A Metastable-Point-Shifting I/O Sense Amplifier for VEQ-Precharged Global I/O Readout in High-Speed DRAM

Hoyeon Jang[1], Do Han Kim[1], Ungbin Oh[1], Seungyeon Hong[1], Kyomin Sohn[2], and Seong-Ook Jung[1]
[1]*School of Electrical and Electronic Engineering*, *Yonsei University*, Seoul, Republic of Korea
[2]*Samsung Electronics*, Hwaseong-si, Gyeonggi-do, Republic of Korea
Email: jhy001229@yonsei.ac.kr

***Abstract*— This paper presents a metastable-point-shifting I/O sense amplifier (MS-IOSA) for VEQ-precharged global I/O (GIO) readout in high-speed DRAM. The proposed scheme increases the sensing margin during local sense amplifier (LSA) activation by introducing a signal-dependent intentional offset. By removing the sequential timing overhead of prior offset-compensation schemes, the MS-IOSA reduces GIO readout latency. Evaluated in an LPDDR5X environment with RC modeling, the proposed MS-IOSA achieves a minimum readout latency of 2.75 ns and consumes 44.4 fJ at a 0.9 V supply in 28 nm CMOS. Compared with the prior state of the art, it reduces readout latency by 35.2% and energy by 67.4%.**

Keywords*— *DRAM, I/O sense amplifier (IOSA), metastable-point shifting, low-power, high-speed, VEQ precharge

I. Introduction

Modern data-intensive workloads such as large language model (LLM) inference, real-time video analytics, and graph processing require high DRAM bandwidth. This demand increases the frequency of column accesses, which account for a substantial fraction of DRAM power consumption. Prior work has focused on reducing IOSA energy during column accesses [1], [2]. In recent DRAM architectures, however, larger bank capacities have lengthened global I/O (GIO) lines and increased GIO capacitance (C_{GIO}) [3]. As a result, GIO precharge after each read has become a major contributor to DRAM power consumption.

Fig. 1 summarizes prior energy-saving approaches and their limitations. The offset-canceling IOSA (OC-IOSA) [4] was proposed to reduce GIO precharge energy by lowering the required ΔV_{GIO}. OC-IOSA achieves this by compensating for threshold-voltage (V_{TH}) mismatch between the input transistors. This benefit comes at the cost of an additional offset-cancellation (OC) phase, which reduces the timing margin for the column-to-column delay (t_{CCD}). The static-current-free presensing IOSA (SCFP-IOSA) [5] shortens the OC phase through presensing but still relies on extra phases and incurs the same fundamental timing penalty. This penalty becomes more severe as C_{GIO} increases. The larger RC delay requires the local sense amplifier (LSA) to remain enabled longer to generate the minimum required ΔV_{GIO}.

Another way to reduce GIO precharge energy is to lower the precharge level (V_{PRE}) from V_{DD} to half-V_{DD} (V_{EQ}) [6]. Because precharge energy is proportional to $C_{GIO} \cdot V_{PRE} \cdot \Delta V_{GIO}$,

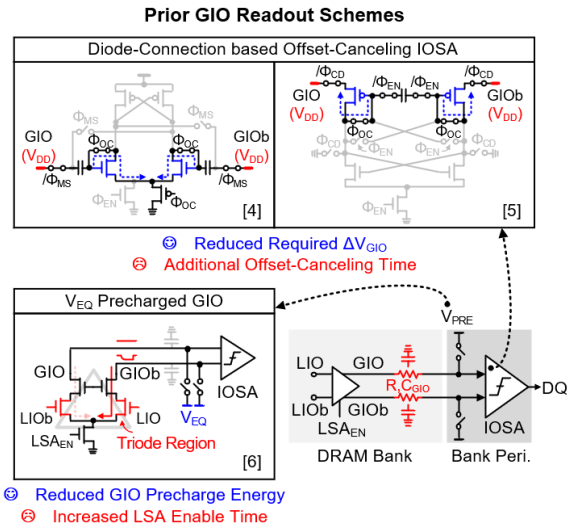

Fig. 1. Prior GIO readout schemes and their limitations in a DRAM bank architecture.

