Nanometer Device Scaling in Subthreshold Logic and SRAM

Scott Hanson, *Student Member, IEEE*, Mingoo Seok, *Student Member, IEEE*, Dennis Sylvester, *Senior Member, IEEE*, and David Blaauw, *Member, IEEE*

***Abstract*—Subthreshold circuit design is promising for future ultralow-energy sensor applications as well as highly parallel high-performance processing. Device scaling has the potential to increase speed in addition to decreasing both energy and cost in subthreshold circuits. However, no study has yet considered whether device scaling to 45 nm and beyond will be beneficial for subthreshold logic. We investigate the implications of device scaling on subthreshold logic and SRAM and find that the slow scaling of gate-oxide thickness leads to a 60% reduction in I_{on}/I_{off} between the 90- and 32-nm device generations. We highlight the effects of this device degradation on noise margins, delay, and energy. We subsequently propose an alternative scaling strategy and demonstrate significant improvements in noise margins, delay, and energy in sub-V_{th} circuits. Using both optimized and unoptimized subthreshold device models, we explore the robustness of scaled subthreshold SRAM. We use a simple variability model and find that even small memories become unstable at advanced technology nodes. However, the simple device optimizations suggested in this paper can be used to improve nominal read noise margins by 64% at the 32-nm node.**

***Index Terms*—Device optimization, low power, subthreshold circuits.**

I. Introduction

INCREASINGLY stringent power budgets have fostered the rapid growth of sophisticated low-power design techniques in recent years. Subthreshold (sub-V_{th}) design has emerged at the forefront of these techniques. In sub-V_{th} design, the supply voltage (V_{dd}) is scaled below the device threshold voltage (V_{th}). A wide range of applications, from radio-frequency-identification tags to cellular phones, demands minute energy budgets and can tolerate the modest switching speeds that are characteristic of sub-V_{th} circuits. Although sub-V_{th} design has not yet gained widespread commercial adoption, recent work has shown that the potential benefits of sub-V_{th} circuits are substantial. Sub-V_{th} processors have been demonstrated with the supply voltage V_{dd} as low as 180 mV [1] and with an energy consumption of only 2.6 pJ/instruction [2]. The speed of sub-V_{th} circuits, which is exponentially dependent on V_{th} and V_{dd}, has generally been reported in the kilohertz and low-megahertz range [1], [2]. In addition to these relatively low data rates, timing variability grows dramatically as V_{dd} reduces, forcing the adoption of pessimistic design practices and large timing margins. Despite poor single-processor performance, sub-V_{th} circuits are also promising for power reduction in high-performance parallel systems [29].

The scaling of transistor dimensions and electrical characteristics represents both an opportunity and a threat for sub-V_{th} circuits. Device scaling offers a reduction in gate capacitance, and at superthreshold (super-V_{th}) voltages, it offers a welcome reduction in switching energy and gate delay. Scaling has also led to a dramatic increase in density (which was an effective cost-reduction measure in the past [9]). At the same time, device scaling has brought about a number of problems in super-V_{th} circuits, including process variability, increased subthreshold leakage, and increased gate leakage. The implications of device scaling on super-V_{th} circuits have been explored previously [7], [8]; however, no such focus has been given to sub-V_{th} circuits. Transistor design is particularly important in the sub-V_{th} regime due to exponential sensitivities to V_{th}, V_{dd}, and inverse subthreshold slope; therefore, it is not immediately clear how sub-V_{th} circuits will fare under device scaling.

Sub-V_{th} device optimizations were considered in [3] and [4], and it was shown that the optimal sub-V_{th} device should minimize the inverse subthreshold slope. Additionally, the use of drain–source underlap was suggested for sub-V_{th} devices in [5]. The use of ultrathin body FinFETs in sub-V_{th} logic was advocated for improved channel control and variability characteristics in [4] and [6]. However, no study has comprehensively studied the effects that device scaling will have on sub-V_{th} circuits. In this paper, we study the evolution of static noise margins (SNMs), performance, and energy in sub-V_{th} circuits as the devices scale deep into the nanometer regime. We place a strong emphasis on understanding the consequences of traditional performance-driven scaling on simple inverters and SRAM cells and also propose an improved scaling strategy targeting the needs of sub-V_{th} circuits.

We first use realistic 2-D device models (in MEDICI) scaled from the 90-nm technology node down to the 32-nm technology node to quantify the device- and gate-level implications of conventional super-V_{th} device scaling. We show that the slow scaling of gate oxide relative to the channel length leads to a 60% reduction in I_{on}/I_{off} between the 90- and 32-nm nodes, which results in SNM degradation of more than 10% between the 90- and 32-nm nodes in a CMOS inverter. We propose a modified scaling strategy that uses increased channel lengths and reduced doping to improve the inverse subthreshold slope. We develop new delay and energy metrics that effectively capture the important effects of device scaling, and we use those to drive device optimization. We find that the noise margins improve by 19% and the energy improves by 23% in 32-nm

Manuscript received June 1, 2007. The review of this paper was arranged by Editor S. Kosonocky.

The authors are with the University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: hansons@umich.edu).

Digital Object Identifier 10.1109/TED.2007.911033

0018-9383/$25.00 © 2008 IEEE

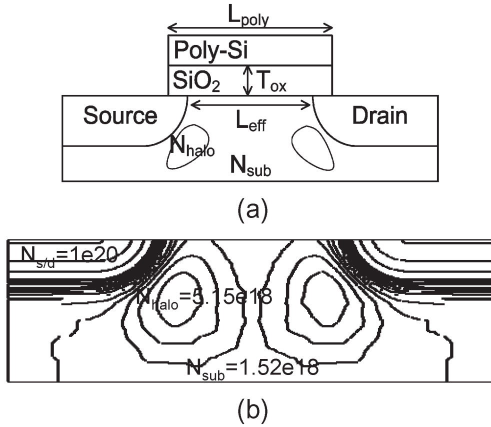


Fig. 1. (a) Device cross section showing the scaling parameters. (b) Doping profile for a 90-nm NFET.

sub-$V_{\rm th}$ circuits when applying our modified device scaling strategy. Our approach is particularly attractive since it requires only simple modifications to the existing device technologies. Following this initial analysis, we look at the problem of scaled sub-$V_{\rm th}$ SRAM, which will likely be the sub-$V_{\rm th}$ circuit that is most sensitive to device scaling. We use both nominal- and corner-based analyses to measure the noise margins in memories targeted at sensor applications, where memory sizes on the order of several kilobits are sufficient. We find that our optimized sub-$V_{\rm th}$ device has a nominal read SNM that is 64% larger than that of the unoptimized super-$V_{\rm th}$ device at the 32-nm generation.

The remainder of this paper is organized as follows. In Section II, we describe the implications of performance-driven scaling in the sub-$V_{\rm th}$ regime. In Section III, we propose an alternative scaling strategy driven by the needs of sub-$V_{\rm th}$ circuits and compare it to a super-$V_{\rm th}$ scaling strategy. In Section IV, we study scaled sub-$V_{\rm th}$ SRAM. Finally, in Section V, we conclude this paper.

II. Super-$V_{\rm th}$ Scaling Trends

In this section, we describe a simple but accurate bulk transistor model, as shown in Fig. 1(a), which captures the important effects of conventional super-$V_{\rm th}$ scaling. Our text and figures will focus on the NFET device for the remainder of this paper, but we use an analogous methodology to describe the PFET device. The device model has four key scaling parameters: physical gate length ($L_{\rm poly}$), gate-oxide thickness ($T_{\rm ox}$), substrate doping ($N_{\rm sub}$), and peak halo doping ($N_{p,\rm halo}$). These parameters receive special attention because they are most important when determining key device characteristics like $V_{\rm th}$, on current, off current, and gate capacitance. In addition to these four parameters, we specify $V_{\rm dd}$ as an additional knob for adjusting performance. All physical dimensions other than $T_{\rm ox}$ (source/drain junction depth, lateral source/drain diffusion, halo dimensions, etc.) scale in proportion to $L_{\rm poly}$.

