34.4 A 256kb Sub-threshold SRAM in 65nm CMOS

Benton H. Calhoun, Anantha Chandrakasan

Massachusetts Institute of Technology, Cambridge, MA

Low-voltage sub-threshold operation has proven to minimize energy per operation for logic [1], and sub-threshold systems will require memories that function at the same low voltages. In this paper, a 65nm SRAM that functions into the sub-threshold region and examines the impact of process variation for low-voltage operation is described.

Previous efforts to reduce SRAM power have included voltage scaling to the edge of sub-threshold [2] or into the sub-threshold region [3], but only for idle cells. Although some published SRAMs operate at the edge of sub-threshold, none function at sub-threshold supply voltages compatible with logic operating at the minimum energy point. The 0.18μm memory in [4] provides one exception. Consisting of latches and using MUX-based read (18T-equivalent bitcell), it operates to 180mV.

Traditional 6T SRAMs face many challenges in deep submicron (DSM) technologies for low V_{DD} operation. Predictions in [5] suggest that process variations will limit standard 90nm SRAMs to around 0.7V operation because of static noise margin (SNM) degradation and write margin, and a V_{DD} of 0.7V is reported for a 65nm SRAM [6]. Measurement results confirm that SNM degradation and inability to write are the two most significant obstacles to sub-threshold SRAM functionality in 65nm. Each of these problems and a bitcell and an architecture that overcomes them, are discussed in this paper.

Figure 34.4.1 shows the impact of local V_T mismatch on the SNM for a standard 6T bitcell in a 65nm process. The Monte-Carlo simulations show that larger channel area decreases the spread of SNM ($\sigma_{VT} \propto 1/\sqrt{WL}$) and that global variation shifts the distribution caused by mismatch [9]. The Hold SNM at 0.3V has roughly the same mean as the Read SNM at 0.5V. However, the 6σ Hold SNM at 0.3V roughly equals the 6σ Read SNM at 0.6V. Likewise, the 6σ Hold SNM at 0.4V and 6σ Read SNM at 0.8V are equivalent. Thus, by eliminating the degraded Read SNM, a bitcell can be operated at 0.3V with the same 6σ stability as a 6T bitcell at 0.6V. A 7T cell avoids Read SNM for above-V_T SRAM [7], but the dynamic storage that it uses is problematic for the longer cycle times of sub-V_T operation.

The 10T bitcell in Fig. 34.4.2 uses transistors M7 to M10 to remove the problem of Read SNM by buffering the stored data during a read access. Thus, the worst-case SNM for this bitcell is the Hold SNM related to M1 to M6, which is the same as the 6T Hold SNM for same-sized M1 to M6. Results from [8] show that single-ended read offers competitive speed for the same area efficiency in DSM. This 10T bitcell uses a full-swing single-ended read that can be 'sensed' using an inverter. Clearly, the extra FETs increase the area by ~66% and also consume leakage power. M10 significantly reduces leakage power relative to the case where it is excluded. In unaccessed cells, M10 prevents node QBB from pulling to '0' even when QB='1'. In this technology, the PMOS sub-threshold current is stronger than NMOS, so node QBB floats close to V_{DD} and decreases sub-threshold current through M8. Also, when QB='0', leakage through M7 is reduced by the stack that M10 creates. Specifically, for iso-V_{DD}, the 10T cell without M10 (a 9T cell) has 50% higher leakage current than the 6T, but adding M10 drops the overhead to 16%. This overhead in leakage current is more than compensated by decreasing V_{DD} by 300mV relative to the 6T bitcell. In simulation, the 10T bitcell at 300mV consumes 2.25× less leakage power than the 6T bitcell at 0.6V (1.75× less relative to 0.5V).

The reduction in sub-threshold leakage through M8 reduces the impact of leakage from unaccessed cells and gives the additional advantage of allowing more cells on a BL during read. Figure 34.4.3 shows the impact of BL leakage on the steady-state voltages while reading a '1' (solid lines) or '0' (dotted lines). For the same number of cells on a BL, the 10T bitcell shows larger BL separation than the 6T (or 9T) bitcells, and 'sensing' with an inverter (whose switching threshold, V_M, is shown) works in simulation from 0°C to 100°C at all corners for 256 cells on a BL. For the 6T cell (or 9T), BL leakage limits the number of cells on a BL to 16 at several process corners for 0.3V. The higher level of integration allowed by the 10T cell reduces the peripheral circuits and slightly mitigates the bitcell area overhead. In order to combat the impact of local V_T mismatch, the WL voltage is boosted relative to the array V_{DD} by 100mV.

