14.1 A 64Mb SRAM in 32nm High-k Metal-Gate SOI Technology with 0.7V Operation Enabled by Stability, Write-Ability and Read-Ability Enhancements

Harold Pilo[1], Igor Arsovski[1], Kevin Batson[1], Geordie Braceras[1], John Gabric[1], Robert Houle[1], Steve Lamphier[1], Frank Pavlik[1], Adnan Seferagic[1], Liang-Yu Chen[2], Shang-Bin Ko[2], Carl Radens[2]

[1]IBM Systems and Technology Group, Essex Junction, VT,
[2]IBM Systems and Technology Group, Hopewell Junction, NY

A 64Mb SRAM macro is fabricated in a 32nm high-k metal-gate (HKMG) SOI technology [1]. Figure 14.1.1 shows the 0.154μm² bitcell (BC). A 2× size reduction from the previous 45nm design [2] is enabled by an equal 2× reduction in BC area. No corner rounding of BC gates allows tighter overlay of gate electrode and active area. The introduction of HKMG provides a significant reduction in the equivalent oxide thickness, thereby reducing the Vt mismatch. This reduction allows aggressive scaling of device dimensions needed to achieve the small area footprint. A 0.7V VDD_{MIN} operation is enabled by three assist features. Stability is improved by a bitline (BL) regulation scheme. Enhancements to the write path include an increase of 40% of BL boost voltage. Finally, a BC-tracking delay circuit improves both performance and yield across the process space.

Stability assist by reducing the pass-gate (PG) strength is an effective method for decreasing failure rate. However, several of these methods [4,5] interfere with other operations and require external modulation to preserve the balance between stability and write-ability. Figure 14.1.2 shows a BL regulation system that does not degrade the write margin. To improve stability, BLs are precharged to a reduced level (VBLH) [3]. VBLH lowers the bump level of the internal "low" node to improve the noise margin of the cross-coupled inverter's trippoint. Figure 14.1.2 shows the improvement in failure rate as a function of VBLH level, plotted as a fraction of VDD supply. The regulator is designed to operate with a VBLH range of 68 to 78% of VDD. An increase in failure rate as VBLH is lowered beyond the operating range is caused by reverse stability fails; the PG connected to the "high" side of the internal BC node begins to conduct and is discharged from VCS (BC power supply) to VBLH. The voltage reference, Ref is PVT compensated by body-contacted device T0. Diode-connected PFET, T1 modulates the body of T0 to compensate for changes in V_T and maintain a near constant reference. The distributed regulator output device, Treg manages the high-current demand of the BL restore. Treg is physically embedded with the BL and Sense-Amplifier (SA) pre-charge FETs. BC leakage causes the VBLH supply to drift towards VDD, which decreases the stability advantage. To prevent VBLH from drifting, a single pull-down device, TL is driven by the push-pull regulator (Gnbias). Overlap current between Treg and TL is minimized with a 40mV dead zone built into the regulator. The voltage translation from the VBLH to the VDD domain occurs in the transition from the SA data lines DLT/DLC to the global read data lines GBLT/GBLC.

Negative BL boosting is an effective technique to improve write margin [4]. The BL boost increases the V_{GS} of the PG, facilitating the discharge of the internal BC node. Limitation in the boost voltage in previous work [4] is caused by partial capacitor discharge, which reduces the charge transfer into the BLs. Another limitation is the leakage from the unselected transistors in the write path that begin to conduct when source voltages are boosted below GND. Figure 14.1.3 shows a schematic of the write driver with boost control that features a 40% improvement in boost voltage compared to the previous design [2]. Boosted node Nboost connects to eight physical BL pairs, segmented into upper (Ntu/Ncu) and lower half (Ntl/Ncl) partitions. Nboost is pre-charged to GND by Nd at the end of the write cycle. Boost capacitor, Cboost is also charged during this time by the transition of WS1n to VDD. To write a "1" into the BC, BLT is discharged to GND through bit-switch device Nt0 and segment device Ntu. Shortly after BLT reaches GND, the gate of Nd is shut off and WS1n transitions to GND to boost BLT below GND. Ntu is selected by the combination of true write data, WDTn and upper write select, WSELn. Boost voltage is increased as the gates of the three unselected segment devices (Gcu/Gtl/Gcl) are also boosted below GND. A 0V V_{GS} across the unselected segment devices guarantees full isolation and no loss of charge.

