On Mitigating Sense Amplifier Offset Voltage Degradation

Daniël Kraak Innocent Agbo Mottaqiallah Taouil Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and CS
Mekelweg 4, 2628 CD Delft, The Netherlands
{D.H.P.Kraak, I.O.Agbo, M.Taouil, S.Hamdioui}@tudelft.nl

Pieter Weckx[1,2] Stefan Cosemans[1] Francky Catthoor[1,2] Wim Dehaene[2]
[1]imec vzw., Kapeldreef 75, B-3001, Leuven, Belgium
[2]Katholieke Universiteit Leuven, ESAT, Belgium
{Pieter.Weckx, Francky.Catthoor}@imec.be
wim.dehaene@esat.kuleuven.be

***Abstract*—The continuous downscaling of CMOS technology causes increased impact of time-zero (due to process variation) and time-dependent variability (due to e.g. Bias Temperature Instability). This causes major challenges w.r.t. the device reliability, which is critical for industries, such as the automotive one. This paper uses an accurate method to quantify the impact of time-zero and time-dependent variability on the offset voltage specification of memory Sense Amplifiers. Furthermore, it presents a run-time design-for-reliability scheme to mitigate the impact of time-dependent variability on the offset voltage specification. The simulation results show that the mitigation scheme improves the offset voltage specification up to ∼40% compared to the standard design after 10 years of aging.**

***Index Terms*—Mitigation, Offset voltage, zero-time variability, run-time variability, SRAM sense amplifier, sensing delay**

I. Introduction

The downscaling of CMOS technology over the past decades has significantly improved the feature size and performance of integrated circuits. However, this downscaling invokes major challenges w.r.t. the device *reliability* [1], [2], which is a critical concern for the automotive, avionics, and aerospace industry. The unreliability is mainly caused by two sources: manufacturing and operational usage [1]. Imperfections in the manufacturing process cause process variations. As a result, the characteristics of devices differ from the intended ones. This process variation is referred to as *time-zero variability*. Variations that occur during the lifetime include *environmental variations* (e.g., supply voltage fluctations and temperature variations) and *aging variations* due to, for instance, *Bias Temperature Instability* [3]; these variations are referred to as *time-dependent variability*. The impact of time-zero and time-dependent variability becomes more severe with CMOS scaling [1], [4]. Traditionally, designers add extra margins to the device, so the circuit is guaranteed to fuction correctly under worst-case conditions; this is referred to as *guardbanding*. However, due to the increased impact of time-zero and time-dependent variability more margin is required, which affects the area, speed, power consumption, and/or yield of the design negatively. As a consequence, advanced technology nodes are currently unsuitable for long-lifetime applications with high reliability requirements, such as the automotive industry. Currently, *fault-tolerant* and *mitigation* schemes are being investigated to counteract the variability in order to make it possible use advanced technology nodes for high reliability applications. This paper focuses on the impact of time-dependent variability on the offset voltage of the memory's Sense Amplifier (SA) and the mitigation of this impact. The SA is very important for high performance memories, as it forms an integral and critical part of the read path delay. The SA behaviour influences the memory delay in two ways: first, a larger SA *offset* requires a larger bitline swing, which means more time must be allocated for the bitline discharge; failing to provide a sufficient swing results in failures in the field. Second, the *delay* from SA trigger to SA output (sensing delay) is on the cricital path. Note that the SA offset voltage is *at least* as important as the SA sensing delay. Therefore, understanding the impact of workload-dependent aging on the memory SA offset voltage and providing appropriate mitigation schemes is an important part of designing a robust and reliable memory system.

A lot of work has been published on the impact of aging and their countermeasures for memory cell arrays [5]–[9]. However, only limited work has been done on the characterization and mitigation of aging in memory peripheral circuits such as SAs [10]–[13]. In [14], a tunable SA is presented to compensate for within-die variations. In [15], the offset voltage is monitored using an on-chip circuit to estimate the yield. Prior work mainly focused on mitigating the SA offset voltage for time-zero variability. The impact and mitigation of time-dependent variability have not been researched.

This paper investigates the impact of time-zero and time-dependent variability on the SA offset voltage. Furthermore, it also proposes a scheme to mitigate the impact of time-dependent variability on the SA offset voltage. The contribution compared to our previous work is that it investigates the impact of aging over 10 years and proposes a mitigation scheme [16].

