Tri-Gate Bulk CMOS Technology for Improved SRAM Scalability

Changhwan Shin, Borivoje Nikolić, Tsu-Jae King Liu
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94720-1770 USA
{shinch, bora, tking}@eecs.berkeley.edu

Chen Hua Tsai, Mei Hsuan Wu, Chung Fu Chang, You Ren Liu, Chih Yang Kao, Guan Shyan Lin, Kai Ling Chiu, Chuan-Shian Fu, Cheng-tzung Tsai, Chia Wen Liang
United Microelectronics Corporation
Hsinchu, Taiwan R.O.C.

Abstract — **A simple approach for manufacturing quasi-planar tri-gate bulk MOSFET structures is demonstrated and shown to be effective for reducing variation in 6T-SRAM read and write margins, in an early 28nm CMOS technology. With optimization of the pocket implant doses, quasi-planar bulk CMOS technology can facilitate voltage scaling. It also provides a means to achieve high yield with a notch-less 6T-SRAM cell layout, to facilitate area scaling.**

I. Introduction

A challenge for continued SRAM scaling is threshold voltage (V_T) mismatch due to process-induced variations [1], which degrades the minimum operating voltage (V_{min}) of an SRAM array [2]. To address this challenge, an improved transistor design that provides for reduced short-channel effects (*i.e.*, improved gate control over the channel potential) is required. The quasi-planar tri-gate bulk MOSFET [3] is an example of such a design; it utilizes a gate electrode that is physically wrapped around the top portion of the channel region to provide for greater capacitive coupling between the gate and the channel region [4]. In this work, a timed etch in dilute hydrofluoric (HF) acid solution is used to recess the isolation oxide prior to gate-stack formation, to form quasi-planar bulk MOSFETs using an otherwise conventional fabrication process flow. Improvements in transistor performance and SRAM yield are demonstrated in an early 28nm CMOS technology. The benefits of quasi-planar bulk MOSFET technology for voltage and area scaling are then assessed using three-dimensional (3-D) device simulations with atomistic doping profiles and analytical modeling to estimate 6T-SRAM cell yield for 22nm CMOS technology.

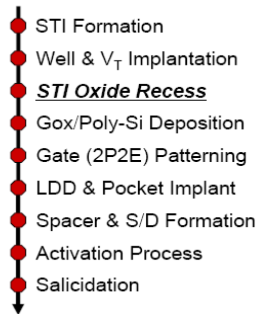


Figure 1. Front-end-of-line fabrication process steps used in this work.

II. Device Fabrication and Results

Individual logic transistors and 6T-SRAM arrays (~2500 cells per DUT) were fabricated using a standard test-chip mask set with an early 28nm low power CMOS technology which incorporates only dual stress liners for performance enhancement.

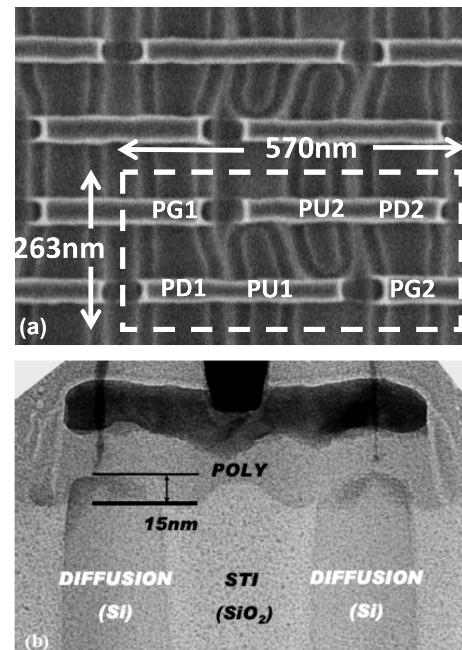


Figure 2. (a) 0.149um^2 SRAM cell plan-view CDSEM image after gate patterning. **(b)** XTEM taken along a poly-Si gate electrode in an SRAM array, for 15nm nominal STI recess depth.

The sequence of front-end-of-line fabrication process steps is illustrated in **Fig. 1**. After conventional shallow trench isolation (STI), well formation and V_T-adjust ion implantation, the STI oxide was recessed by a small amount (15nm) just prior to gate-stack formation. **Fig. 2** shows plan-view scanning electron microscopy and cross-sectional transmission electron microscopy images of a fabricated SRAM cell. Due to improved gate control, the quasi-planar structure achieves higher drive current (I_{ON}) for comparable off-state leakage current (I_{OFF}), as shown in **Fig. 3**. On average, I_{ON} is improved by 82%, 50%, and 79% for the pass-gate (PG), pull-

down (PD), and pull-up (PU) devices, respectively. The standard compact model can be well fit to quasi-planar bulk MOSFET characteristics, including the body effect (**Fig. 4**).