To minimize the gate-dielectric stress of the write path, a boost control scheme reduces the boost voltage at higher VDD where full assist is not required. The amount of boost is determined by the separation of WS0n and WS1n that control Nd and Cboost, respectively. An array edge-cell BL is used as a write path load to accurately time the initiation of the boost after BL is discharged to GND. A write bias generator, Wbias sets the V_{GS} of T0 (Wbias – NS) to modulate the delay of WS0n with respect to WS1n. At high VDD, Wbias is lowered by the increased strength of the four-NFET stack and WS0n is delayed compared to WS1n. The transition of WS1n before WS0n depletes the charge on Cboost by on-device Nd and the boost is attenuated. At lower VDD, WS0n switches low prior to WS1n to prevent the charge of Cboost to drain across Nd and the boost is maximized. Figure 14.1.4 compares waveforms at 0.7V (-198mV boost) and 1.0V (-66mV boost). The timing relationship between WS0n/WS1n for the two VDD cases is shown. The maximum BL boost as a function of VDD is also plotted in Fig. 14.1.4. The dotted lines represent the BL boost level without attenuation. Voltage stress is reduced by 200mV at 1.2V/-40°C.

BC optimization requires unique V_T implants that are decoupled from logic transistors. For improved yield and performance a BC-aware delay controls critical SRAM timings (see Fig. 14.1.5). A small memory array is configured to discharge a capacitor load (node Nbitcell). The WL activation is driven by the macro clock. To reduce the effects from random variations, 16 BCs are activated in parallel. The BC terminals are biased at higher metal levels to guarantee the correct state of the internal nodes at power-up. An edge-cell BL from an adjacent array is connected to the output of the 16-bit array to capture BL capacitance characteristics. A logic delay is added in parallel to the 16-bit delay element; this prevents the overall delay from becoming too fast and guarantees worst-case signal margin at the slow corners. Figure 14.1.5 shows WL to SET time as a function of VDD. At 0.7V, the required SET delay times, t_{SET} are indicated for two BCs (typical and +50mV V_T on PG). Delay times for logic-only delay elements (dotted lines) are compared with BC-tracking elements. Considerable performance improvements are gained at higher VDD for the BC-aware delay element.

Figure 14.1.6 shows hardware results at 85°C. The fail-count is compared with assist features disabled (left) and enabled (right). The VCS supply (array and WL-driver) is plotted against the VDD supply (periphery). SRAM operation is shown down to 0.7V. An overall improvement of 400mV of VCS is observed when these features are enabled compared to the default state. The area overhead for the stability and write assist is 1.5% and 1.2%, respectively. An overall array efficiency of 71.6% is achieved with a 128-BL sub-array configuration. Figure 14.1.7 shows the micrograph of the test chip and design features. The 64Mb SRAM is built from 128 512Kb macros used as the principal building block for high-performance ASIC SoC.

References:

[1] Greene, B., et al., "High Performance 32nm SOI CMOS with High-k/Metal Gate and 0.149μm² SRAM and Ultra Low-k Back End with Eleven Levels of Copper," *Symposium on VLSI Technology*, June 2009.

[2] Pilo, H., et al., "A 450ps Access-Time SRAM Macro in 45nm SOI Featuring a Two-Stage Sensing-Scheme and Dynamic Power Management," *ISSCC Digest of Technical Papers*, pp. 378-379, Feb. 2008.

[3] Bhavnagarwala, A., et al., "A Sub-600mV, Fluctuation Tolerant 65nm CMOS SRAM Array with Dynamic Cell Biasing," *Symposium on VLSI Circuits*, pp. 78-79, June 2007.

