Design and sensitivity analysis of a new current-mode sense amplifier for low-power SRAM

Do, Anh Tuan; Kong, Zhi Hui; Yeo, Kiat Seng; Low, Jeremy Yung Shern

2009

Do, A. T., Kong, Z. H., Yeo, K. S. & Low, Y. S. (2009). Design and Sensitivity Analysis of a New Current-Mode Sense Amplifier for Low-Power SRAM. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 9, 1-1.

<https://hdl.handle.net/10356/91836>

<https://doi.org/10.1109/TVLSI.2009.2033110>

© 2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. <http://www.ieee.org/portal/site> This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Design and Sensitivity Analysis of a New Current-Mode Sense Amplifier for Low-Power SRAM

Anh-Tuan Do, Zhi-Hui Kong, Kiat-Seng Yeo, and Jeremy Yung Shern Low

Abstract—A new current-mode sense amplifier is presented. It extensively utilizes the cross-coupled inverters for both local and global sensing stages, hence achieving ultra low-power and ultra high-speed properties simultaneously. Its sensing delay and power consumption are almost independent of the bit- and data-line capacitances. Extensive post-layout simulations, based on an industry standard 1 V/65-nm CMOS technology, have verified that the new design outperforms other designs in comparison by at least 27% in terms of speed and 30% in terms of power consumption. Sensitivity analysis has proven that the new design offers the best reliability with the smallest standard deviation and bit-error-rate (BER). Four 32×32 -bit SRAM macros have been used to validate the proposed design, in comparison with three other circuit topologies. The new design can operate at a maximum frequency of 1.25 GHz at 1 V supply voltage and a minimum supply voltage of 0.2 V. These attributes of the proposed circuit make it a wise choice for contemporary high-complexity systems where reliability and power consumption are of major concerns

Index Terms—Current mode and sense amplifier, low power, low voltage SRAM.

I. INTRODUCTION

SRAM-BASED cache which is responsible for increasing the speed of the data flows, and hence the speed of the system, is one of the most important components of state-of-the-art VLSI systems. It is prevalently presented in the design of modern microprocessors for bridging the widening divergence between the performances of the Central Processing Unit (CPU) and the DRAM-based main memory [1]. This trend is accentuated by the never-ending market demand for sophisticated communication and multimedia applications, which require high-tech portable electronic gadgets with high-performance as their requisite feature. As on-chip memory will occupy a large portion of the chip area, the power dissipated within the memory, both active and standby, will become a dominant part of the total power consumption of the chip [2]–[4]. In view of the

above, there is invariably an apparent urgency to address these two often-conflicting power and performance requirements [5], [6]. While there are a lot of sources of power consumption (for instance, leakages, memory cells, Sense Amplifier (SA) and I/O circuits), the total delay is mainly determined by the significant capacitances attributed by the long-wire paths routed in close proximity (commonly known as C_{BL} and C_{DL}) [7]. These highly capacitive wires are also important factors that drastically increase the total power dissipation during the read and write operations [8], [9]. The current-mode SA, which has the ability to quickly amplify a small differential signal on the bit-lines (BLs) and data-lines (DLs) to the full CMOS logic level without requiring a large input voltage swing, is widely used as one of the most effective ways to reduce both sensing delay and power consumption of the SRAM [8]–[22].

In this paper, we propose a current-mode SA that improves the sensing speed and reliability of the previously published designs and at the same time reduce the power consumption. It was extensively simulated and graphically presented in comparison with other three widely used SA topologies, namely the high-speed [12], decoupled latch [18], [19], and the alpha latch [20] designs.

II. EXISTING DESIGNS

This section briefly describes the operations of three existing designs studied in this work. The gists of these designs are depicted in Fig. 1.

A. Current-Conveyor-Based Sense Amplifier

The first conveyor-based sense amplifier was proposed by E. Seevinck *et al.* in [8]. It consists of four identical pMOS transistors [P1–P4 in Fig. 1(a)] connected in a feedback structure. It is assumed that the complementary bit-lines (BL and $\overline{BL}$) are precharged to V_{DD} and all four pMOS transistors operate in saturation region during the read cycles. The current conveyor is enabled by triggering the column select (CS) signal low. Since all four transistors are in saturation, their source-to-drain currents are only dependent on their gate-to-source voltages. As a result, voltage at the bit-line terminals (V_{BL} and $V_{\overline{BL}}$) are the same and equal to $(v_1 + v_2)$. The current conveyor therefore has the ability to convey the differential current from the bit-lines to the data-line without waiting for the discharging of the highly capacitive bit-lines. Thus, this design achieved both higher sensing speed and lower power consumption when compared to the conventional voltage mode designs in which large voltage difference must be developed between the bit-lines [8].

Manuscript received May 24, 2009; revised August 05, 2009.

A. T. Do and J. Y. S. Low are with the Center for Integrated Circuits and Systems (CICS), School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore (e-mail: atdo@ntu.edu.sg; z050041@ntu.edu.sg).

