5.6 Mb/mm^2 1R1W 8T SRAM Arrays Operating Down to 560 mV Utilizing Small-Signal Sensing With Charge Shared Bitline and Asymmetric Sense Amplifier in 14 nm FinFET CMOS Technology

Jaydeep P. Kulkarni, *Senior Member, IEEE*, John Keane, Kyung-Hoae Koo, Satyanand Nalam, *Member, IEEE*, Zheng Guo, *Member, IEEE*, Eric Karl, *Member, IEEE*, and Kevin Zhang, *Fellow, IEEE*

***Abstract*— Multiported high-performance on-die memories occupy significantly more die area than a comparable single-port memory. Among various multiport memory topologies, the 1-read (R), 1-write (W) 8-transistor (T) Static Random Access Memory (SRAM) with a decoupled read port allows separate optimization of the read and write ports when organized without interleaved logical columns. This enables a lower minimum operating voltage (V_{min}) compared with other dual-port SRAMs that require ports optimized for read stability and write operations. However, the 1R1W 8T SRAM often employs large signal, hierarchical bitline sensing to achieve high performance due to the nondifferential read bitline. This large-signal read architecture necessitates frequently placed local bitline sensing circuits, degrading the array bit density. In this paper, we present two sense amplifier (SA) techniques for small-signal pseudo-differential sensing to facilitate 256 bits per bitline achieving an 8T SRAM array density of 5.6 Mb/mm^2 in 14 nm FinFET CMOS. The first design employs a charge sharing SA scheme to generate a reference voltage (V_{REF}) by leveraging the capacitance of otherwise unused metal tracks over the bitcell column. The second design utilizes an asymmetric SA in which the read bitline precharged to V_{CC} in the unselected sector acts as a reference voltage and the active bitline side is intentionally upsized to skew the SA. High volume measurement results demonstrate 560 mV V_{min} at 400 MHz/−10 °C and reaches 2.21 GHz at 1 V supply.**

***Index Terms*— 1R (read) 1W (write) 8T (transistor) static random access memory (SRAM), asymmetric sense amplifier (ASA), charge share sense amplifier (CSSA), dual-port SRAMs, large signal sensing, single-ended sensing, small-signal sensing.**

I. Introduction

SYSTEM-ON-CHIP designs contain a variety of IP blocks such as media, graphics processing units, and compute cores that use multiport memories to improve the performance and/or bandwidth by enabling multiple simultaneous operations in the same memory bank. Multiport memories require significantly more die area compared with the equivalent single-ported memories due to larger bitcells and multiple wordline decoder and I/O sensing circuits. Fig. 1 shows the dual-port 8-transistor (T) static random access memory (SRAM) bitcell supporting two read and/or write (2RW) operations by including an additional pair of access transistors driven from a second wordline (WL2) to the 6T SRAM bitcell. Additional bitline pair (BL2 and BL2_B) enables a differential sensing mechanism similar to the 6T SRAM bitcell. However, each port needs to be optimized for both read and write operations. It also suffers from simultaneous read and write disturb events when both wordlines (WL1 and WL2) in a row are activated at the same time. This translates into a higher minimum operating voltage (called V_{min}) compared with other dual-port bitcell alternatives. The 1-read and 1-write (1R1W) dual-port 8T SRAM with a decoupled read port eliminates the read disturb scenario by preventing charge sharing with internal storage nodes when the read wordline (RDWL) is activated (Fig. 1). Dummy-read disturb can also be prevented in 1R1W 8T arrays using a noninterleaved column design. Furthermore, a decoupled read port enables separate read and write port optimization to achieve lower V_{min} compared with 2RW 8T SRAMs. The high-performance single-ended read mechanism in 1R1W 8T SRAM is realized using a full swing hierarchical bitline design as shown in Fig. 2. The hierarchical bitline scheme adopts short local bitlines (LBLs) that minimize the capacitive

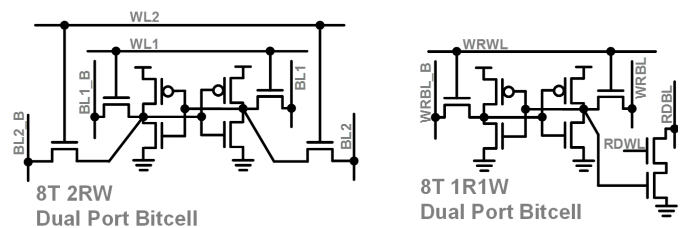


Fig. 1. Dual-port SRAMs: 2RW port bitcell coupled read operation and 1R1W port bitcell with decoupled read and write operations.

Manuscript received April 26, 2016; revised June 28, 2016; accepted August 23, 2016. Date of publication September 29, 2016; date of current version January 4, 2017. This paper was approved by Guest Editor Atsushi Kawasumi.

J. P. Kulkarni is with Circuit Research Lab, Intel Corporation, Hillsboro, OR 97124 USA (e-mail: jaydeep.p.kulkarni@intel.com).

J. Keane, K.-H. Koo, S. Nalam, Z. Guo, E. Karl, and K. Zhang are with the Advanced Design Group, Logic Technology Development, Intel Corporation, Hillsboro, OR 97124 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2016.2607219

0018-9200 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

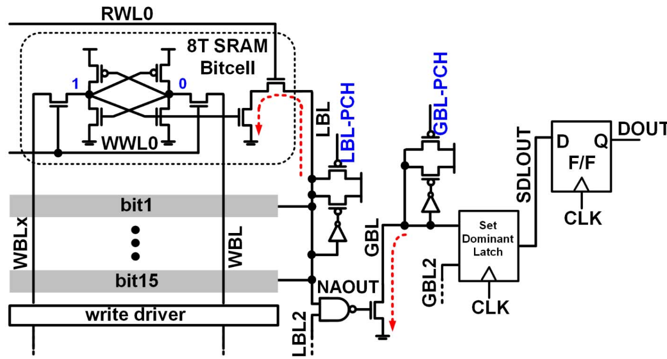


Fig. 2. Conventional high-performance 1R1W 8T SRAM array organization with hierarchical, large signal sensing domino bitline read path with 16 b per LBL.

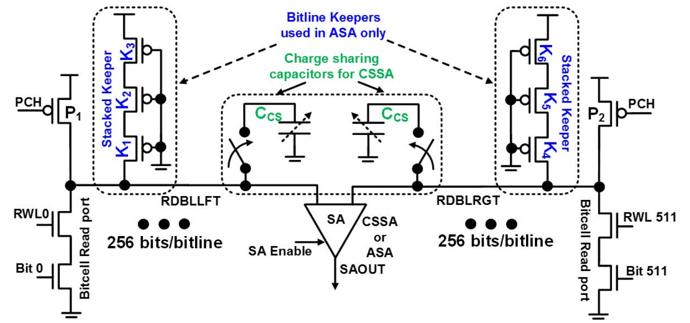


Fig. 3. Small-signal sensing read path for 1R1W SRAM array utilizing the proposed CSSA or ASA designs supporting 256 b per bitline.

load discharged during read operations as well as the noise impact due to leakage from inactive read ports. Although full swing, hierarchical local and global bitline design allows higher performance at the cost of multiple LBL sensing circuits degrading the array density.

To address this bit-density degradation, various single-ended small-signal sensing [1] and pseudo-differential sensing techniques [2], [3] have been proposed for 1R1W 8T arrays and read-only memories to improve the array density by supporting a higher number of bits per read bitline (RDBL) while achieving high performance. The ac coupled sense amplifier (SA) in [1] relies on precharging the input of the SA PMOS to a threshold voltage (V_t) drop below the V_{CC}. Any change on the bitline is capacitively coupled to this PMOS, thereby rapidly modulating its gate overdrive and resolving the RDBL transition. Biasing the sensing PMOS closer toward its V_t improves the read-1 performance but degrades the noise immunity during read-0 operation. This increases the sensitivity of the read operation to coupling noise events and variations in the sense of PMOS Vt and the coupling capacitor.