[4] Fujimura, Y., et al., "A Configurable SRAM with Constant-Negative-Level Write Buffer for Low-Voltage Operation with 0.149μm² Cell in 32nm High-k Metal-Gate CMOS," *ISSCC Digest of Technical Papers*, pp. 348-349, Feb. 2010.

[5] Kolar, P., et al., "A 32nm High-k Metal Gate SRAM with Adaptive Dynamic Stability Enhancement for Low-Voltage Operation," *ISSCC Digest of Technical Papers*, pp. 346-347, Feb. 2010.

 978-1-61284-302-5/11/$26.00

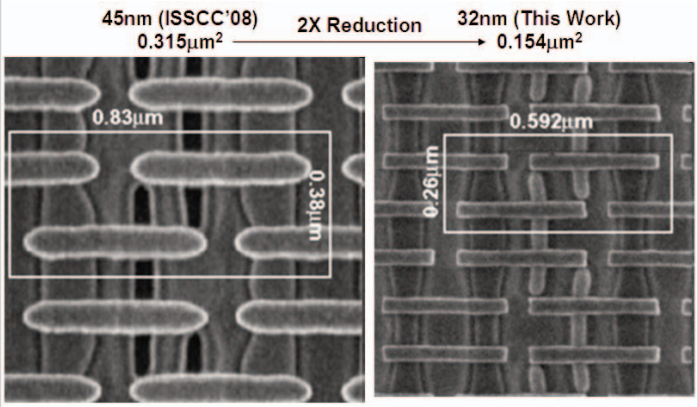


Figure 14.1.1: 45nm to 32nm technology scaling of 6T SRAM bit-cell.

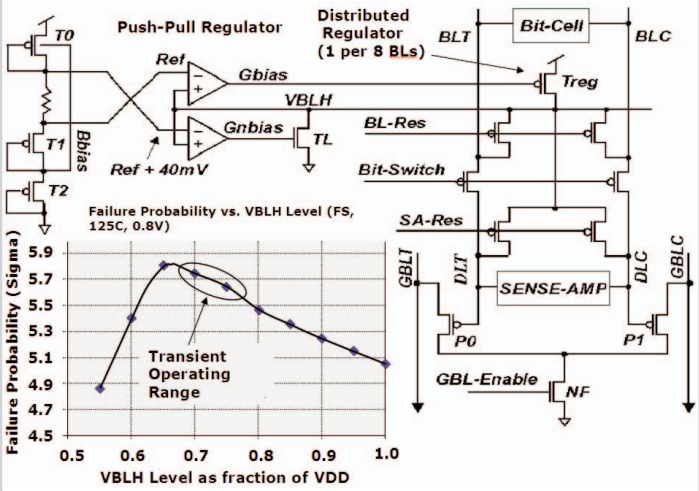


Figure 14.1.2: VBLH Bit-Line Regulation system and yield improvement.

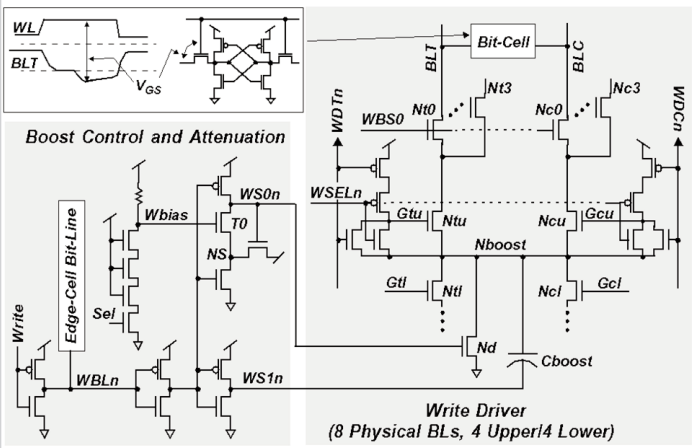


Figure 14.1.3: Write driver with Boost control and attenuation.

