Circuit Design in Nanoscale FDSOI Technologies

B. Nikolić[1], M. Blagojević[1,2,3], O. Thomas[4], P. Flatresse[2], A. Vladimirescu[3]

Abstract - Planar fully-depleted SOI technology with ultra-thin body and buried oxide presents a platform for an energy-efficient design in deeply scaled technologies without major changes in the bulk-CMOS design infrastructure. Good control of short-channel effects with thin transistor body offers a possibility to reduce the supply voltage. Thin buried oxide provides threshold tuning via body bias. Overall design optimality is achieved through sensitivity-based optimization by selecting optimal supplies and thresholds.

I. INTRODUCTION

CMOS technology scaling has lasted for over four decades. To continue the scaling trend, CMOS technology is migrating from the traditional bulk to thin-body device structures. Thin-body structures, finFETs and fully-depleted SOI devices (FDSOI) offer much better control over the charge in the channel and therefore much better off characteristics of the device. Furthermore, the use of ultra-thin buried oxide (BOX) in FDSOI devices allows operation in a very wide voltage range, and provides a substantial range of threshold voltage adjustments.

During the past decade the chip performance has been constrained by its power dissipation. Although power limits vary with the application domain, they, however, dictate the choices of technology and architecture, and necessitate the use of implementation techniques that optimally trade off performance for power savings. While power dissipation is generally managed through appropriate selection of the architecture and circuit design, the ability to continually change the supply voltage and the ratio of active and leakage power presents an opportunity for an additional design optimization. The optimal design is achieved when no additional energy can be saved by adjusting the accessible design variables.

Device parameter tolerances have not been able to track the reduction in feature sizes, and as a result, variation in device performance has been increasing, resulting in added design margins, which affect the design efficiency and its ability to operate at very low voltages. Particularly challenging has been the scaling of SRAM, which relies on very small devices, and is therefore susceptible to manufacturing variations. As a result, scaling of the supply voltage in SRAM has been slowed down.

B. Nikolić is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, e-mail: bora@eecs.berkeley.edu. M. Blagojevć, is with the University of California, Berkeley, Berkeley, STMIcroelectronics,Crolles, France and ISEP, Paris, France. O. Thomas is with CEA-Leti, Grenoble, France, P. Flatresse is with STMicroelectronics, Crolles, France, A Vladimirescu is with ISEP, Paris France,

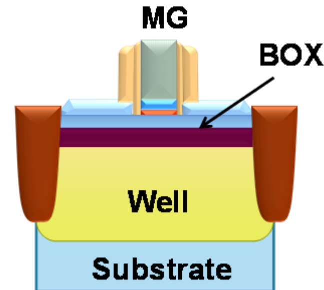

Figure 1: Cross-sectional view of the UTBB-FDSOI device.

This paper examines the ultra-thin body and box (UTBB) FDSOI technology, and its features for power optimization. It analyzes how these features map into a sensitivity-based optimization for energy and performance and provides examples of logic and memory design.

II. UTBB SOI

Undoped thin-film planar FDSOI devices have entered volume production at the 28nm node, as an alternative to bulk CMOS leveraging their excellent short-channel electrostatic control, low leakage currents, and reduced random dopant fluctuations (RDF) [1-2]. Some of the features include: (1) Back-plane (BP) doping underneath the buried oxide (BOX); (2) high body effect combined with low short-channel effects (SCEs); (3) low V_{Th} variability. In undoped-channel FDSOI technology, transistor threshold, V_{Th}, is primarily set by the metal-gate (MG) stack work function. UTBB FDSOI offers additional flexibility by setting the well doping type to be either n or p, as illustrated in Figure 1. In a practical 28nm process [1], minimum channel lengths are 24nm, and the silicon film thickness is 7nm.

BOX thickness is 25nm, which is a compromise between an increased parasitic capacitance and enhanced body effect. By the choice of the well doping and the twin-MG process it is possible to set the device thresholds. By placing an n-well underneath the PMOS, a regular-V_T (RVT) device is formed, while by chosing the p-well, a low-V_T (LVT) device is obtained. Similarly, p-well underneath NMOS results in RVT, while an n-well produces an LVT device. In addition, the BOX dielectric electrically isolates the well from the source and drain of the transistors, which expands the range of possible well

978-1-4799-5296-0/14/$31.00 © 2014 IEEE

bias voltages (V_B) and therefore improves the range of possible V_{Th} adjustments, through a high body factor. V_B is only limited by pn-well junctions. FDSOI achieves low V_{Th} variability because of its immunity to RDF even under forward body bias (FBB) [5], in contrast to bulk technology.

