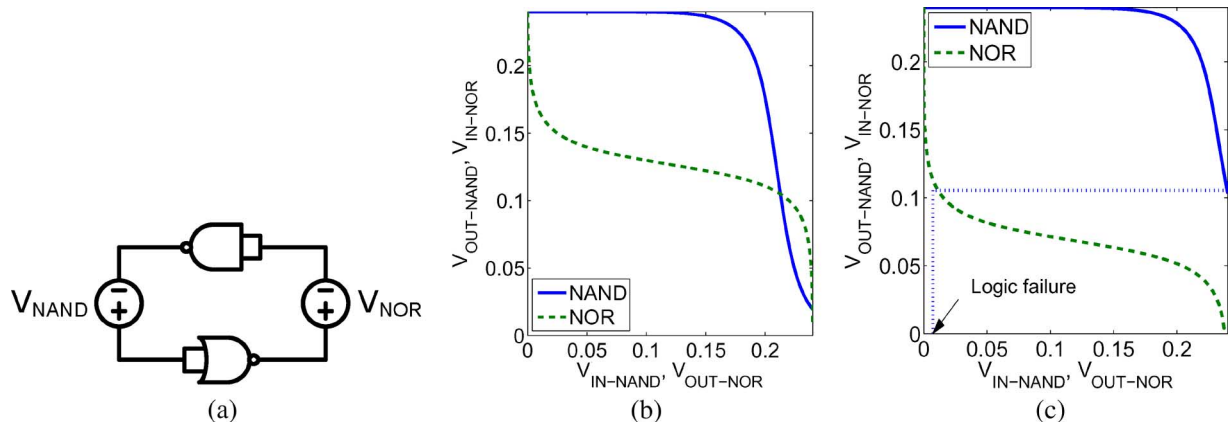


Fig. 5. Butterfly plots for NAND and NOR gates in (a) back-to-back configuration with (b) functional and (c) nonfunctional logic levels.

65 nm, a $\pm 4\sigma$ variation from RDF alone results in a range of drain-current spanning over three orders of magnitude.

An additional consideration is geometric variation, particularly in effective channel length, which impacts the drift mechanism as well as short-channel characteristics like DIBL. In sub- V_t , however, the reduced V_{DS} mitigates the strength of DIBL [23]. Consequently, RDF, due to its exponential impact through V_t , is the dominating source of variability affecting functionality, performance, and energy efficiency.

The exponential variation in sub- V_t drain current is particularly problematic in the face of severely reduced I_{ON}/I_{OFF} . Nominally, the I_{ON}/I_{OFF} of devices in a circuit operating at the minimum energy voltage is between 10^3 – 10^4 , whereas that in strong inversion is approximately 10^7 [24]. Degradation in drain current, due to variation, however, can severely reduce this ratio even further. This introduces a very relevant failure mechanism in both logic and SRAMs and is described further in Sections III-A and IV-D.

III. SUB- V_t LOGIC DESIGN

In sub- V_t , the degradation in I_{ON}/I_{OFF} and the extreme effect of variation necessitates the use of nonratioed static

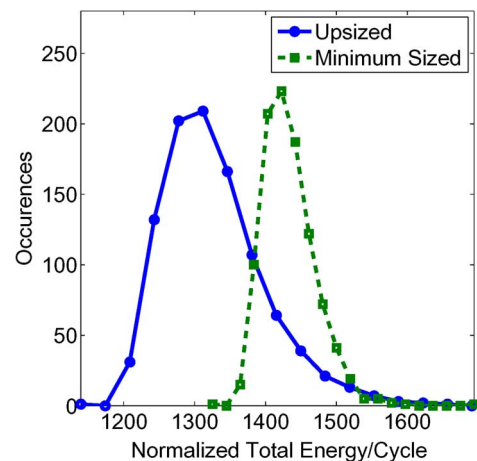


Fig. 6. Total energy distribution of two 32-bit adders, where one was upsized to function at 300 mV and one has a minimum size operating at 340 mV. Energy is normalized to that of a minimum-size inverter.

logic styles. In the presence of V_t shifts, for instance, relative device strengths cannot reliably be set by sizing. Similarly, the magnitude of I_{OFF} approaches that of I_{ON} . As a result, the time constants associated with leakage paths are comparable to the actual gate delays, greatly compromising charge storage

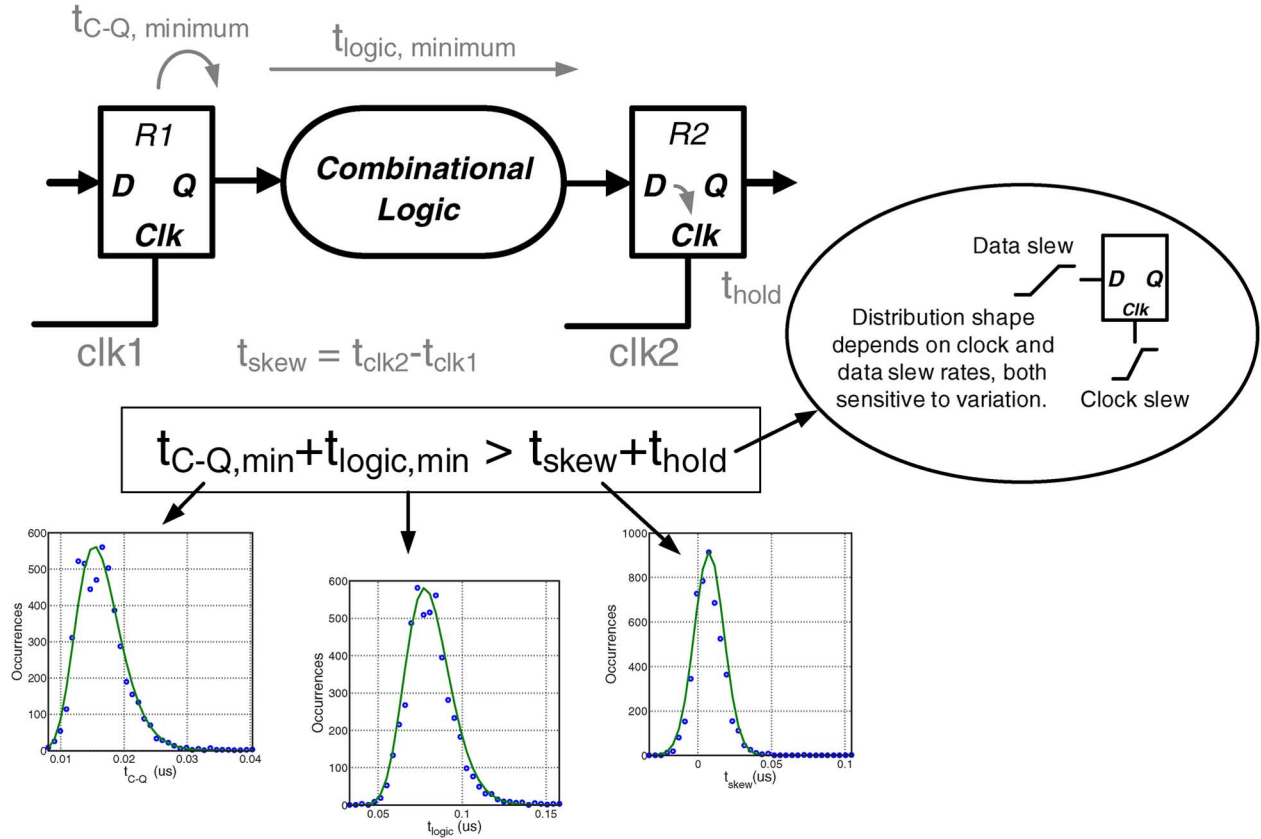


Fig. 7. Logic path timing constraints in sub- V_t .

on dynamic logic nodes. In fact, even actively driven nodes can have degraded logic levels due to the presence of opposing parallel leakage paths.

