SRAM Read/Write Margin Enhancements Using FinFETs

Andrew Carlson, Member, IEEE, Zheng Guo, Student Member, IEEE, Sriram Balasubramanian, Member, IEEE, Radu Zlatanovici, Member, IEEE, Tsu-Jae King Liu, Fellow, IEEE, and Borivoje Nikolić, Senior Member, IEEE

Abstract—Process-induced variations and sub-threshold leakage in bulk-Si technology limit the scaling of SRAM into sub-32 nm nodes. New device architectures are being considered to improve \( V_T \) control and reduce short channel effects. Among the likely candidates, FinFETs are the most attractive option because of their good scalability and possibilities for further SRAM performance and yield enhancement through independent gating. The enhancements to read/write margins and yield are investigated in detail for two cell designs employing independently gated FinFETs. It is shown that FinFET-based 6-T SRAM cells designed with pass-gate feedback (PGFB) achieve significant improvements in the cell read stability without area penalty. The write-ability of the cell can be improved through the use of pull-up write gating (PUWG) with a separate write word line (WWL). The benefits of these two approaches are complementary and additive, allowing for simultaneous read and write yield enhancements when the PGFB and PUWG designs are used in combination.

Index Terms—FinFET, SRAM, variation, pass-gate feedback, pull-up write gating.

I. INTRODUCTION

SRAM needs to track the scaling of digital logic to maintain the continued scaling of CMOS technology. With scaling of linear dimensions by a factor of 0.7, SRAM cell area needs to scale with a factor of 0.5 with each new technology node. Traditionally, overhead needed for decoding, column circuitry and redundancy represented 30% of the area array, usually expressed as an array efficiency of 70%. SRAM design in deeply scaled technologies faces major challenges in overcoming increasing variability as device dimensions scale down. In particular, random dopant fluctuation is a significant cause of device threshold voltage (\( V_T \)) variation, especially for SRAM devices, because its magnitude is inversely proportional to channel area [1]. Accurate \( V_T \) control is essential for high read stability. Similarly, variability and device leakage affect the writeability of the cell. To maintain both desired writeability and read stability of the SRAM arrays, several radical departures from the conventional design have been considered as follows.

1) Scaling of the traditional six-transistor (6-T) SRAM cell at a slower pace, since a transistor with a larger area is more immune to variations. This is a common approach in 65- and 45-nm technology nodes; while it still might be applicable to small arrays in future, it fundamentally undermines the objective of technology scaling.

2) Use of assist techniques to enhance read and write margins. Examples of these techniques include the use of lower column supply voltages during write, bitline and wordline bias, pulsed bit lines, read-, and write-assist column circuitry [2], [3]. These techniques aim to increase the array robustness with smaller cells, but necessarily lower array efficiency, resulting in larger area.

3) Departure from the conventional 6-T SRAM cell design. By using a 7- or 8-T cell structure the read and write requirements can be decoupled. There is a 20%-30% area penalty as compared to a similarly sized 6-T cell; however this approach may yield smaller cell areas when transistor upsizing is needed to maintain stability of the 6-T cell.

4) The use of alternate device technologies to obtain robust 6-T SRAM. The use of an alternate device structure that enables SRAM scaling at the traditional rate would result in the smallest die sizes, but possibly at the cost of increased process complexity.

This paper focuses on the use of alternative device architectures for SRAM in which \( V_T \) control can be achieved without the use of channel dopants, thereby greatly reducing device susceptibility to random dopant fluctuation. Such architectures include fully depleted silicon-on-insulator (FDSOI), FinFETs (vertical double gate), triple-gate, and gate-all-around devices [4]–[7]. Although each of these device architectures provides improved scalability relative to current bulk-Si or partially depleted SOI technologies, the transition to a new architecture has been continually put off in favor of incremental enhancements such as the use of process-induced mechanical strain or, most recently, high-permittivity gate dielectrics. In part this reflects the enormity of the investment and risk associated with the development of the design infrastructure necessary for a new device architecture. The ability to extend scaling SRAM at the traditional pace may become a sufficient motivating factor for absorbing the increased processing costs for reduced die sizes, however. This transition would become attractive particularly if such a device architecture can be integrated with conventional planar CMOS devices.

In this work, the SRAM yield benefits associated with a new device architecture are analyzed. First, the aforementioned architecture candidates are compared, and an argument is
made for FinFETs on the grounds of their scalability, ease of integration into current processes, and potential for further technological enhancements, such as independent gate control. Two possible, complementary SRAM designs exploiting independent gate control are investigated in detail. It is shown that built-in feedback can be used to achieve dramatic improvements in the cell read margin and offers a more favorable tradeoff with writeability than conventional gate work function tuning. Back-gating of the pull-up (PU) devices with a separate WWL can be used to enhance writeability, allowing for simultaneous read and write margin enhancements and further yield improvements. Tradeoffs in cell read currents and architectural constraints are also discussed. The improvements to SRAM yield which it offers make the FinFET a compelling choice for a future device architecture.

II. 6-T SRAM METRICS AND DESIGN TRADEOFFS

The yield and density of a memory array are its most important properties. High yield is guaranteed for large memory arrays by providing sufficiently large design margins for each operation: reading a cell’s state without disturbing it, holding the cell’s state, writing a new state into a cell, and achieving these within a specified timeframe.

The read static noise margin (RSNM), measured from the voltage transfer characteristics [8], is typically used as a metric for read stability. It is highly sensitive to the relative drive strength of the pull-down (PD) and pass-gate (PG) transistors (the cell beta ratio), and it can be increased by upsizing the PD transistors, which results in an area penalty, and/or by increasing the gate length of the PG transistors, which increases the word line (WL) delay and decreases the writeability of the cell.

During a write operation, PG3 and PU5 form a resistive voltage divider for the falling BL_L and node CH (see Fig. 1). If the voltage divider pulls $V_{CH}$ below the trip point of the inverter formed by PU6 and PD2, a successful write operation occurs. Writeability is measured using “N-curves,” which effectively measure PG3 current minus PU5 current [9]. In the writeability measurement, $V_{CH}$ is swept with BL_L at $V_{DD}$, and the current externally sourced into the CH node is measured. This metric has a zero-crossing (the point of zero noise margin) that corresponds to alternative writeability metrics, such as the maximum bit line voltage allowing a write, or the write noise margin as defined in [10]. Larger writeability current, $I_W$ (i.e., the valley current of the “N-curve”), corresponds to a more writeable cell, while $I_W < 0$ represents a write failure. The writeability can be improved by strengthening the PG device relative to the PU device. This is often achieved by keeping the PU device minimum-sized and upsizing the PG transistor at the cost of cell area and RSNM.