Write functionality is the second major obstacle to sub-threshold SRAM, as in this 65nm technology, a 6T bitcell cannot write in the traditional fashion below 0.6V. The plot in Fig. 34.4.4 shows the write margin for the 6T cell under typical and worst-case process corner and temperature. In both cases, the write fails as evident by continued bistability in the cell. Sizing alone cannot correct this problem, because the exponential dependence of sub-threshold drive current on V_T overwhelms the impact of sizing. To achieve write in sub-threshold, the virtual supply (VV_{DD}) to the selected cells floats during the write operation (e.g. [5]). The plot shows that, even for the worst-case, this method provides ample negative noise margin for ensuring a write. Clearly, the side of the bitcell holding a '1' is degraded in voltage due to the collapsing virtual supply. Figure 34.4.4 also shows the essential timing required for the write operation to bring this value to full V_{DD}. The VV_{DD} floats as $\overline{\text{VDDon}}$ is asserted along with WL_WR. The crucial transition in the diagram occurs when $\overline{\text{VDDon}}$ goes low before WL_WR, allowing positive feedback to restore the '1' to full V_{DD}. In the test chip, each row contains a single 128b word that is written at the same time and shares the same VV_{DD}. The block diagram in Fig. 34.4.4 shows how the row is 'folded' so that its cells share a VV_{DD} line.

A 256kb 65nm test chip (Fig. 34.4.7) uses the 10T bitcell and the architecture shown in Fig. 34.4.5. The decoders and other periphery use static CMOS logic for robust sub-threshold operation. The entire array functions at one V_{DD}, and the WL and write drivers operate at 100mV above that supply.

Assuming one redundant row and column are allocated per block, this implementation of the SRAM functions to below 400mV. At 400mV, it consumes 3.28μW and works up to 475kHz. No bit errors for holding data occur in the SRAM until V_{DD} scales below 250mV. Reading works without error at 320mV and writing at 380mV at 27°C. At 85°C, the SRAM writes without error at 350mV and reads without error at 360mV. The measurements on the chip are performed down to 300mV (Fig. 34.4.6 shows correct operation), however at this low voltage mismatch results in bit errors in ~1% of the bits. One type of bit error occurs when a bit holding a '1' is read as a '0' (non-destructive read). This occurs along columns whose I_{RD} has a high V_M due to mismatch. For rows whose M_P is stronger due to mismatch, the write operation fails to overpower M_P sufficiently to flip the contents of the cell, even when VV_{DD} is floating. Both of these problems can be fixed by minor changes to the peripheral circuits, allowing further V_{DD} reduction. Leakage power reduction from V_{DD} scaling is 2.4× and 3.8× relative to 0.6V operation at 0.4V and 0.3V, respectively (Fig. 34.4.6), and active energy savings are 2.25× and 4×.

Acknowledgements:
This work was funded by DARPA and Texas Instruments. We thank Texas Instruments for chip fabrication.

References:
[1] A. Wang, A. Chandrakasan, and S. Kosonocky, "Optimal Supply and Threshold Scaling for Sub-threshold CMOS Circuits," *IEEE Computer Society Annual Symp. on VLSI*, pp. 5-9, Apr., 2002.
[2] N. Kim et al., "Circuit and Microarchitectural Techniques for Reducing Cache Leakage Power," *IEEE Trans. VLSI Systems*, vol. 12, no. 2, pp. 167-184, Feb., 2004.
[3] H. Qin et al., "SRAM Leakage Suppression by Minimizing Standby Supply Voltage," *ISQED,* pp. 55-60, 2004.
[4] A. Wang and A. Chandrakasan, "A 180mV FFT Processor Using Subthreshold Circuit Techniques," *ISSCC Dig. Tech. Papers*, pp. 292-293, Feb., 2004.
[5] M. Yamaoka et al., "Low-Power Embedded SRAM Modules with Expanded Margins for Writing," *ISSCC Dig. Tech. Papers*, pp. 480-481, Feb., 2005.
[6] K. Zhang et al., "A SRAM Design on 65nm CMOS Technology with Integrated Leakage Reduction Scheme," *Symp. VLSI Circuits*, pp. 294-295, June, 2004.
[7] K. Takeda et al., "A Read-Static-Noise-Margin-Free SRAM Cell for Low-Vdd and High-Speed Applications," *ISSCC Dig. Tech. Papers*, pp. 478-479, Feb., 2005.
[8] K. Zhang et al., "The Scaling of Data Sensing Schemes for High Speed Cache Design in Sub-0.18μm Technologies," *Symp. VLSI Circuits*, pp. 226-227, June, 2000.
[9] B. Calhoun and A. Chandrakasan, "Analyzing Static Noise Margin for Sub-threshold SRAM in 65nm CMOS," *ESSCIRC*, pp. 363-366, Sept., 2005.