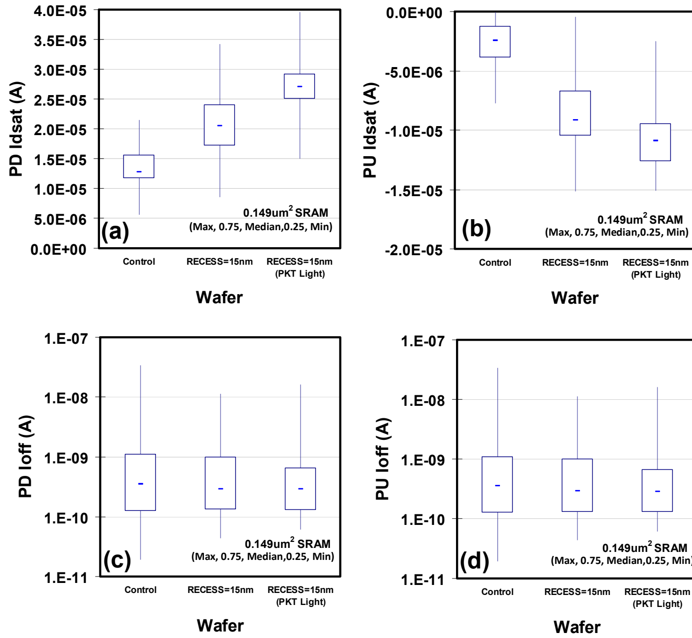


Figure 3. Comparison of on/off current statistics for planar (Control) *vs.* quasi-planar (RECESS=15nm) bulk MOSFETs in SRAM cells. (a) pull-down NMOS I_{ON} (b) pull-up PMOS I_{ON} (c) pull-down NMOS I_{OFF} (d) pull-up PMOS I_{OFF}.

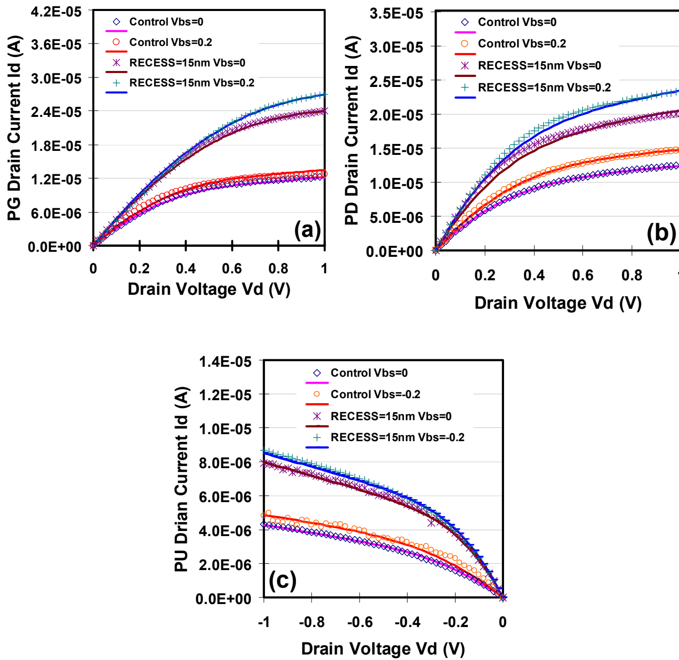


Figure 4. Comparison of measured output characteristics for planar (Control) vs. quasi-planar (RECESS=15nm) bulk MOSFETs in SRAM cells, for $|V_{GS}|$ = 1.0V. The effect of forward body biasing is also shown. **(a)** pass-gate NMOS, **(b)** pull-down NMOS, **(c)** pull-up PMOS. The symbols are measured data; the lines show the fitted compact model.

V_T statistics are shown in **Fig. 5** for the devices. Due to improved gate control (resulting in steeper subthreshold swing) V_T is lower for the quasi-planar devices. For the NMOS devices, variation in V_T increases slightly due to relatively heavy pocket doping which results in more significant impact of random dopant fluctuations (RDF) for the gated sidewalls. This undesirable effect is eliminated by using a lighter pocket implant dose, as shown in **Figs. 5a and 5b** (PKT Light). This further lowers V_T and increases I_{ON} (Figs. 3a and 3b) without increasing I_{OFF} (Figs. 3c and 3d).