Z. H. Kong and K. S. Yeo are with the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore (e-mail: eksyeo@ntu.edu.sg; zhkong@ntu.edu.sg).

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/TVLSI.2009.2033110

Fig. 1. Local sensing stage of existing SRAM sense amplifier. a) Current-conveyor. b) Alpha latch. c) Bit-line decoupled latch.

Based on this basis structure, several improved versions of this design have been reported, mainly by adding current-mirrors to the feet of the current-conveyor to enhance its current driveability [5], [10], [12]. In this paper, we will compare our work with the high-speed design [12] which consists of four additional nMOS transistors, also shown in Fig. 1(a). These nMOS devices form two current-mirrors to intensify the output currents I_1 and I_2 to the data-lines. This design will be used as the benchmark to evaluate the performance of the proposed design, the alpha latch and the decoupled latch sense amplifiers mentioned below. However, because of its current-mode nature, we do not study its input-offset voltage. As a result, input-off set analysis (see Fig. 7) and latching delay analysis (see Fig. 5) are not applicable to this design.

B. Alpha-Latch Sense Amplifier

The alpha latch [20] is depicted in Fig. 1(b). The nMOS transistor N5 is used to turn the amplifier off during standby, thus save power. When the sense amplifier is activated by the enable signal (EN), the differential input from the complementary bit-lines induces a differential transconductance in N3 and N4. As a result, voltage and current differences will appear at the drains of N3 and N4, i.e., the sources of N1 and N2. Since the CS signal turns off N6, the flip-flop structure will latch and full swing voltages will be available at nodes A and B, turning one

Fig. 2. Proposed design coupled with a simplified read-cycle-only memory system.

of the transistors N7 and N8 on while the other is off. During standby, $\overline{EN}$ is kept high to turn P3 and P4 off. During operation, both P3 and P4 are turned on but one of N7 and N8 is turned off, thus only one current will flow to the data-lines [i.e., I_1 or I_2 in Fig. 2(b)]. A global sense amplifier is also used to quickly amplify the voltage difference on the data-lines to the output of the SRAM.

C. Decoupled Latch Sense Amplifier

The decoupled-latch consists of six nMOS and two pMOS transistors, as shown in Fig. 1(c). Similar to the alpha-latch, its N3 is used to save power. The reason we use a tail nMOS device in Fig. 1(b) and Fig. 1(c) is because it gives a smaller area comparing to a pMOS with the same current strength. Furthermore, BLs are precharged to V_{DD} and hence nMOS tail device is required. It is in contrast with our proposed design in Fig. 2 where a tail pMOS device must be used because DL and $\overline{DL}$ are precharged to ground. To tackle the heavily loaded bit-lines issue, these bit-line signals are tapped to the input ports of the amplifier through two decoupled devices, i.e., P3 and P4. Once

the bit-line differential signal is induced at nodes C and D, the latch is enabled by turning off N4 but turning on N3. Concurrently, P3 and P4 are turned off to decouple the bit-lines from the high-swing output nodes. The use of P3 and P4 helps reducing the impact of the bit-line capacitances on the switching activity, hence significantly reducing both sensing delay and power consumption [18], [19]. Similar to the alpha latch design, full swing voltage at nodes C and D is transferred to the data-line differential voltage by the means of a pair of nMOS transistors, as shown in Fig. 1(c).

III. PROPOSED SA

The proposed SA, coupled with a simplified read-cycle-only memory system, is presented in Fig. 2. It consists of two sensing stages: local and global. The local sensing stage is formed by four pMOS (P3–P6) and three nMOS (N1, N2, and N7) transistors. While P3 and P4 act as a column switch, the rest of the transistors establish the local cross-coupled inverters, which are responsible for generating the BL differential currents and transferring them to the DLs. The global sensing stage consists of three pMOS (P7–P9) and five nMOS (N3–N6 and N8) transistors. In Fig. 2, two output inverters, which serve as buffers to drive the potentially large output loads to full CMOS logic output levels, are also included. The operation of the proposed SA is described as follows.

During the standby period, P3 and P4 are turned off to block any BL currents. The Column Select and Global Enable (CS and GEN) signals turn on N7 and N8 respectively to equalize nodes A, B and C, D to the same potential, respectively. Meanwhile, two pre-charge transistors N5 and N6 are turned on to pull both DLs to ground. At the same time, P9 is turned off to save power. Since P9 is off and the DLs are precharged to ground, C and D are also at a low potential (near V_{th}) during standby. The two output inverters are also cutoff by P9, as shown in Fig. 2. This topology ensures that the standby current of the circuit, and thus the power dissipation are minimized.