The rest of the paper is organized as follows: Section II provides a background and discusses Bias Temperature Instability, the targeted standard-latch type sense amplifier, and the used method to calculate the offset voltage specification. Section III discusses the simulation results, where the impact of time-dependent and time-zero variability on the SA offset

voltage is investigated. Section IV discusses the mitigation scheme and the obtained results. Finally, Section V concludes the paper.

II. Background

A. Bias Temperature Instability

Several aging mechanisms exist; e.g., Bias Temperature Instability (BTI) [3], Hot Carrier Injection [17], and Time Dependent Dielectric Breakdown [18]. BTI is considered to be the most important of them and, therefore, it is the focus of this paper [19], [20]. BTI takes place inside the MOS transistors and causes an increment in the threshold voltage (V_{th}). The V_{th} increase happens under *negative gate stress* for PMOS transistors, which is referred to as Negative BTI (NBTI). For NMOS transistors it happens under *positive gate stress*, which is referred to as Positive BTI (PBTI).

To model BTI, this paper uses the atomistic model presented in [21]; it incorporates the dependency on the workload (based on waveform shape), and is based on the capture of traps during stress and relaxation phases of BTI. Each trap contributes to the threshold voltage shift ΔV_{th}. The ΔV_{th} of the transistor is the accumulated result of all gate oxide defect traps. The probabilities of the defects capture P_C and emssion P_E are defined by [22] as follows:

$$P_C(t_{STRESS}) = \frac{\tau_e}{\tau_c+\tau_e}\left\{1 - exp\left[-(\frac{1}{\tau_e}+\frac{1}{\tau_c})t_{STRESS}\right]\right\} \quad (1)$$

$$P_E(t_{RELAX}) = \frac{\tau_c}{\tau_c+\tau_e}\left\{1 - exp\left[-(\frac{1}{\tau_e}+\frac{1}{\tau_c})t_{RELAX}\right]\right\} \quad (2)$$

Here, τ_c and τ_e are the mean capture and emission time constants, T_{STRESS} the stress period, and T_{RELAX} the relaxation period. The impact of temperature is also included in the model [21], [23].

B. Sense Amplifier

The Standard Latch-Type SA shown in Figure 1 will be used as a case study in this work; in principle, the proposed method to calculate the offset voltage specification and mitigation scheme can be applied to other types of SAs, such as look-ahead type sense amplifier [24], double-tail latch-type sense amplifier [25], etc. The working principal of the SA of Figure 1 is as follows: during read operations it amplifies a small voltage difference between bitlines BL and BLBar. Its operation consists of two stages. In the first stage, when SAenable is low, the voltage swing on BL and BLBar is passed to the internal nodes S and Sbar of the sense amplifier. In the second phase, when SAenable is high, the amplification takes place by the cross-coupled inverters. During this amplification the pass transistors disconnect the internal nodes from BL and BLBar. The cross-coupled inverters get current through Mtop and Mbottom and produce the outputs Out and Outbar.

C. Offset Voltage Specification

The method to calculate the offset voltage specification is taken from [16]. This method performs Monte Carlo simulations, where time-zero (i.e., local process variation) and time-dependent variability (i.e., variation due to aging and temperature) are simulated. During each simulation the offset voltage

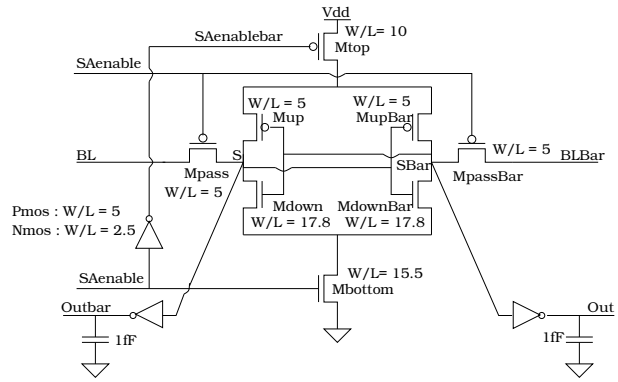

Fig. 1: Standard Latch-Type Sense Amplifier

of one specific sample is determined using a binary search on its inputs. From the Monte Carlo simulations the average and standard deviations of the samples are determined. The offset voltage of SAs typically follows a normal distribution and the relation between this distribution and failure rate is as follows [16]:

$$\int_{V_{in}=-V_{Offset}}^{V_{Offset}} \mathcal{N}(\mu_{MC}, \sigma_{MC}) = 1 - f_r \quad (3)$$