 1-4244-0079-1/06/$20.00 ©2006 IEEE

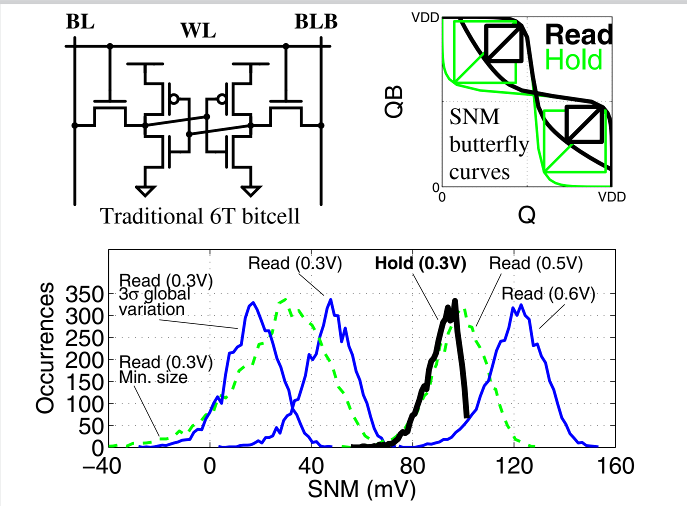

Figure 34.4.1: Impact of local mismatch on 6T SNM in 65nm. Read SNM has larger standard deviation. Hold SNM at 0.3V has roughly the same mean as Read SNM at 0.5V and same 6σ SNM as Read SNM at 0.6V.

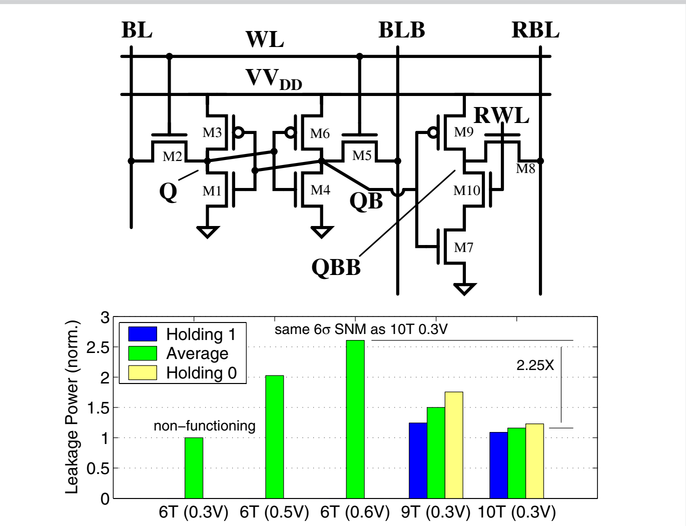

Figure 34.4.2: 10T bitcell for sub-threshold operation. Removing Read SNM allows operation at 0.3V, which leads to 2.25× reduction in leakage power.

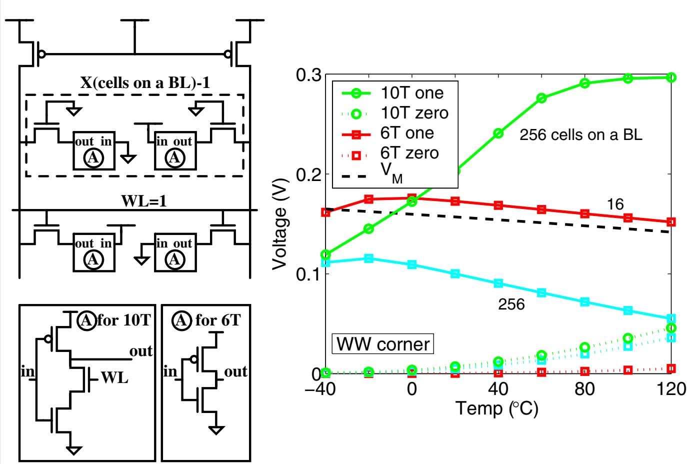

Figure 34.4.3: BL leakage limits the number of cells on a BL. The 10T bitcell can sustain 256 cells/BL at 0.3V compared to 16 without M10 (6T or 9T).