During any read/write access, the word line (WL) voltage is raised only for a limited amount of time specified by the cell access time. If either the read or the write operation cannot be successfully carried out before WL voltage is lowered, access failure occurs. Access time depends upon a number of factors, including the drive strength of the PG transistor and bit line (BL) capacitances. It can be reduced by upsizing the PG transistors, again at the cost of area and RSNM. In this work, the dc read current of the cell is used as a proxy for access time.

The above metrics are often quoted for a nominal cell design, that is, one that does not consider parametric variations. It is also of interest to estimate the yield for these metrics. Yield is determined not only by the nominal cell metric but also by the amount of variation in the metric, which is caused by device parameter variations in the cell. It is useful to compare amounts of variation between different device parameters (such as $V_{TH}$ or gate length, $L_C$) in terms of their respective standard deviations (sigma), and to compute total variation vectorially. “Cell sigma,” the yield figure of merit of the SRAM, is defined as the minimum amount of total variation necessary to cause a failure (RSNM = 0 or $I_W = 0$). A higher cell sigma corresponds to higher yield.

III. DEVICE ARCHITECTURES

Scaling the classical bulk-Si MOSFET structure down into the sub-20 nm $L_C$ regime presents several challenges. Suppression of short channel effects in bulk-Si requires heavy channel doping ($> 10^{19}$ cm$^{-3}$) or heavy super-halo implants to control sub-surface leakage currents. As a result, carrier mobilities are severely degraded due to impurity scattering and a high transverse electric field in the on state. Furthermore, the increased depletion charge density results in a larger depletion capacitance and hence a larger sub-threshold slope. Thus, for a given off-state leakage current specification, on-state drive current is degraded. Off-state leakage current is also enhanced due to band-to-band tunneling between the body and drain.

However, continued SRAM scaling with bulk-Si devices will be ultimately limited by yield considerations. Among the various sources of device parameter variation, random dopant fluctuation is especially detrimental to SRAM yield, especially in RSNM. Sensitivity analyses of SRAM read and write metrics identify the PD and PG transistor threshold voltages as the most significant parameters for variation [11]. RSNM is especially sensitive to $V_{TH}$ mismatch between the two PD transistors [11]–[13]. Table I summarizes the predicted increase in $V_{TH}$ variation due to random dopant fluctuation, following ITRS scaling predictions [14].

In order to address the issue of increasing $V_{TH}$ variability in bulk-Si MOSFETs, adaptive body biasing techniques have been introduced [15]. By segmenting the SRAM array into groups with only a few blocks of cells, the collective nMOS or pMOS $V_{TH}$ values can be adjusted by changing the well potential. The correction coarsely affects all transistors sharing the well, among multiple cells, but it has been shown to reduce frequency variation in logic circuits by a factor of seven.
This technique incurs an area penalty that increases with the resolution of the $V_T$ adjustment. Furthermore, the range of achievable adjustment decreases with the body factor as transistor dimensions scale, limiting the scalability of this form of feedback [16]. In other words, as channel dimensions scale down, $V_T$ control via doping or body biasing decreases as well. Maintaining tight control of $V_T$ will require a device architecture in which $V_T$ is set by parameters with relatively low variability, such as the physical dimensions of the channel and the work function of the gate. FDSOI, FinFET, triple-gate, and gate-all-around devices with light channel doping have been proposed to reduce $V_T$ variability and thereby enable continued CMOS scaling.

In a FDSOI device, the depletion region extends throughout the thickness of the channel layer. Scaled FDSOI designs can eliminate the need for channel dopants, enabling higher carrier mobilities and reducing drain-to-body capacitance, which provide for improved circuit performance with lower dynamic power consumption. Devices with undoped channels have negligible depletion charge and capacitance, and hence a steep subthreshold slope. As a planar, single-gate-material technology, FDSOI can accommodate a wide and continuous range of device widths, enabling optimal ($\beta$) ratios in SRAM designs. Existing bulk-Si designs could be ported to FDSOI technology with the least amount of design effort, relative to other new device technologies. However, the problem with FDSOI is its scalability. Silicon film thicknesses of approximately $T_{Si} < L_G/4$ have been shown to be necessary for good short channel behavior, for gate lengths down to 18 nm [17]. In addition to being expensive to manufacture uniformly, channel thicknesses smaller than five nanometers are expected to suffer from quantum confinement effects [18], making $V_T$ a sensitive function of the channel thickness. These effects will make it difficult to scale FDSOI technology much beyond the 22 nm node.

Double gate devices, such as FinFETs, achieve good short channel behavior with a less stringent body thickness requirement of $T_{Si} < 2L_G/3$. They enjoy the same improvements in carrier mobility and subthreshold slope when an undoped channel is used. The vertical fin of the FinFET can be manufactured with conventional lithography and etching processes. Although low $V_T$ values are difficult to achieve simultaneously for nMOS and pMOS logic devices, a single gate material with mid-gap work function can be used to achieve symmetric and high $V_T$ values for low-leakage applications such as SRAM. In addition, FinFETs have lower parasitic device capacitance because both depletion and junction capacitances are effectively eliminated, which reduces the BL capacitive load.

Further improvements in short channel control can be achieved with triple-gate or gate-all-around devices [6], [7]. The body thickness requirement is relaxed to $T_{Si} < L_G$ for these architectures. Triple-gate devices have relatively poor layout efficiency as compared to double gate devices, however [19]. Gate-all-around devices offer near-ideal channel control [7], but are expensive to manufacture.

