

A 45nm 0.5V 8T Column-Interleaved SRAM with on-Chip Reference Selection Loop for Sense-Amplifier

Mahmut E. Sinangil*, Naveen Verma, Anantha P. Chandrakasan
Massachusetts Institute of Technology, Cambridge, MA
E-mail*: sinangil@mit.edu

Abstract—8T bit-cells hold great promise for overcoming device variability in deeply scaled SRAMs and enabling aggressive voltage scaling for ultra-low-power. This paper presents an array architecture and circuits with minimal area overhead to allow column-interleaving while eliminating the half-select problem. This enables sense-amplifier sharing and soft-error immunity. A reference selection loop is designed and implemented in the column circuitry. By choosing one of the two reference voltages for each sense-amplifier in a pseudo-differential scheme, selection loop effectively reduces input offset. 8T test array fabricated in 45nm CMOS achieves functionality from 1.1V to below 0.5V. Test chip operates at 450MHz at 1.1V and 5.8MHz at 0.5V while consuming 12.9mW and 46 μ W respectively.

I. INTRODUCTION

On-chip caches are widely used in modern ICs due to their high transistor density and low activity factor. Consequently, SRAMs are becoming one of the decisive components of total chip area, power, performance and yield. Design of low-power and high-performance SRAMs with minimum area is a significant problem in today's deep-sub-micron technologies. Conventional 6T SRAMs fail to operate at low voltages due to bit-cell stability problems. 8T bit-cell is proposed as a low-voltage alternative [1]. However, if used with column-interleaving, 8T bit-cell suffers from *half-select* problem. During a write operation, un-accessed bits on the accessed row experience a condition that is equivalent to a read disturb on a 6T cell. Hence, 8T designs (e.g. [2], [3]) often prefer "one-word-per-row" architectures to avoid the half-select conditions. The work in [4] addresses this problem by applying a "read-then-write-back" scheme whereas the work in [5] proposes a 10T bit-cell that can be column-interleaved by vertical and horizontal word-lines (WL). Column-interleaved architectures are almost always preferred due to i) increased soft-error immunity and ii) better sense-amplifier (sense-amp) area utilization. First, in a column-interleaved architecture, bits of a word are spatially separated in the row and simple single-bit error correction coding (ECC) can be used to address soft-errors. Non-interleaved architectures, however, require more complex multi-bit ECC schemes which can be costly in terms of area and delay. Second, only one of many columns are active in a column-interleaved architecture which enables sense-amp sharing across multiple columns. In contrast, non-interleaved SRAMs require a sense-amp for each column to read all bits of a word at the same time.

Maximum attainable SRAM performance is limited by the weakest bit-cell and its ability to create a voltage differential on the bit-lines (BL) that is larger than the input offset of the worst sense-amp. In deep-sub-micron technologies, transistor mismatches are becoming more prominent. Hence, to keep its input offset at an acceptable level, sense-amp area cannot scale down as rapidly as the bit-cell area. This introduces a significant area problem with regards to the sense-amp and the problem is exacerbated in 8T designs without column-interleaving.

Fig. 1 demonstrates the trade-off between array efficiency (AE) and sense-amp offset with and without column-interleaving. For simplicity, AE is given by:

$$AE = \frac{\text{Cell Array Area}}{\text{Cell Array Area} + \text{Total Sense-Amp Area}}$$

For constant array size of 64x64, a higher array-efficiency demands a smaller sense-amp area and consequently a larger input offset voltage. Increasing column-interleaving (clmn-int) ratio, however, provides better array efficiency for the same offset voltage by amortizing sense-amp area over multiple columns.

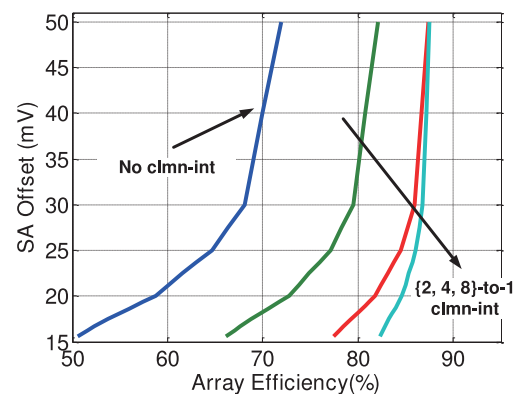


Fig. 1. For constant array size, larger array efficiency (AE) requires a smaller sense-amp area and, consequently, a larger input offset. Column-interleaving provides higher AE for an input offset by amortizing sense-amp area over multiple columns.