Note that halo-doping regions are located near the source and drain edges. Halo doping is used to control $V_{\rm th}$ rolloff observed at short channels and large drain biases, and it has become indispensable for super-$V_{\rm th}$ devices. The $V_{\rm th}$ of a short-channel device with halo doping may be represented as the sum of three components: intrinsic (long-channel) threshold voltage ($V_{\rm th0}$), rolloff due to short-channel effects and drain-induced barrier lowering ($\Delta V_{\rm th,SCE}$), and roll-up due to the halo doping ($\Delta V_{\rm th,halo}$) [11]. In a well-optimized device, the halo regions increase the effective channel doping at short channel lengths such that $-\Delta V_{\rm th,SCE} = \Delta V_{\rm th,halo}$ and that $V_{\rm th}$ remains flat as a function of both $L_{\rm poly}$ and $V_{\rm ds}$. We model the halo regions as a pair of 2-D Gaussian distributions superimposed on a uniformly doped substrate similar to [3] and [12]. The doping contours of a representative 90-nm device are shown for illustrative purposes in Fig. 1(b). The net halo doping $N_{\rm halo}$ is the sum of $N_{\rm sub}$ and $N_{p,\rm halo}$.

For our purposes, describing a device at a particular technology node only requires that the four key parameters and the $V_{\rm dd}$ are specified. We use the iterative process described in [28] to optimize device parameters at a given technology node. $L_{\rm poly}$ and $T_{\rm ox}$ are first determined based on published industry data. $V_{\rm dd}$ and $V_{\rm th}$ (through $N_{\rm sub}$ and $N_{p,\rm halo}$) are then chosen to optimize the delay under leakage constraints. We describe the selection of each parameter in the remainder of this section.

The aggressive scaling of $L_{\rm poly}$ has been one of the primary drivers of performance improvement in MOSFETs. Note that $L_{\rm poly}$ represents the length of the bottom of the poly-Si gate after etching. For example, a gate with a designed length of 90 nm might have an $L_{\rm poly} = 65$ nm after etching. Throughout this paper, we assume that the minimum $L_{\rm poly}$ is reduced by 30% per generation, which agrees well with the recent $L_{\rm poly}$ scaling trends.

Selecting a realistic value for $T_{\rm ox}$ plays a critical role in determining the sub-$V_{\rm th}$ characteristics of a device. As suggested in the previous section, $T_{\rm ox}$ has actually scaled more slowly than $L_{\rm poly}$ due to oxide-reliability and gate-leakage concerns. A survey of recent industrial publications in [13] shows that $T_{\rm ox}$ has been reduced by $\sim$10% per generation below the 130-nm technology node. In this paper, we make the simple assumption that $T_{\rm ox}$ reduces by 10% per generation. Note that the oxide-scaling problem may be even worse than our assumption of 10%. High-κ dielectrics may be the only solution since conventional gate stacks may be limited to a minimum of $\sim$1-nm thickness [20].

With $L_{\rm poly}$ and $T_{\rm ox}$ fixed for each generation, the remaining three parameters ($N_{\rm sub}$, $N_{p,\rm halo}$, and $V_{\rm dd}$) may be tuned to match the delay and leakage requirements. As in [28], our optimization uses delay (τ) as an objective and leakage ($I_{\rm leak,max}$) as a constraint. Note that $N_{\rm sub}$ is treated as a function of the long-channel device (where the halo doping is largely unnecessary) and that $N_{p,\rm halo}$ is treated as a function of the short-channel device. While the approach described in [28] may not converge on the optimal solution, it is a systematic simple heuristic that produces realistic scaled devices.

The selection of $I_{\rm leak,max}$ is a complex topic since every new technology provides a range of devices optimized for different power-delay points. For example, the 65-nm technology described in [14] offers low- and high-power devices, with each device having three different $V_{\rm th}$ variants. The International Technology Roadmap for Semiconductors (ITRS) [15], which maps out near- and long-term goals for the semiconductor

TABLE I
NFET PARAMETERS UNDER SUPER-$V_{\rm th}$ SCALING

Node	90nm	65nm	45nm	32nm
L_{poly} (nm)	65	46	32	22
T_{ox} (nm)	2.10	1.89	1.70	1.53
N_{sub} (cm^{-3})	1.52e18	1.97e18	2.52e18	3.31e18
N_{halo} (cm^{-3})	3.63e18	5.17e18	7.83e18	12.0e18
$N_{ch,avg}$ (cm^{-3})	2.82e18	3.84e18	5.27e18	7.38e18
V_{dd}	1.2	1.1	1.0	0.9
$V_{th,sat}$ (mV)	403	420	438	461
I_{off} (pA/µm)	100	125	156	195
C_gV_{dd}/I_{on} (ps)	1.3	0.97	0.75	0.62

industry, describes three different devices with different power-delay tradeoffs: high performance, low operating power (LOP), and low standby power (LSTP). The LOP and LSTP devices are optimized in a similar manner, although the LSTP device has more stringent leakage constraints. In this paper, we use a super-$V_{\rm th}$ scaling strategy similar to that of the LSTP device. The ITRS predictions rely on the introduction of advanced technologies like the high-κ gate stacks to meet the stringent leakage constraints. Since we are studying the effects of current scaling trends (rather than projected scaling goals that require the introduction of advanced technologies), we relax the leakage constraints slightly. We set a maximum leakage current of 100 pA/μm at the 90-nm node and allow the leakage to grow by 25% in each generation. We reduce $V_{\rm dd}$ regularly at each generation to control dynamic energy, and we optimize the device for minimum delay under the leakage constraint. Table I shows values for the NFET model parameters generated for the 90–32-nm nodes using the scaling approach described in this section. Throughout this paper, we refer to the results in Table I as the "super-$V_{\rm th}$ scaling strategy."

The intrinsic delay of a device may be quantified as $\tau = C_g V_{\rm dd}/I_{\rm on}$, where C_g is the gate capacitance including gate/drain-source overlap, and $I_{\rm on}$ is the drain–current at $V_{\rm gs} = V_{\rm ds} = V_{\rm dd}$. This metric, which has been shown to correlate well with CMOS gate delay [10], is shown for reference in Table I.

A. Device-Level Implications

The device models from the previous section have been simulated in MEDICI—a 2-D device simulator. We begin with a focus on device characteristics and then look at gate-level characteristics. The current in the sub-$V_{\rm th}$ circuit may be described by the well-known weak-inversion current expression shown in [19, eq. (1)], where m is the subthreshold slope factor, and $C_{\rm dep}$ is the depletion capacitance. Note the exponential dependence on m and $V_{\rm th}$

$$I_{\rm sub} = \frac{W}{L_{\rm eff}} \cdot \mu_{\rm eff} \cdot C_{\rm dep} \cdot \nu_T^2 \cdot e^{\left(\frac{V_{\rm gs}-V_{\rm th}}{m \cdot \nu_T}\right)} \cdot \left(1 - e^{-\frac{V_{\rm ds}}{\nu_T}}\right). \tag{1}$$

The inverse subthreshold slope (S_S), which is an excellent measure of the channel control, may be expressed for short-channel MOSFETs as [19]

$$S_S = 2.3 \cdot \nu_T \cdot m \tag{2a}$$

$$S_S = 2.3 \cdot \nu_T \cdot \left(1 + \frac{3 \cdot T_{\rm ox}}{W_{\rm dep}}\right) \times \left(1 + \frac{11 T_{\rm ox}}{W_{\rm dep}} e^{-\frac{\pi \cdot L_{\rm eff}}{2(W_{\rm dep}+3T_{\rm ox})}}\right) \tag{2b}$$

where $W_{\rm dep} \propto 1/\sqrt{N_{\rm eff}}$ is the depletion width with an effective channel doping $N_{\rm eff}$. The value of S_S, which is theoretically limited to values larger than ~60 mV/dec at $T = 300$ K, should be as small as possible to ensure the steepest sub-$V_{\rm th}$ characteristic. As shown in (2b), the final exponential term forces S_S to increase as $L_{\rm poly}$ (and, consequently, $L_{\rm eff}$) reduces relative to $T_{\rm ox}$ and $W_{\rm dep}$. Fig. 2 shows the simulated S_S for an NFET device at different technology nodes. Between the 90- and 32-nm nodes, S_S degrades by 11%, which corresponds to a 60% reduction in the on- to off-current ratio ($I_{\rm on}/I_{\rm off}$) at $V_{\rm dd} = 250$ mV. $I_{\rm on}$ is measured at $V_{\rm gs} = V_{\rm ds} = V_{\rm dd}$. Note in Table I that all devices have $V_{\rm th} > 400$ mV; therefore, $V_{\rm dd} = 250$ mV is well within the sub-$V_{\rm th}$ regime. We will show later in this section that the dramatic reduction in $I_{\rm on}/I_{\rm off}$ leads to serious problems for noise margins and energy efficiency.