Table II lists the silicon thickness constraints and approximate $V_T$ sensitivities to variations in device dimensions for each device architecture, as reported or estimated from simulation studies [11], [17], [20]. Since the short channel control is in large part determined by the geometries of the gates and channel regions, the thickness requirements on $T_{Si}$ are expected to remain valid with $L_G$ scaling. All of the options offer significant improvements in $V_T$ sensitivity over doped-channel devices, but the improvements diminish beyond the FinFET. FinFET-based SRAMs have been demonstrated in silicon to have excellent stability and leakage control, down to $L_G = 20$ nm [21]–[25]. The unique structure of the FinFET enables independent gate operation that is difficult or impossible to achieve with the other device architectures. Independent gate operation is achieved by selectively removing the gate material directly on top of the fin, leaving the gates electrically isolated [26]. In addition to enabling additional connectivity within the SRAM cell, FinFETs provide an alternative direction for future technology development beyond gate length scaling. Several independently gated FinFET SRAM designs have demonstrated improved performance and yield via $V_T$ adjustment [27]–[29], cell-specific feedback [30], [31], or write-assist lines [31]. FinFETs therefore are the most promising device architecture for continued 6-T SRAM scaling, due to robust $V_T$ control and the enhancements achievable with independent gating.

### IV. FinFET-Based SRAM Designs

#### A. Methodology

Mixed-mode device simulation using the drift-diffusion model for carrier transport and the density gradient model to account for quantum-mechanical effects in nanometer-scale MOSFETs is employed to simulate the dc transfer characteristics of SRAM cells under different biasing conditions [32]. Because the high-field transient velocity overshoot effects are ignored, the drain current values may be underestimated. However, the trends and differences between device technologies and their impact on SRAM noise margins should still be valid because they depend on the relative strengths of transistors and not their absolute $I_{ON}$. Actual $I_{ON}$ values may exhibit small deviations from reported values, but the $V_T$ trends and relative relationships of zero-crossings are expected to remain the same. Similarly, in the simulations of access time may deviate from actual values, due to errors in estimating the $I_{ON}$ together with unknown interconnect; however, they are expected to accurately represent relative performance.
It is expected that the effect of parasitic resistances and capacitances will limit circuit performance in deeply scaled CMOS technologies. Series resistance and extrinsic contact resistance are included in this work, which lessens the improvements associated with the intrinsic device structure. With control of short-channel effects in bulk-Si devices becoming increasingly difficult at shorter gate lengths, FinFET devices offer increasing performance improvement over bulk-Si MOSFETs with technology scaling.

The transistor structures used in this study are shown in Fig. 2 and the key design parameters are summarized in Table III. Device dimensions such as $T_{SD}$ and $L_{SD}$ were optimized for the FinFET (consistent with the thickness requirement for good short channel behavior), and parameters such as $V_{th}$ and the S/D doping gradient are estimated from scaling trends. Because completely undoped silicon substrates are expensive and challenging to obtain, a low but realistic doping of $10^{16}$ cm$^{-3}$ is assumed for the FinFET. The FinFETs in this study are chosen to be symmetric, with identical front- and back-gate oxide thicknesses and work functions. The motivation for this choice is the relatively high process complexity of asymmetric devices, which require either precise lithographic alignment (less than $T_{SI}/2$) or tilted implantations. The high aspect ratios of tall, densely packed fins are likely to make asymmetric FinFETs even more challenging. The results are expected to apply to both bulk-Si and SOI FinFETs, since the body effect has been observed to be negligible in fully depleted devices [33]. FinFETs fabricated on a standard (100) wafer have channels on the fin sidewalls that are oriented along (110) planes, for standard layouts. To capture the effect of fin-sidewall surface orientation on FinFET performance, the carrier mobilities in Taurus [32] are calibrated using experimental data for the (110) surface [34]. For the independently gated FinFETs, the front and back gates each have significant control over the channel. Simulated current values for a few bias conditions are presented in Table IV.

FinFET-based SRAMs can be simulated just like planar SRAMs, by using the device $I$-$V$ curves to solve for node voltages. Several researchers have already fabricated FinFET-based SRAMs in silicon, with similar voltage transfer characteristics to planar SRAMs [21]–[25]. Independently gated FinFETs also have been demonstrated in silicon [26], showing good agreement with the simulated $I$-$V$ characteristics. A selective top-gate-removal process has been reported to fabricate circuits using both kinds of FinFETs [26]. Such a process can be used to fabricate FinFET SRAM cells with some gates connected and others separated, as illustrated by Fig. 3.

SRAM failure, as defined by letting the RSNM = 0 or $I_W$ = 0, is caused by a combination of $L_G$ and $T_{SI}$ variations in a FinFET process. Such variations can arise from a combination of systematic and truly random sources. Variations that depend on particular process conditions, such as the uniformity of an etch or an anneal, or on particular aspects of the layout, such as the orientation or the proximity, will tend to systematically affect all devices or cells on a chip. They can be modeled as random variables to account for process fluctuations; however, the high sensitivity to process or layout makes their distributions difficult to predict in a general analysis. On the other hand, variations from uncorrelated random sources such as line edge roughness or (in doped devices) random dopant fluctuations are inherent to semiconductor processing and therefore more suitable for a general analysis. They are also the more significant cause of SRAM failure. Symmetric cells are more easily disturbed by mismatch, particularly in the PD devices. In our analysis, mismatch variations contribute approximately 75% of the read margin cell sigma, regardless of the device or cell design investigated.

In this work, only the uncorrelated random variations are considered, in order to provide the worst-case estimate of the cell yield. For FinFETs, line-width variations are assumed to be independent, Gaussian random variables with zero mean and $\sigma_{L_G} = \sigma_{TSi} = 1.54$ nm. If similar processing techniques are
used to define $L_G$ and $T_{Si}$, it is reasonable to expect that the standard deviations will be equal. Although their value will be process-dependent, the cell sigmas estimated with these numbers (7% of $L_G$) can be scaled accordingly. The cell sigma is estimated as an equivalent number of standard deviations of combined $T_{Si}$ and $L_G$ variations from the nominal design point to the most probable point of failure (RSNM = 0) or $I_W = 0$ using an iterative, sensitivities-based approach. By focusing on the zero-crossings, accurate yield estimates can be made without assuming Gaussian statistics for the cell metrics. In order to determine the most probable point of failure, the simulated SRAM characteristics were translated into a pseudo-analytical model based on seven transistor $I$-$V$ targets. These $I$-$V$ targets were derived from Taurus simulations for nMOS and pMOS FinFETs with various gate work functions ($\Phi_{ni}$), gate lengths ($L_G$), and fin widths ($T_{Si}$). RSNM was extracted from the noise margin curves with WL and BL voltages at $V_{DD}$, except where otherwise noted. The results for RSNM and $I_W$ were validated against mixed-mode Taurus simulations.