Here, V_{in} represents the input voltage of the SA, V_{offset} the offset voltage specification, $\mathcal{N}$ a normal distribution of the offset voltage with mean μ_{MC} and standard deviation σ_{MC}. In this equation, all SA instantiations that require an input offset voltage outside the range [-V_{offset}, V_{offset}] result in a failure. Using Equation 3, it is possible to calculate the offset voltage specification of the SA for a certain failure rate. In this work, a failure rate target f_r=10^{-9} is assumed; thus, targeting an application with *high reliability* requirement. This required failure rate leads to a $V_{offset} = 6.1 \cdot \sigma_{MC}$ (roughly 6σ) for a distribution with a mean of 0 [16]. Besides process variation, the distribution of the offset voltage depends on temperature, supply voltage, and the used workload and stress time. Therefore, the offset voltage specification will be different when these conditions vary.

III. Impact

In [16], we analyzed the impact of time-zero and time-dependent variability on the offset voltage specification of the SA. The experiments were performed for three workloads; unbalanced workloads {r0} (all reads are 0) and {r1} (all reads are 1) and balanced workload {r0r1} (50% {r0}, 50% {r1}); 80% of the executed instructions (e.g., by a CPU) are read instructions. It was determined that temperature variations have the highest impact on the offset voltage specification for time-dependent variability. Especially the unbalanced workloads showed a significant increase in the offset voltage specification. In [16], the effect of time-dependent variability was analyzed after 3 years of operation. In this section the impact of temperature variations is investigated for 10 years of operation.

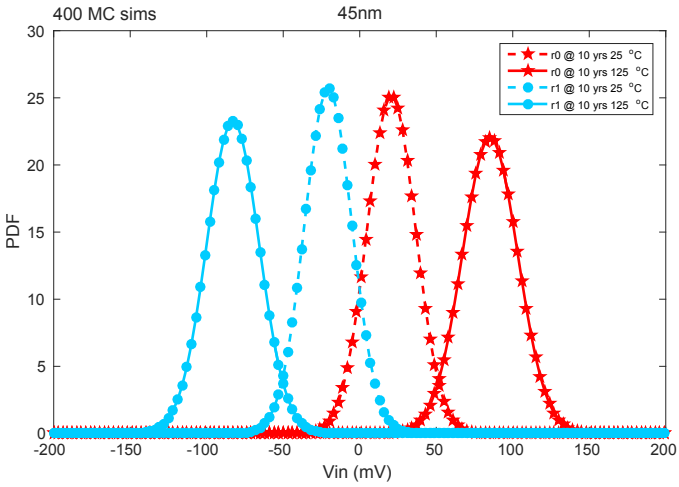

Fig. 2: Temperature impact at run-time for different workloads.

TABLE I: Temperature variation at run-time.

Aging (years)	Workload	Temp. (°C)	μ (mV)	σ (mV)	offset (mV)
0	-	25	0.12	15.7	96.06
10	{r0}	25	20.7	15.9	115.8
10	{r0}	75	51.8	16.8	152.6
10	{r0}	125	85.2	18.1	193.8
10	{r1}	25	-20.6	15.5	113.8
10	{r1}	75	-50.8	16.5	149.6
10	{r1}	125	-82.3	17.1	185.6

Figure 2 shows the temperature dependency of the offset voltage specification at nominal V_{dd}. The figure shows 4 plots: 2 temperatures (25°C and 125°C) $\times$ 2 workloads; the plots for 75°C are omitted for the sake of clarity. However, the μ and σ of its distribution and the offset voltage specification have been added in Table I.