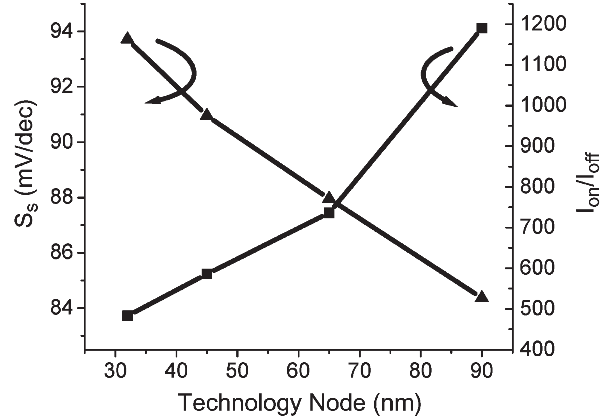


Fig. 2. NFET inverse sub-$V_{\rm th}$ slope and on- to off-current ratio.

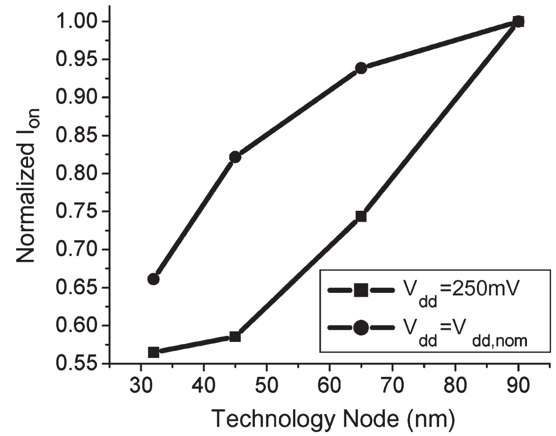


Fig. 3. NFET on current.

Fig. 3 shows the behavior of $I_{\rm on}$ at both nominal $V_{\rm dd}$ (with values taken from Table I) and $V_{\rm dd} = 250$ mV. Under

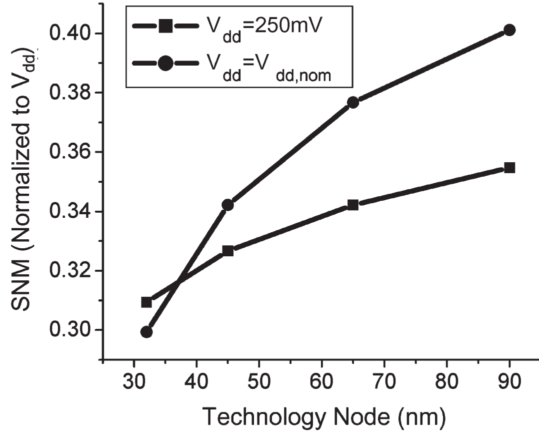


Fig. 4. Simulated SNM for a scaled inverter.

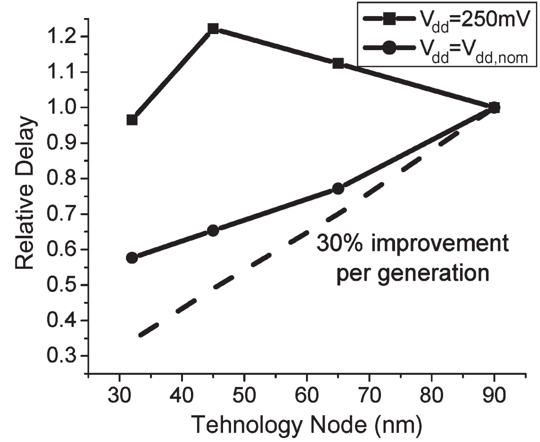


Fig. 5. Simulated delay for a scaled inverter.

our leakage-constrained-scaling scenario, $I_{\rm on}$ reduces between technology generations in the super-$V_{\rm th}$ region. Note that our choice of leakage constraint (100 pA plus 25% per generation) affects this outcome. A more aggressive technology, particularly one leveraging strain in the channel, would likely achieve an increased drain–current with scaling. However, in this paper, we are concerned with low-power devices. Note that the reduction in current is more dramatic for the device measured in the sub-$V_{\rm th}$ region. This loss of drain–current has important delay implications that will be discussed later in this section.

B. Static Noise Margins

Consider the SNMs of a CMOS inverter. The voltage transfer characteristic of a sub-$V_{\rm th}$ inverter is computed by equating the drain–current (1) through the NFET and PFET devices, as shown in (3a) at the bottom of the page. $I_{o,N}$ and $I_{o,P}$ are the NFET and PFET currents at $V_{\rm gs} = V_{\rm th}$ with $V_{\rm ds} \gg \nu_T$. $V_{\rm in}$ and $V_{\rm out}$ are the voltages at the input and the output of the inverter. We can relate $V_{\rm in}$ and $V_{\rm out}$ using (3b), as shown at the bottom of the page. We can further simplify the expression by assuming that $I_{o,N} = I_{o,P}$, $V_{{\rm th},N} = V_{{\rm th},P} = V_{\rm th}$, and $m_N = m_P = m$, as shown in (3c) at the bottom of the page.

The important role of S_S (through m) in determining the voltage transfer characteristic (and, consequently, the SNM) is obvious, particularly in (3c). Fig. 4 shows the evolution of the SNM for a CMOS inverter simulated at nominal $V_{\rm dd}$ (Table I) and $V_{\rm dd} = 250$ mV. We define the SNM at the points where the gain in the voltage transfer characteristic is equal to negative one. The increase in S_S with scaling results in SNM degradation of more than 10% between the 90- and 32-nm nodes. This is a serious concern for sub-$V_{\rm th}$ designers since absolute noise margins are already dramatically reduced compared with high-voltage operation. It is particularly concerning for SRAM, where the noise margins are paramount and a small $I_{\rm on}/I_{\rm off}$ in sub-$V_{\rm th}$ circuits already places tight limits on the maximum number of bits/line [16].

C. Delay

The delay of a CMOS gate may be expressed as

$$t_p = \frac{k_d \cdot C_L \cdot V_{\rm dd}}{I_{\rm on}} \tag{4}$$

where C_L is the load capacitance, and k_d is a fitting parameter. The sub-$V_{\rm th}$ delay may be found by substituting (1) into (4)

$$t_p = \frac{k_d \cdot C_L \cdot V_{\rm dd}}{I_{\rm on}} = \frac{k_d \cdot C_L \cdot V_{\rm dd}}{I_{o,N} \cdot e^{\frac{V_{\rm dd} - V_{\rm th}}{m \cdot \nu_T}}}. \tag{5}$$

The $V_{\rm ds}$ dependence of $I_{\rm on}$ [shown in (1)] has been ignored since it is negligible for $V_{\rm gs} = V_{\rm dd} \gg \nu_T$. The delay expression is clearly dominated by an exponential dependence on $V_{\rm dd}$, $V_{\rm th}$, and m.

The simulated delay of a CMOS inverter with FO1 loading is shown in Fig. 5 at nominal $V_{\rm dd}$ (Table I) and at 250 mV. As

$$I_{o,N} \cdot e^{\frac{V_{\rm in} - V_{\rm th}}{m \cdot \nu_T}} \left(1 - e^{-\frac{V_{\rm out}}{\nu_T}}\right) = I_{o,P} \cdot e^{\frac{V_{\rm dd} - V_{\rm in} - V_{\rm th}}{m \cdot \nu_T}} \left(1 - e^{-\frac{V_{\rm dd} - V_{\rm out}}{\nu_T}}\right) \tag{3a}$$

$$V_{\rm in} = \frac{m_n(V_{\rm dd} - V_{{\rm th},p}) + m_p V_{{\rm th},n} + m_n m_p \nu_T \ln\left(\frac{I_{o,P}}{I_{o,N}} \cdot \frac{1 - e^{-\frac{V_{\rm dd} - V_{\rm out}}{\nu_T}}}{1 - e^{-\frac{V_{\rm out}}{\nu_T}}}\right)}{m_n + m_p} \tag{3b}$$

$$V_{\rm in} = \frac{V_{\rm dd}}{2} + \frac{m \cdot \nu_T}{2} \ln\left(\frac{1 - e^{-\frac{V_{\rm dd} - V_{\rm out}}{\nu_T}}}{1 - e^{-\frac{V_{\rm out}}{\nu_T}}}\right) \tag{3c}$$

expected, the delay at nominal V_{dd} improves with L_{poly}, although at a rate that is slower than the target of 30% per generation under generalized scaling (assuming that $1/\alpha = 0.7$). In contrast, the delay increases between the 90- and 45-nm nodes at $V_{\mathrm{dd}} = 250$ mV. Due to the relaxed I_{off} constraint imposed at advanced technology nodes (a 25% increase is allowed per generation), one might expect that the delay would decrease with scaling (since I_{on} increases with I_{off}). However, S_S degrades over the same region, dramatically reducing the I_{on} and increasing the delay. Between the 45- and 32-nm nodes, the increase in I_{off} begins to dominate any degradation in S_S and causes a reduction in delay. A more stringent leakage constraint during the device optimization would yield a monotonic delay increase. The important lesson is that the sub-V_{th} delay is exponentially sensitive to V_{th} and S_S and only linearly sensitive to L_{poly}. Even small changes to a super-V_{th} device to control the leakage and short-channel effects may result in large fluctuations in the sub-V_{th} delay. It is likely that the scaling V_{th} and S_S, not L_{poly} scaling, will control the performance of future sub-V_{th} circuits. Strict attention to V_{th} selection and S_S control will be an important part of any technology optimized for sub-V_{th} use.