The small-signal sensing scheme in [2] utilizes bitline charge sharing mechanism using two equal-sized additional capacitors (C_1 and C_2) and a three-phase read operation. In the first phase, C_1 and C_{bitline} are precharged to a common shared voltage (V_{CS}). In the second phase, C_1 and C_2 share charge to develop a reference voltage that is half the common shared voltage ($V_{\mathrm{REF}} = V_{\mathrm{CS}}/2$). In the third phase, this V_{REF} is compared with the RBDL voltage (precharged to V_{CS}) to resolve the bitcell data. Three-phase charge sharing mechanism is difficult to implement across a wide frequency range in a two-phase 50% duty cycle precharge/evaluate domino logic used predominantly in 1R1W 8T SRAMs. The V_{REF} generation circuit also requires careful matching of two additional capacitors (C_1 and C_2) to generate a precise V_{REF} of $V_{\mathrm{CS}}/2$.

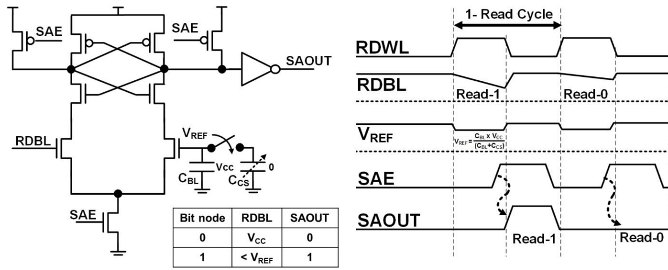
The differential SA in [3] implements a single-ended, small-signal sensing scheme by generating V_{REF} by discharging the unselected bitline with half-sized reference read port (having half the read current compared with the bitcell read current). The V_{REF} generation circuit needs to be optimized carefully to make sure the reference bitline discharge rate is not affected by the random variations of reference read port transistors. In addition, the V_{REF} is generated in every read operation, increasing the switched dynamic capacitance (C_{DYN}) due to unselected bitline discharge. Low swing SA design in [4] utilizes unaccessed bitline to generate a V_{REF} by charge sharing with a dedicated predischarged capacitor. The charge sharing capacitor needs to be of a similar magnitude as the bitline capacitance and incurs area overhead.

In this paper, we present two SA techniques for realizing small-signal pseudo-differential sensing for high-density 1R1W 8T SRAM arrays without using any active dedicated capacitors compared with the earlier proposed V_{REF} generation schemes and without using any V_{REF} (for the second SA) [5]. Fig. 3 shows the read path of the proposed SA designs supporting 256 b/bitline and targeted for improved array density. Single SA across 512 bits eliminates the need for hierarchical bitlines and the corresponding local sensing circuits. The proposed charge share sense amplifier (CSSA) technique utilizes the charge share capacitor (C_{CS}) implemented in metal lines above the bitcell column to generate the V_{REF}. It does not implement bitline keepers that are otherwise used to improve the noise immunity of the domino bitline design. To improve the bitline noise immunity, an asymmetric sense amplifier (ASA) design supporting bitline keepers is proposed. It uses unselected bitline precharged to V_{CC} as the reference voltage and intentionally upsizes the SA side connected to the read bitline.

This paper is organized as follows. Section II describes the proposed CSSA design concept, detailed circuit implementation and simulation results including the effect of bitline keepers, and the statistical $V_{\min}$ analysis. Section III introduces the proposed ASA design concept, detailed circuit implementation and simulation results with statistical $V_{\min}$ analysis showing the sensitivity of the SA asymmetry, and the bitline keeper sizing. Section IV presents the measurement results from a 14 nm FinFET CMOS test-chip and compares this work with the recently proposed dual-port SRAMs [5]–[9]. Section V concludes the paper by highlighting the key benefits and tradeoffs of the proposed SA techniques.

II. Charge Share Sense Amplifier Design

Fig. 4 shows the concept of the proposed CSSA design. The RDBL voltage is compared with the internally generated



Bit node	RDBL	SAOUT
0	V_{CC}	0
1	< V_{REF}	1

Fig. 4. CSSA design concept: active bitline (RDBL) is connected to one of the inputs of a symmetric SA; the unselected precharged bitline shares its charge with a predischarged capacitor (C_{CS}) to generate the reference voltage (V_{REF}).

V_{REF} using a symmetric SA. The V_{REF} is generated in every read cycle using the charge sharing mechanism between the unselected RDBL capacitance (C_{BL}), which is precharged to V_{CC} and a charge share capacitor (C_{CS}) that is predischarged to V_{SS}. In the designs presented in this paper, the active RDBL is discharged during read-1 operation and is maintained high during read-0 operation. When the selected bitcell contains a '0,' the active RDBL voltage is higher than the V_{REF} and the SA output (SAOUT) is resolved as '0' when the SA enable (SAE) signal is asserted. If the selected bitcell contains a '1,' the active bitline (RDBL) is discharged sufficiently below the V_{REF} and the SAOUT is resolved as '1' when the SAE is asserted.

Fig. 5 shows the detailed circuit implementation and the read timing diagram for the proposed CSSA design. Transistors N1-N6 and P1-P2 form the core of the CSSA. Three capacitors CAPLFT [2:0] and CAPRGT [2:0] are used for sharing the charge on the unselected precharged RDBL to generate the V_{REF}. They are opportunistically implemented as metal capacitors routed on each side of SA to generate seven reference levels (R1-R7) without increasing the silicon area. The simulation waveforms in Fig. 5 illustrate two-phase read operations on the active sector of 256 rows, which in this case is on the right side, so the reference is created on the left. Charge share capacitors CAPLFT/CAPRGT [2:0] are discharged before the read operation using transistors N8 [2:0] and N10 [2:0] controlled by PREDISLFT [2:0] and PREDISRGT [2:0] signals. These control signals are selectively set low depending on the CSSHARE [2:0] bits (tuned for optimum R1-R7 V_{REF} generation) as well as the active sector enable (determined by LFT_EN or RGT_EN signals). During a read operation, when read wordline RDWL[n] is asserted in the right sector, a combination of CSHLFT_B [2:0] based on CSSHARE [2:0] dc control bits is also set low to enable charge sharing through N7 [2:0] PMOS devices and to pull charge off of left bitline (RDBLLFT) and create a V_{REF} in between the high and low levels of active right bitline RDBLRGT. CSHRGT_B [2:0] control signals are maintained high to avoid charge sharing on the active bitline. When the SA is enabled by asserting the SAENABLE signal, the SA output nodes resolve depending on the bitline differential.

RDBL keepers are used in the conventional large signal arrays (LSAs) to hold the domino RDBLs high for improved noise immunity. Such bitline keepers cannot be directly used for the proposed CSSA design as they would affect the V_{REF} level set by the charge share ratio. Fig. 6 shows the bitline waveforms with and without RDBL keepers for low-frequency (200 MHz) read-0 operation. Without bitline keepers, both active and reference RDBLs discharge based on the data-dependent read port leakage. This may result in narrowing the bitline differential for wider wordline pulse widths. This low-frequency V_{REF} degradation can be mitigated by tuning the RDWL to SA enable delay and thereby limiting the delay between the end of precharge and the start of SA evaluation. In the presence of RDBL keepers, the noise immunity of the active bitline improves by lowering the droop on active RDBL. However, it would charge the V_{REF} node toward V_{CC} resulting in a frequency-dependent V_{REF} increase. Thus V_{REF} would vary as a function of wordline pulse width and would develop negative bitline differential resulting in wrong SAOUT evaluation.