### B. FinFET SRAM Cell Designs

1) Conventional Double-Gated (DG) Designs: The conventional DG design [29] is first investigated; its schematic and layout are shown in Fig. 4. The layout was generated using generalized 45-nm node logic design rules. The dark outline indicates the memory cell boundary. It should be noted that SOI FinFET-based SRAM cells will generally be denser than similarly sized bulk-Si SRAM cells, because they can avoid the p-channel doping) might not have decreased leakage, due to band-to-band tunneling.

Fig. 4 plots the RSNM curves for both the 6-T bulk-Si NMOS device (with higher $V_{TH}$) and the 6-T FinFET-based SRAM devices achieves a 30% improvement in RSNM as compared to its bulk-Si-based counterpart with a $\beta$ ratio of 1.5. Moreover, a further 37% improvement in RSNM, with 16.6% area penalty, can be achieved by upsizing the PD FinFETs each by 1 fin. High-$V_{TH}$ devices were implemented in the FinFET designs by utilizing a gate material with 4.75 eV work function for both the nMOS and pMOS devices. This improves read/write margins and also suppresses leakage. Using a single gate material also improves manufacturability since it is challenging to implement different gate work functions for closely spaced p-channel and n-channel fins. (The high aspect ratio of the FinFETs makes it difficult to selectively tune the gate work functions along the sidewalls of the fins, e.g., by masked ion implantation.) In contrast, a higher $V_{TH}$ bulk-Si NMOS device (with higher channel doping) might not have decreased leakage, due to band-to-band tunneling.

When the PD FinFETs are strengthened by adding fins, the cell write margin shrinks—primarily due to the reduction in

<table>
<thead>
<tr>
<th>TABLE IV</th>
<th>nMOS FINFET I-V TARGETS UNDER DIFFERENT BACK-GATE (BG) BIASES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double-Gated ($V_{BG} = V_{FG}$)</td>
<td>$V_{DS}$ (V)</td>
</tr>
<tr>
<td>Saturation on current</td>
<td>1.0</td>
</tr>
<tr>
<td>Linear on current</td>
<td>0.1</td>
</tr>
<tr>
<td>Off current</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent-Gated (BG off)</th>
<th>High $V_{DS}$</th>
<th>Low $V_{DS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{DS}$</td>
<td>1.0</td>
<td>0.1</td>
</tr>
<tr>
<td>$V_{DS}$</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$I_{DS}$</td>
<td>0.0</td>
<td>7 $\mu$A</td>
</tr>
</tbody>
</table>
the write trip voltage. The dependences of the read and write margins on the number of fins in the PD devices are shown in Fig. 6(c). Gate-work-function tuning can also be used to balance read and write margins by changing the relative strengths of the nMOS and pMOS devices, but at the cost of increased process complexity.

2) Pass-Gate Feedback (PGFB): Whereas adaptive body biasing becomes less effective with bulk-Si MOSFET scaling, back-gate (BG) biasing of a thin-body MOSFET remains effective for dynamic control of $V_T$ with transistor scaling, and can provide improved control of short-channel effects as well [38]. The strong BG biasing effect can thus be leveraged [39] to optimize the performance of FinFET-based SRAMs through a dynamic adjustment of the effective cell $\beta$-ratio.

By connecting the storage node to the BG of the PG transistor, as shown in Fig. 7, the strength of the PG transistor can be selectively decreased [30]. For example, if the stored bit is a “0,” the BG of the corresponding PG transistor is biased at 0 V, decreasing its strength. This effectively increases the $\beta$-ratio during a read operation, allowing the PD transistor to keep the storage node at a lower voltage. Because the cell retains its state during a read operation or a half-select condition, the storage node at a lower voltage. Because the cell retains its state during a read operation, allowing the PD transistor to keep the storage node at a lower voltage. Because the cell retains its state during a read operation, allowing the PD transistor to keep the storage node at a lower voltage.

This simple BG connection scheme incurs no area penalty over the conventional DG 6-T SRAM cell design. The cell area is actually reduced by 2% due to the elimination of the 80 nm gate extension beyond the active region (fin) that the DG PG device required [see Fig. 7(b)]. In conventional DG SRAM designs, gate-work-function adjustment can be used to trade off the read and write margins. A higher gate work function strengthens the pMOS devices and weakens the nMOS devices. This improves the RSNM by increasing the trip point of the inverter, but doubly decreases writeability by weakening the PG and strengthening the PU device. The PGFB SRAM design offers a more favorable tradeoff. It enables higher RSNM at high $V_{DD}$ than is achievable with gate-work-function tuning alone, and enhanced writeability at a matched RSNM. In addition to enhancing nominal margins, the PGFB design exhibits reduced sensitivities to process variations, resulting in higher-yielding cells.

Fig. 9 illustrates nominal RSNM over a range of $V_{DD}$ for a conventional DG 6-T SRAM cell versus a cell with PGFB. Gate-work-function tuning is used to make the conventional design as stable as the PGFB design at $V_{DD} = 0.7$ V. To match the large RSNM enhancement of PGFB, a higher gate work function ($\Phi_m = 4.82$ eV) is required for the conventional design. Although the two designs have comparable RSNMs up to 0.7 V, at higher $V_{DD}$ tuning is less effective than PGFB. This is because the increasing effects of drain-induced barrier lowering at higher $V_{DD}$ values lower the gain of the inverter and reduce the benefit to RSNM. For $V_{DD} > 1$ V, the PGFB design achieves very high RSNM, exceeding 250 mV, which is not obtainable with $\Phi_m$ tuning alone.

Though the effect of increasing $\Phi_m$ is limited for high supply voltages and introduces process complexity, its largest draw-
back is in the tradeoff with writeability. Fig. 10 illustrates the writeabilities of the conventional DG and PGFB SRAM designs with matched RSNM at $V_{DD} = 0.7$ V. The inset illustrates the N-curves used to find the writeability current $I_{W}$. At low $V_{DD}$, the conventional design exhibits reduced writeability because its larger $\Phi_{m}$ raises the nMOS $V_T$ and thus keeps the PG device in subthreshold operation. The PGFB design, which results in relatively stronger nMOS and weaker pMOS transistors, pulls down the internal node of the cell faster, allowing for a more rapid write operation. On the half of the cell initially storing a low voltage, the pull-up behavior of the PG device is weakened by initially having a low BG bias, but this is partially compensated by the lower $\Phi_{m}$. The net effect of the much stronger pull down and the slightly weaker pull-up is a greater writeability for the PGFB cell. Thus, the inherently improved read stability of the PGFB design enables a better read/write tradeoff by allowing for lower $\Phi_{m}$ and therefore higher $I_{W}$.

