A 512kb 8T SRAM macro operating down to 0.57V with an AC-coupled sense amplifier and embedded data-retention-voltage sensor in 45nm SOI CMOS

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

As Published
http://dx.doi.org/10.1109/ISSCC.2010.5433818

Publisher
Institute of Electrical and Electronics Engineers

Version
Final published version

Accessed
Mon Dec 04 15:40:46 EST 2017

Citable Link
http://hdl.handle.net/1721.1/62194

Terms of Use
Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.

Detailed Terms
generally the minimum supply voltage is determined by the one memory cell out
For idle banks, lowering the supply to the DRV—measured to lie between
power scales down 9.4x over this range. The read path benefits from CMOS scal-
edge of CLK and the rising edge of ACLK (Fig. 19.5.2). Each measurement cor-
4 track on the memory cell pitch.
the conventional pull down NMOS. The RGBS scales linearly with distance, pre-
ing on device Mgp in bank1 and so on until the data traverses the 1.7mm glob-
ized 50mV above the tipping point of the ACSA through Cgd coupling and charge
voltage (DRV) sensor to determine the mismatch-limited minimum standby sup-
lenges related to area efficiency and process variation with three contributions:
There is a need for large embedded memory that operates over a wide range of
supplies of the DRV sensor cells are used to program and skew without access-
main memory.
Given a 1x measurement of Vt variation in memory cell devices, the DRV sen-
ker related to area efficiency and process variation with three contributions:
the conventional pull down NMOS. The RGBS scales linearly with distance, pre-
ing on device Mgp in bank1 and so on until the data traverses the 1.7mm glob-
ized 50mV above the tipping point of the ACSA through Cgd coupling and charge
voltage (DRV) sensor to determine the mismatch-limited minimum standby sup-
power scales down 9.4x over this range. The read path benefits from CMOS scal-
edge of CLK and the rising edge of ACLK (Fig. 19.5.2). Each measurement cor-
4 track on the memory cell pitch;
3) a data retention voltage (DRV) sensor to determine the mismatch-limited minimum standby sup-
voltage without corrupting the contents of the memory.

The ACSA is shown in Fig. 19.5.1. During precharge (q=0), device M5 diode-
connects the sensing PMOS M3, charging the input capacitor to the Vt of M3. Loading
from device M6-M8 and stacking from M4 establishes a steep trans-
characteristic near the tipping current level at the dynamic output Z. A thick
oxide MOS capacitor achieves a coupling ratio close to 1 while keeping the area
of the entire ACSA under 9um2. During sensing (q=1, RWL=1), node X is initial-
ized 50mV above the tipping point of the ACSA through Gp coupling and charge
injection from M5. When reading a 1, devices M2 and M1 discharge the local bit-
line, and after 100mV of signal development the output current quickly charges
node Z to VDD. Waveform plots B and C in Fig. 19.5.3 show amplification when
reading 1 and stability when reading 0.

The conventional domino read path [3], which replaces the ACSA by an optimal-
ized PMOS device of the same Vt, is 1.8x slower (Fig. 19.5.1). Monte Carlo simula-
tion of the 500 timing window reveals that additional variation of the dynamic
PMOS also restricts the minimum operating VDD to 0.8V; whereas, the ACSA works
down to 0.53V (FF corner, 85°C). Small-signal, single-ended sensing
with offset compensation has also been employed in [5] and [6]. By AC cou-
pling, this work avoids the unconditional full swing on the bitlines and the
requirement for interlock between wordline access and precharge found in [6].

By exploiting the sharp cutoff characteristics of a PMOS stack M3-M4 to sepa-
rate 1 from 0, this work avoids the 2 capacitors, 13 transistors, and power penal-
ty related to regenerative feedback in [5]. Hence, 128 parallel ACSAs can be laid out
on the bitline pitch, satisfying the non-interleaved column constraint of low-

When accessing bank0, device Mgp needs only to charge up the GBL input to the
neighbor bank1 above the Vt of NMOS Mx/My to trigger the next ACSA, turn-
ing on device Mgp in bank1 and so on until the data traverses the 1.7mm glob-
line, and after 100mV of signal development the output current quickly charges
node Z to VDD. Waveform plots B and C in Fig. 19.5.3 show amplification when
reading 1 and stability when reading 0.

After local sensing, data is forwarded across eight 64kb banks in the read cycle
via the regenerative global bitline scheme (RGBS) shown in Fig. 19.5.2. Two
ACSAs with shorted outputs are buffered to the global pull-up PMOS Mgp. When
accessing bank0, device Mgp needs only to charge up the GBL input to the neighbor
bank1 above the Vt of NMOS Mx/My to trigger the next ACSA, turning
on device Mgp in bank1 and so on until the data traverses the 1.7mm global
bitline (plot D Fig. 19.5.3). The simulated delay of the RGBS is 40% faster than
the conventional pull down NMOS. The RGBS scales linearly with distance, pre-
serving sharp transitions and performance comparable to a buffering daisy chain
[7]; yet the RGBS avoids cross-over currents and enables bidirectional signal-
—preventing routing congestion of the 128b word—through a single metal
4 track on the memory cell pitch.