In sub-V_{th} applications, V_{dd} is typically set at the energy optimal value $V_{\min}$; therefore, the scaling of delay at $V_{\mathrm{dd}} = V_{\min}$ is of interest. The value of $V_{\min}$ was found in [17] and [18] to be proportional to S_S. If we ignore the dependence of $V_{\min}$ on the slope of the input waveform, then we can set $V_{\mathrm{dd}} = V_{\min} = K_{V_{\min}} \cdot S_S$, where $K_{V_{\min}}$ is a parameter that depends only on the structure of the circuit (and not on the scaling parameters) [17]. By using this new relation and by recognizing that $S_S = V_{\mathrm{dd}}/\log(I_{\mathrm{on}}/I_{\mathrm{off}})$, we can express (4) and (5) in terms of only the scaling-dependent parameters (6). The simple expression in (6) suggests that we can predict the scaling behavior of the sub-V_{th} delay simply by understanding the scaling of C_L, S_S, and I_{off}. We develop a similar expression for energy in the next section.

$$t_p = \frac{k_d \cdot C_L \cdot K_{V_{\min}} \cdot S_S}{I_{\mathrm{off}} \cdot 10^{\frac{K_{V_{\min}} S_S}{S_S}}} \propto \frac{C_L \cdot S_S}{I_{\mathrm{off}}}. \tag{6}$$

D. Energy

The energy of a single inverter driving an identical inverter can nominally be separated into two components: dynamic (E_{dyn}) and leakage (E_{leak})

$$E_{\mathrm{dyn}} = C_L \cdot V_{\mathrm{dd}}^2 \cdot \alpha \tag{7a}$$

$$\begin{aligned} E_{\mathrm{leak}} &= I_{\mathrm{off}} \cdot V_{\mathrm{dd}} \cdot t_p \\ &= I_{\mathrm{off}} \cdot V_{\mathrm{dd}} \cdot \frac{k_d \cdot C_L \cdot V_{\mathrm{dd}}}{I_{\mathrm{on}}} \\ &= C_L \cdot V_{\mathrm{dd}}^2 \cdot k_d \cdot \frac{I_{\mathrm{off}}}{I_{\mathrm{on}}}. \end{aligned} \tag{7b}$$

The term α is the activity factor, and all other terms are previously defined. If we again assume that the operation only occurs at the energy optimal $V_{\mathrm{dd}} = V_{\min}$, then we can simplify (7a) and (7b) as follows:

$$E_{\mathrm{dyn}} = C_L \cdot (K_{V_{\min}} \cdot S_S)^2 \cdot \alpha \propto C_L \cdot S_S^2 \tag{8a}$$

$$E_{\mathrm{leak}} = C_L \cdot (K_{V_{\min}} \cdot S_S)^2 \cdot k_d \cdot 10^{-K_{V_{\min}}} \propto C_L \cdot S_S^2. \tag{8b}$$

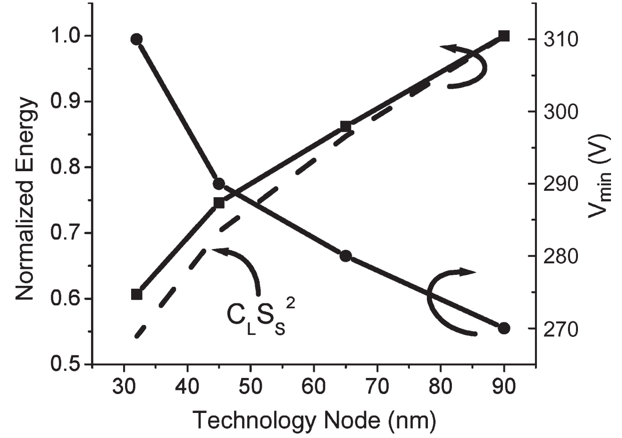


Fig. 6. Simulated energy/cycle and $V_{\min}$ for a chain of 30 inverters with $\alpha = 0.1$.

The only parameters that change as a result of the device scaling are C_L and S_S. Equation (8) suggests the interesting result that the dynamic and leakage energies in sub-V_{th} circuits have an identical dependence on the scaling parameters and that the ratio $E_{\mathrm{dyn}}/E_{\mathrm{leak}}$ is insensitive to scaling when operating at $V_{\mathrm{dd}} = V_{\min}$.

The simulated energy consumed per cycle by a chain of 30 inverters with $\alpha = 0.1$ and $V_{\mathrm{dd}} = V_{\min}$ is shown in Fig. 6. There is a substantial energy reduction as the devices are scaled from the 90- to 32-nm nodes. However, note that $V_{\min}$ increases by 40 mV for this simple circuit between the 90- and 32-nm nodes. Recall that $V_{\min}$ is proportional to S_S; thus, this trend is not surprising. It was shown in [6] that an increase in $V_{\min}$ is generally not beneficial for energy efficiency. An increase in $V_{\min}$ essentially equates to a dynamic-energy $(C_L V_{\mathrm{dd}}^2)$ penalty. Ideally, a scaled sub-V_{th} device should experience a reduction in capacitance while maintaining $V_{\min}$. The factor $C_L \cdot S_S^2$, which is also shown in Fig. 6, matches very closely to the energy measurements, thus confirming the validity of (8).

III. Sub-V_{th} Scaling Trends

It became clear in the previous section that the degradation of S_S with device scaling will be problematic for robust energy efficient sub-V_{th} operation. Moreover, the scaling of L_{poly} to improve the delay characteristics of super-V_{th} devices is not relevant in sub-V_{th} circuits since the delay is largely controlled by V_{th} and S_S. Ideally, we would like a sub-V_{th} transistor with a very small S_S to address the noise-margin and energy concerns. This device should be available in multiple well-controlled thresholds in order to provide a wide range of performance points. In this section, we describe such a device and develop a scaling strategy for this device.

A. Sub-V_{th} Device Optimization

The degradation of S_S with scaling is driven by two related factors. The first factor has already been made clear: The ratio $L_{\mathrm{eff}}/T_{\mathrm{ox}}$ reduces with each technology generation due to the slow scaling of T_{ox}, and it worsens the V_{th} rolloff problem. This suggests that longer channel lengths should be used to

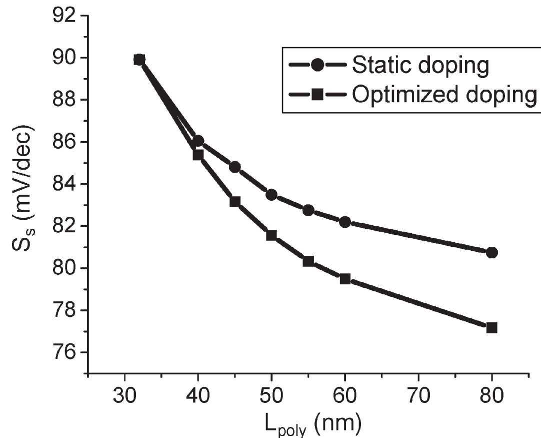


Fig. 7. S_S as a function of the gate length for a 45-nm device.