At high $V_{DD}$, the writeability comparison is more complex. Nominal writeability for the PGFB design saturates as the effect of the lower $\Phi_{m}$ becomes less significant than the reduced drive on the BG. The bias on the BG, indicated by $V_{CH}$ at the $I_{W}$ point in the inset of Fig. 10, is lowered in the PGFB design and can be approximated as $V_{DD}/2$. This is an indirect effect of the lower $\Phi_{m}$, which reduces the trip point of the inverter, but it has an interesting benefit to the cell yield.

The projected cell yield in the presence of statistical variations (random $L_G$ and $T_{Si}$ variations) for all six devices is illustrated in Fig. 11. With matched RSNM at $V_{DD} = 0.7$ V, the conventional DG design and the PGFB design show comparable read cell sigmas across all values of $V_{DD}$. The yield saturates at ten standard deviations, which corresponds to the fin thickness variation. The PGFB design shows a significantly better write cell sigma at low $V_{DD}$. Much of this benefit is due to the improved nominal writeability current seen in Fig. 10; however, the write yield also exhibits a low sensitivity to $V_{DD}$. This enables a wider range of operating voltages for the cell in an array and is in contrast to DG-FinFET and bulk-Si MOSFET SRAM designs, which are often write-limited at low supply voltages. The reason for the low sensitivity is the reduced bias on the back gate at the writeability point. Whereas the PG device in the conventional cell sees a variation in $V_{DD}$ on both its front and back gates, in the PGFB cell it sees only the variation on its back gate, which reduces the sensitivity to $V_{DD}$. This is an easily overlooked tradeoff of conventional gate-work-function tuning: increasing $\Phi_{m}$ not only degrades writeability, but also degrades the yield faster at low $V_{DD}$. Although PGFB enables a low $\Phi_{m}$ if a higher $\Phi_{m}$ were used (perhaps for further RSNM enhancement) the same degradation would be seen at low $V_{DD}$.

The read/write tradeoff of PGFB can be further explored by comparing RSNM at matched writeability. In Fig. 12, $\Phi_{m}$ tuning (from 4.75 to 4.85 eV) was used on a conventional DG design to match the PGFB writeability at each point over a large $V_{DD}$ range. For $V_{DD} > 0.4$ V, the RSNM of the PGFB design is consistently higher by approximately 20%. These results confirm that the improved read/write tradeoff from using PGFB is valid over a wide range of gate work functions.

The PGFB design therefore offers an inherently more stable read operation, with enough margin to lower the gate work function and improve writeability. Its biggest drawback is that the weakened PG can degrade read performance. In this work, read...
chosen such that write-ability currents are matched at each $V_{DD}$. The PGFB design has $\sim 20\%$ greater RSNM in this case for most values of $V_{DD}$.

Fig. 12. Nominal RSNM with $\Phi_{tn}$ chosen such that RSNM is matched at each $V_{DD}$. The PGFB design has $\sim 15\%$ less read current than the conventional design at low $V_{DD}$. Above $0.8 \, V$, the conventional design cannot match PGFB RSNM with any amount of gate-work-function tuning.

performance is measured by the DC read current, that is, the current flowing through the PG transistor on the “0” side of the cell when the WL and BL voltages are high. In some sense, this is a fundamental tradeoff; the reduced current that degrades read performance also decreases the charge that helps destabilize the cell. Fortunately, the degradation is not severe, especially when compared to a conventional DG design with matched RSNM at each $V_{DD}$ up to $0.8 \, V$ (see Fig. 13). Although the PGFB design has only a single gate inverting the channel, the lower $\Phi_{tn}$ reduces $V_T$ and increases the drive current. Furthermore, the “0” storage node in the PGFB design stays closer to $V_{SS}$ than in the conventional DG design (see Fig. 8), thus giving the BG PG transistors more gate overdrive. The net result is only a 15% degradation in read current. It should be noted that for $V_{DD} > 0.8 \, V$, the conventional design was unable to match PGFB RSNM with any amount of gate work function tuning.

Other techniques can be used with the PGFB design to improve the write margin, in addition to gate work function adjustment. Without major impact on RSNM, the pMOS load devices can be made weaker by adjusting their gate lengths. However, this technique will only yield a marginal improvement in the write margin; a much more significant improvement can be attained by lowering the supply voltage during write, while maintaining the WL voltage [39]. This is made possible by adopting a long-aspect-ratio cell layout, which is typical in today’s designs for better manufacturability [2], [40]–[42], since the cell supply can be routed vertically for each column and can be exploited to break the contention between read and write optimization. With the ability for column-based biasing, cell supply voltage can be selectively lowered only for the column containing the cell under write access [2]. This keeps the cell stability high for all other cells connected to the same WL. Thus, high read- and write-margins can be independently achieved, which is important for the half-select condition. Essentially, the contention between read- and write-margins has been replaced by a contention between hold- and write margins, which offers a much bigger window for optimization. Fig. 14 summarizes the enhancement in write margin due to reduced cell supply voltage and the corresponding impact on the hold SNM. The downside to this method is the need to generate and distribute two different voltages, which is otherwise not needed with PGFB.

3) Pull-Up Write Gating (PUWG): A better approach to enhance writeability is to selectively weaken the PU devices. Just as feedback can be used to weaken the PG transistor during a read operation, it is possible to increase writeability by weakening the PU transistors during a write operation [31]. This can be achieved using independently gated FinFETs for the PU devices and connecting their BG to a write word line (WWL) (see Fig. 15). During a write operation, setting $V_{WWL} = V_{DD}$ reverse-biases the BG of the PU devices and thereby weakens them, hence increasing the writeability current. At all other times, $V_{WWL}$ can be set to $V_{SS}$ or an intermediate value ($V_{SS} < V_{WWL} < V_{DD}$) to enable large RSNM and hold margins.

Both PUWG and PGFB can be implemented simultaneously with no cell area penalty as illustrated in Fig. 15. The WWL contacts are located next to the word line contacts and are shared between adjacent cells. The WWL is routed horizontally, but must interleave with the WL to make all the contacts. This requires routing in an additional metal layer, but the cell area is not increased.