Additionally, logic delays exhibit increased variation in sub- V_t , extending the time required to complete the circuit operation. Accordingly, while the leakage currents from the critical path might be reduced, those from the entire circuit are not, and they integrate over a longer time. Furthermore, timing uncertainty in synchronous logic paths increases, leading to possible functional failures. This section describes how these effects can be mitigated by careful device sizing and logic path design, with consideration to path depth.

A. Logic Functionality

The shape of the voltage transfer characteristic (VTC) of a logic gate is important for signal regeneration down the logic path and is, thus, a key indicator of the gate’s functionality. Using the setup in Fig. 3(a), Fig. 3(b) shows the VTC of an inverter in sub- V_t . In the case shown, global variation strengthens the NMOS relative to the PMOS and causes the VTC to shift left. Additional local V_t variation imposes random shifts on the VTC, ultimately degrading the noise margins severely. In particular, the worst-case behavior can be analyzed by considering two-input NAND and NOR gates. Here, the active pull-down and pull-up paths, respectively, are weakened by stacked devices, and they must also fight against parallel leakage paths. Fig. 4(a) and (b) show the resulting distributions of NAND V_{OL} and NOR V_{OH} , in which some samples are clearly

nonfunctional. Accordingly, gates with more than two inputs must be avoided in ultralow-voltage designs.

The butterfly plots shown in Fig. 5 are useful in modeling the effect of variation on proper logic functionality. The plot is obtained by envisioning two logic gates back-to-back [as shown in Fig. 5(a)], which therefore corresponds to plotting the direct VTC of one superimposed on the inverse VTC of the other; intersection points then represent physically stable voltage levels. In particular, three intersection points mean that two stable logic levels are supported by the structure, and one metastable point is supported. This is shown in Fig. 5(b) for NAND and NOR gates with no local variation. Although the pull-up and pull-down networks can be sized to center these VTCs, the overwhelming effect of global variation results in the inherently skewed characteristics shown. Nonetheless, the bistable nature implies that a logic path comprising a cascade of these gates supports both logic “1” and logic “0” levels. Local V_t variation, however, can be modeled as series voltage noise sources (V_{NAND} and V_{NOR}), which, in the worst case, have opposite polarity. Now, the resulting VTCs, shown in Fig. 5(c), have only a monostable point, which implies such severe V_t variation that both required logic levels cannot properly propagate through the path.

The observation that variation can compromise logic functionality gives rise to a design tradeoff. Device up-sizing to increase channel areas reduces V_t variation and improves functional yield, but also increases energy consumption, by virtue of increased CV_{DD}^2 . Raising the supply voltage similarly increases energy, but improves the nominal signal levels

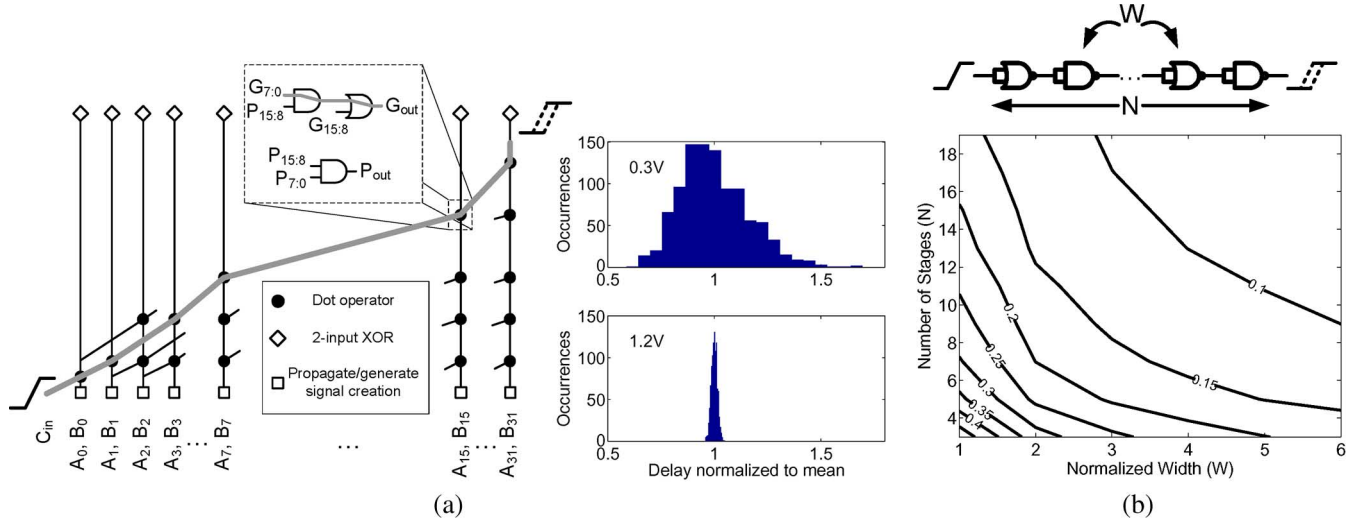


Fig. 8. (a) Delay variability through a critical path of 32-bit Kogge Stone adder at 0.3 and 1.2 V. (b) Constant- σ/μ contours for NAND-NOR chain delay.

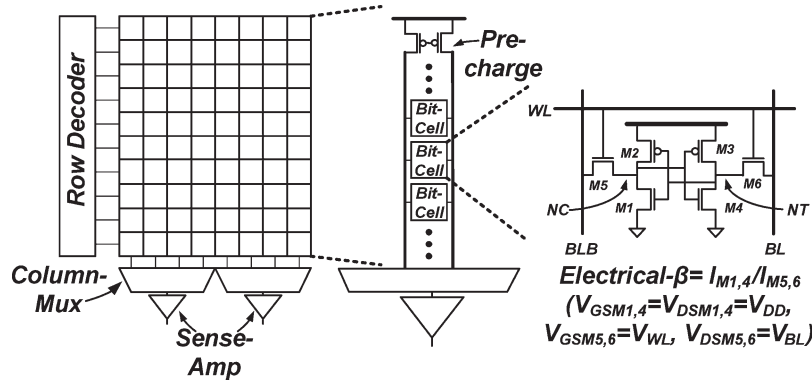


Fig. 9. Typical structure of a modern SRAM. The 6T bit-cell has an electrical β -ratio defined by the ON-current of the driver and pass devices.

and reduces variability [23], providing an alternate option for maintaining yield. Using the butterfly plot as a functional metric for upsizing, however, it can be shown that increasing device widths to operate in deep sub- V_t still provides energy savings [25]. As an example, Fig. 6 compares the total energy distributions of two 32-bit Kogge Stone adders. One is suitably upsized to function at the minimum energy V_{DD} (300 mV), while the other is strictly minimum-sized. Accordingly, it can only operate down to 340 mV at the required yield constraint. Here, it can be seen that proper device sizing leads to improved energy efficiency.