Fig. 7 shows 6-sigma statistical simulation results for read-0 and read-1 V_{min} (operating at 400 MHz, worst case data conditions, process and temperature skew for each read data scenario) as a function of V_{REF} (expressed as a fraction of the supply voltage). Reed-1 V_{min}, which captures the read bitline evaluation delay improves with increasing V_{REF} as the active RDBL does not need to discharge significantly for the correct SA operation. On the other hand, read-0 V_{min}, which captures the functional failures due to noise events degrades with increasing V_{REF} as both RDBLs may droop differently, reducing the bitline differential smaller than the required SA offset voltage.

III. Asymmetric Sense Amplifier Design

As seen in the previous section, the CSSA technique cannot implement the RDBL keepers as it would raise the V_{REF} with time and would result in a time-dependent read-0 operation and degraded noise immunity. The active RDBL, when floated, would be subject to various noise events such as charge sharing noise, the data-dependent leakage of the unselected bitcells, ground bounce, and the capacitive coupling noise due to neighboring switching metal lines. Another sensing approach, enabling RDBL keepers on for improved noise immunity, is devised. Fig. 8 shows the conceptual design of the proposed decoupled input ASA for small-signal single-ended sensing without requiring any dedicated V_{REF}. In this case, the unselected RDBL precharged to V_{CC} acts as V_{REF}. The SA side connected to the active RDBL node is intentionally upsized by creating the asymmetry in the SA pull-down path. During the read operation, when the corresponding RDWL is asserted, the active RDBL is not discharged if the data stored in bitcell are '0.' The RDBL voltage is maintained by the bitline keepers. For the ASA, since both pull-down inputs are at the same voltage (V_{CC}), the upsized SA input side connected to the active RDBL evaluates the SAOUT to '0' once the sense amplifier enable (SAE) is asserted. On the other hand, if the data stored in the bitcell are '1,' the active RDBL starts to discharge when the corresponding RDWL is asserted. If the active RDBL is discharged sufficiently low ($V_{CC}-\Delta$), the effective strength of the upsized SA side

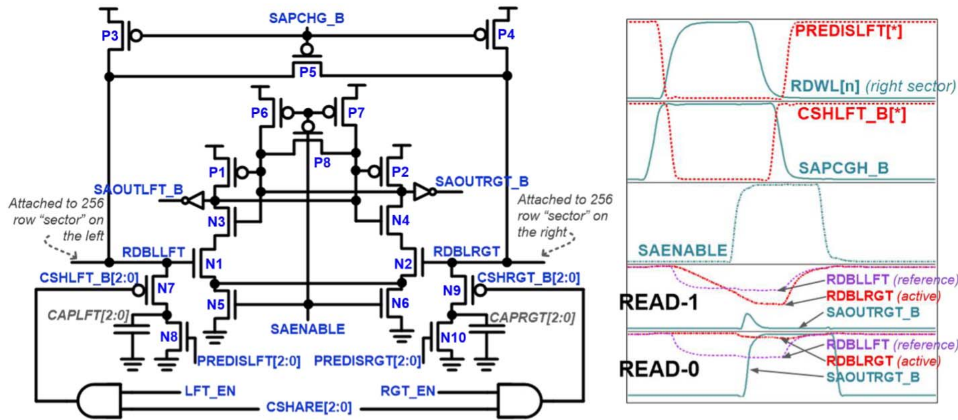


Fig. 5. CSSA circuit schematics and read operation simulation results: CSHARE [2:0] determines the amount of charge sharing for generating the reference voltage depending on the sector select signal (LFT_EN or RGT_EN).

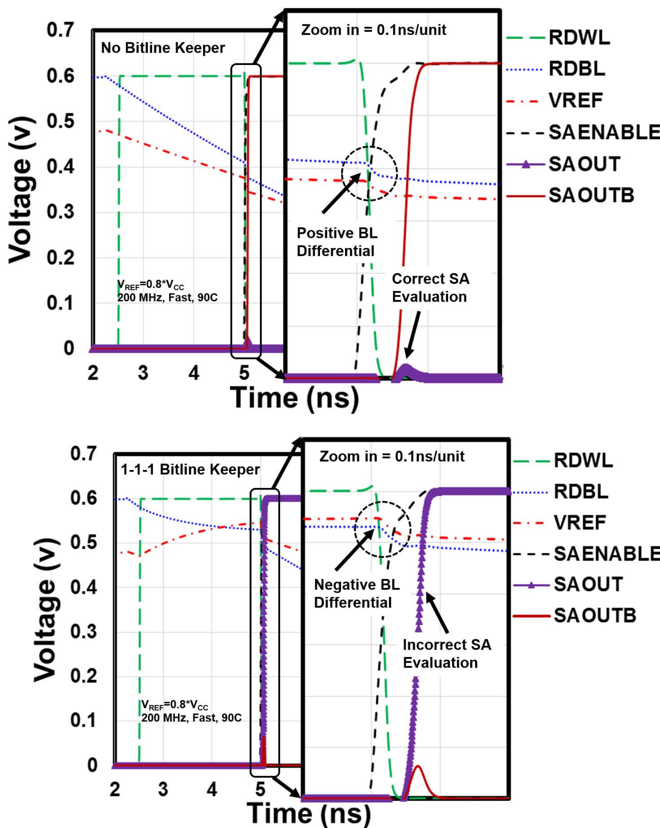


Fig. 6. CSSA read-0 operation with and without bitline keepers at low frequency, high temperature, and fast process corner: without bitline keepers, the bitline differential reduces, yet SA output is resolved correctly. In the presence of bitline keeper, V_{REF} continues to charge toward V_{CC} resulting in negative bitline differential and incorrect SA output.

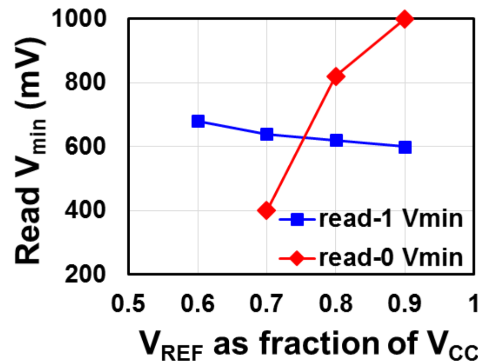


Fig. 7. CSSA without bitline keeper V_{min} analysis at iso-frequency (400 MHz) using 6-sigma statistical simulations: read-0 V_{min} is limited by the noise due to unselected bitcell read port leakage, coupling noise, and V_{REF} fluctuations.

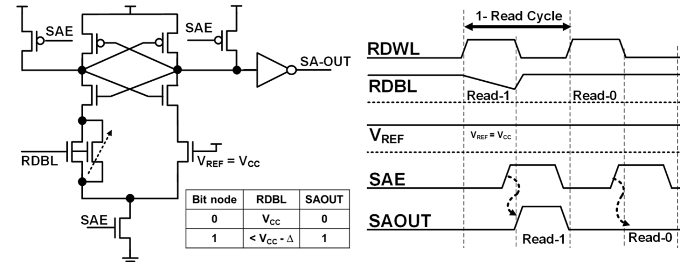


Fig. 8. ASA design concept: active bitline (RDBL) is connected to upsized input; the nominally sized unselected precharged bitline (at V_{CC}) acts as a reference. For read-0 condition, with RDBL at $V_{CC,}$ RDBL side resolves the SAOUT node to 0. For read-1 condition, with RDBL developed sufficient bitline differential (Δ), the unselected reference bitline (at V_{CC}) resolves SAOUT node to '1.'

connected to the active RDBL is weaker than the nominally sized V_{REF} side precharged to V_{CC}. Once the SAE is triggered, the SA evaluates in the other direction and resolves SAOUT as '1.'