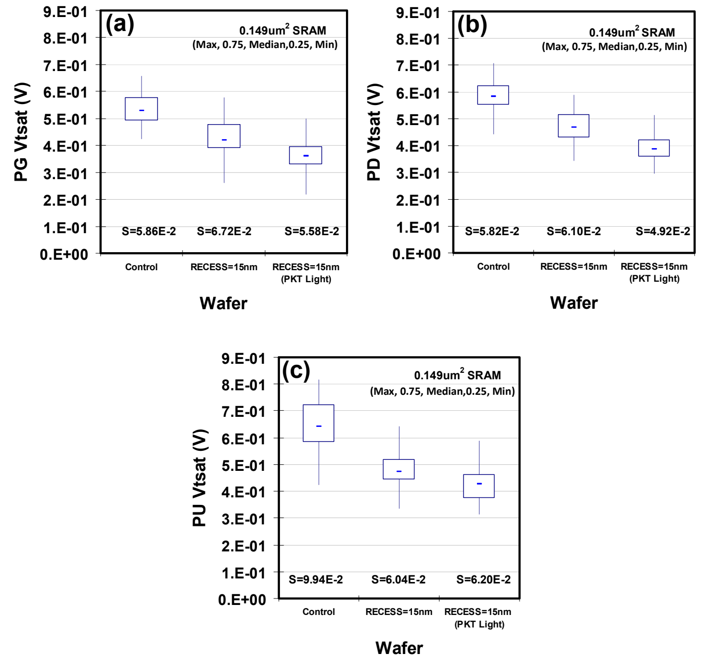


Figure 5. Comparison of saturation V_T statistics for planar (Control) vs. quasi-planar (RECESS=15nm) bulk MOSFETs in SRAM cells: (a) pass-gate NMOS, (b) pull-down NMOS, (c) pull-up PMOS.

Good short-channel control is maintained with the quasi-planar structure despite the lower pocket doping (**Fig. 6a**). In this early 28nm technology, the PMOS devices have lighter pocket doping than the NMOS devices, so that variation in V_T decreases when the STI oxide is recessed, due to the improved electrostatic integrity of the quasi-planar structure. However, if an even lighter pocket implant dose is used, then V_T variation increases slightly, as shown in **Fig. 5c**, due to slightly degraded short-channel control (**Fig. 6b**).

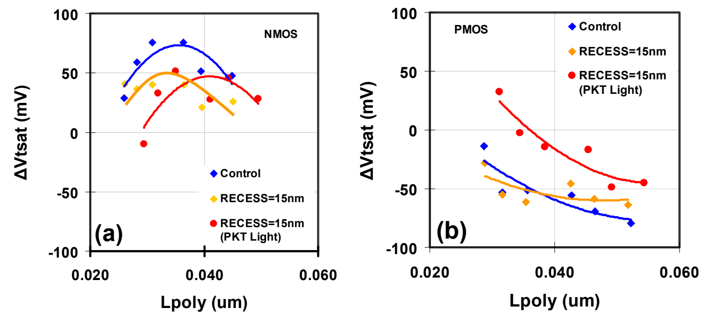


Figure 6. Change in saturation threshold voltage with decreasing gate length, for logic devices with 0.25μm drawn width. **(a)** NMOS **(b)** PMOS.

III. MEASURED 28NM SRAM RESULTS

With lighter pocket doping to reduce the impact of RDF, the quasi-planar structure results in reduced variability in SRAM read margin (SNM) and write margin (WRM) when the STI oxide is recessed, as shown in **Fig. 7**.

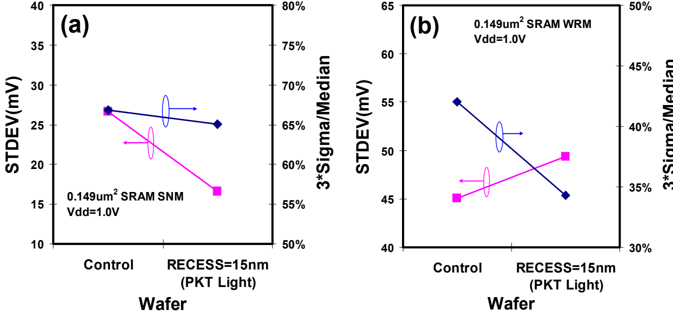


Figure 7. Sigma and 3sigma/median values for **(a)** read margin (SNM) and **(b)** write margin (WRM). $V_{dd} = 1.0V$.