B. Logic Path Delays

Local variation also causes uncertainty in all components of a logic path. Although systems with low-speed requirements have relaxed setup time constraints, the hold time constraint merits close attention, since these violations can cause functional failures independent of the clock frequency. Thus, it is critical to characterize the minimum delay through a logic path. Fig. 7 shows a typical synchronous logic path with the relevant parameters: t_{C-Q} , t_{logic} , t_{skew} , and t_{hold} , each with a corresponding statistical distribution. The example distributions, plotted in Fig. 7, are derived from simulations of a

relatively short timing path that is susceptible to hold violation in a 16-bit sub- V_t microcontroller. Logic delay t_{logic} and clock skew t_{skew} both consist of delays through combinational gates, and hence, their distributions are well-modeled by canonical distributions. However, sequential timing parameters such as clock-to-Q t_{C-Q} and hold times t_{hold} are less predictable. For instance, hold time distribution heavily depends on the clock and data slew rates, which are themselves subject to variation, thus making timing analysis more complex.

The wide variation in the delay of sub- V_t logic path can be seen in the distributions in Fig. 7 and is another consequence of the exponential dependence on V_t variation. To compare with strong inversion, for instance, Fig. 8(a) highlights the critical path of a 32-bit Kogge Stone adder, which passes from the carry input, through several stages of AND and OR gates, to the output. The delay distribution at 300 mV displays a variability that is an order of magnitude higher compared to the distribution at 1.2 V. Previously, it was shown how device up-sizing can mitigate variation to achieve functional digital levels in logic gates. Similarly, device up-sizing mitigates delay variation. However, an additional mechanism to reduce delay variation involves increasing logic path depth to take advantage of statistical averaging across gates. In Fig. 8(b), a uniformly sized NAND-NOR chain characterizes delay variation in generic

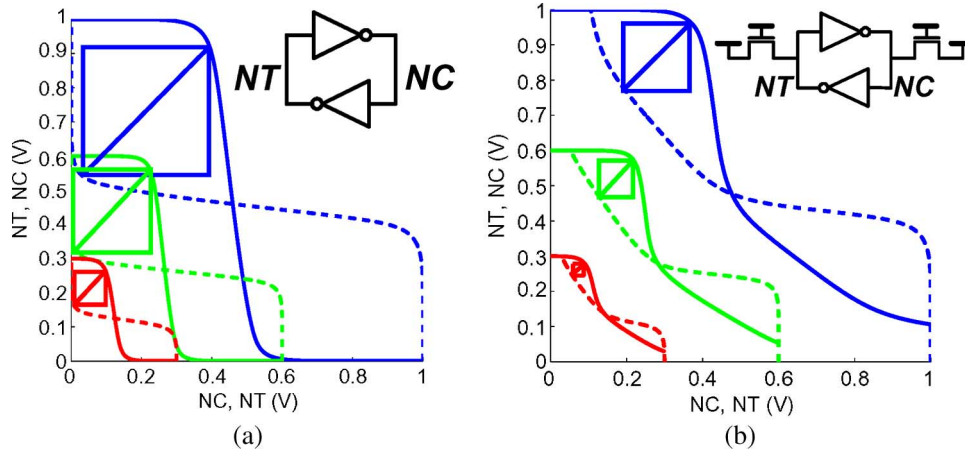


Fig. 10. 6T bit-cell butterfly curves showing bistable behaviors during (a) hold, where pass devices are OFF, and during (b) read, where pass devices are ON and bit-lines are clamped to V_{DD} .

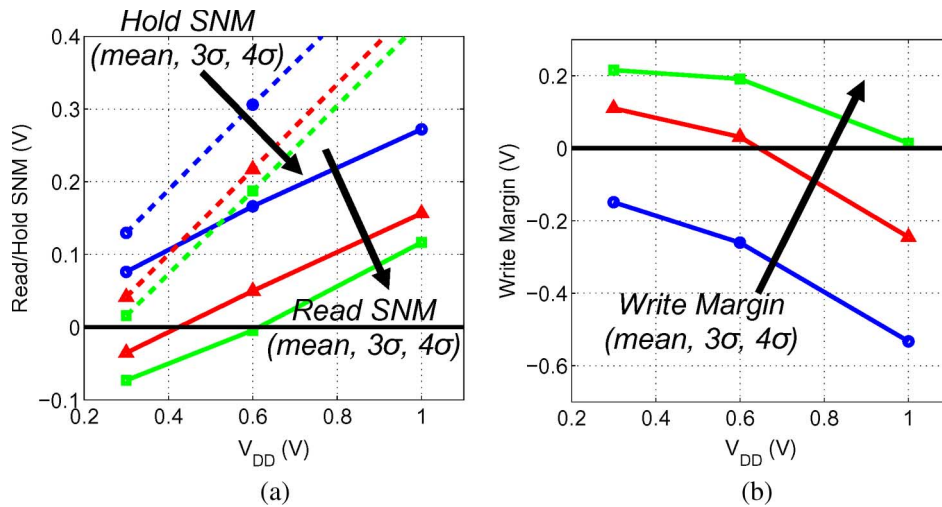


Fig. 11. 6T (a) read/hold SNM and (b) write margin.

logic paths involving gates with stacked devices. The contours of equal σ/μ variability illustrate the similarity between increased logic depth and device sizing. Importantly, however, the left-hand and bottom edges of the plot show diminishing returns. This implies that a small increase in one parameter can be traded off for a large decrease in the other while maintaining the same variability.

IV. SUB- V_t SRAM DESIGN

The design prescriptions that ensure robust sub- V_t logic, namely, static nonratioed CMOS topologies with minimal parallel leakage paths, are not practical for SRAMs. A critical metric for SRAMs, which retains its importance in sub- V_t , is density. As a result, device optimizations and circuit topologies that favor density are employed, and often, these rely heavily on device characteristics that, in sub- V_t , are severely degraded.

A. 6T SRAM Operation

Fig. 9 shows the architecture of a modern SRAM. A combination of row decoders and column multiplexers is used to access the bit-cells. Data-retention circuits for logic, like flip-

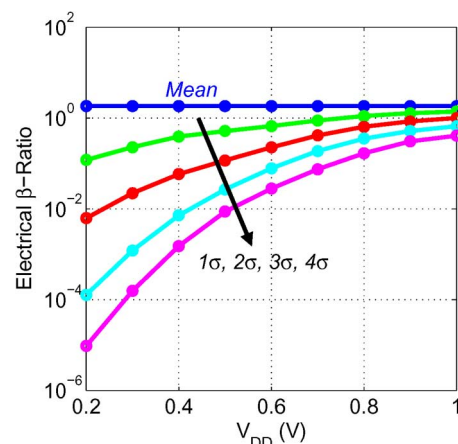


Fig. 12. Electrical β -ratio, in the presence of variation, is severely degraded, by over four orders of magnitude, in sub- V_t .

flops and latches, typically employ between 10 and 20 devices, but the six-transistor (6T) bit-cell shown relies on ratioed operation to achieve the required functionality with very high density; 6T CMOS bit-cells in the 65- and 45-nm nodes occupy between $0.4\text{--}0.5 \mu\text{m}^2$ [26], [27] and $0.24\text{--}0.33 \mu\text{m}^2$ [28], respectively.