Consider both RS1 and CS2 being activated during a read operation. The precharge signal (PRE) turns N5 and N6 off, allowing the DL voltages to change freely. The memory cell at the upper row and right column will be selected, resulting in a small cell current I_{cell} flowing from the $\overline{BL}$ into the cell as shown in Fig. 2 and discharges the $\overline{BL}$ to a voltage level lower than that of the BL. As CS2 is triggered low, P3 and P4 are turned on to transfer the BL potentials and BL currents to the inputs of the local cross-coupled inverters. At the same time, N7 is turned off to activate the local cross-coupled inverters. This building block senses the voltage and current difference at the source terminals of P5 and P6 and quickly finishes its latching process. Hence, node A is pulled to V_{DD} while node B is discharged to the same potential of the $\overline{DL}$, which is near ground, as shown in Fig. 3 [18]. More importantly, during this latching process, the pulsing current flowing from N2 to $\overline{DL}$, i.e., I_2 , is much higher than that from the N1 to the DL, i.e., I_1 , as shown in Fig. 3. This phenomenon can be intuitively explained as follows. During standby, nodes A and B reside at a low potential near V_{th} . Once the sense amplifier is activated, both node potentials will slightly rise and then quickly start to deviate. For example, in Fig. 3, node A approaches near V_{DD} while node B

Fig. 3. Waveforms at several nodes of the proposed SA during a read cycle.

plunges to near ground. Thus, transistor N1 is in cutoff most of the time. On the other hand, transistor N2 operates in triode region and then moves to saturation region, resulting in a much larger pulsing current when compared to that of N1. Integrating these two currents over time we get the total charges flowing to DL and $\overline{DL}$, respectively.

These differential currents flow to the DLs and induce a voltage difference on the global data-lines. Similarly, this voltage difference is amplified by the global sensing stage to the intermediate outputs V_C and V_D , also shown in Fig. 2. These two voltages are then fed to the output buffers to get the full CMOS logic levels. It is worth mentioning that the global sensing stage can only be activated after the latching process of the local amplifier has completed. The waveforms of several nodes of the proposed SA during a read cycle are also shown in Fig. 3. This hierarchical two-level sensing scheme helps reducing both power consumption and sensing delay imposed by the bit-lines and the data-lines on high density SRAM designs. Furthermore, although nodes A and B have a near-full-swing during a read operation, they can not be tapped directly to the data-lines. Otherwise, the total power consumption and sensing delay will be increased dramatically. As a result, a global sensing stage is required to amplify the small differential signal on the data-lines to a full CMOS logic level at the output of the SRAM.

The total active power dissipated in the proposed SA is limited by the cell current flowing from one of the BLs to the node of the cell where a “0” is stored (which solely depends on the cell design) and the switching currents of the sensing stages. After latching, the cross-coupled configuration stays at one of its bi-stable stages and no additional current is consumed and hence, power dissipated on the BLs and DLs is optimized. Fig. 4 below shows the prelayout transient waveforms of several nodes of the proposed design during a read cycle at 1 GHz. Sensing delay is defined from the time when CS signal reaches half- V_{DD} to the time when the differential output reaches half- V_{DD} .

Since the global data-lines are shared among many columns, their parasitic capacitances are significant and have an impor-

Fig. 4. Output waveforms at 1 GHz.

tant impact on the input margin of the global sensing stage. The voltage difference on the data-lines must be larger than the input offset voltage of the global sense amplifier in order to perform a correct readout. Thus, number of columns sharing the data-lines should be considered carefully to maintain a reasonable input margin. It is determined by the size of the MOS transistors in the local sense amplifier (i.e., N1–N2 and P5–P6) and the layout dimension of memory cell (as it affect the length of the data-lines and hence their parasitic capacitances). This number does not depend on the technology as it can be adjusted by changing the size of the transistor in the local sense amplifier. Our analysis indicated that to maintain an input of at least 100 mV to the global sense amplifier at 1 V voltage supply and 1.25 GHz operating frequency (as will be mentioned in Section VI-C), number of columns sharing the data-lines must not exceed 164.

IV. SIMULATION AND DESIGN METHODOLOGY

A. Test Structure

All the sense amplifiers in comparison, i.e., [12], [18]–[20] and the proposed circuit have been extensively simulated using four identical 32×32 -bit SRAM cores. Each column of the core has one local sense amplifier which transfers the signal to the data-lines for global sensing. The orders in which the memory cells are activated are identical for all four designs. Furthermore, lump-sum C_{BL} and C_{DL} are connected to the bit- and data-lines to model additional parasitic capacitance in bigger SRAM macro. As a simple approximation, each row contributes 1 fF to the bit-line capacitance and each column contributes 1 fF to the data-line. It means that if $C_{BL} = 100$ fF and $C_{DL} = 150$ fF, our structure is equivalent to a SRAM macro of 132 rows and 182 columns. This facilitates the needs to vary both C_{BL} and C_{DL} for investigation. It also reduces the simulation time with reasonable accuracy. Detailed investigations for various C_{BL} and C_{DL} parasitic conditions and supply voltage V_{DD} have also been performed to gauge the robustness of the designs. C_{DL} and C_{BL} are swept from 100 to 200 fF simultaneously while V_{DD} is swept from 0.2 to 1 V. Besides the sensing delay and the average power consumption, power-delay product (PDP) is used

Fig. 5. Latching delay distributions of the three designs using Monte Carlo simulation at room temperature, 1 V supply voltage, 100 mV differential input. Number of iteration is 1000.

as the main performance indicator which takes both entities into consideration. The transistor sizes of different designs of SAs have also been fully optimized to achieve the minimum PDP.