Supply-voltage (V_{dd}) reduction is desirable to reduce power density and/or facilitate increased transistor density. Generally, however, relative variability increases as the gate overdrive (V_{dd}-V_T) decreases, so that yield (gauged by 3-sigma/median) is degraded. **Fig. 8** shows that the degradation in SNM yield with V_{dd} scaling can be dramatically reduced for quasi-planar bulk CMOS technology. With optimized pocket implant doses for both NMOS and PMOS devices (not achieved here), similar improvement is expected for WRM yield.

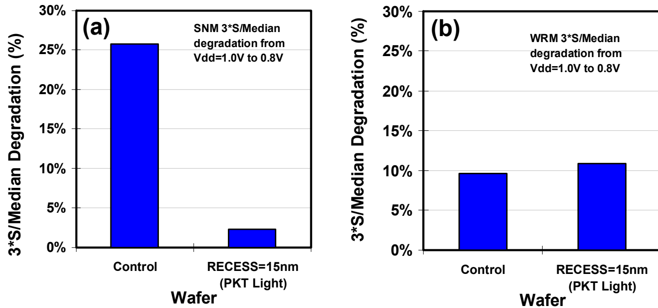


Figure 8. Degradation in 3-sigma/median for (a) SNM and (b) WRM as V_{dd} is reduced from 1.0V to 0.8V.

The reverse narrow width effect, *i.e.*, V_T reduction with decreasing channel width (W), stems from increased gate control for narrower channel width (due to fringing electric fields between the gate electrode and channel sidewalls). This effect is slightly worsened when the STI oxide is recessed, *i.e.*, quasi-planar devices show slightly increased sensitivity of V_T to variations in W, as shown in **Fig. 9**. Overall, however, variability is reduced for quasi-planar devices due to improved short channel control, which provides for the improved SRAM yield seen in Figs. 7 and 8.

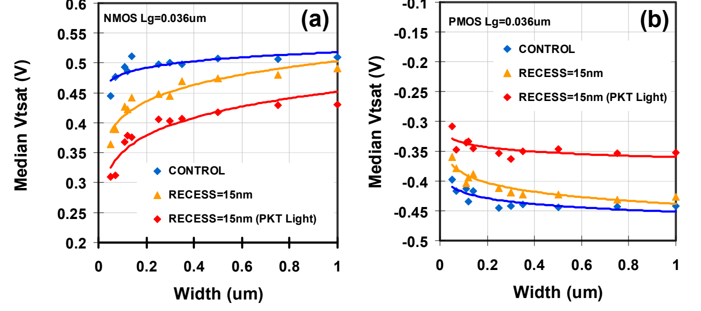
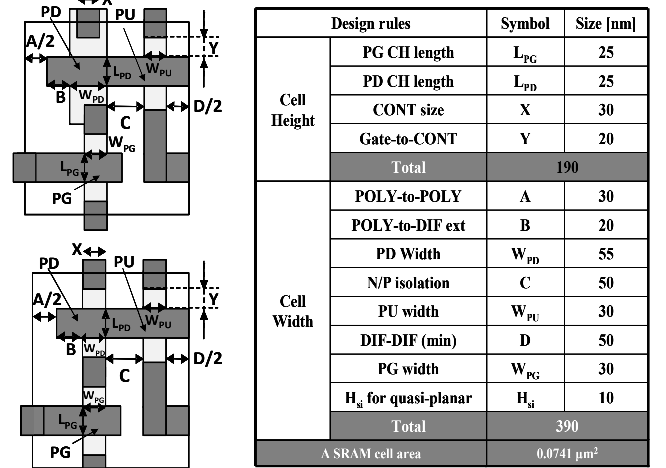


Figure 9. Measured reverse narrow width effect for devices with 36nm gate length: (a) NMOS (b) PMOS. Median V_T is lower when the STI oxide is recessed, due to improved gate control over the channel potential.

IV. 22NM BULK SRAM CELL DESIGN STUDY

Layout dimensions for 22nm (25nm drawn gate length) 6-T SRAM cells were selected based on recent publications [6-10] and are summarized in **Table I** for a conventional notched cell layout.

Table I. 6T- SRAM cell layout dimensions. Schematic half-cell plan views are shown for a conventional notched layout (upper left) and a smaller notch-less layout (lower left) with $W_{PD} = W_{PG} = W_{PU} = 35nm$ (area = 0.0684 μm²).