Fig. 16 illustrates nominal writeability currents for a conventional DG design, a design with PGFB, and a design with both PGFB and PUWG. (For the PGFB + PUWG design, $\Phi_{tn}$ is...
chosen to be the same as that for the PGFB design, 4.65 eV.) The combination PGFB + PUWG design provides significantly higher writeability currents than the conventional design at all supply voltages. For a given threshold of writeability current, the PUWG design can be written at approximately a 200-mV lower \( V_{DD} \) than the conventional design. Whereas the PGFB design alone saturates in the range of 50 mV lower \( V_{DD} \), together with PUWG it achieves higher writeabilities with increasing \( V_{DD} \).

During a read operation, the WWL voltage is lowered to \( V_{SS} \) or an intermediate bias value. The choice of this bias value affects the voltage transfer characteristics of the cell. Fig. 17 shows the impact of the WWL bias on the SRAM RSNM curves with PGFB and PUWG, at \( V_{DD} = 0.7 \) V. Setting \( V_{WWL} = V_{DD} \) causes the PU to weaken, thereby lowering the trip point of the inverter. On the other hand, setting \( V_{WWL} = V_{SS} \) fully turns on the PU of the PU device, thereby pushing out the upper shoulder of the RSNM curve (increasing the maximum \( V_{CH} \) for \( V_{CH} \approx V_{DD} \)). The effect complements that of PGFB in increasing the RSNM of the cell: while PGFB boosts RSNM by lowering the node settling voltage during a read, WWL biasing can increase the trip point of the inverter and achieve an increase in the RSNM as well. In this case, the RSNM is increased to 170 mV with \( V_{WWL} = V_{SS} \), from 100 mV with \( V_{WWL} = V_{DD} \). The largest RSNM is obtained when the trip point of the inverter is close to \( V_{DD}/2 \). There is an optimal WWL bias that maximizes RSNM near this point.

The bias value of WWL during a read operation determines the range of \( \phi_m \) that meets a given RSNM target (Fig. 18). At \( 0.7 \) V, RSNM \( > 170 \) mV can be obtained with a high \( V_{WWL} \) and a wide range of moderate to high \( \phi_m \). Decreasing \( V_{WWL} \) enables a lower \( \phi_m \) to achieve the same RSNM, such that at \( V_{WWL} = 0 \) V, \( \phi_m = 4.65 \) eV. As for the PGFB and DG designs, a lower work-function improves writeability. A particular advantage of choosing \( V_{WWL} = V_{SS} \) is that it does not require an additional voltage source. The disadvantage of a low \( V_{WWL} \) is that it increases leakage through the PU device. In particular, as \( \Delta V_{WWL} = V_{DD} - V_{WWL} \) becomes large (\( > 0.7 \) V), the leakage current can reach a level such that it degrades RSNM, in addition to increasing static power. Therefore in order to minimize these effects, and maintain high yield at low \( V_{DD} \), \( \Delta V_{WWL} \) should be kept to a moderate value.

During the write operation, WWL is raised for all the bits on the same WL. If a partial word is written, those bits that are not being written are subject to half-select stress. This condition is similar to the read stress and these cells will have a lower RSNM. The right edge of Fig. 18 illustrates the RSNM for these cells. For \( \phi_m < 4.85 \) eV, increasing \( \phi_m \) will increase RSNM, behaving as DG or PGFB cells would with weak PU transistors. Depending on the read bias of the WWL, though, this trend could be opposite that for the read condition. If this tradeoff can be avoided by limiting write operations to complete words, then the opportunity to modulate the PU strength dynamically allows for an improved tradeoff between read stability and writeability.

Fig. 19 compares RSNM for PGFB + PUWG in two cases: with \( V_{WWL} = 0 \) V to represent the case without an additional voltage source (dotted line), and with an optimal \( \Delta V_{WWL} \).
0.48 V (or $V_{\text{WWL}} = V_{\text{SS}}$, for $V_{\text{DD}} < 0.48$ V, thick line). (Note that the writeability is the same as shown in Fig. 16, since for both cases $V_{\text{WWL}} = V_{\text{DD}}$ during the write.) The optimal $\Delta V_{\text{WWL}}$ of 0.48 V is small enough such that the leakage current through the PU devices is very small, 4 nA at 0.7 V. At very low $V_{\text{DD}}$ (< 0.7 V), the highest RSNM is obtained for the combination of PGFB + PUWG, due to their complementary effects on the butterfly curves. The inset of Fig. 19 compares butterfly curves at $V_{\text{DD}} = 0.7$ V for the combination (bold curves) with the conventional design (thin curves). The inverter trip point for PGFB + PUWG is closer to $V_{\text{DD}}/2$, due to a lower gate work function and the effects of PUWG. The lower shoulders of the butterfly curves exhibit less linearization from the PG, due to PGFB. The slope of the curves near the trip point is somewhat degraded due to the reduced PMOS gain of PUWG; however, the effect on RSNM is small. The PGFB + PUWG design with $\Delta V_{\text{WWL}} = 0.48$ V has higher RSNM than the DG design for all $V_{\text{DD}}$, and the highest of all designs in intermediate voltages (0.7 V < $V_{\text{DD}}$ < 0.9 V). Above 0.9 V, RSNM in the PGFB + PUWG design is limited by the low gain of the independent-gated PU devices. For the special case of PGFB + PUWG with $V_{\text{WWL}} = 0$ V, RSNM saturates at about 170 mV, due to increasing leakage current through the BG device.

A high nominal RSNM increases the read yield for the PGFB + PUWG combination, particularly at low $V_{\text{DD}}$ (see Fig. 20). The read yields for the PGFB and DG designs are comparable to that of the PGFB + PUWG combination with $\Delta V_{\text{WWL}} = 0.48$ V, but is slightly lower for the PGFB + PUWG design with $V_{\text{WWL}} = V_{\text{SS}}$ at high $V_{\text{DD}}$. For this design, the large back-gate bias on the pMOS transistors during the read increases the sensitivity to device parameter variations than in other designs. The yield therefore decreases as $V_{\text{DD}}$ increases; however, over six sigma yield is still achievable in the range of 0.45 V < $V_{\text{DD}}$ < 1.2 V.