Fig. 9 shows a topology for implementing the ASA design and simulation waveforms for read operation. Transistors N1-N6 and P1-P2 form the core of the ASA. Transistors N7, N8, and N9 form one pull-down path that is driven either by the left bitline (RDBLLFT) or V_{SS} depending on the SKEWLEFT [1:0] settings. When the left sector is accessed, this additional SA pull-down branch is driven by the RDBLLFT and skews the SA by upsizing the left side. Similarly, transistors N10, N11, and N12 form another additional SA pull-down branch that is driven either by the right bitline (RDBLRGT) or V_{SS} depending on the SKEWRGT [1:0] setting. When the right sector is accessed, this branch is driven by the RDBLRGT and skews the SA by upsizing the right side. Note that the ASA layout is physically symmetric but electrically asymmetric by dynamically skewing left or right, depending on the selected sector.

As shown in the simulation waveforms in Fig. 9 during the precharge phase SAPCHG_B is asserted low and both the

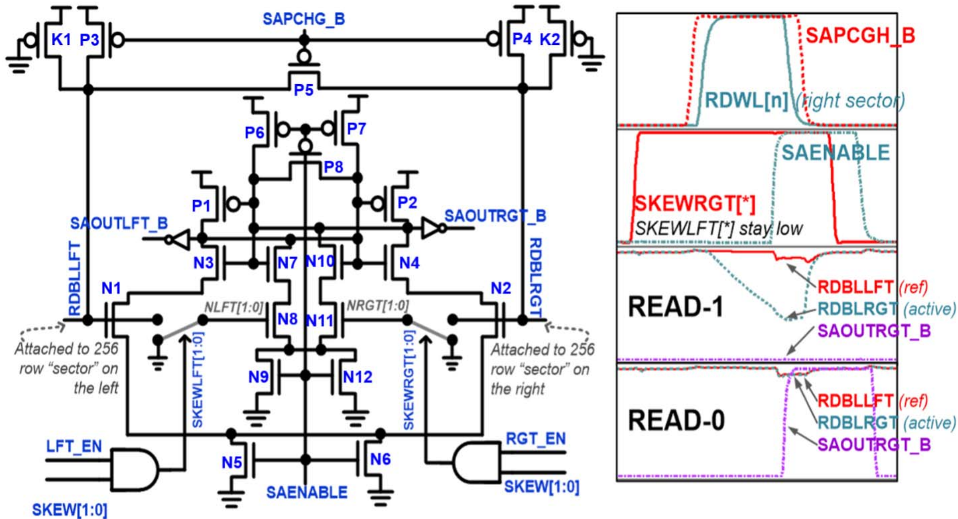


Fig. 9. ASA circuit schematics and read operation simulation results: SKEW [1:0] signals determine the SA skew depending on the sector enable (LFT_EN or RGT_EN) signal.

RDBLs are precharged to V_{CC} through P3 and P4 devices. Transistor P5 equalizes the two bitlines and mitigates any bitline precharge differential. Similarly, transistors P6, P7, and P8 driven by SAENABLE precharge and equalize the internal SA nodes. In this simulation, the right sector is accessed during a read operation. The sector select signal RGT_EN is asserted while LFT_EN stays low. Only one pull-down branch (N10-N11-N12 or N7-N8-N9) is shown in Fig. 9 for illustrative purposes. The complete design involves two pull-down branches on each side of SA for skew selection. SKEW [1:0] are dc control signals that enable either or both pull-down branches of either side to realize three skew settings (S1-S3). In this case, SKEWRGT [1:0] signals connect the NRGT [1:0] nodes to the RDBLRGT and skew the right side of the SA. SKEWLFT [1:0] signals connect NLFT [1:0] nodes to '0' disabling the left side N7-N8-N9 pull-down branches. Note that the additional tail current due to N9 and N12 is steered to one of the sides depending on the selected sector. The precharged inactive RDBLLFT is held at V_{CC} by a PMOS keeper K1 and acts as the reference bitline. During a read-0 operation, the active RDBLRGT and reference RDBLLFT bitline voltage levels are equivalent, but the SA pull-down on the right (active) side is skewed up so SAOUTRGT_B is pulled high. During a read-1 operation, the active RDBLRGT bitline must be sufficiently discharged to compensate for the upsized SA pull-down path. Thus read-1 involves contention between the upsized SA pull-down path with a lower gate drive on the active bitline side, and the nominally sized SA pull-down path with a larger gate drive (V_{CC}) on the reference side.

Fig. 10 shows the 6-sigma 14 nm statistical simulation results (operating at 400 MHz, worst case data conditions, temperature, and process skew for each read data scenario) characterizing the ASA for various amounts of mismatch and bitline keeper sizing. Three stacked bitline keepers of size 1-1-1 fin or 2-2-2 fins are used to improve the noise immunity of the bitlines against leakage and coupling events. Read-0 V_{min}, which captures the functional failures due to noise events improves with increasing SA pull-down

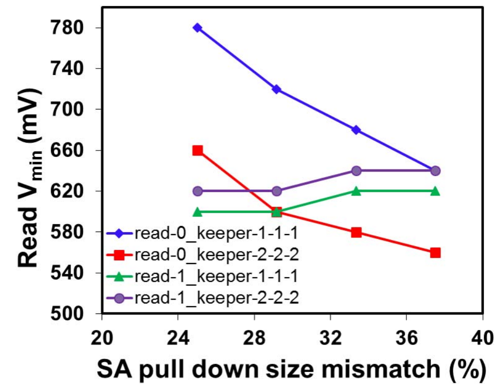
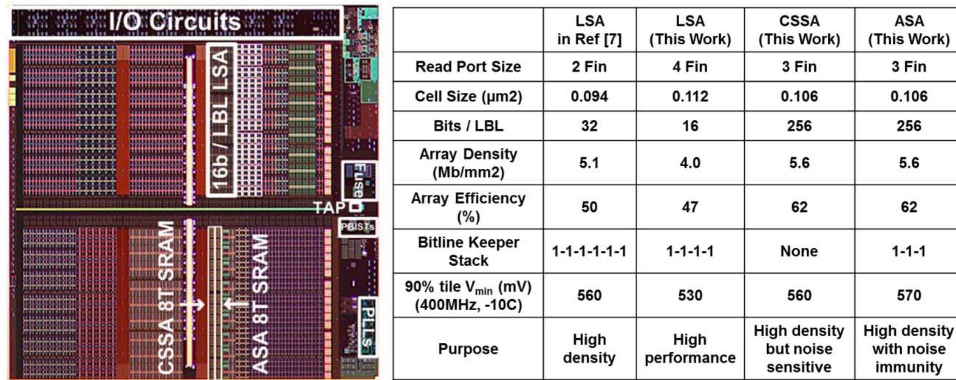


Fig. 10. ASA V_{min} sensitivity analysis for various SA skew and bitline keeper sizing: 6-sigma statistical simulations showing lower noise-induced read-0 V_{min} for increasing SA mismatch and increasing keeper size; read-1 V_{min} which is determined by the bitline differential development delay, increases with higher SA mismatch and stronger bitline keeper.

mismatch. Increased mismatch skews the SA toward the active bitline. Hence a higher droop on the active bitline due to noise coupling events is compensated for by the upsized SA pull-down improving the noise immunity. Similarly, upsizing the bitline keepers improves the read-0 V_{min}. On the other hand, read-1 V_{min} that captures the read bitline evaluation delay degrades with increasing SA pull-down mismatch. The active bitline needs to develop more differential to compensate for the upsized SA pull-down so that the reference side, which is nominally sized and precharged to V_{CC}, evaluates the SA in the other direction. Read-1 V_{min} also degrades with the increasing keeper size due to the increased bitline contention. Hence, an optimum SA pull-down mismatch and bitline keeper sizing needs to be chosen for balancing the read-0 versus read-1 V_{min}.