Table I shows that the impact of temperature is significant on the offset voltage specification; for example, after 10 years of operation at 25°C the offset voltage specification is 115.8mV, while this is 193.8mV at 125°C (which is up to $\sim$ 67%) for {r0} workload. This increase in offset voltage specification can be attributed to both a shift of the mean μ and an increase in spread σ of the distributions at higher temperatures. The figure and table show that the impact of temperature on the mean μ of the distributions is significant. For example, for {r0} workload at 25°C, the average shift is 20.7mV, while this is 85.2mV for 125°C. The shift of the mean μ happens due to the fact that only transistors Mdown (MdownBar) and MupBar (Mup) are stressed for sequence {r0} ({r1}); this results in an increased V_{th} for these transistors, leading to a shift in the required offset voltage. Moreover, the standard deviation σ increases, as the temperature rises, irrespective of the workload; e.g., the standard deviation increases from 15.7mV to 18.1mV at 125°C and for {r0} workloads when considering time-dependent variability; however, compared to the impact on the mean, the impact on the standard deviation is considerably lower.

Note that higher offset voltage specification means that more time is needed for the bitline discharging, which makes the overall memory system slower; failing to provide this timing margin causes *unreliability* of the SA design. Therefore, *mitigation schemes* are crucial to enhance the reliability and the performance of the sense amplifier under time-dependent variability.

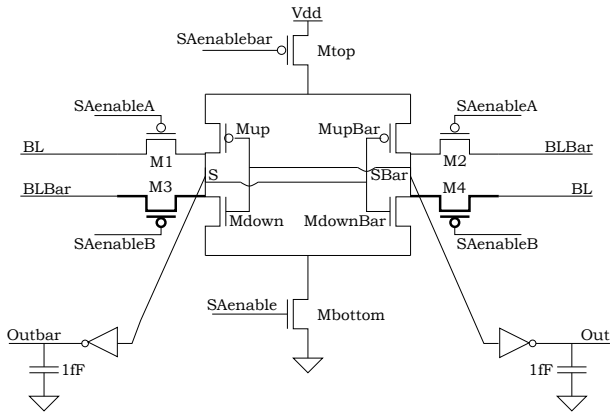

Fig. 3: Input Switching Sense Amplifier

IV. Mitigation

The previous section showed that the SA offset voltage specification increases significantly when an *unbalanced* workload is applied. The increased offset voltage negatively impacts the time needed to produce the read result as more time is needed to discharge the bitlines. In [16], it is shown that a *balanced* workload (reading 50% zeros, 50% ones) has the lowest impact on the offset voltage specification. Therefore, as a mitigation scheme we propose the Input Switching Sense Amplifier (ISSA), where the SA switches its inputs periodically in order to create an on-line control-based balanced workload at its internal nodes. Next, the ISSA design and its control logic are presented.

A. Input Switching Sense Amplifier

Figure 3 depicts the structure of the ISSA. A second pair of pass transistors, M3 and M4, is added compared to the Standard Latch-Type Sense Amplifier. Pass transistor M3 (M4) connects BLBar (BL) to S (SBar). This makes it possible to switch the inputs of the SA, and connect BLBar to S and BL to SBar. This requires, however, additional control circuitry. Pass transistors M1 and M2 are controlled by signal SAenableA and pass transistors M3 and M4 by signal SAEnableB. When SAenableA is low/enabled, pass transistors M1 and M2 forward the voltage level on BL and BLBar to the internal nodes, similar as in Figure 1. Note that in this case the SA operates in the same way as that of Figure 1; the signal SAenableB is disabled/high (i.e., M3 and M4 are off).

When the inputs of the SA are switched (i.e., SAenableB is low/enabled and SAenableA high/disabled), the SA will effectively read the opposite value. Hence, by controlling this switching, it is possible to balance the amount of zeros and ones read by the internal nodes of the SA. This will lead

TABLE II: Truth table for SAenableA and SAenableB

Switch	SAenableBar	SAenableA	SAenableB
0	0	1	1
0	1	0	1
1	0	1	1
1	1	1	0

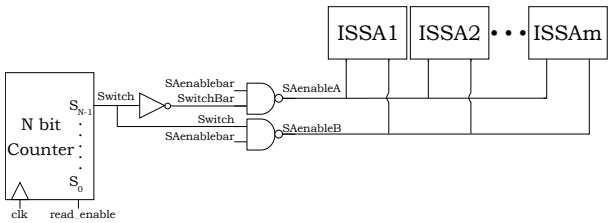

Fig. 4: Control Logic for ISSAs

to a more balanced workload for the SA and mitigate the degradation of the offset voltage. It is worth noting that in the case that the inputs of the SA are switched, the final read value should be inverted (e.g., using additional circuitry).