	Design rules	Symbol	Size [nm]
Cell Height	PG CH length	L_{PG}	25
	PD CH length	L_{PD}	25
	CONT size	X	30
	Gate-to-CONT	Y	20
	Total		190
Cell Width	POLY-to-POLY	A	30
	POLY-to-DIF ext	B	20
	PD Width	W_{PD}	55
	N/P isolation	C	50
	PU width	W_{PU}	30
	DIF-DIF (min)	D	50
	PG width	W_{PG}	30
	H_{si} for quasi-planar	H_{si}	10
	Total		390
A SRAM cell area			0.0741 μm²

The quasi-planar bulk MOSFET design was optimized via 3-D device simulations to achieve the highest I_{ON} for I_{OFF} = 3nA/um, at V_{dd} = 1V: electrical channel length (distance between the points where the source/drain doping profiles fall to $2\times10^{19}cm^{-3}$) L_{eff} = 27nm; effective oxide thickness EOT = 9Å; source/drain extension junction depth $X_{J,ext}$ = 10nm. Near-band-edge gate work functions (4.2eV for NMOS, 5.1eV for PMOS) are assumed. The STI recess depth (H_{si}) is 10 nm, to maintain a low aspect ratio for ease of fabrication. The retrograde channel doping profile is assumed to have a gradient of 4nm/dec and peak doping concentration = $10^{19}cm^{-3}$ at a depth T_{si} below the top channel surface. The planar bulk MOSFET design (for comparison) was optimized in the same manner.

The benefit of quasi-planar bulk CMOS technology for improved SRAM cell yield is assessed using the concept of cell sigma, defined as the minimum amount of variation for read/write failure [11-12]. Random V_T variation due to gate

line-edge-roughness (LER) and RDF was estimated from 3-D Monte Carlo device simulations with realistic gate profiles and atomistic doping profiles, as described in [12]. Random V_T variation due to gate work function variations was estimated from [13]. Variations in transistor performance due to systematic variations in L_{eff}, W, EOT, and H_{si} (each assumed to have Gaussian distributions, with ±10% corresponding to 3-sigma variation) as well as random V_T variations are considered in estimating SRAM cell yield.

As shown in **Fig. 10**, quasi-planar SRAM cells are projected to provide for >1 sigma improvement in read and write yields as compared with the planar SRAM cell, across a wide range of V_{dd}, primarily due to reduced V_T variation. As described in [14], the quasi-planar bulk MOSFET offers a new method of V_T adjustment, via tuning of the retrograde channel doping depth, to mitigate the tradeoff between reduced RDF-induced random V_T variation and improved short-channel control. This feature is leveraged in the notch-less quasi-planar SRAM cell design, in which the PD and PU devices have deeper retrograde channel doping profiles (such that $T_{Si} > H_{Si}$) and hence lower V_T [14], to achieve cell beta ratio > 1. The optimal T_{si} values to satisfy the 6-sigma (read and write) yield requirement were determined to be 14nm/10nm/14nm for the PD/PG/PU devices. For V_{dd} down to ~0.8V, a notch-less quasi-planar cell design with $W_{PD} = W_{PG} = W_{PU}$ = 35nm can meet the 6-sigma yield requirement while providing for significant (~10%) cell area savings as compared to the conventional notched cell design. (Here W_{PU} is constrained to be equal to W_{PD} and W_{PG} so as to be compatible with a regularly corrugated starting substrate [5] for improved W control.) Interestingly, the nominal SNM for the notch-less quasi-planar cell design is less sensitive to V_{dd}: it decreases by only 52 mV (from 174 mV to 122 mV) as V_{dd} is decreased from 0.9V to 0.5V, whereas the nominal SNM for the notched planar cell design decreases by 92 mV (from 180mV to 88mV) over the same V_{dd} range. This is because the benefit of the quasi-planar structure (improved sub-threshold swing) is greater for the narrower PD devices used in the notch-less cell design, so that they operate in the linear regime down to lower V_{dd}. Since variability decreases with decreasing V_{dd} (due to reduced short-channel effect and drain-induced barrier lowering), the SNM cell sigma stays relatively constant with decreasing V_{dd}, down to 0.6V, for the notch-less quasi-planar cell design. It should be noted that the notch-less cell design is expected to provide for reduced transistor mismatch arising from systematic variations in W, but that this benefit is not presumed for the SRAM cell yield analysis herein.