IV. Measurement Results

Three 1.75 Mb macros were implemented to characterize the baseline LSA and the small-signal arrays (SSA) with two proposed SA techniques on the same die in a 14 nm FinFET



	LSA in Ref [7]	LSA (This Work)	CSSA (This Work)	ASA (This Work)
Read Port Size	2 Fin	4 Fin	3 Fin	3 Fin
Cell Size (μm2)	0.094	0.112	0.106	0.106
Bits / LBL	32	16	256	256
Array Density (Mb/mm2)	5.1	4.0	5.6	5.6
Array Efficiency (%)	50	47	62	62
Bitline Keeper Stack	1-1-1-1-1-1	1-1-1-1	None	1-1-1
90% tile V_{min} (mV) (400MHz, -10C)	560	530	560	570
Purpose	High density	High performance	High density but noise sensitive	High density with noise immunity

Fig. 11. 14 nm FinFET CMOS test chip die photo and summary table containing bitcell read port size, bitcell area, array density, array efficiency, keeper sizing, measured V_{min}, and target application area of various topologies.

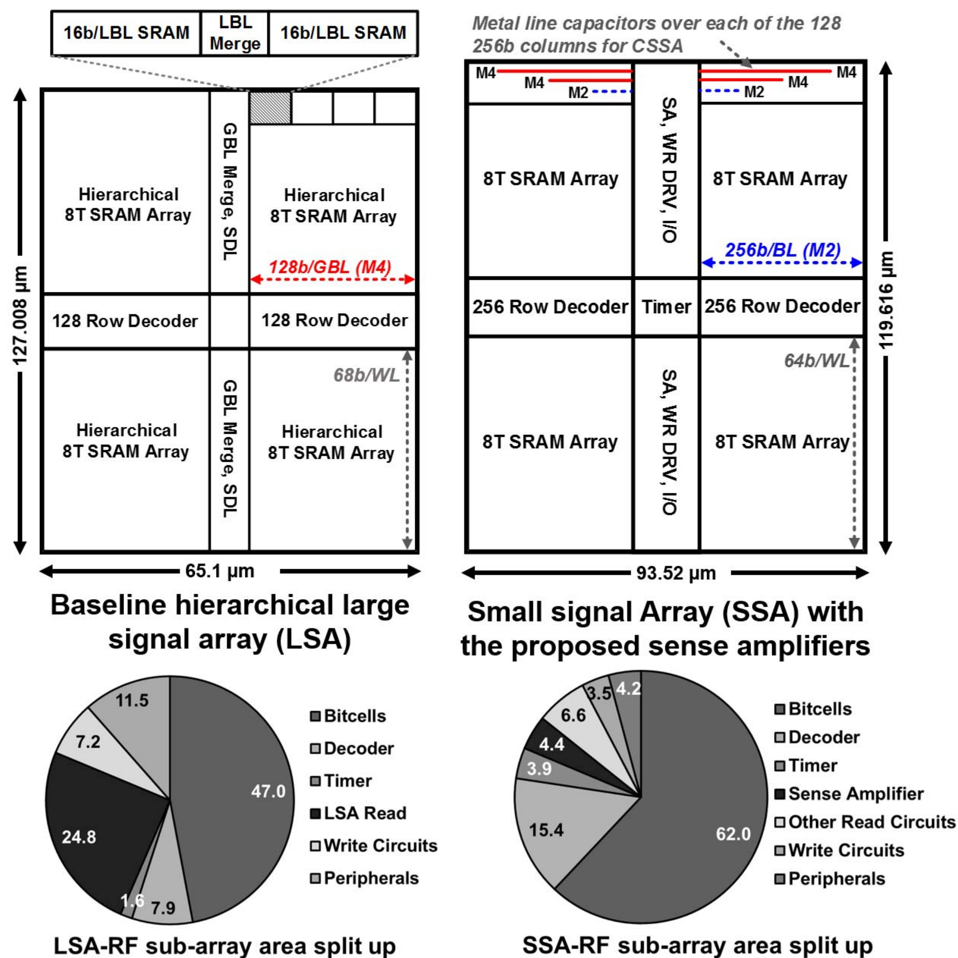
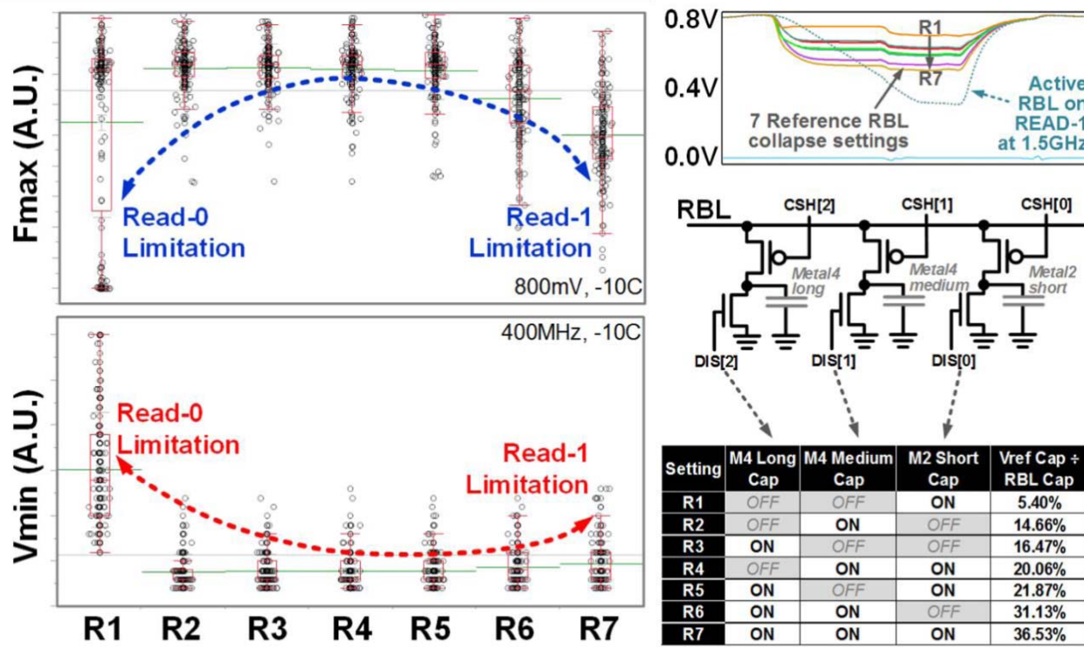


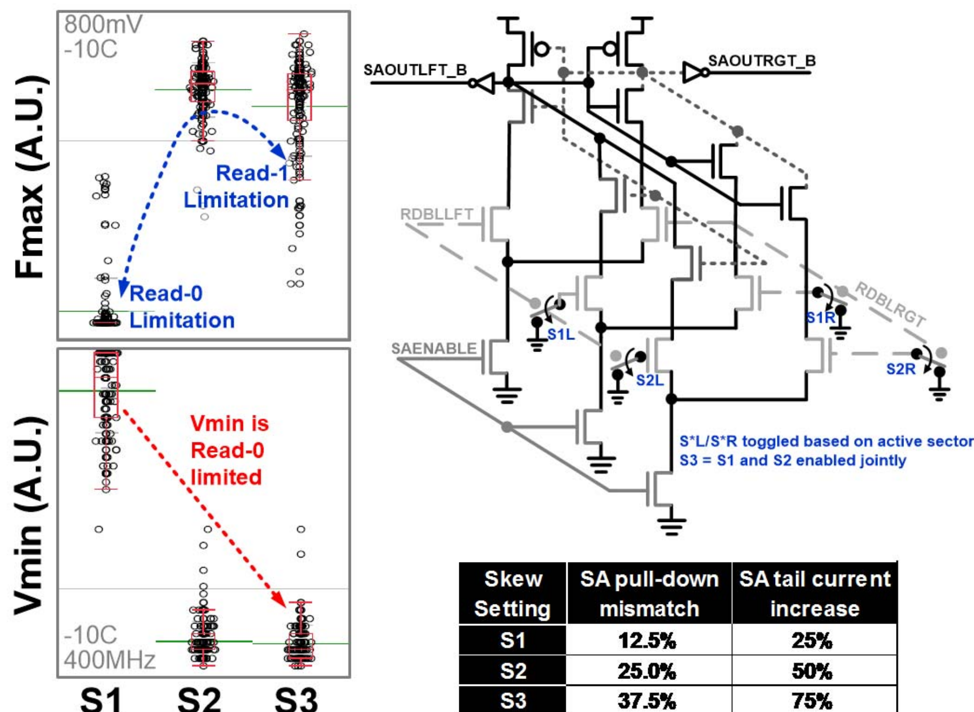
Fig. 12. Large signal and small-signal sensing subarray floorplans, dimensions implemented on the test-chip, and the subarray area split-up.