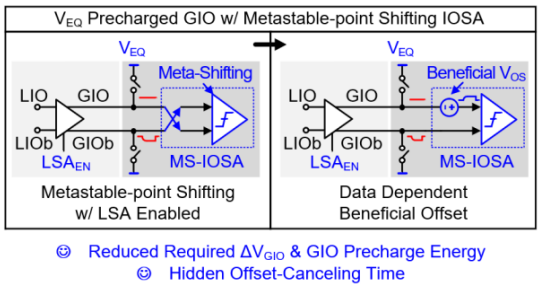

Fig. 2. Proposed VEQ-precharged GIO readout scheme with MS-IOSA.

V_{EQ} precharge reduces the energy directly. Setting V_{PRE} to V_{EQ}, however, drives the input transistors of the LSA into the triode region, which weakens the LSA drive strength and slows ΔV_{GIO} development. Prior OC schemes [4], [5] could in principle mitigate this slowdown by reducing the required ΔV_{GIO} for reliable sensing. These schemes, however, rely on diode-connected OC with an input common-mode level of V_{DD}. When the common-mode level is lowered to V_{EQ}, the reduced voltage headroom degrades OC effectiveness. Under V_{EQ} precharge, prior OC schemes cannot be effectively applied, and

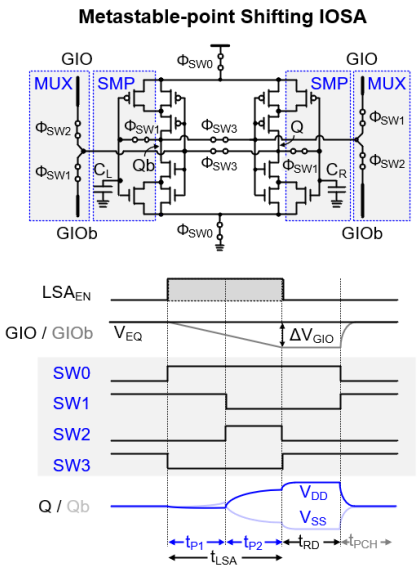

Fig. 3. Proposed MS-IOSA schematic and operating waveforms. P1 and P2 overlap with LSA activation.

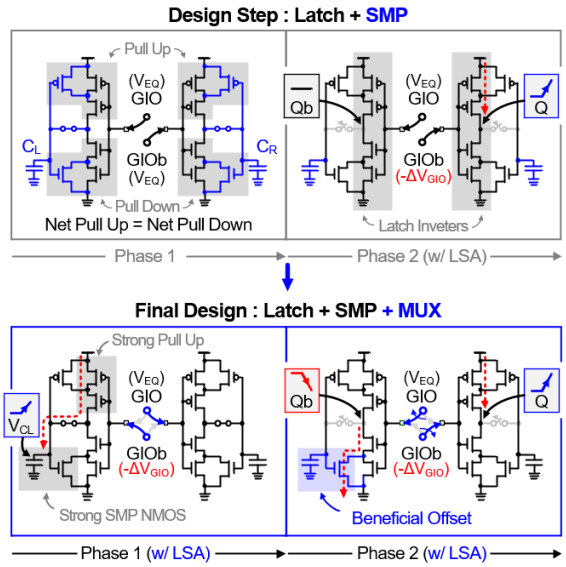

Fig. 4. Design evolution of the proposed MS-IOSA. (Top) Latch with SMP only. (Bottom) Latch with SMP and MUX.

