Current-Based Data-Retention-Time Characterization of Gain-Cell Embedded DRAMs Across the Design and Variations Space

Robert Giterman, Andrea Bonetti, Ester Vicario Bravo, Tzachi Noy, Adam Teman, and Andreas Burg, *Member, IEEE*

***Abstract*— The rise of data-intensive applications has resulted in an increasing demand for high-density and low-power on-chip embedded memories. Gain-cell embedded DRAM (GC-eDRAM) is a logic-compatible alternative to conventional static random access memory (SRAM) which offers higher density, lower leakage power, and two-ported operation. However, in order to maintain the stored data, GC-eDRAM requires periodic refresh cycles, which are determined according to the worst-case data retention time (DRT) across process, voltage and temperature (PVT) variations. Even though several DRT characterization methodologies have been reported in literature, they often require unfeasible run-times for accurate DRT evaluation, or they result in highly pessimistic design margins due to their inaccuracy. In this work, we propose an current-based DRT (IDRT) characterization methodology that enables accurate DRT evaluation across process variations without the need for a large number of costly electronic design automation (EDA) software licenses. The presented approach is compared with other DRT characterization methodologies for both accuracy and run-time across several gain-cell structures at different process technologies, providing less than a 4% DRT error and over 100× shorter run-time compared to a conventional DRT evaluation methodology.**

***Index Terms*— Embedded dynamic random access memory (eDRAM), gain-cells (GCs), retention time, embedded memory.**

I. Introduction

With an increasing interest in data-intensive applications, such as neural networks, the silicon area and power consumption of modern system-on-chips (SoCs) is often dominated by embedded memories [1]–[5]. Traditionally, on-chip embedded memories are implemented using the conventional 6-transistor (6T) static random access memory (SRAM) due to its successful scaling across process nodes, high performance, and reliable operation under nominal supply voltages. However, 6T SRAM requires a big area footprint due to the large memory bitcell, it dissipates significant static power due to its high leakage, and it fails to operate reliably under scaled supply voltages. Therefore, when higher memory densities are required, a one-transistor one-capacitor (1T-1C) embedded dynamic random access memory (eDRAM) is preferred over SRAM as it can be implemented with more dense 1T-1C storage cells [1], [6]–[9]. However, the 1T-1C eDRAM requires dedicated process steps to fabricate the cell capacitor resulting in additional complexity and cost, and it suffers from a destructive read operation, which increases power consumption.

Gain-cell embedded DRAM (GC-eDRAM) is an interesting alternative to conventional SRAM and 1T-1C eDRAM [10]–[23]. Compared to SRAM, GC-eDRAM offers higher density due to its lower transistor count, low-leakage, and inherent two-ported operation. Compared to 1T-1C eDRAM, it is fully logic-compatible, thus it does not require additional process steps, and its read operation is non-destructive. The majority of gain cell (GC) implementations reported in literature [10] comprise a write port, a storage node (SN) holding the data on a parasitic SN capacitor, and a read port. In this paper, 3-transistor (3T) and 4T NMOS-only GCs, depicted in Fig. 1, are considered as working examples, as they feature the basic GC characteristics that are common to most GC implementations. Nevertheless, the contributions of this paper immediately apply also to other GC implementations. For the considered GCs, the write port, which is implemented with either a single (Fig. 1(a)) or two (Fig. 1(b)) write transistors (MW and MB) starts conducting when the write word line (WWL) is asserted, allowing to write the data provided on the write bit line (WBL) into the SN. When the GC is not accessed, the data is dynamically stored on the parasitic capacitance associated with the SN and, for the GCs in Fig. 1, the read bit line (RBL) is precharged to a default voltage, "V_{DD}". The read operation is performed once the read word line (RWL) is driven to GND (4T) or V_{DD} (3T), allowing the

Manuscript received May 17, 2019; revised September 3, 2019 and December 16, 2019; accepted January 26, 2020. This article was recommended by Associate Editor R. Azarderakhsh. *(Corresponding author: Robert Giterman.)*

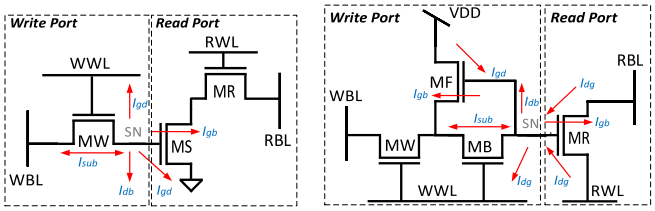
Robert Giterman, Andrea Bonetti, and Andreas Burg are with the Telecommunications Circuits Laboratory (TCL), Institute of Electrical Engineering, EPFL, 1015 Lausanne, Switzerland (e-mail: robert.giterman@epfl.ch).