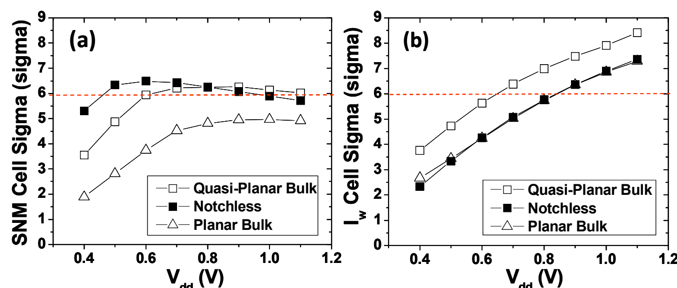


Figure 10. Comparison of cell sigma *vs.* V_{dd} for (a) read static noise margin (SNM) and (b) writeability current (I_W).

V. CONCLUSION

A simple process for achieving quasi-planar tri-gate bulk MOSFET structures is demonstrated in an early 28nm CMOS technology. With appropriate adjustments in the pocket implant doses, quasi-planar bulk CMOS technology can provide for improved performance and reduced variability, and thus can facilitate the scaling of SRAM operating voltage. Since the tradeoff between reduced RDF-induced random V_T variation and improved short-channel control is mitigated for quasi-planar CMOS technology, V_T (rather than W) adjustment can be used to tune the cell beta ratio and meet the 6-sigma yield requirements with a smaller cell.

ACKNOWLEDGMENTS

This work was supported in part by the Center for Circuit & System Solutions (C2S2) Focus Center, one of six research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program. C. Shin appreciates the support of the Korea Foundation for Advanced Studies (KFAS), and would like to thank Dr. Yasumasa Tsukamoto (Renesas Electronics, Japan) for his helpful discussion.

REFERENCES

[1] E. Josse *et al.*, "A cost-effective low power platform for the 45-nm technology node," *IEDM Tech. Dig.,* pp. 693-696, 2006.

[2] K. Nii *et al.*, "A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment," *Symp. VLSI Circuit Dig.*, pp. 212-213, 2008.

[3] X. Sun *et al.*, "Tri-gate bulk MOSFET design for CMOS scaling to the end of the roadmap," *IEEE Electron Device Letters*, Vol. 29, No. 5, pp. 491-493, 2008.

[4] M. Kito *et al.*, "Vertex channel array transistor (VCAT) featuring sub-60nm high performance and highly manufacturable trench capacitor DRAM," *Symp. VLSI Tech. Dig.,* pp. 32-33, 2005.

[5] U.S. Patent 7,190,050.

[6] H.S. Yang *et al.*, "Scaling of 32nm low power SRAM with high-K metal gate," *IEDM Tech. Dig*., pp. 233-236, 2008.

[7] H. Kawasaki *et al.*, "Demonstration of highly scaled FinFET SRAM cells with high-K/metal gate and investigation of characteristic variability for the 32nm node and beyond," *IEDM Tech. Dig*., pp. 237-240, 2008.

[8] B.S. Haran *et al.*, "22nm technology compatible fully functional 0.1μm^2 6T-SRAM cell," *IEDM Tech. Dig*., pp. 625-628, 2008.

[9] C.H. Diaz *et al.*, "32nm gate-first high-k/metal-gate technology for high performance low power applications," *IEDM Tech. Dig*., pp. 629-632, 2008.

[10] F. Arnaud *et al.*, "32nm general purpose bulk CMOS technology for high performance applications at low voltage," *IEDM Tech. Dig*., pp. 633-636, 2008.

[11] A.E. Carlson, "Device and Circuit Techniques for Reducing Variation in Nanoscale SRAM," Ph.D. dissertation, Univ. California Berkeley, 2008.

[12] C. Shin, *et al.*, "Performance and area scaling benefits of FD-SOI technology for 6-T SRAM cells at the 22nm node," *IEEE Trans. Electron Devices*, Vol. 57, No. 6, pp. 1301-1309, 2010.

[13] H. Dadgour, K. Endo, V. De, and K. Banerjee, "Modeling and analysis of grain-orientation effects in emerging metal-gate devices and implications for SRAM reliability," *IEDM Tech. Dig*., pp. 705-708, 2008.

[14] C. Shin, X. Sun, and T.-J. King Liu, "Study of random-dopant-fluctuation (RDF) effects for the tri-gate bulk MOSFET," *IEEE Trans. Electron Devices*, Vol. 56, No. 7, pp. 1538-1542, 2009.