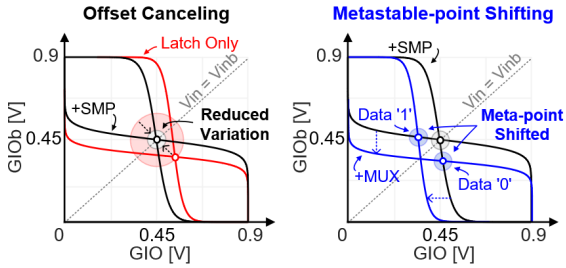

Fig. 5. Transfer curves of the two designs in Fig. 4. (Left) SMP only. (Right) SMP + MUX.

the required ΔV_{GIO} for reliable sensing remains large. As a result, GIO readout latency increases, making it difficult to meet t_{CCD} in high-speed DRAM.

Fig. 2 shows the proposed V_{EQ}-precharged GIO readout scheme based on a metastable-point-shifting IOSA (MS-IOSA). MS-IOSA improves the sensing margin by shifting the metastable point during LSA operation. Unlike prior OC-based approaches, it does not require an additional phase. MS-IOSA therefore avoids the associated latency overhead and enables faster GIO readout.

II. Proposed MS-IOSA

A. Schematic and Operating Waveforms of MS-IOSA

Fig. 3 shows the schematic and operating waveforms of the proposed MS-IOSA. The MS-IOSA consists of three main components: (i) a cross-coupled latch for data sensing, (ii) sampling transistors and capacitors (SMP) that introduce an intentional latch offset voltage, and (iii) multiplexer transistors (MUX) for input selection. During phase 1 (P1) and phase 2 (P2), the MS-IOSA improves the sensing margin through metastable-point shifting and presensing while the LSA remains enabled. Once presensing is sufficient, the latch senses the data.

B. Detailed Operation of SMP and MUX

The operation of the proposed MS-IOSA is described in two steps to separately highlight the roles of the two proposed components, the SMP and the MUX. Fig. 4 presents the design evolution of the proposed IOSA. Fig. 4 (top) shows the proposed latch with SMP only to isolate the contribution of the SMP, whereas Fig. 4 (bottom) shows the complete MS-IOSA with both SMP and MUX.

Fig. 4 (top) shows the latch with SMP only. During P1, SW1 is turned on to form a negative-feedback path, and both inverters receive the common-mode input V_{EQ}. Due to V_{TH} mismatch, each inverter exhibits pull-up and pull-down imbalance, resulting in an input-referred offset. The SMP compensates for this mismatch by sampling the latch balance point. If one inverter exhibits a stronger pull-up path, the SMP stores a higher voltage through the SW1 path. The SMP NMOS then discharges the internal node, thereby weakening the effective pull-up dominance. This action rebalances the pull-up and pull-down strengths and reduces the input-referred offset. In P2, SW1 is turned off, and the LSA is enabled to provide the input for presensing while the inverter amplifies the input. Because the balance point is stored on the SMP, the input-referred offset is reduced and the sensing margin improves. A similar approach was reported in [7], where the trip voltage is stored on a capacitor and the input is applied through capacitive coupling. In contrast, the proposed SMP balances the latch with external transistors, which avoids the input attenuation caused by parasitic capacitance.

Fig. 4 (bottom) shows the complete MS-IOSA with the MUX added to the SMP-based latch. During P1, the MUX reverses the input polarity while the LSA is enabled. One inverter receives V_{EQ}, whereas the other receives a lower voltage from the LSA. Under this condition, the sampling

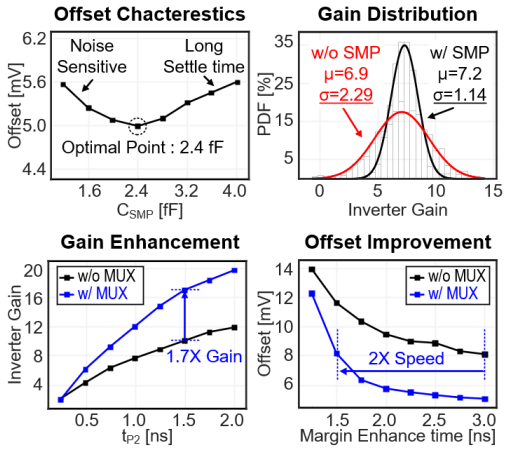

Fig. 6. Simulated characteristics of the proposed MS-IOSA. (Top) SMP design point and presensing-gain variation. (Bottom) MUX effect on gain and input-referred offset.