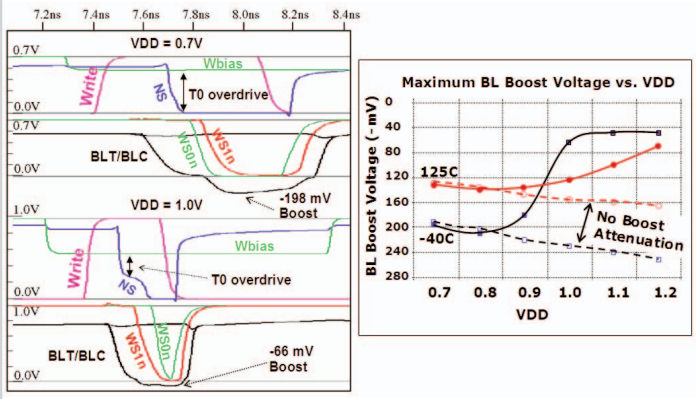


Figure 14.1.4: Write cycle simulation waveforms and attenuation results.

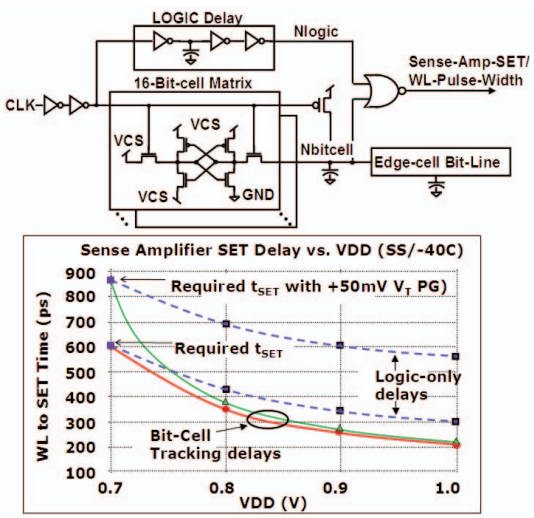


Figure 14.1.5: Bit-cell tracking circuit for critical timings.

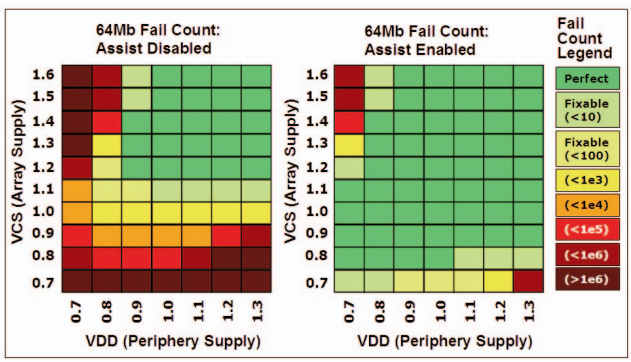
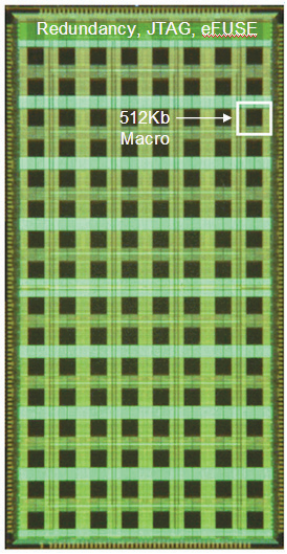


Figure 14.1.6: 64Mb Hardware results: 85°C VCS/VDD Voltage shmoo.



Technology	32nm PD SOI with High-k Metal Gate
Cell Size	0.154μm²
512Kb Macro Size	331μ x 339μ
Sub-Array Configuration	128 Word-Line x 128 Bit-Line
Operating Voltage	Core: 0.7V – 1.0V (0.9V typ.) SRAM: 0.7V – 1.0V (0.9V typ.)
Performance Target	1.4GHz (Slow process, 0.8V, 0C)

Figure 14.1.7: Micrograph of 64Mb Test Chip and Features.

978-1-61284-302-5/11/$26.00