CMOS test-vehicle. Fig. 11 shows a die photo including a 4.5 Mb LSA macro, of which only 1.75 Mb was used for comparison. The baseline LSA 1.75 Mb macro consists of 56 32 Kb subarrays. Fig. 12 shows the floorplan and layout dimensions for the LSA and SSA subarrays. Each LSA subarray is organized as 256 entries $*$ 136 data I/O. The LSA design employs hierarchical bitlines with 16 bits per LBL for read operation as shown in Fig. 2. The LSA bitcell read port employs stacked 4 fin read port devices which is a 33% larger read port than the corresponding SSA bitcell read port to improve the performance of the full swing design. Each 1.75 Mb SSA macro with the proposed SAs consists of 28 64 Kb subarrays with 128 noninterleaved columns divided into two sectors of 256 rows. In this design, the area used for CSSA capacitor control devices and ASA skews is equivalent at 4.4% of the subarray area, but can be



Setting	M4 Long Cap	M4 Medium Cap	M2 Short Cap	Vref Cap ÷ RBL Cap
R1	OFF	OFF	ON	5.40%
R2	OFF	ON	OFF	14.66%
R3	ON	OFF	OFF	16.47%
R4	OFF	ON	ON	20.06%
R5	ON	OFF	ON	21.87%
R6	ON	ON	OFF	31.13%
R7	ON	ON	ON	36.53%

Fig. 13. Measurement results for CSSA $V_{\min}$ and $F_{\max}$ variation with V_{REF} setting: higher V_{REF} limits read-0 operation while lower V_{REF} (larger charge sharing) limits read-1 operation due to higher bitline differential requirement.



Skew Setting	SA pull-down mismatch	SA tail current increase
S1	12.5%	25%
S2	25.0%	50%
S3	37.5%	75%

Fig. 14. Measurement results for ASA $V_{\min}$ and $F_{\max}$ variation with skew setting: lower skew results in smaller SA mismatch and limits read-0 operation; higher skew results in larger SA mismatch and limits read-1 operation due to larger bitline differential requirement.

reduced further by 20% by optimizing the number of V_{REF} or ASA skew settings. The LSA design with a bigger read port and frequently placed LBL sensing circuits results in a 28% lower bit density and a 24% lower array efficiency compared with the SSA design as shown in Fig. 11.

Measurements were performed on 4400 subarrays for each sensing scheme. The measured read and write failure characterization results showed that overall $V_{\min}$ is limited by the read operation and not by the write operation. Fig. 13 shows the measured CSSA array maximum frequency ($F_{\max}$) and

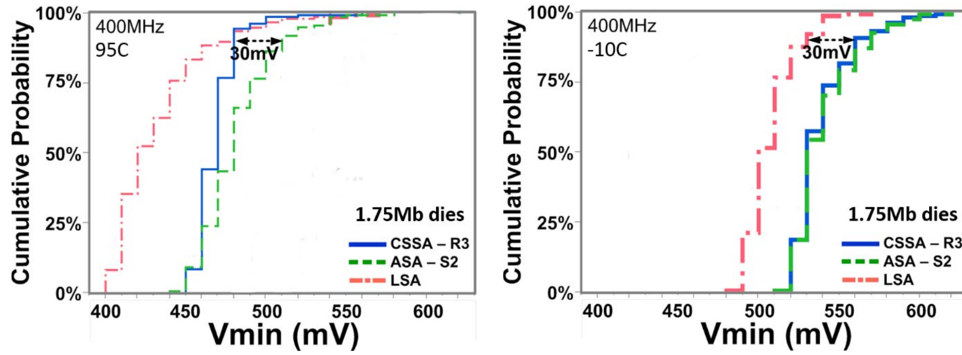


Fig. 15. Measurement results for cumulative $V_{\min}$ distribution at 95 °C and −10 °C temperature: CSSA with optimal R3 setting shows the same $V_{\min}$ as LSA at 95 °C and 30 mV higher $V_{\min}$ at −10 °C 90th percentile; ASA with optimal S2 setting shows 40 mV higher $V_{\min}$ than LSA at 95 °C and 30 mV higher $V_{\min}$ at −10 °C at 90th percentile.

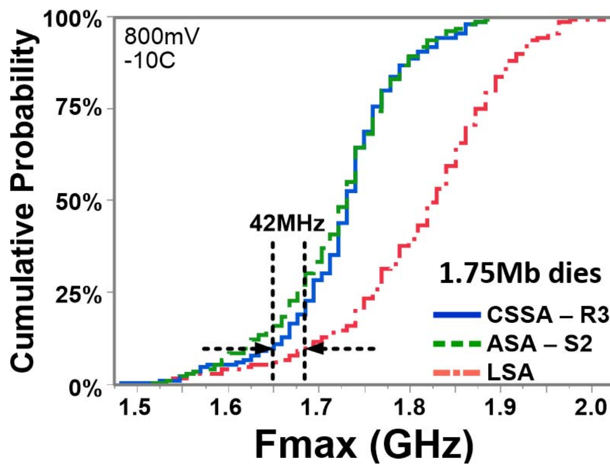


Fig. 16. Measurement results for cumulative $F_{\max}$ distribution at 800 mV, −10 °C: 10th percentile $F_{\max}$ for LSA, CSSA, and ASA is 1.68, 1.64, and 1.62 GHz respectively; LSA design with 33% larger bitcell read port shows ∼3% better $F_{\max}$ compared with the SSA designs; ASA $F_{\max}$ is marginally worse than CSSA case due to increased contention in the internal SA nodes.

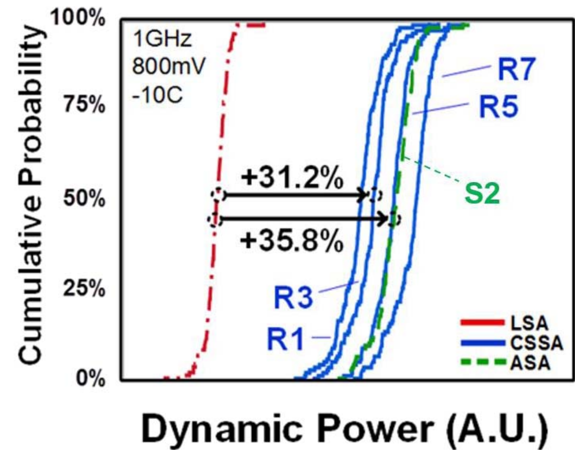


Fig. 17. Measurement results for cumulative dynamic power distribution from a median die at 1 GHz, 800 mV, and −10 °C.