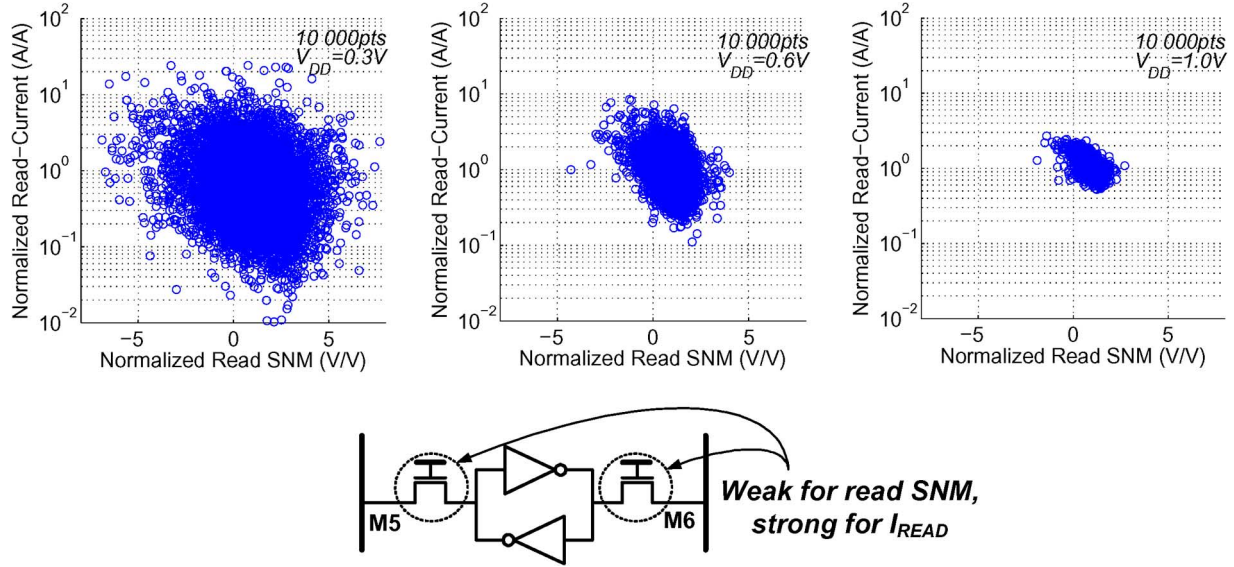


Fig. 13. Read current shows a strong correlation with read SNM, and the extent of the correlation increases drastically at reduced voltages.

Data is held in the 6T cell by the cross-coupled inverter structure, whose butterfly curves, shown in Fig. 10(a), have the bistable nature supporting logic “0” and “1” data retention to very low voltages. Strictly speaking, read access is a nonratioed operation where the bit-lines (BL/BLB) are precharged, and after word-line (WL) assertion, the cell read current I_{READ} causes a droop on one bit-line, which can be sensed with respect to the other to quickly decipher the accessed data. However, the worst-case transient behavior on the critical storage nodes (NT/NC) can result in loss of the bistable characteristic and can be analyzed by assuming that BL/BLB are clamped at V_{DD} . The butterfly curves in Fig. 10(b) now show precariously degraded lobes, quantified by the static noise margin (SNM), which measures the edge length of the largest embedded square [29]. In this scenario, proper operation is contingent on the driver devices $M1/4$, being stronger than the pass devices $M5/6$, and the critical ratio of their ON-currents is defined as the electrical β -ratio, as in Fig. 9. Typical β -ratios of 1.2–3 are required for proper operation.

Data is written to the 6T cell by pulling the appropriate bit-line low. The cell is made monostable at only the desired data value, and after WL gets deasserted, the local feedback regenerates to the correct state. Write operation is explicitly ratioed since the NMOS pass devices are required to overpower the PMOS load devices $M2/3$ in order to overwrite new data.

The ratioed operation, both during read and write, leaves the 6T SRAM highly susceptible to both variation and manufacturing defects. In particular, since a typical SRAM is composed of bit-cell arrays of hundreds of kilobits to several megabits, extreme worst-case behavior at the 4 or 5σ level must be considered.

B. 6T Bit-Cell Failure Mechanisms

Fig. 11 shows the Monte Carlo simulations of a bit-cell in 65-nm CMOS, considering the effect of RDF and gate length variation on the read/hold SNM and the write margin [17]. At low voltages, the read SNM is negative, indicating loss of

bistability, and the write margin is positive, indicating inability to achieve monostability; both conditions represent failures at sub- V_t supply levels. Generally speaking, the failures arise both because of the reduced signal levels at the reduced voltage, as shown in Fig. 10, and also because of the exponential effect of V_t variation, as discussed in Section II-A. The electrical β -ratio isolates the severe contribution from variation, and as shown in Fig. 12, it is degraded by over four orders of magnitude in deep sub- V_t .

An additional effect limiting the minimum supply voltage is gate-oxide soft breakdown, resulting in extremely high gate-leakage in the driver devices $M1/4$ [30]. In 65 nm and beyond, even with very high-quality oxide, soft breakdown unfavorably distorts the read butterfly curves, limiting the minimum voltage for read stability similar to RDF.

Study of the design tradeoff between the electrical β -ratio and I_{READ} suggests that the 6T topology imposes inherent restrictions to sub- V_t operation. Fig. 13 shows a strong inverse-correlation between I_{READ} and read SNM; read SNM requires a high β -ratio, implying weak pass devices $M5/6$, and, accordingly, a reduced read current. Consequently, strategies to improve read SNM by increasing β , through pass device down-sizing or reduction in WL bias, negatively affect I_{READ} , severely limiting not just performance but, more importantly, functionality, as discussed in Section IV-D. The electrical β -ratio can also be increased by upsizing the driver devices $M1/4$. However, the upsizing required to overcome the degradation in Fig. 12 is too drastic to achieve through sizing alone, particularly since the resulting effect on density would be too costly. Furthermore, a large increase in gate area can exacerbate the limiting effect of gate-oxide soft breakdown [31], opposing the read SNM improvement.

C. Sub- V_t Read and Write Stabilizing Circuits

Circuit assists to incrementally improve the read SNM of a 6T cell are insufficient for sub- V_t operation. Significant improvement is afforded, instead, by the use of a read-buffer,