Ester Vicario Bravo is with TU Delft University, 2628 CD Delft, The Netherlands.

Tzachi Noy and Adam Teman are with the Emerging Nanoscaled Integrated Circuits and Systems (EnICS) Labs, Faculty of Engineering, Bar-Ilan University, Ramat Gan 5290002, Israel (e-mail: adam.teman@biu.ac.il).

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

Digital Object Identifier 10.1109/TCSI.2020.2971695



(a) 3T all-NMOS gain cell. (b) 4T all-NMOS gain cell.

Fig. 1. (a) 3T and (b) 4T all-NMOS gain-cell schematics with the main leakage mechanisms affecting the storage node.

read transistors (MR and MS) to conditionally discharge the RBL depending on value (charge) stored on the SN.

As GCs rely on dynamic storage, the data stored on the SN deteriorates over time due to leakage currents. Therefore, it requires periodic refresh operations to avoid the loss of data. The refresh period for GC-eDRAM is determined by its data retention time (DRT), which is defined as the maximum time interval between a write operation and a successful read operation. The DRT depends on various parameters, such as the stored data, the temperature and operating voltage, the amount of global process variations (GPVs) and local process variations (LPVs), and the operating frequency, among others. Therefore, the memory content is maintained only when the refresh period is set to a smaller value than the lowest DRT of the memory array and according to worst-case operating conditions, which maximize the leakage through the SN.

With technology and voltage scaling, the increased leakage currents and process variations have been shown to reduce the DRT of GC-eDRAMs [14], [20], resulting in increased power consumption and reduced memory availability, which become even more significant when taking into account various design margins to guarantee reliable operation under process, voltage and temperature (PVT) variations. Therefore, an accurate and practical evaluation of the DRT across the design and variation space is crucial for the integration of reliable GC-eDRAMs in nanometer nodes.

The most straightforward and accurate method to evaluate the DRT of GCs is referred to as effective DRT (EDRT) characterization methodology [10]. In this approach, a write operation is applied to the GC, followed by an idle phase, and a subsequent trial-and-error read operations are performed to verify the integrity of the stored data. This test is repeated for increasing idle phases and the DRT is defined as the longest idle phase that still enables to correctly read the stored data. While the EDRT approach provides the most accurate evaluation of the DRT, it becomes impractical to evaluate the DRT of large memory arrays across LPVs, as many trial-and-error transient simulations need to be repeated for each Monte Carlo (MC) run. This time consuming analysis either leads to a prohibitive run-time or it requires a large number of costly electronic design automation (EDA) software licenses. In order to avoid the long run-times of multiple EDRT simulations, modern Spice simulators provide optimization methods which can be used instead of trial-and-error simulations in order to find the maximum read-after-write delay while successfully maintaining the correctness of the output data. While this approach is useful in determining the DRT under a given operating condition, extensive Monte-Carlo simulations would still result in unfeasible run-times.

In order to reduce the DRT evaluation run-time and to avoid the need for high license costs, a voltage-based DRT (VDRT) characterization methodology has been proposed in [10], where the voltage of the SN is monitored. According to the VDRT methodology, the DRT is determined as the time interval after a write operation at which the voltage difference between stored '1' and '0' crosses a critical threshold at which its polarity can no longer be distinguished reliably using a read operation. The simulation time of the VDRT approach is significantly shorter than the time required by the EDRT methodology as it does not require trial-and-error read accesses. Hence, it can be easily used for LPV analyses when executing a high number of MC runs. However, the consideration of the SN voltage only does not account for the impact of LPVs on the GC read port. Thus, the accuracy of the VDRT methodology is limited.