The access times in Fig. 19.5.3 equal the minimum difference between the falling
edge of CLK and the rising edge of ACLK (Fig. 19.5.2). Each measurement corre-

For idle banks, lowering the supply to the DRV—measured to lie between
0.375V and 0.4V—reduces leakage power by 29x going from 1.2V to 0.4V.
Generally the minimum supply voltage is determined by the one memory cell out
of 524,288 with the largest DRV, resulting from local mismatch variation and its
functioning relation to the static noise margin (SNM) of the memory cells. This
relation changes with process corner, temperature, and end-of-life degradation.
Prior work [8,9] recognizes the importance of tracking the DRV but relies on
apriori characterization of the main memory. The primary challenge remains:
how to determine the DRV without corrupting the contents of the main

In this work, a separate column of 256 cells is employed in conjunction with sta-
tistical techniques as a DRV sensor to predict the 5σ failure rate of the main
memory (Fig. 19.5.4). The test chip contains 4096 DRV sensor cells but only one
segment of 256 is measured to demonstrate the area efficiency of the technique.
The layout of DRV sensor cells is identical to the main memory cells except for
one modification: the metal 2 wire supplying VDDR is split into two separate
VDDL and VDDR supply lines. Along with VSSL and VSSR, the four independent
supplies of the DRV sensor cells are used to program and skew without access-

Given a 1x measurement of Vt variation in memory cell devices, the DRV sen-
ker is skewed in multiple directions to reconstruct the SNM as a linear function
of Vt. The resulting coefficients are combined with αp to estimate failure. The
detailed algorithm in Fig. 19.5.5 searches the magnitude of three linear inde-
vendent voltage skews that collapse the nominal SNM, as observed by 50% fail-
ure in the 256 memory cells. The application of voltage skews emulates an effect-
ive shift in threshold voltage—through a transformation represented by T in the
matrix equation of Fig. 19.5.4—since the drain current depends on (VSD − Vt) and
the voltage skews directly add to or detract from VSD. By simplifying coeffi-
cients (assuming M5 and M6 do not influence retention and M3 and M2 have
no influence on the SNM), a system of three equations and three unknowns can
be solved. The measurement results in Fig. 19.5.5 show that the predicted
retention failure matches the actual retention failure in the main memory. A con-
servative estimate is also produced by replacing the failure criterion of 50% with
25% in the prediction algorithm, introducing a measured margin of 40mV. Thus,
for real-time embedded operation, the DRV sensor can permit aggressive reduc-
tion of retention standby voltage with negligible risk of losing data. This tech-
nique is relevant to state-of-the-art embedded SRAM that requires the retention
voltage as an input to standby power regulation circuits [10].

Figure 19.5.6 lists key features of the fabricated 8T SRAM, a viable candidate for
voltage scalable operation down to 0.57V. The die photo in Fig. 19.5.7 illustrates
the 2.0mm x 0.35mm chip floor plan containing 8 banks of 64kb along with on-
chip testability circuits.

Acknowledgements:
This work was funded in part by the FCPR Focus Center for Circuit & System
Solutions (C252), under contract 2003-CT-888. The authors thank Robert
Montoye for technical discussion and Keith Jenkins for test support.

References:
2006.
ance Challenges for Sub-70nm Microprocessor Circuits,” ESSCIRC Dig. Tech.
[3] L. Chang, et al., “A 5.3GHz 8T-SRAM with Operation Down to 0.41V in 65nm
Design in Sub-0.18um Technologies,” Dig. Symp. VLSI Circuits, pp. 226-227,
380-381, Feb., 2008.
variation Immune Sensing in High-performance Content-Addressable Memories,”
SRAM on 90nm CMOS Technology for Logic Application,” Dig. Symp. VLSI
Suppression by Minimizing Standby Supply Voltage,” ISQED Dig. Tech.
High-k Metal-Gate CMOS with Integrated Power Management,” ISSCC Dig. Tech.
Figure 19.5.1: The AC coupled sense amplifier supports a bitline of 256 cells.

Figure 19.5.2: The regenerative global bitline scheme reuses the ACSA.

Figure 19.5.3: Read access time measurement and simulated operational waveforms at 0.6V.

Figure 19.5.4: The DRV sensor with voltage skews illustrated.

Figure 19.5.5: DRV Sensor algorithm and measurement results from test chip.

Figure 19.5.6: Summary of chip characteristics and 64kb bank floor plan.
Figure 19.5.7: Die photo of the 512kb SRAM test chip in 45nm SOI CMOS along with a close-up snapshot of the bank layout. Global bit lines run 1.7mm long from left to right across 8 banks of 64kb each.