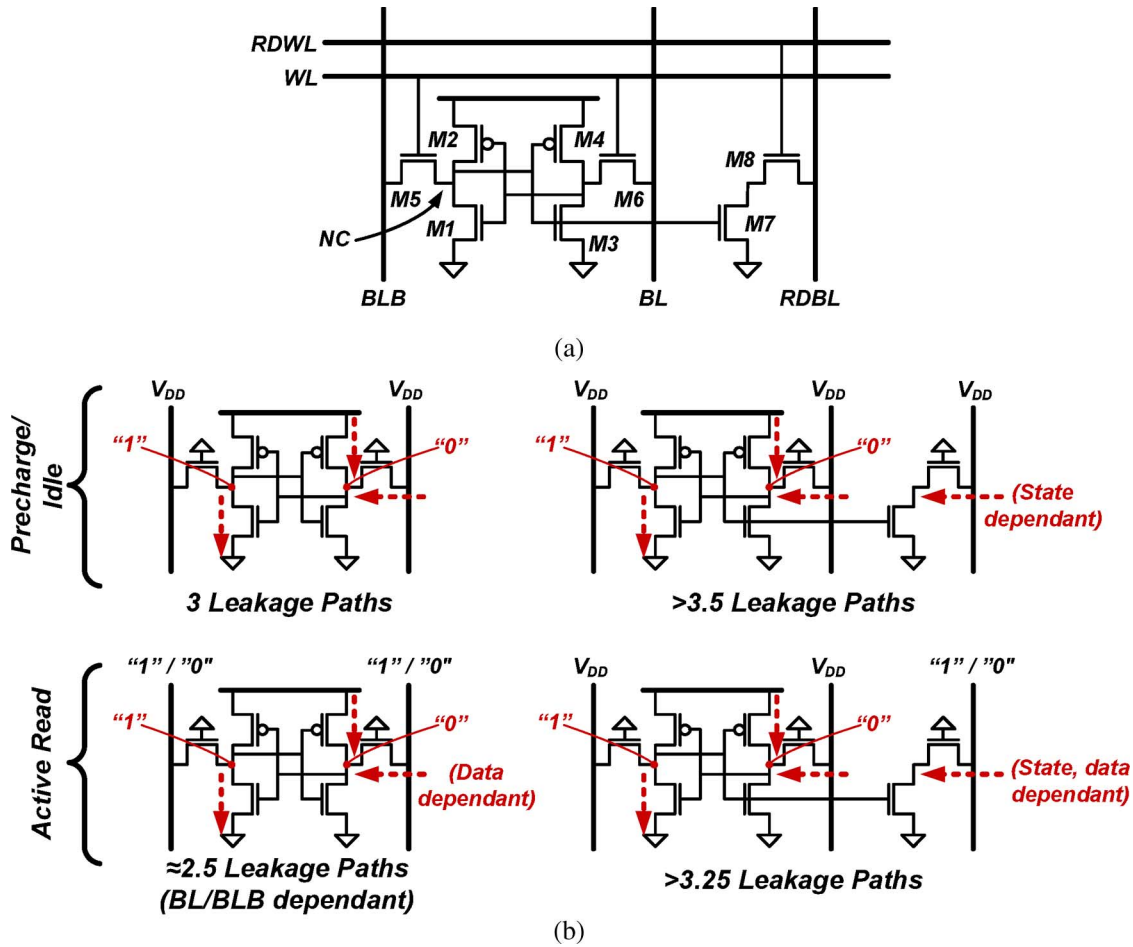


Fig. 14. Bit-cell with (a) 8T topology uses a read-buffer for significantly improved operating margins and I_{READ} , but (b) introduces additional leakage paths.

which isolates the storage nodes (NT/NC). The resulting structure is shown in Fig. 14(a) and can be free of the read SNM limitation. The operating margin, which is thus greatly improved, can instead be set by the hold SNM and the write margin. Fig. 11(a) indicates that, in the presence of RDF, the hold SNM enables operation deep into sub- V_t , and additionally, the increased margin affords significant immunity against soft breakdown failure mechanisms.

Although the read-buffer increases the size of the bit-cell by 25%–40% [14], the overwhelming improvement in I_{READ} and stability justifies the additional overhead. Unfortunately, however, the extra devices do introduce additional leakage, resulting in an increase of over 20% in leakage power, which, as mentioned in Section II, is a critical metric for sub- V_t SRAMs. Specifically, as shown in Fig. 14(b), during the precharge portion of the read-cycle, both branches of the cross-coupled inverters and one of the pass devices pose leakage paths. Additionally, a leakage current of half the magnitude on average (due to the dependence on the stored data) passes through the read-buffer. The other half of the time, the leakage current is reduced somewhat owing to the stacked effect of the two OFF NMOS devices in series [32]. The net result is a leakage current increase of over 17%. However, during the bit-line discharge portion of the read-cycle, the pass device leakage current in the 6T cell can actually contribute to the read current depending on the accessed data. In that case, it can be disregarded. The

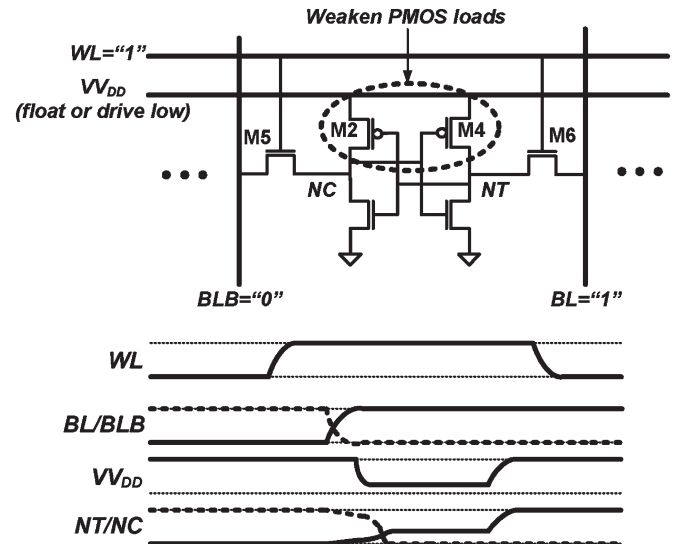


Fig. 15. Virtual V_{DD} allows internal cell feedback to be weakened during write operation to ensure that pass devices can overwrite new data.

leakage currents in the storage element of the 8T cell remain unchanged, but depending on both the stored and accessed data, the read-buffer contributes an additional leakage path. The net increase in this case is over 30%.

As mentioned in Section IV-A, write operation depends on the ability of the NMOS pass devices $M5/6$ to overpower