Figure 4 illustrates the control logic. Two NAND gates are used to generate SAenableA and SAenableB from the original SAenable(bar) and the Switch signal. The Switch signal is generated by an N-bit counter (updated only during reads, controlled by read_enable) and used to decide when the inputs of the sense amplifier should be swapped; for example each 2^{N-1} reads. Table II contains the truth table for SAenableA and SAenableB. When Switch is low (high), only SAenableA (SAenableB) is able to change its value. SAenableB (SAenableA) is always high in this case, to make sure the corresponding pass transistors M3 and M4 (M1 and M2) are off.

B. Simulation Results

Figure 5 shows the temperature impact on the offset voltage at nominal V_{dd} for the ISSA. The results are shown for two temperatures (25°C and 125°C) after 10 years of operation. For comparison, the results of the normal SA from the previous section are also added. Note that for the ISSA the workloads {r0} and {r1} are compiled by the design-for-reliability scheme into the same balanced workload; hence, it is unnecessary to denote the workload. In Table III, the corresponding μ and σ of the distributions and the offset voltage specification can be found.

It can be seen that the ISSA significantly reduces the shift of the distributions compared to the normal SA; all distributions of the ISSA have a mean (μ) of ∼0mV. This happens due to the balanced workload at the internal nodes of the ISSA, resulting in equal stress for transistors M1, M2, M3, and M4. As a result, the offset voltage specification is significantly lower for the ISSA; e.g., the offset voltage specification is 193.8mV for the sequence {r0} at 125°C after 10 years for the normal SA (see Table I), while this is only 114.8mV for the ISSA; a reduction of ∼40%. Hence, less time is needed for the bitline discharge, which makes the overall memory system

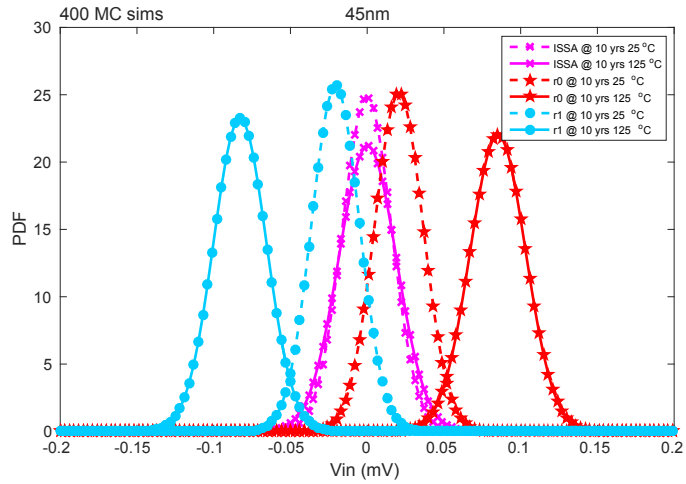

Fig. 5: Temperature variation at run-time for ISSA and normal SA.

TABLE III: Temperature variation at run-time for the ISSA.

Aging (years)	Temp. (°C)	μ (mV)	σ (mV)	offset (mV)
0	25	0.1	14.7	89.9
10	25	-0.1	16.1	98.1
10	75	0.1	17.5	107.2
10	125	0.3	18.8	114.8

faster. From this, we can conclude that the ISSA performs significantly better under high stress (unbalanced workloads, high temperature) for long lifetime-applications.

V. Conclusion

This work clearly shows that dedicated, well-tuned run-time mitigation schemes can be a very good alternative to the traditional guardbanded designs. Not only can they provide optimal design solutions, but they can even extend the lifetime of the devices by balancing the workload. This is extremely important for cutting edge technology as they suffer from reduced lifetime and increased failure rate.

References

[1] S. Borkar, "Microarchitecture and design challenges for gigascale integration," in *Intl. Sympos. on Microarchitecture*, Dec 2004, pp. 3–3.

[2] S. Hamdioui *et al.*, "Reliability challenges of real-time systems in forthcoming technology nodes," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2013*, March 2013, pp. 129–134.

[3] B. Kaczer *et al.*, "Atomistic approach to variability of bias-temperature instability in circuit simulations," in *IRPS*, April 2011, pp. XT.3.1–XT.3.5.

[4] J. Srinivasan *et al.*, "The impact of technology scaling on lifetime reliability," in *DSN*, June 2004, pp. 177–186.