This work presents an array architecture and supporting circuits to allow column-interleaving for 8T bit-cells. The

architecture exploits the dynamic nature of half-select problem and uses very short internal-BLs that are shared by interleaved cells of the same row. This minimizes disturbances on half-selected bit-cells. To keep area overhead very low, extra transistors are embedded into the bit-cell array. Moreover, an area efficient sense-amp offset reduction technique is proposed to improve SRAM performance. An on-chip reference selection loop automatically selects one of the two reference voltages to compensate for the offset of each sense-amp. A 128kbit test array demonstrating these ideas is fabricated in a low-power 45nm CMOS process and achieves functionality from 1.1V down to 0.5V.

II. HALF-SELECT PROBLEM FOR 8T CELL

Static-noise-margin is the conventional way of analyzing stability of a bit-cell. However, it cannot capture the dynamic behaviour of the BLs and this provides pessimistic results. Recent works in [6], [7], [8] investigate the dynamic nature of read-stability and analyze it by a transient simulation with BLs represented by finite capacitances (C_{BL}). Since, BLs with a small capacitance discharge very quickly, bit-cell storage nodes experience minimal disturbance and stability can be maintained. Fig. 2 plots failure probability due to half-selection of an 8T cell at different voltage levels. Only the data denoted with “DC” is simulated with conventional simulation method whereas transient analysis is used to capture the dynamic nature due to different BL capacitances. A shorter BL translates into a smaller capacitance which results in an improvement of half-select problem.

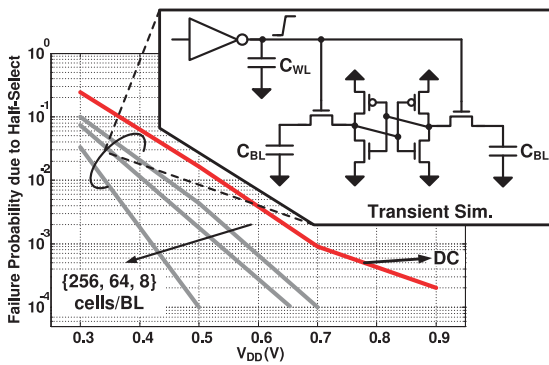


Fig. 2. Failure probability due to half-selection of an 8T cell. DC simulation provides a pessimistic result whereas transient simulation shows smaller BL capacitance improves half-select problem significantly.

Although using smaller number of bit-cells on a BL improves half-selection and enables operation at lower voltages, the area overhead of the column circuitry significantly reduces array efficiency. Previous publications address this problem by using hierarchical-BL architecture and simple local-sense circuits requiring full-swing inputs. The next section proposes a column-interleaved array architecture allowing very short internal-BLs (only 8bits/BL) but needing sensing circuitry over 256 rows which are shared among 4 columns.

III. COLUMN-INTERLEAVED ARRAY ARCHITECTURE FOR 8T BIT-CELL

Conventional bit-cell layout has an aspect ratio close to three i.e. its width is three times longer than its height. Consequently, efficient placement of additional transistors into bit-cell pitch in horizontal direction can be easier compared to vertical placement.

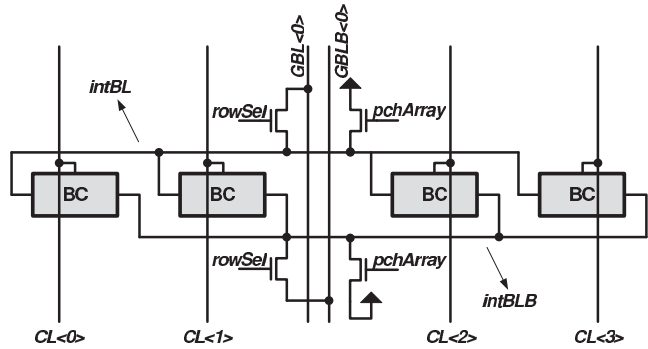


Fig. 3. Schematic illustration of the proposed architecture suitable for column-interleaving. BL/BLB-ports of four bit-cells are shared in horizontal direction and Column-Line (CL)s are routed in vertical direction. rowSel selects the active row.

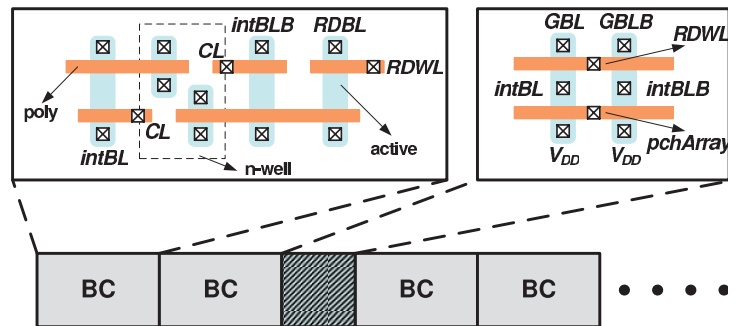


Fig. 4. In layout, additional NMOS transistors fit in the bit-cell pitch providing area efficient implementation. RDWL is used instead of rowSel which allows shorting of poly-layer between bit-cells and row-select NMOSs.