B. Circuit Optimization

All transistors in the readout circuits of the four designs have a constant channel length of 65 nm and parameterized channel widths. Each circuit is then optimized using a systematic parameter sweeping methodology. To ensure the fairness of the comparison, transistor widths are set to obtain the minimum PDP at 1 V supply voltage and $C_{DL} = C_{BL} = 100$ fF. Parasitic capacitances are extracted and back-annotated from the layout view to the schematic view to perform post-layout simulations. All results presented in Figs. 5–13 are based on post-layout simulation results.

C. Speed Deviations

In digital and memory circuit, time matching is vital since it ensures that sufficient input voltage is available to be amplified. If the output signal of one stage is slowed down, the input of the next stage may be smaller than the input-offset voltage, resulting in a wrong sensing. This issue is even more critical in highly compact SRAM macros, due to their heavily loaded bit- and data-lines, which are likely to cause signal mismatches. Therefore, each sensing stage should have a very stable sensing delay to minimize the above-mentioned mismatches. Thus, speed deviations due to inter-die variations of the circuits in comparisons must be evaluated. These are done with the SA alone as well as in the context of 32×32 -bit SRAM macro. Monte Carlo simulations are performed with inter-die variations to monitor the stability of the circuits and simulation results are presented in Figs. 5 and 6. All circuits are simulated at a power supply of 1 V, $C_{DL} = 100$ fF, $C_{BL} = 100$ fF, $C_L = 20$ fF and clock frequency of 250 MHz. The latching delay is defined as the interval

Fig. 6. Total sensing delay distributions of the designs in comparison using Monte Carlo simulations at room temperature. Number of iterations is 200. The numbers in the brackets explain the mean and standard deviation in sensing delay of each design.

from $0.5 V_{DD}$ of the enable signal of the sense amplifier to the time when the differential output of the sense amplifier is $0.5 V_{DD}$. The total sensing delay is measured from $0.5 V_{DD}$ of the CS signal to the time when the final differential output of the SRAM reaches $0.5 V_{DD}$, as illustrated in Fig. 4.

D. BER Consideration

In this work, we investigate the input-offset quality of the sense amplifier designs. Therefore, our BER investigations are only performed on the sense amplifiers alone. In this work, BER refers to the failure rate of the sense amplifier at some specific condition, not the memory cell. Since the input offset voltage is the main cause of read failure and is more critical to the cross-coupled based sense amplifiers; only three designs investigated, namely the proposed, the decoupled latch and the alpha latch. The BL voltage is set to V_{DD} . The input voltage is defined as the difference between BL and $\overline{BL}$. All simulations are performed using Monte Carlo simulations, taking both process variations (inter-die) and device mismatches (intra-die) into considerations. Device variations are from foundry-given data with all parameters considered simultaneously (i.e., doping level, V_{th} , W , etc.). Number of iterations is 35 000. Simulation results are shown in Fig. 7.

E. Maximum Operating Frequencies at Various Supply Voltages

As the supply voltage scales down, the maximum operating frequency of the SRAM also reduces. For each supply voltage from 1 V down to 0.2 V, we consider the maximum frequency at which the sense amplifiers are able to work correctly. Performance comparisons are also carried by monitoring the sensing delay and power consumption per MHz. All transistor sizes are kept unchanged, as obtained in Section IV-B.

Fig. 7. BER of the three cross-coupled based sense amplifier using Monte Carlo simulations with 35 000 iterations. a) BER versus supply voltage, input voltage equals to $0.1 V_{DD}$. b) BER versus input voltage at $V_{DD} = 1$ V.

V. SENSITIVITY ANALYSIS

A. Process Variations

As CMOS technology scales down, process variations are becoming predominant concerns in designing VLSI system, especially in SRAM where device geometries are especially small. It is therefore critical for a SA to work properly not only under power supply fluctuations but also process variations.

In this work, a detailed sensitivity analysis has been carried out to investigate the operation of the four designs by using the process data from the foundry. While the latching delay analysis is only performed on the three cross-coupled based sense amplifiers, the total sensing delay analysis is carried out on all four designs, with the current-conveyor based high-speed sense amplifier used as a reference circuit. Circuit setups for these simulations are shown in Figs. 1 and 2.

Fig. 5 shows the latching delay distribution of the proposed, the decoupled latch and the alpha latch. It is evident that the proposed design offers the best latching delay with the smallest mean value (161 ps) and a standard variation (13 ps) similar to

Fig. 8. Sensing delay versus C_{BL} and C_{DL} variations for the circuits in comparison $C_L = 20$ fF.