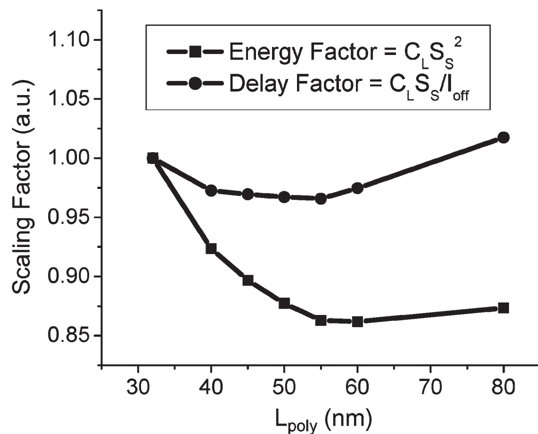


Fig. 8. Energy and delay factors for a 45-nm device.

accommodate the gate oxide. The second factor causing the S_S degradation, which was also covered in [3], is more subtle. To compensate for the $V_{\rm th}$ rolloff problem, the channel doping is effectively increased through aggressive use of the halo doping. For long-channel devices, the halo doping is less critical and actually degrades the S_S. Therefore, to fully optimize S_S with device scaling, it is not sufficient to simply lengthen the $L_{\rm poly}$ without considering the doping. Instead, the $L_{\rm poly}$ and the doping must be optimized simultaneously. This notion is confirmed in Fig. 7, which shows S_S for a 45-nm device with a fixed doping profile and for a 45-nm device with a doping profile optimized for each value of $L_{\rm poly}$.

Increasing the $L_{\rm poly}$ and reducing the doping improve the S_S at the cost of an increased gate capacitance. The cost of this optimization can be quantified in terms of energy and delay. Equation (6) shows us that the sub-$V_{\rm th}$ delay is proportional to $C_L \cdot S_S/I_{\rm off}$ at $V_{\rm dd} = V_{\rm min}$. Similarly, (8a) and (8b) show that the energy in a sub-$V_{\rm th}$ circuit is proportional to $C_L \cdot S_S^2$. These expressions are useful since they are simple functions of device parameters and offer a quick estimation of energy and delay in a prospective technology. Fig. 8 shows these energy and delay factors as functions of $L_{\rm poly}$ for the optimized 45-nm device, which is originally shown in Fig. 7. Both reach a minimum, suggesting that there is both a delay optimal and an energy optimal $L_{\rm poly}$. However, since the delay minimum is very shallow, we can select the energy minimal $L_{\rm poly}$ (60 nm in Fig. 8) for a negligible penalty. Note that the delay typically degrades as $\sim 1/L_{\rm poly}$, but we are able to avoid this problem by also optimizing the doping.

TABLE II
NFET PARAMETERS UNDER SUB-$V_{\rm th}$ SCALING

Node	90nm	65nm	45nm	32nm
L_{poly} (nm)	95	75	60	45
T_{ox} (nm)	2.10	1.89	1.70	1.53
N_{sub} (cm^{-3})	1.61e18	1.99e18	2.53e18	3.19e18
N_{halo} (cm^{-3})	2.02e18	2.73e18	2.93e18	4.89e18
$N_{ch,avg}$ (cm^{-3})	2.01e18	2.45e18	2.93e18	3.55e18
$C_L \cdot S_S^2$ (a.u.)	1	0.80	0.65	0.51
$C_L \cdot S_S$ (a.u.)	1	0.80	0.65	0.50

B. Sub-$V_{\rm th}$ Scaling Model

Given the important role that the S_S plays in determining energy efficiency, performance, and noise margins, we propose a scaling strategy that reduces the S_S by targeting the energy optimal $L_{\rm poly}$ at each technology node. The proposed strategy uses longer channel lengths that scale more slowly than the rate of 30% assumed in Section II. As we will see, one consequence of this strategy is that the S_S remains approximately constant with device scaling. For this paper, we maintain a constant $I_{\rm off}$ of 100 pA/μm across all device generations. Fixing $I_{\rm off}$ yields a more predictable delay-scaling characteristic and avoids the problems shown in Fig. 5. Just as in the super-$V_{\rm th}$ technologies, different performance levels can be targeted by offering multiple thresholds.

We begin with a 90-nm device identical to the 90-nm device in Section II, but the $L_{\rm poly}$ and the doping have been optimized for minimum energy using (8a) and (8b). We again assume that $T_{\rm ox}$ reduces by 10% and all other physical dimensions, excluding $L_{\rm poly}$, reduce by 30% in each generation. We find the optimal $L_{\rm poly}$, $N_{\rm sub}$, and $N_{p,\rm halo}$ at each generation, as described in Section III-A. The resulting NFET device parameters are listed in Table II. The delay (6) and energy (8) factors are also listed in Table II. Note that the delay factor simplifies to $C_L \cdot S_S$ since $I_{\rm off}$ is constant with scaling. A similar set of values is derived for the PFET devices. We find that the energy optimal $L_{\rm poly}$ for the PFET device is almost identical to that of the NFET; therefore, we use the $L_{\rm poly}$ values in Table II during the PFET doping optimization. For the remainder of this paper, we refer to the results in Table II as the "sub-$V_{\rm th}$ scaling strategy."

C. Device- and Circuit-Level Implications

The primary purpose of our revised scaling strategy is to maintain strong channel control, even at very small dimensions. Fig. 9 shows how $L_{\rm poly}$ and S_S scale under our proposed scaling strategy and under the original super-$V_{\rm th}$ scaling strategy. $L_{\rm poly}$ is larger than in the super-$V_{\rm th}$ scaling scheme and also scales at a slower rate (20%–25% per generation) than the $L_{\rm poly}$ in the super-$V_{\rm th}$ scaling scheme (30%). Note that S_S stays very

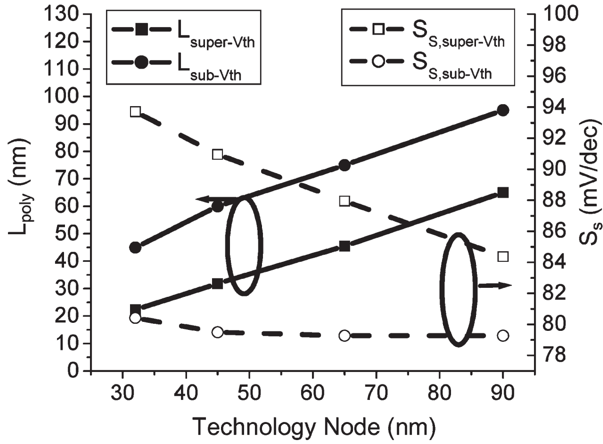


Fig. 9. NFET $L_{\rm poly}$ and S_S for the sub-$V_{\rm th}$ and super-$V_{\rm th}$ scaling strategies.

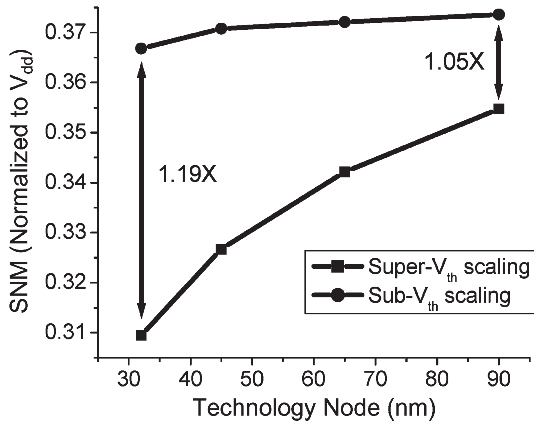


Fig. 10. Simulated SNM for an inverter under the super-$V_{\rm th}$ and sub-$V_{\rm th}$ scaling strategies.

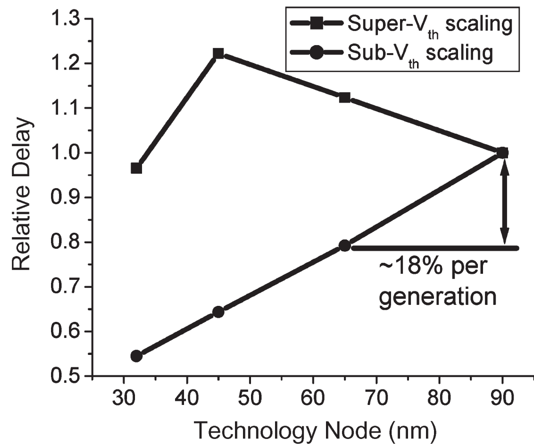


Fig. 11. Simulated delay for an inverter at $V_{\rm dd} = 250$ mV under the super-$V_{\rm th}$ and sub-$V_{\rm th}$ scaling strategies.

close to ∼80 mV/dec under our proposed strategy, varying by only 1.2 mV/dec between the 90- and 32-nm nodes. As a result, the SNM remains nearly constant as well (Fig. 10). At the 32-nm node, the optimized sub-$V_{\rm th}$ scaling strategy yields an SNM that is 19% larger than that observed under the super-$V_{\rm th}$ scaling strategy.