Fig. 3 shows a simple schematic illustration of a unit portion of the proposed array architecture. Four new transistors are inserted between four bit-cells. BL-ports of adjacent cells are shared horizontally (intBL/intBLB) and column-lines (CL) are routed vertically and connected to the access devices of the 8T bit-cells. Only the CLs of the active columns are asserted during a write access. rowSel signal selects active row and drives internal-bit-lines (intBL/intBLB) to global-bit-lines (GBL/GBLB) through NMOS pass-gates. Although these pass-gates cause a degradation on logic levels, bit-cells are sized to ensure write-ability under this condition. For un-selected rows (i.e. rowSel=’0’), intBL/intBLBs are isolated from GBL/GBLBs. Therefore, an un-selected bit-cell sees a small intBL/intBLB capacitance and resulting half-select-disturbance on this cell is small. At the end of every write-cycle, pchArray is asserted for a short period of time

to pre-charge $intBL/intBLB$. This is necessary because the previous state of $intBL/intBLB$ can cause an elevated level of disturbance on the half-selected bit-cells. Finally, a read-access is done through the read-buffer of the 8T cell in the conventional way.

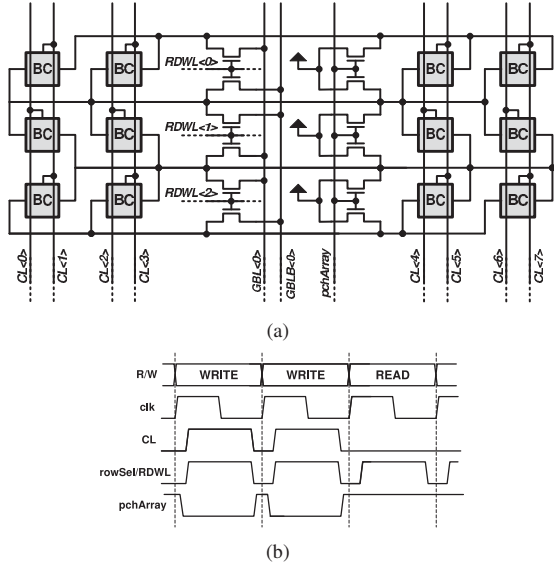


Fig. 5. (a) Schematic of three rows and four columns of bit-cells in proposed architecture and (b) waveforms for important signals during read and write accesses. $RDWL$ is used for row-select during write operation and $pchArray$ signal is asserted at the end of each write cycle.

Layout realization of the schematic illustration in Fig. 3 is shown in Fig. 4. The following steps are taken for efficient layout implementation.

- 1) Devices connected to $pchArray$ are chosen to be NMOS to be able to stack them with row-select devices.
- 2) $RDWL$ for each row is used for $rowSel$ to prevent extra metal routing. During a write-access, $RDWL$ is also asserted to select the active row.
- 3) $intBL$ and $intBLB$ are shared between adjacent rows from above or from below. Hence, effectively eight bit-cells share the same local-bitline. To prevent two bit-cells driving same internal-bitline, two separate CL signals are routed for each column and they are connected to bit-cells on alternating rows ($CL < 0 >$ and $CL < 1 >$) in Fig. 5.

Area overhead of this layout implementation compared to a conventional 8T array with non-interleaved architecture is 12%. This is due to i) four additional NMOS devices and ii) non-overlapping CL contacts between adjacent columns of bit-cells. In layout, minimum sizes allowed by core design rules are used.

Complete schematic of three rows and four columns of the proposed architecture and waveforms of critical signals are shown in Fig. 5. $RDBLs$ (not shown in figure for simplicity) are kept low and only pre-charged to V_{DD} at the beginning of a read cycle to prevent unnecessary $RDBL$ discharge during

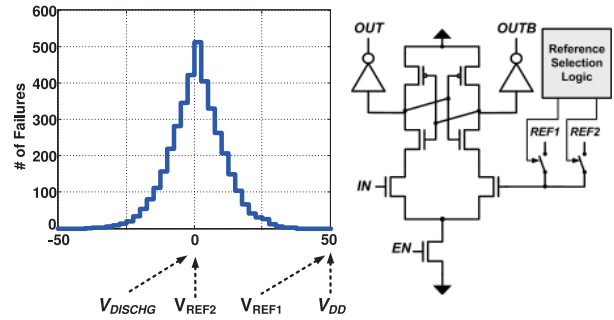


Fig. 6. Sense-amp offset distribution and two-reference voltage scheme. Reference levels can be chosen to reduce the offsets of the sense-amps.

write accesses. $pchArray$ is pulled to high at the end of a write cycle and kept high during a read access.