The largest difference is observed in the writeability yield, for which the PGFB + PUWG design is significantly higher than both the PGFB and DG designs (see Fig. 21). The large $I_{\text{W}}$ cell sigma can be attributed to two factors: a low $\Phi_{\text{m}}$ enabled by the PGFB design, and the weaker PU devices with $V_{\text{WWL}} = V_{\text{DD}}$. Since the pMOS transistors are weaker during the write operation, their current variations are smaller and yield is further enhanced. Extremely high yields of almost 10 sigma are achievable for $V_{\text{DD}} > 0.6$ V using this combination.

Gate work function tuning, PGFB, and PUWG offer three different kinds of read/write tradeoff for SRAM design. Increasing the gate work function increases RSNM by weakening all the
PGFB increases RSNM by dynamically weakening only the PG device, allowing higher \( I_W \) at low \( V_{DD} \), but limiting \( I_W \) at high \( V_{DD} \). PUWG increases \( I_W \) yield by dynamically weakening the PU devices. It can further increase RSNM at low \( V_{DD} \) by dynamically strengthening the PU devices, at some penalty in leakage current, if write operations are restricted to complete words. High RSNM yield and low leakage can be maintained at high \( V_{DD} \) if \( \Delta V_{WWL} \) is kept small, below 0.7 V. Because the PGFB and PUWG tradeoffs are complementary in the \( V_{DD} \) range over which they offer the greatest benefit, the combination of these designs can produce significant yield enhancements at all \( V_{DD} \).

### C. Architectural Considerations

In typical SRAM design, segmentation is used to optimize the access speed of an SRAM array. The number of cells connected to each BL, commonly referred to as the column height, sets the BL capacitance, which in turn determines how fast the BL can charge/discharge. In addition, as transistor off-state leakage current continues to increase with scaling, column height also determines the worst-case effective read current—when \( B_{L_{L}} \) and \( B_{L_{R}} \) are simultaneously discharging through the \( I_{READ} \) of the accessed cell along with the leakage currents of all other cells in the column. In this case, the effective read current is equal to the \( I_{READ} \) minus the sum of the leakage currents for all unaccessed cells in the column. The low off-state leakage of FinFETs should dramatically mitigate this problem.

However, the worst-case read scenario is exacerbated for the PGFB FinFET-based SRAM cell design, wherein the PG transistors are back-biased by the storage node. These BG connections prevent the PG transistors from completely shutting off, and hence prevent the BL from discharging fully. Specifically, the unaccessed PG transistors experience a drive voltage of up to \( V_{DD} - V_{BB} \) at the BG, which sources significant current onto the BL when \( V_{BL} < V_{DD} - V_{T} \). In other words, the current conducted by the BG PG transistors tends to contest the cell read current in discharging the BL below a certain voltage. A sense amplifier is therefore required to generate a zero voltage at the output. A good sense amplifier design typically requires only a small BL differential voltage (usually 10% of \( V_{DD} \)) to determine the cell state, for which the leakage through the BG connections is very small. Therefore, the degradation in BL discharge speed should not be significant. Nevertheless, in both PGFB and conventional DG FinFET SRAM cell designs, the effective worst case read current is reduced with increasing column height.

Fig. 22 compares the worst-case BL discharging patterns for a conventional single-fin DG 6-T SRAM cell and a PGFB 6-T SRAM cell. The gate work function is set to 4.75 eV for the conventional DG cell and 4.65 eV for the PGFB cell, for similar \( I_{READ} \). As expected, the BL is able to discharge all the way to “0” in the DG cell, whereas the BL can only discharge to just under 600 mV in the PGFB cell. However, the BL discharging speed is weakly affected by the feedback connection. Fig. 23 summarizes the impact of PGFB on sensing speed. Delta \( T \) is the difference in BL discharging time for a conventional DG 6-T SRAM design and a PGFB 6-T SRAM design with a column height of 128 cells. \( T \) is the time required to develop a 100-mV bit-line differential in a conventional DG 6-T SRAM, and sense amplifier offset is the tolerable offset voltage in the sense amplifier.

![Fig. 22. Bit-line voltage simulations for a conventional DG 6-T SRAM design and a PGFB 6-T SRAM design.](image)

![Fig. 23. Impact of PGFB on sensing speed; where \( \Delta T \) is the difference in bit-line discharging time for a conventional DG 6-T SRAM design and a PGFB 6-T SRAM design, \( T \) is the time required to develop a 100-mV bit-line differential in a conventional DG 6-T SRAM, and sense amplifier offset is the tolerable offset voltage in the sense amplifier.](image)

![Fig. 24. Bit-line voltage simulations for a PGFB 6-T SRAM design with various column heights (16-bit indicates column of 16 cells). \( B_{L_{R}} \) did not change significantly with column height. Therefore only one waveform is shown.](image)
impact of column height on BL discharging for a PGFB cell is illustrated in Fig. 24. As expected, the maximum BL differential decreases and the BL discharge time increases with increasing column height. Decreasing the column height speeds up the cell access time but incurs more area overhead from the sense amplifiers. Column multiplexing can be used to optimize area efficiency by allowing the read/write circuitry and sense amplifiers to be shared among multiple columns. However, the nonzero resistance of the BL multiplexers degrades column performance.

The BG connections also have a small effect on the stability of other PGFB cells in the column. As current discharges onto the BL from the internal node of an unselected cell, the hold margin of that cell is degraded. Specifically, the leaking PG acts as a resistive voltage divider, which gradually decreases $V_{CL}$ as $V_{CH}$ increases. The effect is analogous to the decrease from hold margin to RSNM in a conventional DG cell. It is mitigated for a read access in the column, since the BL only discharges to an intermediate value. It is most severe for a write access in the column, when the BL is fully discharged. Even under this condition though, the hold margin of the affected side of the PGFB cell is about 250 mV at $V_{DD} = 1.0 \text{ V}$, still 35 mV larger than RSNM in the conventional DG cell, and comparable to RSNM in the PGFB cell. A similar advantage is seen for the PGFB + PUWG cells in an unselected row.

While PGFB does not seem to impose extra restrictions on BL segmentation over the conventional DG design (see Fig. 22), cell disturbs in unselected columns during a write operation may restrict WL segmentation and result in extra area penalty in a PUWG cell. As described above, during a write operation, the WWL is charged to $V_{DD}$ and the PU transistor drive strength is reduced, thus reducing the RSNM of unselected cells in the same row, even when PGFB is used simultaneously. Therefore, cell designs with PUWG should be floor-planned with spatially adjacent words, where word-size may restrict the row length. Moreover, since bits from different words are no longer spatially interleaved, additional area penalty may result from the interleaving of ECC codes [45].