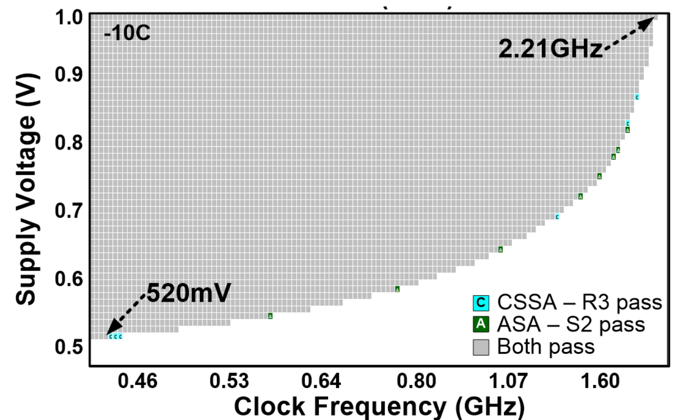


Fig. 18. Measurement results for voltage-frequency Shmoo from a median die: both SSA designs show $V_{\min}$ of 520 mV at 400 MHz, −10 °C, and $F_{\max}$ of 2.21 GHz at 1 V, −10 °C.

$V_{\min}$ results plotted against R1-R7 reference levels, where R7 is the lowest V_{REF} generated by sharing the precharged bitline capacitance with the metal-2 and/or metal-4 capacitor. When using the shallowest reference level R1, read-0 operation limits $V_{\min}$ and $F_{\max}$ because the voltage reference is too close to the active bitline high level, while read-1 operation becomes the limiter with the deepest level R7, as the active bitline needs to discharge sufficiently below the V_{REF}. $V_{\min}$ and $F_{\max}$ are stable across the middle range of V_{REF} where the performance is limited by critical paths not impacted by the bitline differential. The stable reference setting indicates the excellent sensitivity of $V_{\min}$ and $F_{\max}$ to any capacitor mismatch and/or SA offset.

Fig. 14 shows the measured sensitivity of ASA array $V_{\min}$ and $F_{\max}$ to SA skew strengths. There is no bitline differential during read-0 operation, so a sufficiently strong skew that overcomes the SA variability must be selected for the SA to switch in the correct direction. Therefore, both $V_{\min}$ and $F_{\max}$ are read-0 limited at the weak S1 setting. For read-1 operation, the $F_{\max}$ is limited by an S3 setting because the active bitline needs to discharge sufficiently and compensate for the strong SA pull-down so that the nominal sized reference bitline side can resolve the SA output in the other direction. The best $V_{\min}$ and $F_{\max}$ results are obtained with the medium S2 skew with a 25% mismatch in the SA pull-down path.

Fig. 15 shows the measured $V_{\min}$ distributions for both LSA and SSA designs. The 90th percentile $V_{\min}$ of the

TABLE I
DUAL-PORT 8T SRAM COMPARISON

Source	CICC 2013 [3]	VLSI 2013 [6]	ISSCC 2014 [7]	ISSCC 2015 [8]	VLSI 2015 [9]	This Work
Technology	28nm	22nm	20nm	16nm	14nm	14nm
Sub-array Size	128kb	-	16kb	64kb	32kb	64kb
Cell Size (um^2)	-	0.238	-	0.138	0.094	0.106
Density (Mb/mm^2)	3.16	-	3.08	-	5.1	5.6
Sensing Scheme	Small Signal	Large Signal	Small Signal	Large Signal	Large Signal	Small Signal
Port Type	1R1W	1R1W	2RW	1R1W	1R1W	1R1W
# Bits/BL	-	16	-	16	32	256
Performance	1.69GHz (1.0V/125C)	1.6GHz (- /25C)	-	1.67GHz (1.0V/25C)	2.25GHz (1.0V/-10C)	2.21GHz (1.0V/-10C)
Functional	-	-	-	- (0.505V/25C)	400MHz (0.56V/-10C)	400MHz (0.56V/-10C)

CSSA (R3 setting) and ASA (S2 setting) designs at 400 MHz, −10 °C is 560 mV, and 570 mV respectively, while the LSA $V_{\min}$ is 530 mV (the LSA design uses a 33% larger bitcell read port). The 90th percentile 95 °C $V_{\min}$ of the CSSA and LSA designs are matched at 480 mV, while the ASA $V_{\min}$ is 510 mV due to the read-1 contention, which necessitates a larger bitline differential to overcome the stronger SA skew at high temperatures. Note that for both the temperature conditions, the SSA Vmin is within 30–40 mV of the LSA array while using a smaller read port as well and supporting 256 b/bitline.

Fig. 16 shows the measured $F_{\max}$ distributions for the three designs operating at 800 mV, −10 °C. The tenth percentile $F_{\max}$ for ASA and CSSA are 1.62 and 1.64 GHz, respectively, while that of the LSA is 1.68 GHz. LSA design with a 33% larger bitcell read port supporting only 16 b/LBL shows a slightly better (∼3%) $F_{\max}$ compared with the SSA designs supporting 256 b/BL. ASA $F_{\max}$ is marginally worse than the CSSA case due to increased contention in the internal SA nodes.

Fig. 17 shows the measured dynamic power cumulative distributions for the LSA and the SSA designs operating at 800 mV, 1 GHz, and −10 °C. The switched capacitance (C_{DYN}) in the full-swing LSA design employing short LBL (16 b/BL) is smaller than the small-swing SSA design supporting 256 b/BL. The CSSA (ASA) with optimum R3 (S2) setting design median dynamic power is 31.2% (35.8%) higher than the corresponding LSA design. This is due to the large diffusion capacitance connected to the RDBL in the SSA design. In addition, the RDBL swing is not restricted in the current SSA design and can be more than the bitline differential required for correct SA operation. The power overhead of the SSA designs can be lowered by limiting the maximum bitline swing (e.g., a bitline clamp).

The dynamic power consumed by the CSSA array is impacted by the reference level due to the switching capacitance of the reference RDBL while generating V_{REF} for each read operation. The ASA design does not discharge the reference RDBL, but has a higher SA tail current than CSSA due to the intentional sizing mismatch, and the initial droop on the SA internal nodes is larger, which results in more contention across the cross-coupled SA pair. This results in a higher power consumption for the ASA compared with CSSA when low-to-moderate reference levels are used in the latter. The ASA design draws 3.49% more dynamic power at its optimal $V_{\min}/F_{\max}$ setting of S2 compared with the CSSA design at the R3 setting. However, for deep R7 setting in the CSSA design, higher switching capacitance (C_{DYN}) of the reference bitline in every read operation incurs higher power consumption compared with the ASA design.

Fig. 18 shows the Shmoo plot for both the SSA designs from a median 1.75 Mb die operating across a wide voltage/frequency range. Both the designs demonstrate similar behavior at −10 °C, achieving 2.21 GHz at 1 V and 400 MHz at 520 mV.

Table I compares the design parameters for various 2-port SRAMs recently reported. Note that [9] LSA design implements 2 fin read port bitcell with 32b/LBL with a weak 6-stack LBL keeper while LSA design in this test-chip implements a 4 fin read port bitcell with 16b/LBL and a 4-stack LBL keeper. Both LSA designs achieved 560 mV $V_{\min}$ at 90th percentile at 400 MHz/−10 °C. The 1R1W 8T SRAM SSA design utilizing the proposed SA techniques achieves 5.6 Mb/mm^2 bit density, which is 10% higher than the previous best LSA bit density reported [9], in spite of a 13% larger bitcell area (0.106 μm^2 in the proposed SSA vs 0.094 μm^2 in [9] LSA) while achieving equivalent $V_{\min}$ (560–570 mV at 400 MHz, −10 °C) and $F_{\max}$ (2.2 GHz at 1 V, −10 °C). These results illustrate the utility of small-signal pseudo-differential sensing for optimizing the footprint of high performance, low $V_{\min}$ 1R1W 8T SRAM arrays.