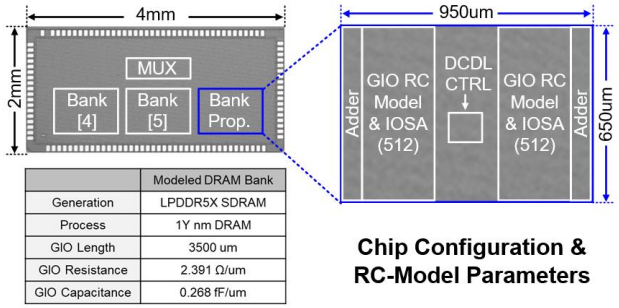

Fig. 7. Test-chip configuration and RC-modeled GIO parameters.

capacitor stores a higher voltage (V_{CL}) than in the SMP-only operation. In P2, the input polarity returns to normal while the LSA remains enabled. The inverter that previously received the lower input is now driven by V_{EQ}. The voltage stored on the SMP then introduces an additional intentional offset, thereby making the effective pull-down strength larger than the pull-up strength. As a result, the output node (Qb) is discharged, which acts as an additional differential input to the latch.

Fig. 5 compares the transfer curves of the two designs. While SMP alone reduces metastable-point variation (Fig. 5, left), the MUX-added design additionally shifts the metastable point toward a favorable sensing direction (Fig. 5, right). This shift further improves the sensing margin.

Fig. 6 (top) shows the design parameters and characteristics of the SMP in the MS-IOSA. Considering the trade-off between noise robustness and settling time, the SMP capacitor (C_{SMP}) is set to 2.4 fF. The SMP reduces presensing-gain variation, indicating that it compensates for pull-up and pull-down strength mismatch within each inverter. Fig. 6 (bottom) shows the effect of the MUX in the MS-IOSA. By allowing the intentional offset voltage to act as an additional input, the MUX increases the inverter gain during P2. As a result, the sensing margin develops faster in the MS-IOSA than in the SMP-only design.

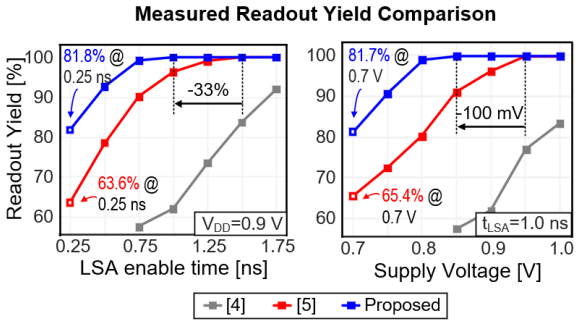

Fig. 8. Measured readout yield as a function of LSA enable time (left) and supply voltage (right).

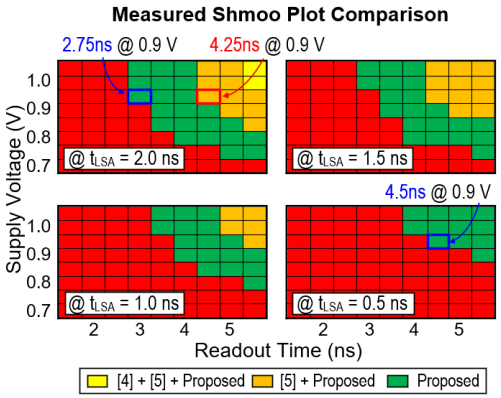

Fig. 9. Measured shmoo plots of readout time versus supply voltage for different LSA enable times.