Fig. 10. PDP versus C_{BL} and C_{DL} variations for the circuits in comparison $C_L = 20$ fF.

Fig. 9. Power versus C_{BL} and C_{DL} variations for the circuits in comparison $C_L = 20$ fF.

that of the decoupled latch (15 ps). This can be explained as the proposed design has the smallest capacitive load at the switching nodes (nodes A and B in Fig. 2) compared to those of the alpha latch [nodes A and b in Fig. 1(b)] and the decoupled-latch [nodes C, D in Fig. 1(c)]. Furthermore, it contains the least number of transistor hence, its variations is smallest.

Fig. 6 illustrates the total sensing delay distribution of the three above-mentioned circuit with the high-speed design added as a reference. It is accordant with the data shown in Fig. 5 where the proposed and the decoupled designs offer the best performance. It is evident that all three cross-coupled based sense amplifiers are more reliable with much smaller mean values and standard deviations, also shown in Fig. 6. For example, the proposed design is $3.6\times$ faster than the high-speed design and its delay standard deviation is almost $10\times$ smaller.

Fig. 11. Layout of four local SA designs in consideration. From left to right: Proposed, high-speed, decoupled latch and alpha latch. V_{SS} signal runs horizontally and is not shown in this figure.

B. Device Mismatches

Device mismatches refer to intra-die variations, which is caused by local random variations during fabrication. In the sensing circuit, this issue is more critical than inter-die variations as it is the main cause of the input offset voltage which in turn leads to a wrong sensing if the input swing is smaller than the required offset value.

Fig. 7(a) and (b) show the BER of the three cross-coupled based caused by the device mismatches in various supply and input conditions, respectively. Both figures show that the proposed circuit has a smaller BER at every condition. For example,

Fig. 12. Leakage currents of the global and local sense amplifiers of four designs versus operating temperature.

Fig. 13. Maximum operating frequency of four circuits in comparison at different the supply voltages. $C_L = 20$ fF, $C_{BL} = 100$ fF, $C_{DL} = 100$ fF. Room temperature.

at 1 V voltage supply and 110 mV input, BER of the proposed, decoupled latch and alpha latch are 171, 20 171, and 75 532 part-per-million (ppm), respectively. This is because the proposed design has the least transistor count (4 versus 6). Although the BER of the proposed design increases drastically when the supply voltage scales down [see Fig. 7(a)], it is still smaller than the other two designs. Furthermore, this trend saturates when V_{DD} approaches 0.5 V and still ensures better performance than its counterparts down to 0.2 V supply voltage.

In contrast of Fig. 7(a) and (b) presents three parallel lines which indicate a predictable behavior of all three designs when input voltage changes. At 1 V supply voltage, the BER of the proposed design is at least $50\times$ better than the other designs. As the proposed design suffers less from the process variations (Figs. 5–7), it scales better with technologies. Therefore it is reasonable to conclude that the proposed design is more reliable than the other latch-based topologies and hence more suitable for applications where reliability is of crucial concern.

VI. PERFORMANCE COMPARISONS

A. Power Consumption and Sensing Delay

Performance indicators (sensing delay, power consumption and PDP) of the above-mentioned circuits are graphically presented in Figs. 8 to 10. Fig. 8 compares the sensing delay of the four designs with respect to C_{BL} and C_{DL} variations, respectively. It is apparent that all four designs are insensitive to both C_{BL} and C_{DL} , manifested by the almost-horizontal surfaces. This is because all switching nodes are isolated from the highly loaded bit-lines and data-lines. However, data-line capacitance has a greater impact on the performance of the circuits with a higher slope along the data-line capacitance axis. This figure also demonstrates the superiority of the proposed design over the other circuits at 1 V supply voltage against C_{BL} and C_{DL} variations, respectively. For example, at $C_{BL} = 100$ fF, $C_{DL} = 100$ fF, and $C_L = 20$ fF, its sensing delay is reduced to 21.3%, 72.8%, and 27.6% of that of the high-speed [12], decoupled latch [18], and alpha latch [20], respectively. This observation is consistent over a wide range of parasitic conditions, also shown in Fig. 8.

A similar observation can be seen in Fig. 9, regarding the power consumptions of the four circuits. For example, at the same working condition as above (i.e., at $V_{DD} = 1$ V, $C_{BL} = 100$ fF, $C_{DL} = 100$ fF, and $C_L = 20$ fF) the power consumption of the new design is reduced to 70.2%, 34.7%, and 64.3% of that of the high-speed [12], decoupled latch [18], and alpha latch [20], respectively. This is because the output of the local sensing stage in our design has very low voltage swing and thus can be tapped directly to the data-lines. Furthermore, after latching, no bit-line current is flowing from the bit-lines to the data-lines. This is in contrast with the other designs in which at least one bit-line current flows from the bit-lines to the data-lines. Thus, the PDP of the proposed design is more than 74% superior as compared to other designs, as shown in Fig. 10. In addition, the proposed circuit achieves the most stable behavior with a total change across the simulated regions (i.e. C_{DL} ranges from 100 to 200 fF and C_{BL} ranges from 100 to 200 fF) of 6.5% whereas that of the high-speed [12], decoupled latch [18] and alpha latch [20] are 10.9%, 17.3%, and 34.2%, respectively. Table I summarizes the comparison of these four designs, including the layout area of each topology. As shown in Table I and Fig. 11, the proposed local design occupies the smallest active area, which is only 79%, 67%, and 64% of that of the high-speed, decoupled latch and alpha latch designs, respectively. All transistor sizes are obtained from the circuit optimization mentioned in Section IV-B.