V. CONCLUSION

FinFETs with an undoped channel, as an alternate device architecture to planar bulk-Si MOSFET, may enable continued scaling of 6-T SRAM in the presence of increasing process variation and short channel effects. Conventional DG FinFET SRAM is expected to provide immediate cell sigma enhancements over bulk-Si SRAM. Independent gating enables further read and write margin enhancement. Cell RSNM can be improved by utilizing PGFB to dynamically adjust the PG transistor strength without area penalty and 15% read-current penalty—achieving significant stability improvements while keeping standby leakage current in check. PGFB is found to offer a better tradeoff than gate-work-function tuning to improve read stability, because of its improved writeability at low supply voltage. PUWG utilizes a separate write word line to allow for up to two sigma of further writeability yield enhancement by weakening the PMOS drive strength during write operation. (Greater cell sigma estimation accuracy can be obtained by including process-specific sources of systematic variation in the analysis.) The PGFB and PUWG designs are shown to be complementary and can be used in combination to simultaneously achieve improved read and write stability over conventional DG designs, making FinFET-based SRAM extremely attractive for high-density, low-power cache memory applications.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of the Center for Circuit & System Solutions (C2S2) Focus Center, one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program.

REFERENCES

Andrew Carlson (S’04–M’08) received the B.S. degree in engineering sciences from Harvard University, Cambridge, MA, in 2003, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley, in 2005 and 2008, respectively. He is currently working at Advanced Micro Devices, Boxborough, MA, on technology development for the 22-nm node.

Zheng Guo received the B.S. degree in computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 2003, and the M.S. degree in electrical engineering from the University of California at Berkeley, in 2005, where he is currently pursuing the Ph.D. degree in electrical engineering.

Upon graduation, he attended the University of California at Berkeley in fall of 2003. There, he joined the Berkeley Wireless Research Center in 2004 as a graduate student researcher (GSR) under the supervision of Prof. B. Nikolic. His research focus is in the area of robust SRAM design in deeply scaled CMOS processes. Dr. Guo was a recipient of the National Defense Science and Engineering Graduate (NDSEG) Fellowship in the fall of 2004, the best student paper award on a paper titled “FinFET-based SRAM Design” at the International Symposium on Low Power Electronics and Design (ISLPED) conference in 2005, and he was chosen as a winner in the 46th DAC/ISSCC Student Design Contest for his work in “Large-scale SRAM Variability Characterization Chip in 45nm CMOS” in 2009.
Sriam Balasubramanian (S’02–M’06) received the B.Tech. degree in metallurgical engineering from the Indian Institute of Technology, Madras, India, in 1998 and the M.S. degree in materials science and engineering and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, in 2002 and 2006, respectively.

From 2006 to 2008, he worked as a Device Engineer with Advanced Micro Devices focused on SRAM bitcell optimization and scaling for AMD’s high-performance technology. He is currently a Member of Technical Staff at GLOBALFOUNDRIES, Sunnyvale CA, and continues to work on SRAM scaling across multiple technologies. His interests include CMOS devices, and SRAM bitcell scaling into the nanoscale regime.

Radu Zlatanovici (S’00–M’06) received the B.S. and M.S. degrees in electronics from Politehnica University Bucharest, Bucharest, Romania, in 1999 and 2000, respectively, and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from University of California at Berkeley, Berkeley, in 2002 and 2006, respectively.

He was on the faculty of Politehnica University Bucharest from 1999 to 2000. In 2002 and 2003, he interned at IBM T.J. Watson Research Center, Yorktown Heights, NY, where he worked on power-performance tradeoffs for pipelined digital circuits. Upon graduation in 2006, he joined Cadence Research Laboratories, Berkeley, CA, as a Research Scientist. His research interests include low power techniques for digital circuits, arithmetic circuits, and clock distribution.

Tsu-Jae King Liu (SM’00–F’07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1984, 1986, and 1994, respectively.

In 1992, she joined the Xerox Palo Alto Research Center, Palo Alto, CA, as a Member of Research Staff, to research and develop polycrystalline-silicon thin-film transistor technologies for high-performance flat-panel displays. In August 1996, she joined the faculty of the University of California, Berkeley, where she is currently the Conexant Systems Distinguished Professor of Electrical Engineering and Computer Sciences and Associate Dean for Research in the College of Engineering. Her research activities are presently in nanometer-scale integrated-circuit devices and technology, as well as materials, processes, and devices for integrated microsystems.

Dr. Liu was a recipient of awards including the DARPA Significant Technical Achievement Award (2000) for development of the FinFET, and the IEEE Kiyo Tomiyasu Award (2010) for contributions to nanoscale MOS transistors, memory devices, and MEMs devices. She has served on committees for many technical conferences, including the International Electron Devices Meeting and the Symposium on VLSI Technology. She was an Editor for the IEEE ELECTRON DEVICES LETTERS from 1999 to 2004.

Borivoje Nikolić (S’93–M’99–SM’05) received the Dipl.Ing. and M.Sc. degrees in electrical engineering from the University of Belgrade, Serbia, in 1992 and 1994, respectively, and the Ph.D. degree from the University of California at Davis, in 1999.

In 1999, he joined the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, where he is now a Professor. Before, he lectured electronics courses at the University of Belgrade from 1992 to 1996. He spent two years with Silicon Systems, Inc., Texas Instruments Storage Products Group, San Jose, CA, working on disk-drive signal processing electronics. His research activities include digital and analog integrated circuit design and VLSI implementation of communications and signal processing algorithms. He is coauthor of the book Digital Integrated Circuits: A Design Perspective, 2nd ed. (Prentice-Hall, 2003).

Dr. Nikolić was a recipient of the NSF CAREER Award in 2003, College of Engineering Best Doctoral Dissertation Prize and Anil K. Jain Prize for the Best Doctoral Dissertation in Electrical and Computer Engineering at University of California at Davis in 1999, as well as the City of Belgrade Award for the Best Diploma Thesis in 1992. For work with his students and colleagues he received the Best Paper Award at the ACM/IEEE International Symposium of Low-Power Electronics in 2005, and the 2004 Jack Kilby Award for the Outstanding Student Paper at the IEEE International Solid-State Circuits Conference.