By creating openings through the thin BOX layer, it is possible to add standard bulk features to the design, such as diodes and passives.

II. Digital Logic

One advantage of the planar FDSOI technology is that the design migration form bulk is relatively straightforward. Early experiments used the same mask set for implementing the design in bulk and FDSOI. By not processing the layer with bulk well taps and adding the layers with definitions under the box a design can be converted from bulk to FDSOI.

From the physical design point of view, the main difference between bulk and FDSOI logic is that FDSOI requires explicit diodes added for protection from antenna-rule violations, even on supply networks. In contrast to bulk, transistor source and drain regions do not present a diode connection to the substrate, and do not provide a natural protection from antenna rule violations.

In logic design, the main difference from bulk is in the way multiple transistor thresholds are being used. In bulk CMOS, each individual gate can be assigned to use either LVT or RVT transistors, since their thresholds are set by channel doping, though a mask. However, the V_{Th} assignment has to be done for larger groups of gates in FDSOI, since the thresholds are set through well doping underneath the box. Spacing between well types is somewhat larger and high layout density requires a large number of digital gates to share the same well. Therefore a more effective method of mitigating leakage at gate granularity is by using poly bias, where transistors with varying channel lengths are being used in the design. This is a common feature of bulk designs as well.

By 'flipping' the wells with appropriate biasing as illustrated in Fig. 2, i.e. by using the N-well underneath the NMOS transistors and the P-well underneath the PMOS transistors the RVT design can be converted to all LVT, thus trading off increase in power for a gain in performance.

Highly-effective tradeoffs between performance and transistor leakage can be made by using back bias applied to the wells below BOX. The range of voltage is not limited by forward biasing the S/D junctions or by the gate-induced drain leakage as in bulk, but simply by the voltage limitations on the diode formed between the substrate wells. In the forward direction, the voltage cannot exceed 0.6V, while in the reverse direction the voltage between the wells can be as high as a 3V, limited by the breakdown.

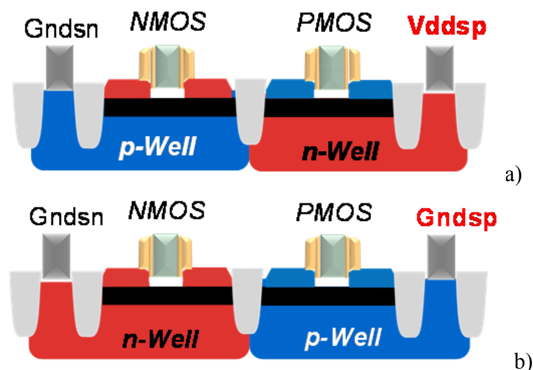

Figure 2: Illustration of the a) standard and b) flipped well FDSOI structure.

III. SRAM

Voltage reduction and increased variability associated with technology scaling compromise margins necessary for robust SRAM operation in bulk CMOS. Inability to effectively scale SRAM into sub-20nm bulk technologies is one of the main motivators for the shift towards the use of thin-body transistors [1]. UTBB FDSOI eliminates channel doping to lower intrinsic transistor variability, and along with multiple threshold voltages provides the ability to lower SRAM operating voltage while maintaining operating margins.

Elimination of channel doping reduces the standard deviation of random dopant fluctuations in SRAM devices by 25-30% compared to bulk, thus allowing for stable operation at lower supply voltages. The choice of the well structure for NMOS and PMOS devices allows for adjustments between read/write access times and stability while the range of back bias allows for tradeoff between performance, stability and leakage power.

Assessment of SRAM functionality is nowadays based on dynamic margins, as opposed to static margins, for better representation of actual failure modes. Dynamic margins are obtained through transient simulations of SRAM arrays and present the timing difference from each mode of failure. Margins against read stability (RS), read access time (RA) and writeability (WA) failure are assessed by using Monte Carlo (MC) based bit error rate (BER) estimates that do not make any assumption about the distribution of each failure metric [6]. The estimation of these margins can be accelerated by using importance sampling methods [7]. RS failures happen when the bitcell changes its value accidentally or when the internal node voltage is less than 80% of V_{DD} at the end of the clock period during the read access. RA failures occur when the bitline difference voltage is less than the offset (100mV) of the sense amplifier at the end of the wordline (WL) pulse width. WA failures appear when the flipped written internal node voltage is less than 80% of V_{DD} at the end of the clock period. The 80% threshold is considered to

prevent bitcell failures on read access consecutive to any operation [9-10].