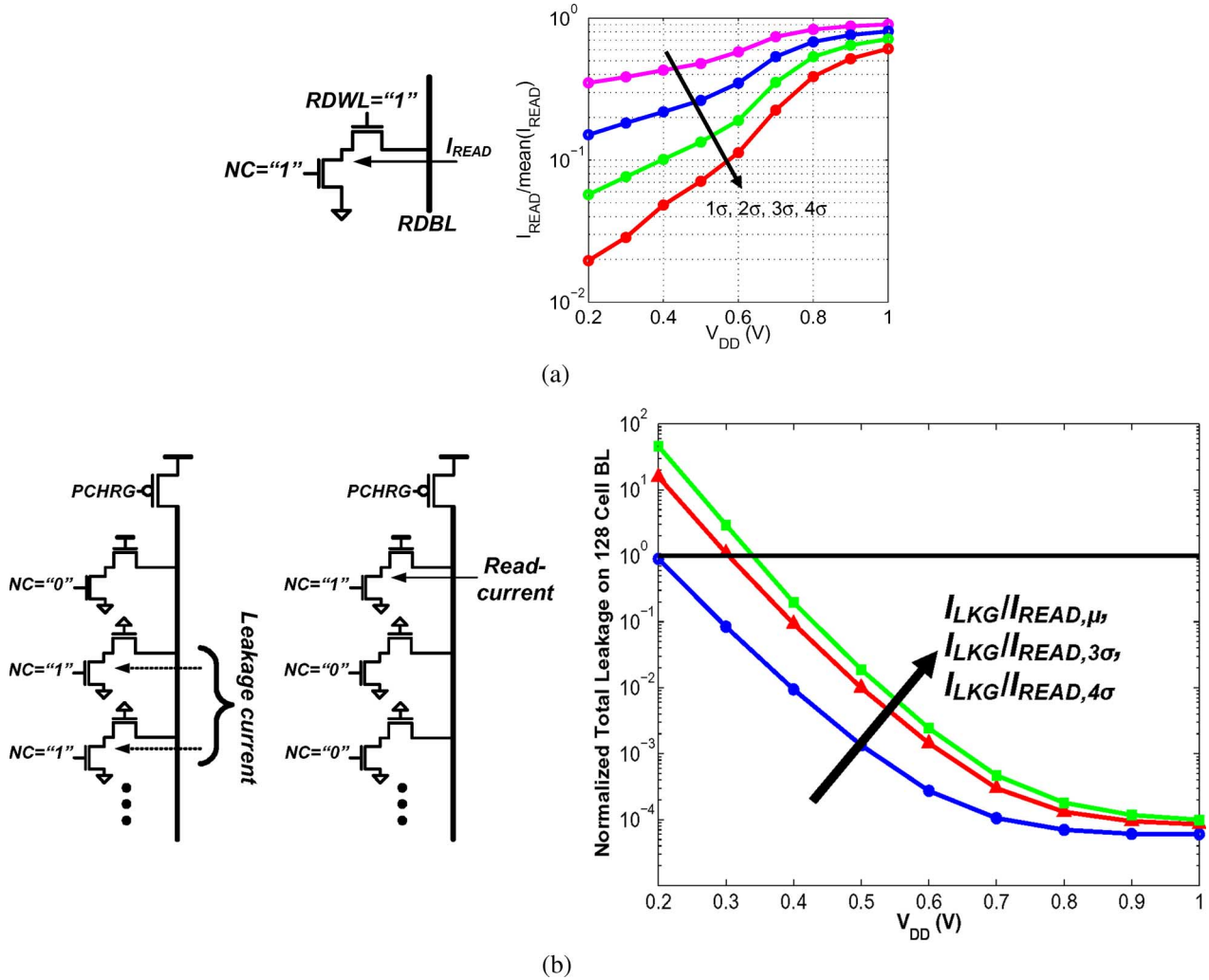


Fig. 16. Read-current degradation only from variation is (a) most severe in sub- V_t , and (b) it can be less than the aggregate leakage currents from the unaccessed cells sharing the bit-lines.

the PMOS load devices $M2/3$. As shown in Fig. 11(b), those relative strengths cannot be guaranteed in the presence of variation. Although a read-buffer eliminates read SNM limitations, allowing pass devices to be upsized or overdriven, sizing in sub- V_t is relatively ineffective due to the severe impact of variation (as explained in Section III), and overdriving the word-line entails significant overhead in boosting a large capacitance beyond the rail voltage. Alternatively, the desired relative strengths can be enforced by weakening the PMOS loads. For instance, the bit-cell shown in Fig. 15 uses a virtual V_{DD} [9], [17], [18]. During a write access, V_{DD} either floats or is actively biased to reduce the strength of the PMOS loads, ensuring that the pass devices can overwrite new data.

D. Sub- V_t Read-Current Degradation

A significantly reduced read current is expected in sub- V_t due to the lower gate drive. However, the exponential impact of variation further degrades I_{READ} . If the statistical I_{READ} 's are normalized by the mean I_{READ} , the effect of variation alone can be isolated, and as shown in Fig. 16(a), this effect is particularly severe in sub- V_t , where the weak-cell read current

can easily be a couple of orders of magnitude worse than the mean current. The combination of variation on top of drastically reduced mean read current implies that the read access time can extend almost arbitrarily. This is undesirable from a performance point of view, but more importantly, it affects the ability to correctly sense data. Specifically, all of the unaccessed cells that share the read bit-line impose a leakage current that depends on their stored data. In Fig. 16(b), the aggregate leakage current is normalized by the statistical read current, assuming 128 cells per bit-line. As shown, the leakage can exceed the read signal, making the accessed data indecipherable [17].

The first solution path focuses on increasing the cell read current. I_{READ} can be increased by increasing the width of the read-buffer devices. Normally, the resulting increase in cell area makes this approach unattractive. However, since the standard deviation of V_t is inversely related to the square root of device areas [33], the variation-induced degradation of I_{READ} is greatly reduced in sub- V_t , where the dependence on V_t is exponential. Accordingly, up-sizing has enhanced appeal in sub- V_t . Fig. 17(a) shows the gain in 4σ read current if the read-buffer devices are upsized by 25% and 50%. Although the

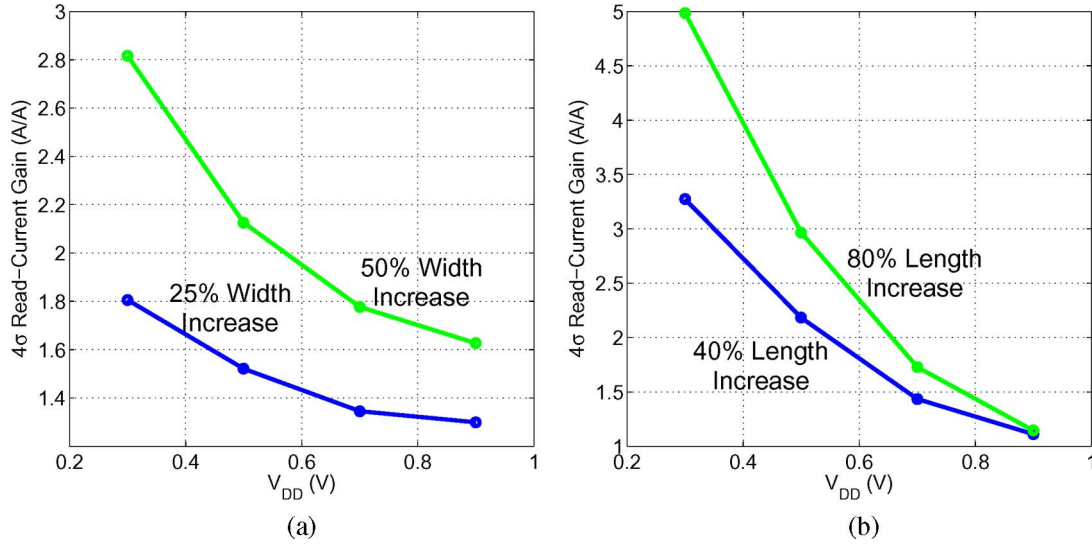


Fig. 17. Read-current gain of 4σ cell from read-buffer device up-sizing of (a) width and (b) length.