Normalized FO1 inverter delay is shown in Fig. 11 for both scaling scenarios. Delay reduces by ∼18% per generation

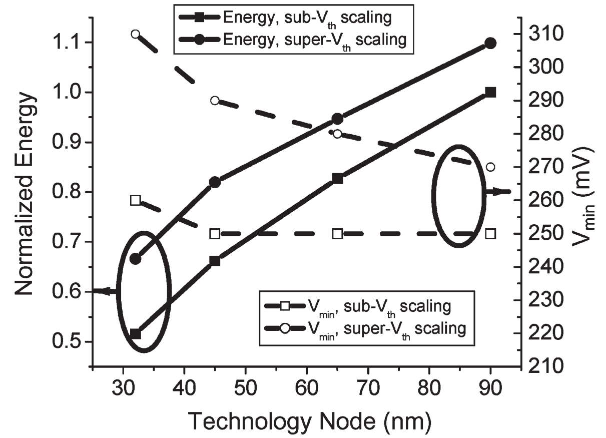


Fig. 12. Simulated energy and $V_{\rm min}$ under the super-$V_{\rm th}$ and sub-$V_{\rm th}$ scaling strategies.

under our proposed strategy. Recall from Section II-C that the delay characteristic for the super-$V_{\rm th}$ scaling strategy is not monotonic due to the scaling of $V_{\rm th}$ and $I_{\rm off}$. It is therefore not fair to directly compare the delay scaling of the two strategies. However, it is clear that the sub-$V_{\rm th}$ scaling strategy exerts much tighter control over $I_{\rm off}$ and S_S than the super-$V_{\rm th}$ strategy; therefore, the delay characteristic scales much more gracefully.

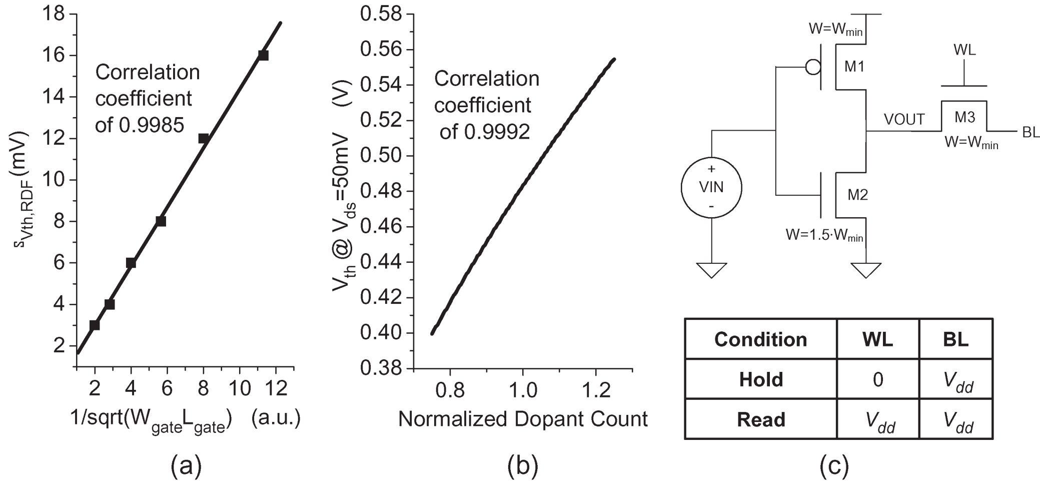
Fig. 12 shows the simulated energy and the $V_{\rm min}$ for a chain of 30 inverters under the conventional super-$V_{\rm th}$ scaling scheme and our proposed scheme. The proposed strategy consumes ∼23% less energy than the super-$V_{\rm th}$ scaling strategy at the 32-nm node (measured at $V_{\rm min}$), with $V_{\rm min}$ changing by only 10 mV from 90-nm node to the 32-nm node. The relatively low $V_{\rm min}$ (which previous work has shown to be a strong function of S_S and leakage energy [17], [18]) is responsible for this energy reduction.

IV. Stability of Scaled Sub-$V_{\rm th}$ SRAM

Robust memory design is the most challenging task facing low-voltage designers. Recent work has demonstrated dramatic improvements in low-voltage operation [16], [21]–[24], but concerns about density and robustness in the face of variability still remain. In this section, we focus on memory design deep in the sub-$V_{\rm th}$ regime (in this case, $V_{\rm dd} = 250$ mV). Such designs typically target sensor-mote applications where memory sizes on the order of several kilobits are sufficient. Although most of the low-voltage memory work has proposed new SRAM cell architectures to improve read/write margins [21]–[24], we will use the traditional 6-transistor (6T) SRAM as a test vehicle since it is the most widely adopted SRAM variant. Before looking at the scalability of the sub-$V_{\rm th}$ SRAM, we first describe a simple variability model that will be used to supplement our observations.

A. Variability in Sub-$V_{\rm th}$ SRAM

Process-induced $V_{\rm th}$ variation makes sub-$V_{\rm th}$ SRAM design extremely challenging. Due to the exponential dependence of subthreshold current (1) on $V_{\rm th}$, even small variations lead



Condition	WL	BL
Hold	0	V_{dd}
Read	V_{dd}	V_{dd}

Fig. 13. (a) RDF $V_{\rm th}$ variability model in a 65-nm device closely matches the gate-area dependence in (9). (b) $V_{\rm th}$ is approximately linear with the dopant count. (c) SRAM test circuit for measuring the SNM and the $I_{\rm read}/I_{\rm leak}$. Node voltages during the hold and read conditions are also shown.

to large strength mismatches between the pull-up [M1 in Fig. 13(c)], pull-down (M2), and pass transistors (M3). Random $V_{\rm th}$ variability due to random dopant fluctuations (RDFs) is largely responsible for within-cell mismatch. Past work has shown that the RDF-induced variation is a formidable problem in super-$V_{\rm th}$ circuits [30], and it has also been shown that the importance of RDF grows relative to other random $L_{\rm poly}$ variation as $V_{\rm dd}$ reduces [31]. The $V_{\rm th}$ variations due to RDF may be modeled as [25]

$$\sigma_{V_{\rm th,RDF}} = 3.19e-8 \cdot \frac{T_{\rm ox}N_A^{0.4}}{\sqrt{L \cdot W}}. \quad (9)$$

The inverse relationship to the square root of gate area suggests that random variation will worsen with device scaling. To model RDF in MEDICI, we first calculate the expected number of dopants in the channel by integrating the continuous doping profile in the box with corners at the source/drain top and bottom edges and multiplying by the width of the device. We then use a Poisson distribution to create perturbations in the dopant count [25]. To map the discrete dopant count back to a continuous distribution, we scale the entire doping profile by a constant. The model is shown in Fig. 13(a) to agree well with the gate-area dependence highlighted in (9).

Rather than running computationally intense Monte Carlo simulations, we skew the SRAM cell to a worst case corner. The read becomes unstable when the pull-down device becomes weak (i.e., M2 has a high $V_{\rm th}$), and the pass transistor becomes strong (i.e., M3 has a low $V_{\rm th}$). We skew each transistor to equally probable corners to achieve a desired joint probability. For example, we may skew the pull-down transistor to a point that is slower than 99.87% of all transistors (i.e., 3σ away from the mean μ on a normal distribution). At the same time, we would skew the pass transistor to a point that is faster than 99.87% of all transistors. If we observe failure at this point, we conclude that failure may occur any time that the pull-down transistor is slower than the $\mu - 3\sigma$ point or the pass transistor is faster than the $\mu + 3\sigma$ point, giving a failure probability of 0.13% (3σ away from the mean on a normal distribution). Note that this is a conservative approach since we make the approximation that failure occurs any time either one of the devices is $> 3\sigma$ away from the mean, whereas the observed failure occurred when both devices were skewed to the 3σ corners.

We simplify our approach by making the following approximations: 1) that the dopant-count distribution is approximately normal; and 2) that the dopant count maps linearly to a $V_{\rm th}$. The first approximation allows us to use σ, 2σ, 3σ, etc., as meaningful metrics, and the second approximation [which is shown to be reasonable in Fig. 13(b)] allows us to say that the σ, 2σ, and 3σ points in the dopant distribution map directly to the σ, 2σ, and 3σ points, respectively, in the $V_{\rm th}$ distribution. We will use this model in the next section to approximately bound the characteristics of sub-$V_{\rm th}$ SRAM. A more accurate RDF model could be obtained by dividing the channel into cells with unique Poisson distributions [26] or by placing each dopant atom individually within the channel [27]. However, our simple model is useful since we only seek to identify the general characteristics that are favorable in scaled sub-$V_{\rm th}$ devices rather than to predict accurate estimates of variation in future sub-$V_{\rm th}$ SRAM.