FDSOI technology offers a unique degree of optimization for the SRAM array. By placing the P-well underneath the NMOS transistors and the N-well underneath the PMOS transistors, a standard, bulk-like cell is obtained with all RVT transistors, in accordance to Figure 2.a. However, by realizing that the bulk-like cell is limited by WA failures at low voltages, an improved cell architecture can be developed as in Fig. 3. By using a single-p-well (SPW) underneath the entire array, the PMOS transistor is operating with a reduced V_{Th} (LVT), enhancing the writeability with a small power penalty [10]. The NMOS transistors' RVT does not change. By increasing V_B, NMOS transistors are forward biased to improve RA and therefore $V_{DD,MIN}$, as the cell becomes read-limited. The PW is isolated from the p-substrate by using a deep n-well (DNW) tied to V_{DDS}. Thanks to the single common well, V_B can be biased up to (or tied to) V_{DDS}, biasing the NMOS transistors in a full forward mode. The use of a single well also reduces the impact of well proximity effects. WA improvement lowers $V_{DD,MIN}$ by 120mV for 64 and 128 bitcell columns, while for 256b RA improvement lowers $V_{DD,MIN}$ by 60mV, as in Figure 5.

The SPW architecture and V_B biasing improve $V_{DD,MIN}$ at the cost of increased leakage at same supply voltage. However, the SPW bitcell leads to the best $V_{DD,MIN}$–leakage current tradeoff, as shown in Fig. 5, in particular for a short bitline architecture for SRAM arrays. By adding SRAM assist techniques [9] it is capable of operation in a wide range of supply voltages, comparable to that of logic, making it a good choice for first level of cache memories.

Figure 3: UTBB-FDSOI 6T SPW SRAM bitcell schematic: PD & PG are RVT, PU is LVT. VB can be biased from a negative voltage up to the deep n-well (DNW) voltage.

Figure 4: $V_{DD,MIN}$ (TT, 27°C) for regular and SPW cells SPW and VB tied to VDD leads to the lowest RA and WA $V_{DD,MIN}$. RA64/128 $V_{DD,MIN}$ < WA $V_{DD,MIN}$ = 650mV; RA256 $V_{DD,MIN}$ = 740mV.

Figure 5: Leakage current vs. $V_{DD,MIN}$ for varying column heights (TT, 27°C). The leakage is normalized to the baseline bitcell at $V_{DD,MIN}$ for a 64b column [10].

IV. Design Optimization

The use of back-bias in FDSOI technology opens another avenue for design optimization unavailable in nanoscale bulk CMOS. All current designs are power-limited, and maximum performance is limited by power dissipation. When optimizing the design it is necessary to trade off excess performance for power savings. There are several design variables that can be adjusted to trade off energy for performance at various levels of design hierarchy. The tradeoff achieved by adjusting a design variable x is given by the energy/delay sensitivity to the variable x:

$$S_x(X) = \left. \frac{\partial E / \partial x}{\partial D / \partial x} \right|_{x=X} \tag{1}$$

This quantity represents the amount of energy that can be traded for delay by tuning variable x, around the design point X. An energy-efficient design is achieved when the relative sensitivities to all the tuning variables are balanced [11-12].

FDSOI enables the optimization through both the selection of threshold voltage and its continuous tuning via back-bias. A strong body effect (~60-80mV/V) with a wide tuning range (2-3 V) is able to trade off almost two orders of magnitude of leakage power for both speedup and extended V_{DD} operating range.

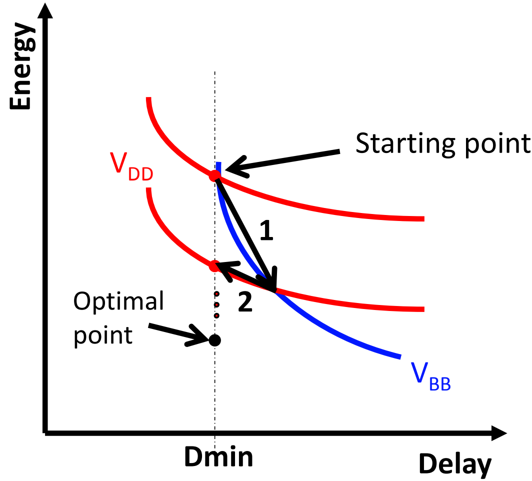

Figure 6: Illustration of energy-delay tradeoff curves and energy optimization.