III. Measurement Results

Fig. 7 shows the test chip configuration for evaluating the proposed GIO readout scheme in 28 nm CMOS logic technology. To emulate an LPDDR5X SDRAM environment, the GIO resistance and capacitance were modeled from the physical interconnect length and metal-layer characteristics [8]. The LSA was sized and trimmed to match the driving strength of a 1Y-nm DRAM process [9]. Each bank includes 1024 IOSAs used in the evaluation. GIO readout performance was measured by sweeping the timing of each operating phase and the supply voltage while monitoring readout yield. For a fair comparison, the latch transistors in all IOSAs were set to the same size (600 nm/30 nm), providing the same baseline mismatch level according to Pelgrom's law. MS-IOSA occupies 20.5 μm^2, which is 56.9% smaller than OC-IOSA and 69.4% larger than SCFP-IOSA. This area penalty remains small at the bank level because IOSAs are typically placed at the bank periphery.

Fig. 8 compares the measured readout yields of the proposed MS-IOSA for V_{EQ}-precharged GIO readout with those of prior V_{DD}-precharged architectures [4], [5]. Fig. 8 (left) shows the yield versus LSA enable time at a 0.9 V supply. As the LSA enable time decreases, prior architectures fail to maintain 100% yield when it is reduced below 1.5 ns. In contrast, the MS-IOSA maintains 100% yield down to 1.0 ns even with V_{EQ}-precharged GIO. Fig. 8 (right) shows the yield as V_{DD} is scaled down with the LSA enable time fixed at 1.0 ns.

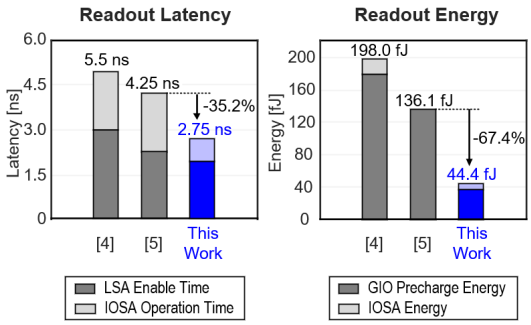

Fig. 10. Measured readout latency (left) and energy (right) at the minimum readout latency.

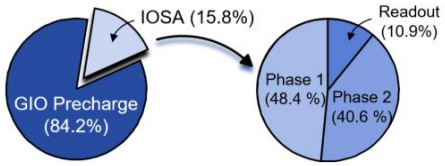

Fig. 11. Energy breakdown of the proposed GIO readout scheme.

TABLE I. PERFORMANCE COMPARISON

	ISSCC'22 [4]	JSSC'25 [5]	This Work
Target	1Y-nm DRAM, LPDDR5X SDRAM (1 Gb bank)		
Latch Transistor Size	NMOS, PMOS (W/L) = 600 nm/30 nm		
Technology [nm]	Samsung 28-nm CMOS		
Supply Voltage [V]	0.9		
GIO Precharge Level	V_{DD}	V_{DD}	V_{EQ}
OC Method	Diode Connection	Diode Connection	Metastable-point Shifting
OC Phase Hidden	X	X	O
Number of Devices	7T+2C+6SW	5T+1C+10SW	14T+2C+8SW
Area [μm²]	47.5	12.1	20.5
Input-referred offset σ 1) [mV] †	17.7	9.5	5.0
Sensing Latency 2) [ns] †	6.7	2.93	2.52
Sensing Energy 2) [fJ] †	14.5	3.75	1.45
Readout Latency 3) [ns]	5.5	4.25	2.75
Readout Energy 3) [fJ]	198.0	136.1	44.4
FoM [(fJ·ns)$^{-1}$]	1.00	1.88	8.91

†: Simulation results. FoM = (Readout Energy × Readout Latency)$^{-1}$
1) Margin enhancement time is set to 3ns.
2) 100% Pass at ΔV_{in} = 50mV over 1,000 Monte Carlo simulations.
3) Fastest readout at V_{DD} =0.9V.