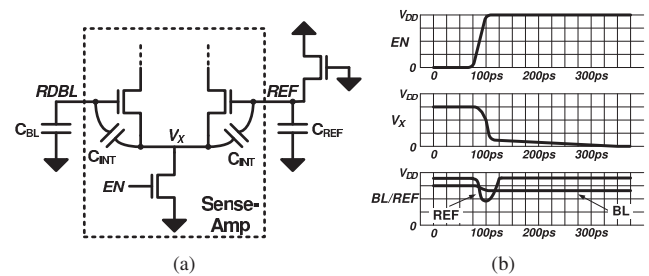


Fig. 7. Effect of coupling to BL and REF nodes with different capacitive divider ratios. Different voltage coupling alter sense-amp inputs and negate the effect of offset compensation.

IV. REFERENCE SELECTION LOOP FOR SENSE-AMPLIFIER

Sense-amp input offset is a critical metric directly impacting the performance of an SRAM. Because of the single-ended $RDBL$, 8T designs often use pseudo-differential sense-amps where one of the inputs is connected to $RDBL$ while the other one is connected to a reference voltage. Fig. 6 shows input offset of a widely-used sense-amp. Reference voltage (V_{REF}) should be selected such that:

- 1) A sense-amp with a large negative offset can output ‘1’ when $RDBL=V_{DD}$ (neglecting leakage from $RDBL$) and
- 2) A sense-amp with a large positive offset can output ‘0’ if $RDBL$ is sufficiently discharged.

For a worst-case input offset of $\pm 50mV$ as shown in Fig. 6, V_{REF} can be placed $50mV$ below V_{DD} . As a result, to sense a logic low, $RDBLs$ should be discharged by at least $V_{DISCH}=100mV$. Alternatively, if two reference voltages are available (V_{REF1} , V_{REF2}), these voltage levels can be selected to compensate for the offset of each sense-amp. Specifically, a sense-amp with a negative offset can be assigned a higher reference and a sense-amp with a positive offset can be assigned a lower reference voltage. This lowers V_{DISCH} and hence improves SRAM performance significantly at low voltages where the ratio of worst-case cell current to nominal cell current is very large due to variation. The work in [9]

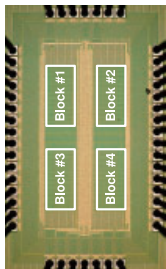
proposes a similar approach with 16 reference voltage levels for a specific SRAM architecture.

In this work, column-circuit is designed with a simple reference selection logic and two reference voltages. To minimize area overhead, only a latch and a few logic gates are used. At the startup, selection loop is triggered by a series of off-chip signals and each sense-amp is tested with $RDBL=V_{DD}$ and $V_{REF}=V_{REFH}$. If the output of the sense-amp is correct ('1'), then V_{REFH} is selected. Otherwise, V_{REFL} is assigned to the sense-amp. The result of the selection loop is stored in a data latch in column-circuit.

Robust operation of the sense-amp and the offset reduction through reference voltage selection relies on stable, low-noise reference levels. In single-reference scheme, large de-coupling capacitors can be placed and their area can be amortized over the entire array since every sense-amp is using the same reference voltage. However, for our scheme, each sense-amp is connected to its assigned reference voltage through a PMOS switch. Fig. 7 plots the effect of signal coupling on sense-amp inputs. After the assertion of EN , node V_x is rapidly pulled down which is coupled to both inputs through capacitor dividers. If divider ratios are significantly different, BL and REF voltages are altered at the beginning of sensing. Although REF is actively driven by a PMOS, the response of the PMOS to this coupling might not be adequate especially at the early stage of sensing. To address this problem, in this work, $GBLB$ for every column is connected to REF during read-accesses. $GBLB$ is designed to have a capacitance very close to C_{BL} so the amount of coupling to both inputs is almost the same.