Figure 6 illustrates the tradeoff procedure applied to the supply voltage, V_{DD} and back-bias, V_{BB} as design variables. A continuous curve is traced by varying either the supply voltage or back bias to maximize the performance under varying energy constraints. The slope of the curves changes in each point. A slack can be created by tracing a more sensitive variable, and the performance can be recovered by tracing a less sensitive one. The procedure can be repeated until the two slopes match. By adjusting the back-bias voltage and the supply, savings of 15% of energy from sizing- and supply-optimized designs have been reported [13-14].

This procedure can be applied to all design variables, including gate widths, poly bias, logic depth or block topologies [11].

V. Conclusion

UTBB FDSOI technology allows for energy savings in memory and logic through back-bias optimization. The use of single-P-well SRAM arrays enables operation at low supply voltages and enables scaling into the next technology nodes.

Acknowledgement

The authors acknowledge students, faculty and members of the Berkeley Wireless Research Center and Soitec.

References

[1] N. Planes et al, "28nm FDSOI technology platform for high-speed low-voltage digital applications ," *Symposium on VLSI Technology*, Honolulu, HI, 2012, pp. 133-134.

[2] E. Karl et al, "A 4.6GHz 162Mb SRAM design in 22nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry," *Proc. IEEE International Solid-State Circuits Conference, ISSCC 2012*, San Francisco, CA, 2012, pp. 230-232.

[2] K. Cheng et al., "Extremely thin SOI (ETSOI) technology: Past, present, and future," *Proc. IEEE Int'l SOI Conf.*, 2010, San Diego, CA, pp.1-4.

[4] J-P. Noel et al., "Multi- VT UTBB FDSOI Device Architectures for Low-Power CMOS Circuit," *IEEE Transactions on Electron Devices*, vol. 58, pp. 2473-2482, 2011.

[5] O. Weber et al., "Work-function engineering in gate first technology for multi-VT dual-gate FDSOI CMOS on UTBOX," *2010 IEEE International Electron Devices Meeting*, IEDM 2010, San Francisco, CA, pp. 3.4.1-3.4.4.

[6] D.E. Khalil, M. Khellah, N.-S. Kim, Y. Ismail, T. Karnik, V.K..De, "Accurate Estimation of SRAM Dynamic Stability, *IEEE Transactions on VLSI*, vol. 16, pp. 1639-1647, 2008.

[7] L. Dolecek, L. Dolecek, M. Qazi, D. Shah, A. Chandrakasan, "Breaking the simulation barrier: SRAM evaluation through norm minimization," *Proc. ICCAD 2008*, San Jose, CA, 2008, pp. 322-329.

[8] O. Rozeau, M. Jaud, T. Poiroux, M. Benosman, "Surface potential based model of ultra-thin fully depleted SOI MOSFET for IC simulations," *Proc. IEEE Int'l SOI Conf.*, 2011, Tempe, AZ, 2011, pp. 1-22.

[9] B.M. Zimmer, et al, "SRAM Assist Techniques for Operation in a Wide Voltage Range in 28-nm CMOS," *IEEE Trans. Circuits and Systems II: Express Briefs*, vol. 59, pp. 853-857.

[10] O. Thomas, et al, "6T SRAM design for wide voltage range in 28nm FDSOI," *Proc. IEEE Int'l SOI Conf. 2012*. Napa, CA, 2012.pp. 1-2.

[11] B. Nikolić, "Design in the power-limited scaling regime," *IEEE Trans. Electron Devices*, vol. 55, pp. 71-83, 2008.

[12] D. Marković, V. Stojanović, B. Nikolić, M.A. Horowitz, R.W. Brodersen, "Methods for True Energy-Performance Optimization," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1282-1293, 2004.

[13] P. Flatresse, et al, "Ultra-wide body-bias range LDPC decoder in 28nm U;TBB FDSOI technology," *IEEE Int'l Solid-State Circuits Conf, ISSCC'13,* San Francisco, CA, 2013, pp. 424-425.

[14] M.G Weiner, et al, "A Scalable 1.5 to 6Gb/s, 6.2 to 38.1mW LDPC decoder for 60GHz wireless networks in 28nm UTBB FDSOI," *IEEE Int'l Solid-State Circuits Conf, ISSCC'14*, San Francisco, CA, 2014, to appear.