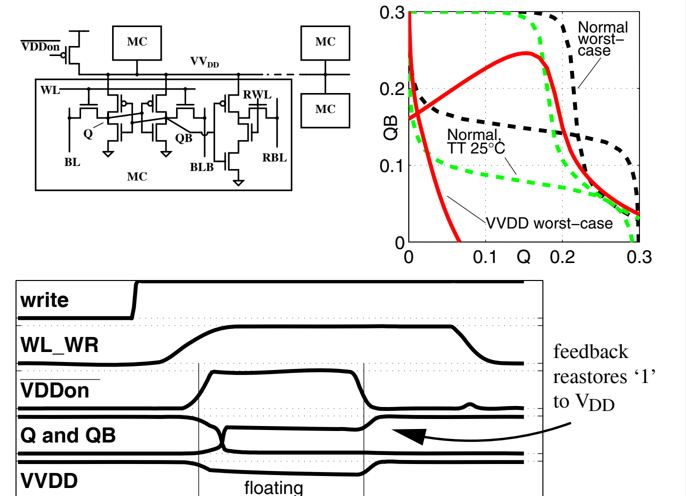

Figure 34.4.4: A virtual supply voltage (VV_{DD}) that floats during write allows robust write operation into sub-V_T (mono-stable butterfly curve). VV_{DD} stops floating while WL_WR remains asserted to restore the '1'value to full V_{DD}.

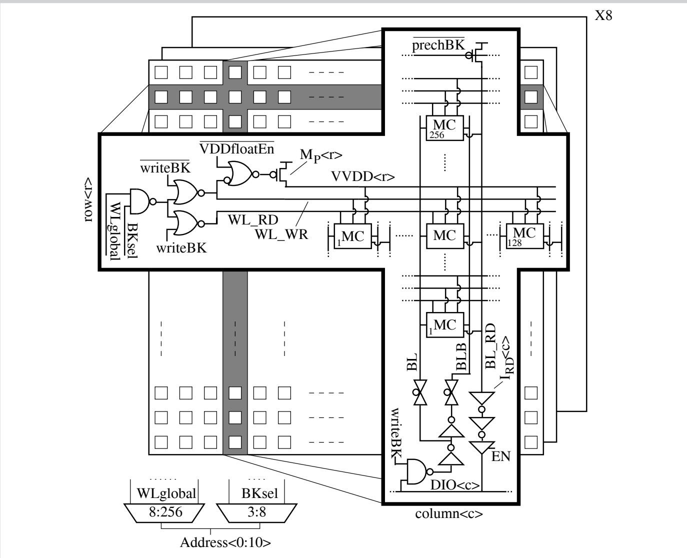

Figure 34.4.5: Architecture of the 256kb test chip.

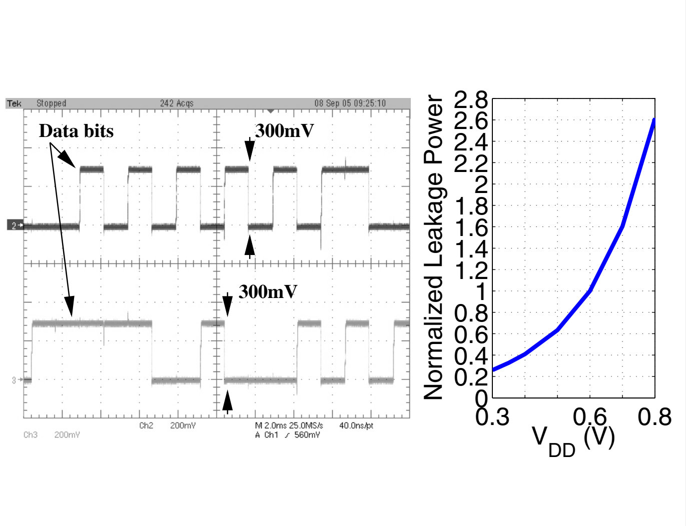

Figure 34.4.6: Chip functioned correctly to below 400mV. Scope plot shows 300mV operation; at this low voltage, some bit errors were observed.

Continued on Page 678

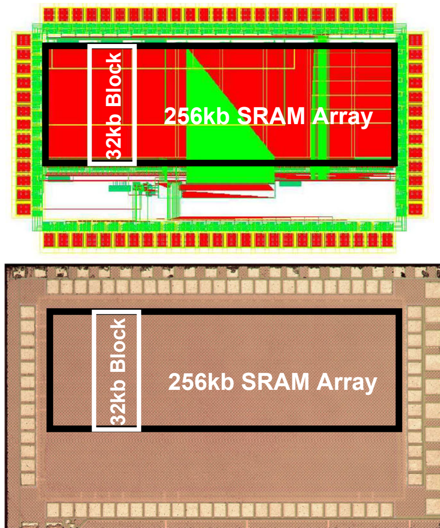

Figure 34.4.7: Annotated die micrograph and layout of 256kb sub-threshold SRAM in 65nm. Die size is 1.88mm × 1.12mm.

 1-4244-0079-1/06/$20.00 ©2006 IEEE