B. Leakage Consideration

Leakage currents of the four sense amplifiers are investigated at various operating temperature using DC analysis. All four sense amplifiers (see Figs. 1 and 2) are turned off by setting their control signals to either V_{DD} or 0 V. At the same time, V_{DD} is kept at 1 V and temperature is swept from 0 °C to 125 °C, to cover with the commercial standard range. Simulation results are shown in Fig. 12. As the proposed local design has only seven transistors cascaded into two branches (see Fig. 2), it has the smallest leakage current, as illustrated by the black

TABLE I

COMPARISON SUMMARY OF THREE CIRCUITS FOR $C_L = 20$ fF, $C_{BL} = 100$ fF, $C_{DL} = 100$ fF AT 65-nm CMOS TECHNOLOGY AND 250 MHz FREQUENCY. ALL DESIGNS HAVE THE SAME LAYOUT WIDTH OF $1.6 \mu\text{m}$ TO FIT ONE COLUMN PITCH

	Sensing delay, ps	Average power, μW	PDP, fJ	Layout area, μm^2
Proposed	156	25.58	3.99	8.64
High-speed [12]	732	36.40	26.64	10.88
Decoupled latch [18]	214	73.69	15.77	12.80
Alpha Latch [20]	566	39.76	22.50	13.44

curves in Fig. 12. For example, at room temperature, leakage currents of the proposed, decoupled latch, alpha latch, and high speed designs are 9, 19, 17, and 18 nA, respectively. Similarly, the proposed global sense amplifier also offers the least leakage although the difference between the four designs is not significant. The reason is because all four designs contain two pairs of output buffers which contribute a large portion to their total leakage. These two observations confirm that the proposed design consumes the least standby power and hence enabling longer battery life of the system.

C. Operating Frequency

We aim to design a new SA that can work with a clock frequency higher than 1 GHz. Furthermore, we also study the maximum frequency of each design at several supply voltages, as shown in Fig. 13. It is noticeable that the high-speed design ceases to work at a supply voltage of 0.3 V. As shown in Fig. 13, the proposed design and the decoupled-latch have similar maximum operating frequency at every supply voltage and about $2\times$ and $4\times$ higher than that of the alpha latch and the high-speed circuits, respectively. This agrees with the data presented in Fig. 14, as the proposed design and the decoupled latch have similar sensing delay. However, power consumption per MHz of the proposed design is smaller than that of the decoupled latch, which is even higher than that of the alpha latch, as both shown in Fig. 14 and Fig. 9. Fig. 14 also clearly indicates that the current-conveyor-based high-speed sense amplifier has the largest sensing delay as well as power consumption. This conclusively proves the superiority of the proposed circuit when both stability and performance are of critical design specifications.

VII. CONCLUSION

A latch-type SA has been presented, offering both speed, and power improvements when compared to the existing circuit topologies. Furthermore, it can operate with clock frequency as high as 1.25 GHz, which is the highest among the circuits in consideration. The sensitivity analysis carried out across process corners has reaffirmed that the new design can tolerate excessive process variations with smallest performance fluctuations. It also provides better reliability with at least $50\times$ BER at 1 V supply voltage. In view of the above, it can be concluded that the new SA is best suited for applications where low-voltage, low-power, high-speed and stability are of crucial design considerations.

Fig. 14. Maximum operating frequency of four circuits in comparison at different the supply voltages. $C_L = 20$ fF, $C_{BL} = 100$ fF, $C_{DL} = 100$ fF. Room temperature.