[5] B. Cheng *et al.*, "Impact of nbti/pbti on sram stability degradation," *IEEE Electron Device Letters*, vol. 32, no. 6, pp. 740–742, June 2011.

[6] A. Gebregiorgis *et al.*, "Aging mitigation in memory arrays using selfcontrolled bit-flipping technique," in *20th Asia and South Pacific Design Automation Conference*, Tokyo, Japan, November 2015.

[7] M. Khan *et al.*, "Bias temperature instability analysis of finfet based sram cells," in *DATE*, Dresden, Germany, March 2014.

[8] S.K. Krishnappa *et al.*, "Comparative bti reliability analysis of sram cell designs in nano-scale cmos technology," in *Quality Electronic Design (ISQED), 2011 12th International Symposium on*, March 2011, pp. 1–6.

[9] P. Pouyan *et al.*, "Process variability-aware proactive reconfiguration technique for mitigating aging effects in nano scale sram lifetime," in *2012 IEEE 30th VLSI Test Symposium (VTS)*, April 2012, pp. 240–245.

[10] I. Agbo *et al.*, "Comparative analysis of rd and atomistic trap-based bti models on sram sense amplifier," in *Design Technology of Integrated Systems in Nanoscale Era (DTIS), 2015 10th International Conference on*, April 2015, pp. 1–6.

[11] ——, "Integral impact of bti and voltage temperature variation on sram sense amplifier," in *2015 IEEE 33rd VLSI Test Symposium (VTS)*, April 2015, pp. 1–6.

[12] ——, "Bti analysis of sram write driver," in *2015 10th International Design Test Symposium (IDT)*, Dec 2015, pp. 100–105.

[13] ——, "Comparative bti analysis for various sense amplifier designs," in *2016 IEEE 19th International Symposium on Design and Diagnostics of Electronic Circuits Systems (DDECS)*, April 2016, pp. 1–6.

[14] M.H. Abu-Rahma *et al.*, "Characterization of sram sense amplifier input offset for yield prediction in 28nm cmos," in *2011 IEEE Custom Integrated Circuits Conference (CICC)*, Sept 2011, pp. 1–4.

[15] J. Vollrath, "Signal margin analysis for dram sense amplifiers," in *Electronic Design, Test and Applications, 2002. Proceedings. The First IEEE International Workshop on*, 2002, pp. 123–127.

[16] I. Agbo *et al.*, "Quantification of sense amplifier offset voltage degradation due to zero- and run-time variability," in *IEEE Computer Society Annual Symposium on VLSI*, Pittsburgh, U.S.A., July 2016.

[17] F. Cacho *et al.*, "Hot carrier injection degradation induced dispersion: Model and circuit-level measurement," in *2011 IEEE International Integrated Reliability Workshop Final Report*, Oct 2011, pp. 137–141.

[18] K.B. Yeap *et al.*, "A realistic method for time-dependent dielectric breakdown reliability analysis for advanced technology node," *IEEE Transactions on Electron Devices*, vol. 63, no. 2, pp. 755–759, Feb 2016.

[19] S. Bhardwaj *et al.*, "Predictive modeling of the nbti effect for reliable design," in *CICC*, Sept 2006, pp. 189–192.

[20] R. Newhart, "Early reliability modeling for aging and variability in silicon system, ibm view, ermavss workshop," in *DATE*, 2016.

[21] B. Kaczer *et al.*, "Origin of nbti variability in deeply scaled pfets," in *IRPS*, May 2010, pp. 26–32.

[22] M. Toledano-Luque *et al.*, "Response of a single trap to ac negative bias temperature stress," in *IRPS*, April 2011, pp. 4A.2.1–4A.2.8.

[23] T. Grasser *et al.*, "Analytic modeling of the bias temperature instability using capture/emission time maps," in *Electron Devices Meeting (IEDM), 2011 IEEE International*, Dec 2011, pp. 27.4.1–27.4.4.

[24] T. Asano *et al.*, "Low-power design approach of 11fo4 256-kbyte embedded sram for the synergistic processor element of a cell processor," *IEEE Micro*, vol. 25, no. 5, pp. 30–38, Sept 2005.

[25] D. Schinkel *et al.*, "A double-tail latch-type voltage sense amplifier with 18ps setup+hold time," in *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, Feb 2007, pp. 314–605.