While conventional designs exhibit yield degradation below 0.95 V, the MS-IOSA maintains 100% yield down to 0.85 V. Fig. 9 shows the measured shmoo plots of readout time versus supply voltage for different LSA enable times. The readout latency of prior schemes was limited by the additional OC phase. In contrast, the proposed MS-IOSA removes this timing overhead by performing OC concurrently with LSA activation. As a result, the MS-IOSA achieves a minimum readout latency of 2.75 ns at 0.9 V with the same 2.0 ns LSA enable time, reducing the readout latency by 35.2% compared to SCFP-IOSA. In addition, owing to the metastable-point-shifting scheme, the MS-IOSA shows no readout failures at an LSA enable time of 0.5 ns and a readout time of 4.5 ns.

Fig. 10 compares the readout latency and energy consumption of each scheme at the minimum readout latency under a 0.9 V supply. The proposed V_{EQ}-precharged MS-IOSA achieves the lowest latency and energy consumption, reaching 2.75 ns and 44.4 fJ. This corresponds to 35.2% lower latency and 67.4% lower energy than SCFP-IOSA. Fig. 11 shows the energy breakdown of the proposed GIO readout scheme. Although 15.8% of the total energy is consumed in the IOSA for margin enhancement, the improved sensing margin reduces GIO precharge energy and yields the lowest overall energy consumption.

IV. CONCLUSION

Table I summarizes the overall performance of the compared GIO readout schemes. At a 0.9 V supply, the proposed VEQ-precharged MS-IOSA achieves 2.75 ns readout latency and 44.4 fJ readout energy while eliminating the extra OC-phase overhead. These results show that the proposed MS-IOSA enables robust sub-1-V operation and offers a practical solution for low-power, high-speed DRAM readout.

ACKNOWLEDGMENT

This work was supported by the Samsung Electronics, under Grant IO250214-11969-01. The chip fabrication and EDA tool were supported by the IC Design Education Center(IDEC), Korea.

REFERENCES

[1] Y. K. Kim et al., "A 1.5 V, 1.6 Gb/s/pin, 1 Gb DDR3 SDRAM with an address queuing scheme and bang-bang jitter-reduced DLL scheme," in *Proc. IEEE Symp. VLSI Technol.*, 2007, pp. 182–183.

[2] Y. Moon et al., "1.2 V, 1.6 Gb/s, 56 nm, 6F², 4 Gb DDR3 SDRAM with hybrid-I/O sense amplifier and segmented sub-array architecture," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2009, pp. 128–129.

[3] D. Kwon et al., "A 1.1 V, 6.4 Gb/s/pin, 24-Gb DDR5 SDRAM with a highly accurate duty corrector and NBTI-tolerant DLL," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2023, pp. 27–29.

[4] D.-H. Kim et al., "A 16 Gb, 9.5 Gb/s/pin LPDDR5X SDRAM with low-power schemes exploiting dynamic voltage-frequency scaling and offset-calibrated readout sense amplifiers in a fourth-generation 10 nm DRAM process," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2022, pp. 448–450.

[5] D.-Y. Lim et al., "A 3× offset, 2.9× power, 1.3× sensing time, and 4× area reduction direct input transfer offset-cancel DRAM IO sense amplifier with static current-free pre-sensing," *IEEE J. Solid-State Circuits*, vol. 61, no. 3, pp. 1216–1227, Mar. 2026.

[6] C. Yoo et al., "A 1.8-V 700-Mb/s/pin 512-Mb DDR-II SDRAM with on-die termination and off-chip driver calibration," *IEEE J. Solid-State Circuits*, vol. 39, no. 6, pp. 941–951, Jun. 2004.

[7] B. Giridhar, N. Pinckney, D. Sylvester and D. Blaauw, "A reconfigurable sense amplifier with auto-zero calibration and pre-amplification in 28 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2014, pp. 242–243.

[8] J. B. Park, W. R. Davis, and P. D. Franzon, "3-D-DATE: A circuit-level three-dimensional DRAM area, timing, and energy model," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 2, pp. 756–768, Feb. 2019.

[9] M. Marazzi et al., "HiFi-DRAM: Enabling high-fidelity DRAM research by uncovering sense amplifiers with IC imaging," in *ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA)*, 2024, pp. 133–149.