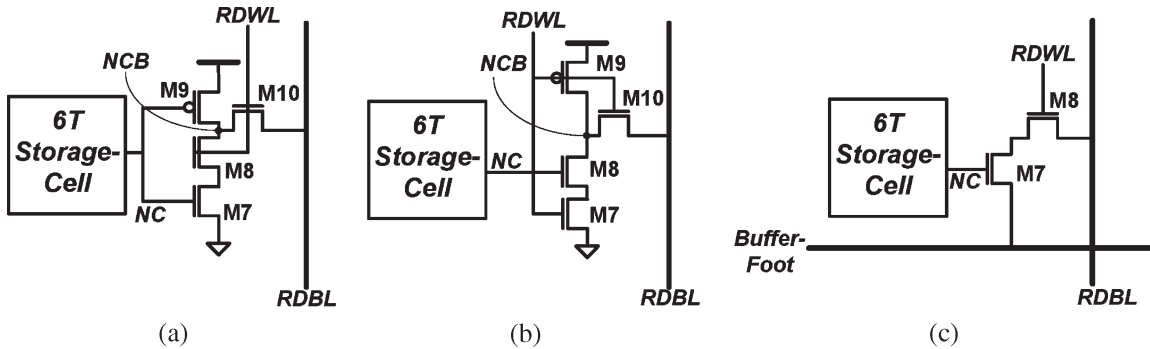


Fig. 18. Read-buffers with no sub- V_t bit-line leakage: (a) 10T cell relying on PMOS/NMOS I_{OFF} ratio [18]. (b) 10T cell independent of I_{OFF} , but with increased leakage power [10]. (c) 8T cell employing peripheral assist to manage bit-line leakage while maintaining 8T density [17].

mean read currents are expected to increase by only $1.25\times$ and $1.5\times$, the weak-cell read currents, which limit the design and performance of the entire array, increase by $1.8\times$ and $2.8\times$ in sub- V_t . Similarly, increasing the device lengths by 40% and 80% increases the weak-cell read currents by $3.3\times$ and $5\times$, respectively, as shown in Fig. 17(b). It is important to note that, in sub- V_t , increasing device lengths can have a significant impact on increasing even the mean read current since the reverse short-channel effect [34] essentially causes a decrease in the effective V_t .

The second solution path focuses on reducing the leakage currents from the unaccessed bit-cells. The read-buffers shown in Fig. 18 ensure that, after RDBL is precharged, there is no voltage drop across the pass devices of the unaccessed cells regardless of the stored data, thereby eliminating any sub- V_t read bit-line leakage [10], [17], [18]. In Fig. 18(a), node NCB is actively driven high when NC is low, and when NC is high, its value is set by the relative leakage currents of M8 and M9. Importantly, however, the threshold voltage of PMOS devices is often engineered to be lower than that of NMOS devices to offset the drive current asymmetry in strong inversion due to mobility differences. As a result, the leakage current of the PMOS is exponentially higher, and NCB tends to approach V_{DD} . In Fig. 18(b), NCB is actively driven high,

independent of NC. As a result, this structure is more robust to skewed process corners where the PMOS strength is reduced relative to the NMOS strength; however, since it does not take advantage of the stacked effect [32], its total leakage current is considerably higher. Finally, in Fig. 18(c), a peripheral assist is used to pull up the Buffer-Foot node in all unaccessed cells, eliminating the leakage path to ground through the read-buffer. In this topology, no additional devices are required in the bit-cell, significantly enhancing the array density; however, the peripheral Buffer-Foot driver does need to sink the read current from all cells in the accessed row, without contributing a large leakage current of its own. Hence, a simple charge pump circuit is used to provide over $500\times$ current gain without any increase in the device sizes. All approaches in Fig. 18 are effective in mitigating bit-line leakage and enable integration of over 256 cells per column while operating in deep sub- V_t .

V. CONCLUSION

This paper describes the value of sub- V_t operation for achieving minimum energy in circuits. Although numerous applications are enabled as a result, the severe variation and current degradation in sub- V_t devices are a critical barrier to achieving robust ultralow-voltage systems. In particular, both logic and

SRAM circuits require special treatment in order to overcome the associated failures. Proper logic styles and device sizing ensure functional logic gates for a desired yield, while long logic paths provide an averaging effect similar to device up-sizing, which reduces the statistical spread of propagation delays. Conventional 6T SRAMs fail to operate in sub- V_t because of both increased variation and reduced signal levels. Furthermore, read-current degradation and soft-oxide breakdown pose limiting tradeoffs in improving bit-cell stability. Instead, bit-cell topologies incorporating a read-buffer provide a viable alternative and can be enhanced to mitigate bit-line leakage mechanisms that prevent reliable data sensing. The prescribed techniques enable circuit operation to below 0.3 V in advanced technologies and are compatible with the energy levels required for emerging portable and self-powered systems.