REFERENCES

- [1] B. D. Yang and L. S. Kim, "A low-power SRAM using hierarchical bit line and local sense amplifier," *IEEE J. Solid-State Circuits*, vol. 40, no. 6, pp. 1366–1376, Jun. 2005.
- [2] E. Grossar, "Technology-aware design of SRAM memory circuits," Ph.D. dissertation, Dept. Electron., Katholieke Univ., Leuven, Belgium, 2007.
- [3] H. I. Yang and M. H. Chang, "A low-power low-swing single-ended multi-port SRAM," in *Proc. Techn. Pap. Int. Symp. VLSI Des., Autom. Test (VLSI-DAT)*, 2007.
- [4] E. A. Ramy and A. B. Magdy, "Low-power cache design using 7T SRAM cell," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 4, pp. 318–322, Apr. 2007.
- [5] S. Sundaram, P. Elakkumanan, and R. Sridhar, "High speed robust current sense amplifier for nanoscale memories: A winner take all approach," in *Proc. 19th Int. Conf. VLSI Design Held Jointly With 5th Int. Conf. Embedded Syst. Des.*, 2006, pp. 569–574.
- [6] K. Itoh, M. Horiguchi, and M. Yamaoka, "Low-voltage limitations of memory-rich nano-scale CMOS LSIs," in *Proc. 33rd Eur. Solid State Circuits Conf. (ESSCIRC)*, 2007, pp. 68–75.
- [7] J. L. Shin, B. Petrick, M. Singh, and A. Leon, "Design and implementation of an embedded 512-KB level-2 cache subsystem," *IEEE J. Solid-State Circuit*, vol. 40, no. 9, pp. 1815–1820, Sep. 2005.
- [8] E. Seevinck, P. J. V. Beers, and H. Ontrop, "Current-mode techniques for high-speed VLSI circuits with application to current SA for CMOS SRAM's," *IEEE J. Solid-State Circuits*, vol. 26, no. 5, pp. 525–536, May 1991.
- [9] S. Katsuro, I. Koichiro, U. Kiyotsugu, K. Kunihiro, H. Naotaka, T. Hiroshi, K. Fumio, Y. Toshiaki, and S. Akihiro, "A 7-ns 140-mW 1-Mb CMOS SRAM with current sense amplifier," *IEEE J. Solid-State Circuits*, vol. 21, no. 11, pp. 1511–1518, Nov. 1992.
- [10] K. S. Yeo, "New current conveyor for high-speed low-power current sensing," *IEE Proc. Circuits Dev. Syst.*, vol. 145, no. 2, pp. 85–89, Apr. 1998.
- [11] A. Hajimiri and R. Heald, "Design issues in cross-coupled inverter sense amplifier," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1998, vol. 2, pp. 149–152.
- [12] K. S. Yeo, W. L. Goh, Z. H. Kong, Q. X. Zhang, and W. G. Yeo, "High-performance, low-power current sense amplifier using a cross-coupled current-mirror configuration," *IEE Proc. Circuits Dev. Syst.*, vol. 149, no. 5–6, pp. 308–314, Oct./Dec. 2002.
- [13] R. Singh and N. Bhat, "An offset compensation technique for latch type sense amplifier in high-speed low-power SRAMs," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 6, pp. 652–657, Jun. 2004.

- [14] H. C. Chow and S. H. Chang, "High performance sense amplifier circuit for low power SRAM applications," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2004, vol. 2, pp. 741–744, pt. 2.
- [15] C. L. Hsu, M. H. Ho, and C. F. Lin, "New current-mirror sense amplifier design for high speed SRAM applications," *IEICE Trans. Fundam. Electron., Commun., Comput. Sci.*, vol. E.89-A, pp. 377–384, 2006.
- [16] M. Golden, J. Tran, B. McGee, and B. Kou, "Sense amp design in SOI," in *Proc. IEEE Int. SOI Conf.*, 2005, pp. 118–120.
- [17] S. Ardalan, D. Chen, M. Sachdev, and A. Kennings, "Current mode sense amplifier," in *Proc. 48th Midw. Symp. Circuits Syst.*, Aug. 2005, vol. 1, pp. 17–20.
- [18] R. Singh and N. Baht, "An offset compensation technique for latch type sense amplifiers in high-speed low-power SRAMs," *IEEE Trans. VLSI Syst. Trans. Briefs*, vol. 12, no. 6, pp. 652–657, Jun. 2004.
- [19] S. J. Lovett, G. A. Cibbs, and A. Pancholy, "Yield and matching implications for static RAM memory array sense-amplifier design," *IEEE J. Solid-State Circuit*, vol. 35, no. 8, pp. 1200–1204, Aug. 2000.
- [20] B. Witch, T. Nirschl, and D. S. Landsiedel, "Yield and speed optimization of a latch-type voltage sense amplifier," *IEEE J. Solid-State Circuit*, vol. 39, no. 7, pp. 1148–1158, Jul. 2000.
- [21] A. Kawasumi, T. Yabe, Y. Takeyama, O. Hirabayashi, K. Kushida, A. Tohata, T. Sasaki, A. Katayama, G. Fukano, Y. Fujimura, and N. Otsuka, "A single-power-supply 0.7 V 1 GHz 45 nm SRAM with an asymmetrical unit- β -ratio memory cell," in *Proc. Solid-State Circuits Conf.*, 2008, pp. 382–383.
- [22] N. Verma and A. P. Chandrakasan, "A high-density 45 nm SRAM using small-signal non-strobed regenerative sensing," in *Proc. Solid-State Circuits Conf.*, 2008, pp. 380–381.

Kiat-Seng Yeo received the B.Eng. (honors) degree in electronics, and the Ph.D. in electrical engineering from Nanyang Technological University, Singapore, in 1993 and 1996, respectively.