V. CONCLUSION

Multiported on-die memory has significant usage in high-performance CPU and GPU IP to provide increased bandwidth. Multiport 1R1W 8T SRAM memory with decoupled read and write ports when organized in a noninterleaved

hierarchical bitlines fashion achieves low V_{min} operation at the expense of significant area cost. In this paper, we propose two SA techniques to realize high-density 1R1W 8T SRAM arrays utilizing small-signal pseudo-differential, single-ended sensing schemes. The charge sharing-based sense amplifier (CSSA) compares the active bitline voltage with an internally generated reference voltage. It is generated during every read operation by charge sharing the unselected bitline precharged to the V_{CC} and a metal capacitor predischarged to Vss. The metal capacitors are realized over the same bitcell column and do not consume additional active silicon area. ASA does not require a dedicated reference voltage generation circuit. It utilizes the unselected bitline precharge voltage (V_{CC}) as the reference voltage. The active bitline side is intentionally upsized to skew the SA. The high-volume measurement results for both the designs implemented on a 14 nm FinFET CMOS test-chip demonstrate successful 560–570 mV V_{min} at 400 MHz/−10 °C, similar to the baseline LSA. Small-signal sensing enables 10% higher bit density compared with the previous best LSA design despite using a bigger bitcell. These results illustrate the utility of small-signal pseudo-differential sensing for optimizing the footprint of 1R1W 8T SRAM arrays.

REFERENCES

[1] M. Qazi, K. Stawiasz, L. Chang, and A. P. Chandrakasan, "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," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2010, pp. 350–351.

[2] B.-D. Yang and L.-S. Kim, "A low-power ROM using charge recycling and charge sharing techniques," *IEEE J. Solid-State Circuits*, vol. 38, no. 4, pp. 641–653, Apr. 2003.

[3] M. Yabuuchi, H. Fujiwara, Y. Tsukamoto, M. Tanaka, S. Tanaka, and K. Nii, "A 28nm high density 1R/1W 8T-SRAM macro with screening circuitry against read disturb failure," in *Proc. Custom Integr. Circuits Conf.*, Sep. 2013, pp. 1–4.

[4] B. Zimmer, "Resilient design techniques for improving cache energy efficiency," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, 2015.

[5] J. Keane *et al.*, "5.6Mb/mm^2 1R1W 8T SRAM arrays operating down to 560mV utilizing small-signal sensing with charge-shared bitline and asymmetric sense amplifier in 14nm FinFET CMOS technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2016, pp. 308–309.

[6] J. Kulkarni *et al.*, "Dual-V_{CC} 8T-bitcell SRAM array in 22nm tri-gate CMOS for energy-efficient operation across wide dynamic voltage range," in *VLSI Technol. Symp. Dig. Tech. Papers*, Jun. 2013, pp. C126–C127.

[7] M. Yabuuchi, Y. Tsukamoto, M. Morimoto, M. Tanaka, and K. Nii, "20nm high-density single-port and dual-port SRAMs with wordline-voltage-adjustment system for read/write assists," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 234–235.

[8] H. Fujiwara *et al.*, "A 64kb 16nm asynchronous disturb current free 2-port SRAM with PMOS pass-gates for FinFET technologies," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 1–3.

[9] K.-H. Koo, L. Wei, J. Keane, U. Bhattacharya, E. A. Karl, and K. Zhang, "A 0.094μm^2 high density and aging resilient 8T SRAM with 14nm FinFET technology featuring 560mV VMIN with read and write assist," in *VLSI Technol. Symp., Dig. Tech. Papers*, Jun. 2015, pp. C266–C267.

Jaydeep P. Kulkarni (M'09–SM'15) received the B.E. degree in electrical engineering from University of Pune, Pune, India, in 2002, the M.Tech. degree in electrical engineering from the Indian Institute of Science, Bengaluru, India, in 2004, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 2009.

From 2004 to 2005, he was a Design Engineer at Cypress Semiconductors, Bengaluru, where he designed I/O circuits for micropower SRAMs. He joined the Circuit Research Laboratory at Intel Corporation, Hillsboro, OR, USA, in 2009, where he is currently a Staff Research Scientist. He filed 30 patents and has published 55 papers in referred journals and conferences. His current research interests include energy efficient integrated circuits and systems.

Dr. Kulkarni was a recipient of the 2004 Best M.Tech. Student Award from IISc Bangalore, the 2008 SRC Inventor Recognition Awards, the 2008 ISLPED Design Contest Award, the 2008 Intel Foundation Ph.D. Fellowship Award, the 2008 SRC TECHCON Best Paper in Session Award, the 2010 Purdue School of ECE Outstanding Doctoral Dissertation Award, the 2015 IEEE Circuits and Systems Society's Transactions on VLSI Systems Best Paper Award, and the 2015 Semiconductor Research Corporation's Outstanding Industrial Liaison Award. He has participated in Technical Program Committees of A-SSCC, ISLPED, ISCAS, and ASQED conferences. He serves as an Associate Editor of the IEEE TRANSACTIONS ON VLSI SYSTEMS, and as an Industrial Liaison for SRC and STARnet research programs.

John Keane received the B.S. degree in computer engineering from the University of Notre Dame, Notre Dame, IN, USA, in 2003, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 2010.

He joined the Advanced Design Organization at Intel Corporation, Hillsboro, OR, USA, where he designed memory test circuits. In 2016, he joined the product engineering group at Medtronic, Mounds View, MN, USA, where he is responsible for the development of implantable cardiac devices. He has co-authored over 20 peer-reviewed publications and is the co-inventor of five U.S. patents.

Kyung-Hoae Koo received the B.S. degree from Korea University, Seoul, South Korea, in 2005, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 2007 and 2011, respectively.

From 2007 to 2011, he was a Research Assistant with the Center for Integrated Systems, Department of Electrical Engineering, Stanford University. From 2011 to 2015, he was with the Intel Corporation Advanced Design Group, where he was involved in SRAM bitcell technology and circuits design. Since 2015, he has been with Apple Inc., Cupertino, CA, USA, where he is involved in display module architecture and technology development for small form factors.

Satyanand Nalam (S'06–M'11) received the B.Tech. degree in electrical engineering from IIT Madras, Chennai, India, in 2004, and the M.S. and Ph.D. degrees in computer engineering from the University of Virginia, Charlottesville, VA, USA, in 2008 and 2011, respectively.

Since 2011, he has been a Design Engineer with the Advanced Design Memory Circuit Technology group at Intel, Hillsboro, OR, USA, where he is involved in memory design and validation on test chips at advanced nanometer nodes, simulation methodology, and simulation-to-silicon correlation.

Zheng Guo (S'03–M'09) received the B.S. degree in computer engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2003, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley, Berkeley, CA, USA, in 2005 and 2009, respectively.

In 2010, he joined the Logic Technology Development Organization at Intel Corporation, where he focuses on memory circuit technology for low-power, high-performance, and high-density applications.

Eric Karl (S'03–M'08) received the B.S.E., M.S.E., and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2002, 2004, and 2008, respectively.

He held positions at the Intel Circuit Research Laboratory, Hillsboro, OR, USA, the IBM T.J. Watson Research Center, Yorktown Heights, NY, USA, and Sun Microsystems, Chelmsford, MA, USA. In 2008, he joined Intel Logic Technology Development, Hillsboro, OR, USA, where he is a Principal Engineer involved in the development of memory technology, technology development testchips, and memory IP power performance and area optimization. He has authored more than 23 conference papers and technical journal articles and has reviewed numerous conferences and journals.

Kevin Zhang (S'93–M'95–F'11) is a Vice President of the Technology and Manufacturing Group and an Intel Fellow at Intel Corporation. He is responsible for advanced circuit technology development for the company's future products. He oversees the development of process design rules, circuit and device modeling, digital circuit libraries, key analog and mixed-signal circuits, high-speed I/O and embedded memories. He has published more than 80 papers at international conferences and in technical journals and is the editor of *Embedded Memory for Nano-Scale VLSIs* (Springer, 2009). He holds more than 55 U.S. patents in the field of integrated circuit technology. He was the 2016 International Solid-State Circuit Conference (ISSCC) Program chair and also serves on IEEE VLSI Executive Committee.