B. Noise Margins in Sub-$V_{\rm th}$ SRAM

The nominal hold and read SNMs of a 6T SRAM at $V_{\rm dd} =$ 250 mV are shown in Fig. 14. For the unoptimized super-$V_{\rm th}$ device (called "super-$V_{\rm th}$ short $L_{\rm poly}$"), the nominal read SNM goes as low as 10% of $V_{\rm dd}$ at the 32-nm node. Note that the read SNM at the 32-nm node is 64% larger for the sub-$V_{\rm th}$ device (called "sub-$V_{\rm th}$") than for the super-$V_{\rm th}$ device. Recall from (3) that the noise margins have a strong dependence on S_S, which is dramatically improved in the device optimized for the sub-$V_{\rm th}$ operation. It is interesting to note that the discrepancy in SNM can be nearly eliminated by increasing the lengths in the super-$V_{\rm th}$ devices to match those of the sub-$V_{\rm th}$ devices (called "super-$V_{\rm th}$ long $L_{\rm poly}$" in Fig. 14). However, the read/write delays will not be optimal since doping has not been reoptimized for the larger gate length.

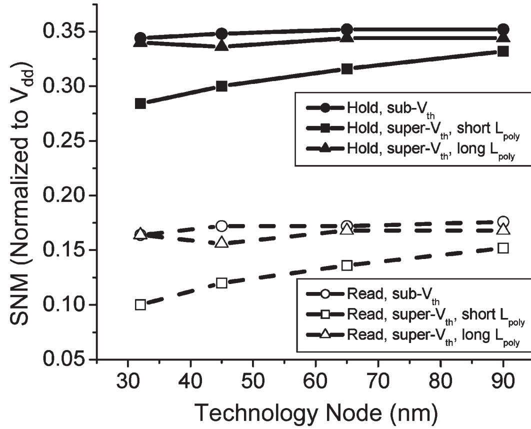


Fig. 14. Simulated SNM in a 6T SRAM cell at $V_{\rm dd} = 250$ mV under three different device-optimization strategies: 1) the sub-$V_{\rm th}$ optimized device; 2) the unoptimized super-$V_{\rm th}$ device with minimum length; and 3) the unoptimized super-$V_{\rm th}$ device with the same length as in case (1).

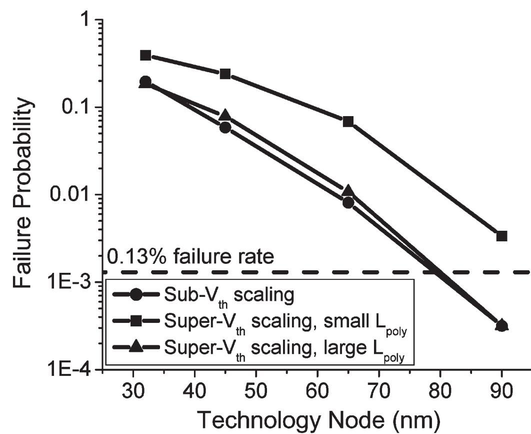


Fig. 15. Read failure probability for a single SRAM cell under different device-optimization strategies at $V_{\rm dd} = 250$ mV. Failure is defined as the point where the read SNM drops below 6% of $V_{\rm dd}$ (15 mV).

Fig. 15 shows the probability of a read failure for a single cell at each technology node for each device type. A failing case has a read SNM below 6% of $V_{\rm dd}$ (15 mV for $V_{\rm dd} = 250$ mV). The probability of the failing case is calculated based on the individual probabilities of the pull-down and pass-transistor corners, as shown in the previous section. Fig. 15 also includes a 3σ failure probability line ($\sim$0.13% failure probability) for reference. We focus on the 3σ point since 99.87% yield should be sufficient for small SRAM arrays in sensor applications with several kilobits and column/row redundancy. The unoptimized device with short $L_{\rm poly}$ exceeds the 3σ failure probability at 90 nm. The optimized and unoptimized devices with long $L_{\rm poly}$ exceed the 3σ failure probability at 65 nm, suggesting that simple device optimizations may extend the lifetime of the sub-$V_{\rm th}$ 6T SRAM by one technology generation. It is interesting to again note that increasing gate length is responsible for most of the improvement in noise margins and could be used as a near-term fix for subthreshold memories using conventional devices.

It is also interesting to consider the implications of the $I_{\rm on}/I_{\rm off}$ reductions shown in Fig. 2. At 32 nm, this ratio may drop below 500 for a device operating at 250 mV. This is particularly challenging when reading data out of an SRAM cell since it becomes very difficult to distinguish between read current and bitline leakage. Fig. 16 shows the ratio of the read-current to pass-gate leakage current in a 6T SRAM cell at 250 mV. This ratio is extremely important since it is proportional to the number of bits allowed on a single bitline. At 90 and 32 nm, the sub-$V_{\rm th}$ device has a 1.7× and 3.7× larger current ratio, respectively, than the super-$V_{\rm th}$ device. These numbers can be reduced to 1.17× and 1.28× by increasing the lengths of the super-$V_{\rm th}$ devices to match those of the sub-$V_{\rm th}$ devices at the cost of increased read/write delay.

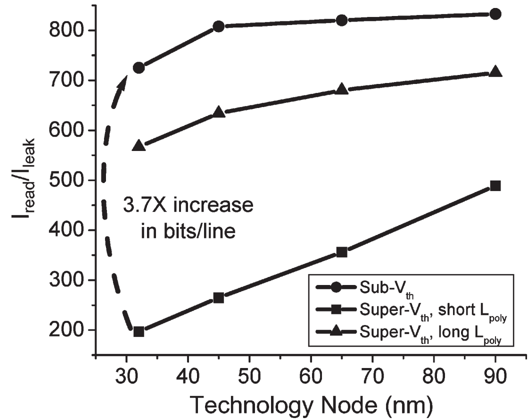


Fig. 16. Ratio of read-current to pass-transistor leakage in a 6T SRAM at $V_{\rm dd} = 250$ mV under the super-$V_{\rm th}$ and sub-$V_{\rm th}$ scalings. $I_{\rm read}/I_{\rm leak}$ is proportional to the maximum number of bits per bitline and is therefore closely tied to the SRAM area.

The data in this section suggest that the future is quite grim for scaled sub-$V_{\rm th}$ memories. Significant help can be offered at advanced technology nodes by the gate length and doping optimizations studied in this paper. Radical device redesign may be required for large SRAM arrays. High-κ gate dielectrics will help improve the channel control (and, subsequently, the noise margins, the delay, and the energy). Multigate devices with lightly doped bodies will offer a superior channel control and will eliminate RDF [6]. In the near term, however, designers need to focus on variability-aware circuit-design techniques in combination with simple device modifications.

V. Conclusion

Sub-$V_{\rm th}$ circuits are promising for future energy efficient applications. In this paper, we investigated the implications of device scaling on sub-$V_{\rm th}$ operation. In particular, we found that the slow scaling of gate oxide leads to 60% $I_{\rm on}/I_{\rm off}$ degradation in the sub-$V_{\rm th}$ regime. We used MEDICI simulations of simple circuits to illustrate the energy, the performance, and the robustness characteristics of the scaled sub-$V_{\rm th}$ devices. We proposed an alternative scaling strategy that uses larger gate lengths and reduced doping to achieve a more improved inverse subthreshold slope. Our proposed strategy maintains an $S_S \sim 80$ mV/dec down to the 32-nm node and offers a robust energy efficient alternative to conventional devices. Our study of scaled sub-$V_{\rm th}$ 6T SRAM suggested that read noise margins will be dangerously small due to the variability at

the 90-nm node, but simple device modifications can push the problem out to the 65-nm node. It is likely that new device geometries will be important for the aggressive scaling of sub-V_{th} circuits, but the simple modifications described in this paper may help sub-V_{th} circuits reliably scale in the near term.