V. MEASUREMENT RESULTS

A column-interleaved architecture suitable for 8T bit-cell and reference selection loop for sense-amplifier ideas are implemented in a 128kbit 45nm SRAM test chip. Die photo and a summary table are shown in Fig. 8. The test chip achieves functionality down to 0.6V with no bit-errors and down to 0.5V with 2×10^{-4} bit-error-rate. Fig. 9 shows measured performance of the test array. When operating at 1.1V, test chip achieves 450MHz read and write functionality. Reference selection loop improves performance by around 10%. Leakage power scales down from $179\mu W$ to $10.7\mu W$ on 1.1-0.5V range. Below 0.5V, write-ability problems begin to emerge and increase bit-error-rate to 10^{-3} at 0.4V.



Process	45nm CMOS
Die Size	0.9mm x 1.5mm
Capacity	128kbit (256 Rows x 128 Clms x 4Blks)
Column-Interleaving	4-to-1
Voltage Range	0.5V – 1.1V
Performance	5.8MHz – 450MHz
Active Power	46 μ W – 12.9mW
Leakage Power	10.7 μ W – 179 μ W

Fig. 8. Die photo and summary of measurement results for 128kbit SRAM test chip fabricated in 45nm CMOS process.

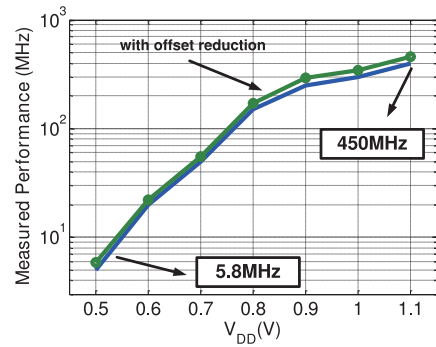


Fig. 9. Measured performance vs. V_{DD} with and without offset reduction. SRAM performance scales from 450MHz down to 5.8MHz over the voltage range.

VI. CONCLUSION

An 8T SRAM fabricated in 45nm CMOS process is operating from 1.1V to below 0.5V. Proposed array architecture addresses half-select problem by decoupling large BL capacitance from half-selected cells. Reference selection loop enables offset reduction by intelligently choosing a reference voltage for each sense-amp depending on offset voltage. At 0.5V, test chip operates at 5.8MHz with $46\mu W$ active power consumption.

ACKNOWLEDGMENT

This work is funded by DARPA and chip fabrication is provided by Texas Instruments.

REFERENCES

- [1] L. Chang and et al., "Stable SRAM Cell Design for the 32nm Node and Beyond," in *Symposium on VLSI Circuits (VLSI) Digest of Technical Papers*, June 2005, pp. 128–129.
- [2] R. Joshi and et al., "6.6+ GHz Low Vmin, read and half select disturb-free 1.2Mb SRAM," in *Symposium on VLSI Circuits (VLSI) Digest of Technical Papers*, June 2007, pp. 250–251.
- [3] N. Verma and A. Chandrakasan, "A 65nm 8t sub-vt sram employing sense-amplifier redundancy," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2006, pp. 328–329.
- [4] Y. Morita and et al., "An Area-Conscious Low-Voltage-Oriented 8T-SRAM Design under DVS Environment," in *Symposium on VLSI Circuits (VLSI) Digest of Technical Papers*, June 2007, pp. 256–257.
- [5] I. J. Chang, J. Kim, S. P. Park, and K. Roy, "A 32kb 10T Subthreshold SRAM Array with Bit-interleaving and Differential Read Scheme in 90nm CMOS," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2008, pp. 388–389.
- [6] A. Kawasumi and et al., "A Single-Power-Supply 0.7V 1GHz 45nm SRAM with An Asymmetrical Unit- β -ratio Memory Cell," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2008, pp. 382–383.
- [7] K. Kushida and et al., "A 0.7V single-supply SRAM with 0.495 μ m² cell in 65nm technology utilizing self-write-back sense amplifier and cascaded bit line scheme," in *Symposium on VLSI Circuits (VLSI) Digest of Technical Papers*, June 2008, pp. 46–47.
- [8] M. Yamaoka, K. Osada, and T. Kawahara, "A cell-activation-time controlled SRAM for low-voltage operation in DVFS SoCs using dynamic stability analysis," in *IEEE European Solid-State Circuits Conference (ESSCIRC) Digest of Technical Papers*, Sept. 2008, pp. 286–289.
- [9] S. Coemans, W. Dehaene, and F. Cathoor, "A 3.6pJ/access 480MHz, 128Kbit on-chip SRAM with 850MHz boost mode in 90nm CMOS with tunable sense amplifiers to cope with variability," in *IEEE European Solid-State Circuits Conference (ESSCIRC) Digest of Technical Papers*, Sept. 2008, pp. 278–281.