He began his academic career as a Lecturer in 1996, and was promoted to Assistant Professor, Associate Professor, and Full Professor in 1999, 2002, and 2009, respectively. He was Sub-Dean (Student Affairs) from 2001 to 2005. During this period, he held several concurrent appointments as Program Manager of the System-on-Chip flagship

project, Coordinator of the Integrated Circuit Design Research Group and Principal Investigator of the Integrated Circuit Technology Research Group at NTU. He is currently a board member of Microelectronics IC Design and Systems Association of Singapore (MIDAS), a member of the Advisory Committee of the Centre for Science Research and Talent Development of Hwa Chong Institution, Chairman of the Advisory Committee of Dazhong Primary School and consultants/advisors to several statutory boards and multinational corporations in the areas of semiconductor devices, electronics, and integrated circuit design. He currently heads the Division of Circuits and Systems and is also the Interim Director of the IC Design Centre of Excellence at NTU. His research interests include device characterization and modeling, RF IC design, and low-voltage low-power IC design. He has successfully completed 10 research projects amounting to more than \$3 million and is currently the Principal Investigator for several ongoing research projects of over \$10 million. He has authored five books: *Intellectual Property for Integrated Circuits* (J. Ross Publishing, International Edition, 2009), *Design of CMOS RF Integrated Circuits and Systems* (World Scientific Publishing, International Edition, 2009), *Low-Voltage, Low-Power VLSI Subsystems* (McGraw-Hill, International Edition, 2005), *Low-Voltage Low-Power Digital BiCMOS Circuits: Circuit Design, Comparative Study, and Sensitivity Analysis* (Prentice-Hall, International Edition, 2000), and *CMOS/BiCMOS ULSI: Low-Voltage, Low-Power* (Prentice-Hall, International Edition, 2002). The latter was translated to a Chinese version and is currently one of the excellent foreign textbooks in China. He has filed/granted 17 international patents and published over 250 articles on CMOS/BiCMOS technology and integrated circuit design in leading technical journals and conferences worldwide.

Prof. Yeo is the General Chair and General Co-Chair of 2009 and 2007 International Symposium on Integrated Circuits, respectively, and Technical Chair of 1999 and 2001 International Symposium on Integrated Circuits, Devices, and Systems. He also served in the program committee of the International Symposium on VLSI Technology, Systems, and Applications (VLSI-TSA), Taiwan, from 1999 to 2005, the International Symposium on Low-Power and High-Speed Chips (COOL Chips) in Japan since 2002, 2006 IEEE Asia Pacific Conference on Circuits and Systems, 2005 and 2006 IEEE International Workshop on Radio-Frequency Integration Technology, and several other international conferences. He is a technical reviewer for several prestigious international journals and was listed in Marquis Who's Who in the World, Marquis Who's Who in Science and Engineering, International Who's Who of Professionals, Madison Who's Who, Leading Engineers of the World 2006, First Edition (2006–2007) of Who's Who in Asia, Academic Keys Who's Who in Engineering Higher Education and Continental Who's Who. He was a recipient of the Public Administration Medal (Bronze) on National Day 2009 and the Nanyang Alumni Achievement Award in 2009.

Jeremy Yung Shern Low originates from Malaysia. He received the B.Eng. (honors) degree in electronics from Nanyang Technological University (NTU), Singapore, in 2009. He is currently pursuing the Ph.D. degree under the division of Circuit and System, EEE.

His area of research is high speed and fault tolerant residue number system (RNS) arithmetic.

Anh-Tuan Do was born in Hanoi, Vietnam, in 1984. He received the B.Eng. (honors) degree in electronics from Nanyang Technological University (NTU), Singapore, in 2007, where he is currently pursuing the Ph.D. degree.

He became a Project Officer with NTU in 2007. His research interests include low-power, high speed SRAM designs, low-leakage and sub-threshold circuits designs, circuit /architecture designs for the emerging probabilistic CMOS (PCMO) technology.

Zhi-Hui Kong received the B.Eng. (honors) degree in electronics from University of Technology, Malaysia, in 2000, and the Ph.D. degree in electrical engineering from Nanyang Technological University (NTU), Singapore, in 2006.

Since Mar 2007, she has been a Teaching Fellow and is currently a Visiting Assistant Professor with the School of Electrical and Electronic Engineering, NTU. From 2000 to 2002, she worked as a Research Engineer with the Institute for Infocomm Research (I2R). She then worked full-time pursuing the Ph.D.

degree from NTU from 2003 to 2004. She became a Project Officer in NTU in 2005 and subsequently converted to Research Fellowship in 2006. Her research interests include digital/mixed-signal circuit designs for low-voltage low-power applications and circuit /architecture designs for the emerging probabilistic CMOS (PCMO) technology.

Dr. Kong was a recipient of a highly competitive research fund as a co-Principal Investigator amounting to more than a quarter million dollars.