REFERENCES

[1] A. Wang and A. Chandrakasan, "A 180 mV FFT processor using subthreshold circuit techniques," in *Proc. Int. Solid-State Circuits Conf.*, 2004, pp. 292–293.
[2] B. Zhai *et al.*, "A 2.60 pJ/Inst subthreshold sensor processor for optimal energy efficiency," in *Proc. Symp. VLSI Circuits*, 2006, pp. 154–155.
[3] B. C. Paul, A. Raychowdhury, and K. Roy, "Device optimization for digital subthreshold logic operation," *IEEE Trans. Electron Devices*, vol. 52, no. 2, pp. 237–247, Feb. 2005.
[4] J. J. Kim and K. Roy, "Double gate-MOSFET subthreshold circuit for ultralow power applications," *IEEE Trans. Electron Devices*, vol. 51, no. 9, pp. 1468–1474, Sep. 2004.
[5] B. Paul, A. Bansal, and K. Roy, "Underlap DGMOS for digital-subthreshold operation," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 910–913, Apr. 2006.
[6] S. Hanson *et al.*, "Ultralow-voltage, minimum-energy CMOS," *IBM J. Res. Develop.*, vol. 50, no. 4/5, pp. 469–490, Jul. 2006.
[7] R. H. Dennard *et al.*, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. SSC-87, no. 4, pp. 256–268, Apr. 1974.
[8] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a 1/4 micrometer MOSFET design," *IEEE Trans. Electron Devices*, vol. ED-31, no. 4, pp. 452–461, Apr. 1984.
[9] W. Haensch *et al.*, "Silicon CMOS devices beyond scaling," *IBM J. Res. Develop.*, vol. 50, no. 4/5, pp. 339–361, Jul. 2006.
[10] E. J. Nowak, "Maintaining the benefits of CMOS scaling when scaling bogs down," *IBM J. Res. Develop.*, vol. 46, no. 2/3, pp. 169–180, Mar.–May 2002.
[11] B. Yu *et al.*, "Short-channel effect improved by lateral channel-engineering in deep-submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 44, no. 4, pp. 627–634, Apr. 1997.
[12] Z. K. Lee, M. B. McIlrath, and D. A. Antoniadis, "Two-dimensional doping profile characterization of MOSFETs by inverse modeling using I–V characteristics in the subthreshold region," *IEEE Trans. Electron Devices*, vol. 46, no. 8, pp. 1640–1649, Aug. 1999.
[13] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm design exploration," in *Proc. Int. Symp. Quality Electron. Des.*, 2006, pp. 585–590.
[14] Z. Luo *et al.*, "High performance and low power transistors integrated in 65 nm bulk CMOS technology," in *IEDM Tech. Dig.*, 2004, pp. 661–664.
[15] *The Int. Technology Roadmap for Semiconductors*, 2005.
[16] B. Zhai, D. Blaauw, D. Sylvester, and S. Hanson, "A sub-200 mV 6T SRAM in 130 nm CMOS," in *Proc. Int. Solid-State Circuits Conf.*, 2007, pp. 332–333.
[17] B. Zhai *et al.*, "The limit of dynamic voltage scaling and insomniac dynamic voltage scaling," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 13, no. 11, pp. 1239–1252, Nov. 2005.
[18] B. Calhoun and A. Chandrakasan, "Characterizing and modeling minimum energy operation for subthreshold circuits," in *Proc. Int. Symp. Low Power Electron. Des.*, 2004, pp. 90–95.
[19] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge, U.K.: Cambridge University Press, 1998.
[20] E. P. Gusev *et al.*, "Advanced high-κ dielectric stacks with PolySi and metal gates: Recent progress and current challenges," *IBM J. Res. Develop.*, vol. 50, no. 4/5, pp. 387–410, 2006.
[21] B. Calhoun and A. Chandrakasan, "A 256 kb sub-threshold SRAM in 65 nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2006, pp. 2592–2601.
[22] N. Verma and A. Chandrakasan, "A 65 nm 8T sub-Vt SRAM employing sense-amplifier redundancy," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2007, pp. 64–65.
[23] T. Kim, J. Liu, J. Keane, 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.*, 2007, pp. 330–606.
[24] L. Chang *et al.*, "Stable SRAM cell design for the 32 nm node and beyond," in *VLSI Symp. Tech. Dig.*, 2005, pp. 128–129.
[25] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1837–1852, Sep. 2003.
[26] P. Stolk, F. Widdershoven, and D. B. M. Klaassen, "Modeling statistical dopant fluctuations in MOS transistors," *IEEE Trans. Electron Devices*, vol. 45, no. 9, pp. 1960–1971, Sep. 1998.
[27] D. J. Frank, Y. Taur, M. Ieong, and H.-S. P. Wong, "Monte Carlo modeling of threshold variation due to dopant fluctuations," in *VLSI Symp. Tech. Dig.*, 1999, pp. 169–170.
[28] S. Hanson, M. Seok, D. Sylvester, and D. Blaauw, "Nanometer device scaling in subthreshold circuits," in *Proc. Des. Autom. Conf.*, 2007, pp. 700–705.
[29] N. Azizi, M. M. Khellah, V. De, and F. N. Najm, "Variations-aware low-power design with voltage scaling," in *Proc. Des. Autom. Conf.*, 2005, pp. 529–534.
[30] A. J. Bhavnagarwala, X. Tang, and J. D. 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.
[31] 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.

Scott Hanson (S'05) received the B.S. and M.S. degrees in electrical engineering from the University of Michigan, Ann Arbor, in 2004 and 2006, respectively, where he is currently working toward the Ph.D. degree in electrical engineering.

His research interests include low-voltage circuit design for ultralow-energy applications, variation-tolerant circuit design, and energy-efficient high-performance circuit design.

Mr. Hanson was the recipient of the Semiconductor Research Corporation Fellowship.

Mingoo Seok (S'06) was born in Seoul, Korea. He received the B.S. degree in electrical engineering from Seoul National University, Seoul, in 2005. He is currently working toward the Ph.D. degree in electrical engineering at the University of Michigan, Ann Arbor.

His research interests include low-voltage/low-power circuit design and physical aspects of submicrometer technology.

Mr. Seok was the recipient of the Korea Foundation for Advanced Studies Fellowship.

Dennis Sylvester (S'96–M'97–SM'04) received the B.S. degree in electrical engineering (*summa cum laude*) from the University of Michigan, Ann Arbor, and the Ph.D. degree in electrical engineering from the University of California, Berkeley, Berkeley, in 1999.

He has held research staff positions with the Advanced Technology Group of Synopsys, Mountain View, CA, and with the Hewlett–Packard Laboratories, Palo Alto, CA. He was also a Visiting Associate Professor of electrical and computer engineering with the National University of Singapore, Singapore. He is currently an Associate Professor of electrical engineering and computer science with the University of Michigan. He has published numerous articles along with one book and several book chapters in his field of research, which includes low-power circuit design and design-automation techniques, design for manufacturability, and on-chip interconnect modeling. He also serves as a Consultant and a Technical Advisory Board Member for several electronic-design-automation firms in these areas.

Dr. Sylvester has served on the technical program committee of numerous design-automation and circuit-design conferences, the steering committee of the Association for Computing Machinery (ACM)/IEEE International Symposium on Physical Design, and was the General Chair of the 2005 ACM/IEEE Workshop on Timing Issues in the Synthesis and Specification of Digital Systems. He is currently an Associate Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN and served as an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS. He is a member of the ACM, the American Society of Engineering Education, and the Eta Kappa Nu. He was the recipient of the NSF CAREER Award, the Beatrice Winner Award at the International Solid State Circuits Conference, the IBM Faculty Award, the Semiconductor Research Corporation Inventor Recognition Award, and several best paper awards and nominations. He was the recipient of the ACM Special Interest Group on Design Automation Outstanding New Faculty Award and the University of Michigan Henry Russel Award for distinguished scholarship. His dissertation research was recognized with the David J. Sakrison Memorial Prize as the most outstanding research in the UC Berkeley Electrical Engineering and Computer Science Department.

David Blaauw (M'94) received the B.S. degree in physics and computer science from Duke University, Durham, NC, in 1986, and the Ph.D. degree in computer science from the University of Illinois at Urbana—Champaign, Urbana, in 1991.

Until August 2001, he was with the Motorola, Inc., Austin, TX, where he was the Manager of the High-Performance Design Technology Group. Since August 2001, he has been an Associate Professor with the University of Michigan, Ann Arbor. His work has focused on very large scale integration and computer-aided designs with a particular emphasis on circuit design and optimization for high-performance and low-power applications.

Dr. Blaauw was the Technical Program Chair and the General Chair of the International Symposium on Low Power Electronic and Design and was a member and the Technical Program Cochair of the Executive Committee of the Association for Computing Machinery/IEEE Design Automation Conference. He is currently a member of the International Solid State Circuits Conference Technical Program Committee.