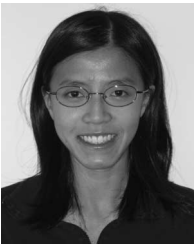
REFERENCES

- [1] B. H. Calhoun, D. C. Daly, N. Verma, D. F. Finchelstein, D. D. Wentzloff, A. Wang, S.-H. Cho, and A. Chandrakasan, "Design considerations for ultra-low energy wireless microsensor nodes," *IEEE Trans. Comput.*, vol. 54, no. 6, pp. 727–740, Jun. 2005.
- [2] B. Calhoun and A. Chandrakasan, "Ultra-dynamic voltage scaling using sub-threshold operation and local voltage dithering in 90 nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2005, pp. 300–301.
- [3] V. Sze, R. Blazquez, M. Bhardwaj, and A. Chandrakasan, "An energy efficient sub-threshold baseband processor architecture for pulsed ultra-wideband communications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2006, pp. 908–911.
- [4] B. H. Calhoun, A. Wang, and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1778–1786, Sep. 2005.
- [5] A. Wang and A. Chandrakasan, "A 180 mV FFT processor using sub-threshold circuit techniques," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2004, pp. 292–293.
- [6] M.-E. 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 *Proc. IEEE Symp. VLSI Circuits*, Jun. 2007, pp. 154–155.
- [7] M. Miyazaki, J. Kao, and A. P. Chandrakasan, "A 175 mV multiply-accumulate unit using an adaptive supply voltage and body bias (ASB) architecture," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2002, pp. 58–59.
- [8] 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 *Proc. IEEE Symp. VLSI Circuits*, Jun. 2006, pp. 154–155.
- [9] B. Zhai, D. Blaauw, D. Sylvester, and S. Hanson, "A sub-200 mV 6T SRAM in 0.13 μm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 332–333.
- [10] T.-H. Kim, J. Liu, J. Kean, and C. H. Kim, "A high-density subthreshold SRAM with data-independent bitline leakage and virtual ground replica scheme," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 330–331.
- [11] T. Suzuki, Y. Yamagami, I. Hatanaka, A. Shibayama, H. Akamatsu, and H. Yamauchi, "0.3 to 1.5 V embedded SRAM with device-fluctuation-tolerant access-control and cosmic-ray-immune hidden-ECC scheme," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2005, pp. 484–485.
- [12] Y. Morita, H. Fujiwara, H. Noguchi, K. Kawakami, J. Miyakoshi, S. Mikami, K. Nii, H. Kawaguchi, and M. Yoshimoto, "A V_{th} -variation-tolerant SRAM with 0.3-V minimum operation voltage for memory-rich SoC under DVS environment," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2006, pp. 13–14.
- [13] V. Sze and A. Chandrakasan, "A 0.4-V UWB baseband processor," in *Proc. Int. Symp. Low Power Electron. Des.*, Aug. 2007, pp. 262–267.
- [14] Y. Morita, H. Fujiwara, H. Noguchi, Y. Iguchi, K. Nii, H. Kawaguchi, and M. Yoshimoto, "An area-conscious low-voltage-oriented 8T-SRAM design under DVS environment," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2007, pp. 256–257.
- [15] K. Takeda, Y. Hagihara, Y. Aimoto, M. Nomura, Y. Nakazawa, T. Ishii, and H. Kobatake, "A read-static-noise-margin-free SRAM cell for low- V_{DD} and high-speed applications," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2005, pp. 478–479.
- [16] Y. Ramadass and A. P. Chandrakasan, "Minimum energy tracking loop with embedded dc–dc converter delivering voltages down to 250 mV in 65 nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 64–65.
- [17] N. Verma and A. Chandrakasan, "A 65 nm 8T sub- V_t SRAM employing sense-amplifier redundancy," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 328–329.
- [18] B. Calhoun and A. Chandrakasan, "A 256 kb sub-threshold SRAM in 65 nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2006, pp. 628–629.
- [19] L. Chang, Y. Nakamura, R. K. Montoye, J. Sawada, A. K. Martin, K. Kinoshita, F. H. Gebara, K. B. Agarwal, D. J. Acharya, W. Haensch, K. Hosokawa, and D. Jamsek, "A 5.3 GHz 8T-SRAM with operation down to 0.41 V in 65 nm CMOS," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2007, pp. 252–253.
- [20] M. Khellah, Y. Ye, N.-S. Kim, D. Somasekhar, G. Pandya, A. Farhang, K. Zhang, C. Webb, and V. De, "Wordline and bitline pulsing schemes for improving SRAM cell stability in low- V_{cc} 65 nm CMOS designs," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2006, pp. 9–10.
- [21] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug. 1995.
- [22] A. Wang, A. Chandrakasan, and S. Kosonocky, "Optimal supply and threshold scaling for sub-threshold CMOS circuits," in *Proc. IEEE Comput. Soc. Annu. Int. Symp. VLSI*, Apr. 2002, pp. 5–9.
- [23] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "Analysis and mitigation of variability in subthreshold design," in *Proc. Int. Symp. Low Power Electron. Des.*, 2005, pp. 20–25.
- [24] J. Chen, L. T. Clark, and Y. Cao, "Ultra-low voltage circuit design in the presence of variations," *IEEE Circuits Devices Mag.*, pp. 12–20, Nov/Dec. 2005.
- [25] J. Kwong and A. Chandrakasan, "Variation-driven device sizing for minimum energy sub-threshold circuits," in *Proc. Int. Symp. Low Power Electron. Des.*, 2006, pp. 8–13.
- [26] K. Utsumi, E. Morifuji, M. Kanda, S. Aota, T. Yoshida, K. Honda, Y. Matsubara, S. Yamada, and F. Matsuoka, "A 65 nm low power CMOS platform with 0.495 μm^2 SRAM for digital processing and mobile applications," in *Proc. IEEE Symp. VLSI Technol.*, Jun. 2005, pp. 216–217.
- [27] A. Chatterjee, J. Yoon, S. Zhao, S. Tang, K. Sadra, S. Crank, H. Mogul, R. Aggarwal, B. Chatterjee, S. Lytle, C. T. Lin, K. D. Lee, J. Kim, L. Olsen, M. Quevedo-Lopez, K. Kirmse, G. Zhang, C. Meek, D. Aldrich, H. Mair, M. Mehrotra, L. Adam, D. Mosher, J. Yang, D. Crenshaw, B. Williams, J. Jacobs, M. Jain, J. Rosal, T. Houston, J. Wu, N. S. Nagaraj, D. Scott, S. Ashburn, and A. Tsao, "A 65 nm CMOS technology for mobile and digital signal processing applications," in *IEDM Tech. Dig. Papers*, Dec. 2004, pp. 665–668.
- [28] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, S. Imaoka, H. Makino, Y. Yamagami, S. Ishikura, T. Terano, T. Oashi, K. Hashimoto, A. Sebe, G. Okazaki, K. Satomi, H. Akamatsu, and H. Shinohara, "A 45 nm low-standby-power embedded SRAM with improved immunity against process and temperature variations," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 326–327.
- [29] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. SSC-22, no. 5, pp. 748–754, Oct. 1987.
- [30] M. Agostinelli, J. Hicks, J. Xu, B. Woolery, K. Mistry, K. Zhang, S. Jacobs, J. Jopling, W. Yang, B. Lee, T. Raz, M. Mehal, P. Kolar, Y. Wang, J. Sandford, D. Pivin, C. Peterson, M. DiBattista, S. Pae, M. Jones, S. Johnson, and G. Subramanian, "Erratic fluctuations of SRAM cache V_{min} at the 90 nm process technology node," in *IEDM Tech. Dig. Papers*, Dec. 2005, pp. 655–658.
- [31] R. Rodriguez, J. H. Stathis, B. P. Linder, S. Kowalczyk, C. T. Chuang, R. V. Joshi, G. Northrop, K. Bernstein, A. J. Bhavnagarwala, and S. Lombardo, "The impact of gate-oxide breakdown on SRAM stability," *IEEE Electron Device Lett.*, vol. 23, no. 9, pp. 559–561, Sep. 2002.
- [32] Y. Ye, S. Borkar, and V. De, "A new technique for standby leakage reduction in high-performance circuits," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 1998, pp. 40–41.
- [33] 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.
- [34] 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.



Naveen Verma (S'04) received the B.A.Sc. degree in electrical and computer engineering from the University of British Columbia, Vancouver, BC, Canada, in 2003 and the M.S. degree from the Massachusetts Institute of Technology, Cambridge, in 2005, where he is currently working toward the Ph.D. degree.

He is the recipient of the Intel Foundation Ph.D. fellowship and the NSERC Postgraduate fellowship. His research interests include low-power mixed signal circuits in the areas of analog-to-digital converters, SRAMs, and implantable biological systems.



Joyce Kwong (S'02) received the B.A.Sc. degree in computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2004 and the M.S. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 2006, where she is currently working toward the Ph.D. degree.

She is the recipient of the Texas Instruments Graduate Woman's Fellowship for Leadership in Microelectronics and the NSERC Postgraduate Fellowship. Her research interests include subthreshold circuit design methodology and system implementation.



Anantha P. Chandrakasan (M'95–SM'01–F'04) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1989, 1990, and 1994, respectively.

Since September 1994, he has been with Massachusetts Institute of Technology (MIT), Cambridge, where he is currently the Joseph F. and Nancy P. Keithley Professor of Electrical Engineering and the Director of the Microsystems Technology Laboratories. His research interests include low-power digital integrated circuit design, wireless microsensors, ultrawideband radios, and emerging technologies. He is a coauthor of *Low Power Digital CMOS Design* (Kluwer, 1995), *Digital Integrated Circuits* (Pearson Prentice-Hall, 2nd ed., 2003), and *Sub-Threshold Design for Ultra-Low Power Systems* (Springer, 2006). He is also a coeditor 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 was a corecipient of several awards, including the 1993 IEEE Communications Society's Best Tutorial Paper Award, the IEEE Electron Devices Society's (EDS) 1997 Paul Rappaport Award for the Best Paper in an EDS publication during 1997, the 1999 Design Automation Conference (DAC) Design Contest Award, the 2004 DAC/International Solid-State Circuits Conference (ISSCC) Student Design Contest Award, and the ISSCC 2007 Beatrice Winner Award for Editorial Excellence. He has served as a Technical Program Cochair for the 1997 International Symposium on Low Power Electronics and Design, VLSI Design 1998, 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 2004–2007. He was an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS from 1998 to 2001. He served on the IEEE Solid-State Circuits Society AdCom from 2000 to 2007, where he was the Meetings Committee Chair from 2004 to 2007. He is the Technology Directions Chair for ISSCC 2008.