To avoid the limitations of existing strategies to determine the DRT, we present a new, current-based DRT (IDRT) method. The proposed method relies on current monitoring through the read port, thereby taking into account the LPV affecting both the SN voltage deterioration and the RBL driver. The IDRT method is compatible with MC runs as it does not rely on time-consuming trial-and-error transient simulations of the read operation, and it enables an accurate evaluation of the DRT across LPVs without the need for a large number of costly EDA software licenses. The IDRT method is compared with state-of-the-art EDRT and VDRT characterization methodologies in terms of accuracy and run-time for different GC-eDRAM bitcell topologies as well as various process technologies. The IDRT method provides less than 3.7% DRT deviation from the reference EDRT approach, which is over 26× lower than the average error of the VDRT method. Furthermore, the DRTs of both a 3T GC and a 4T GC designed in a 28 nm process technology are measured and analyzed across the design and variations space to demonstrate how the EDRT and IDRT methodologies can coexist for an accurate evaluation of the DRT in GC-eDRAMs.

Contributions: The major contributions of this paper are summarized as follows:

1) This work presents a Current-based DRT (IDRT) method, which provides an accurate DRT evaluation tool without relying on time-consuming trial-and-error transient simulations.
2) The IDRT method is compared with the EDRT and VDRT methods for different GC structures and across different process technologies (i.e., 28 nm FD-SOI and 28 nm Bulk CMOS), demonstrating an average DRT error below 3.7%.
3) The IDRT method is used to explore the design and variation space of 3T and 4T GCs in 28 nm technology.
4) The DRT extracted from the IDRT methodology is used to evaluate key GC-eDRAM characteristics, such as retention power and memory availability for different memory sizes under process variations, to demonstrate how the IDRT methodology can be used to reduce overly pessimistic worst-case refresh.

Outline: The rest of this paper is organized as follows. Section II analyzes the main mechanisms which affect the DRT of GC-eDRAM. The proposed IDRT characterization methodology is described in Section III. Section IV evaluates the accuracy of the IDRT method and compares it with state-of-the-art EDRT and VDRT methodologies. The DRT analysis of 3T and 4T GCs across different design choices as well as global and local variations is provided in Section V. Section VI provides analysis and results of key GC-eDRAM characteristics evaluated at a macro level using the extracted DRT. Finally, Section VII concludes the paper.

II. DRT Analysis

This section describes the different mechanisms, which affect the DRT of GC-eDRAM based on the examples of the 3T all-NMOS and 4T all-NMOS bitcells. The DRT is primarily affected by the leakage currents, which deteriorate the SN voltage, and the read port variation, which affects the RBL access latency during read operations.

A. Storage Node Leakage

The write port of GCs has the most significant effect on the SN deterioration. The most simple version of a GC write port consists of a single transistor (1T), implemented with either an NMOS or a PMOS, which is controlled by the WWL and enables data transfer from the WBL to the SN during a write operation. Fig. 1(a) depicts a 3T all-NMOS GC, which contains a 1T write port implemented with an NMOS device (MW). During standby, the WWL is driven to GND and the voltage of the WBL is determined by the data written to other rows of the memory. When the WBL holds an opposite value to the SN, a sub-threshold leakage (I_{sub}) path is created between the drain and source of MW. In addition, when the cell stores a '1', the SN deteriorates even further through gate leakage (I_{gate}) to WWL and bulk leakage (I_{bulk}) to the substrate of MW, resulting in a faster deterioration of '1' compared to '0'. With technology scaling, the impact of I_{gate} and I_{bulk} have become comparable to that of I_{sub} under nominal conditions [24]. However, due to the exponential dependence of I_{sub} on temperature and process variations, it remains the most significant leakage across corners which deteriorates the SN. In order to reduce the effect of I_{sub} on the SN degradation, a 4T GC with internal feedback was proposed in [20]. The 4T GC, depicted in Fig. 1(b), introduces two additional transistors to the original 1T write port implementation, which form an internal feedback that helps to retain the weaker ('1') voltage level by reducing the leakage from the WBL through MF. However, while the 4T GC reduces I_{sub}, it adds two additional gate leakage components through MF.

The leakage mechanisms through the read port of the GC depend on the number of transistors implementing it. In the case of a 1T read port (Fig. 1(a)), both the RWL and RBL are driven to V_{DD} during standby, increasing the deterioration of a stored '0' through I_{gate} leakage from the diffusions of MR to the SN. On the other hand, the source of MS in the 2T read port (Fig. 1(a)) is driven to GND, hence I_{gate} is drawn from the SN when it holds a '1'. The effect of the various leakage mechanisms on the SN deterioration can be derived from Fig. 2, illustrating the SN voltage after write, as extracted from simulations of the 3T GC under worst-case biasing conditions.

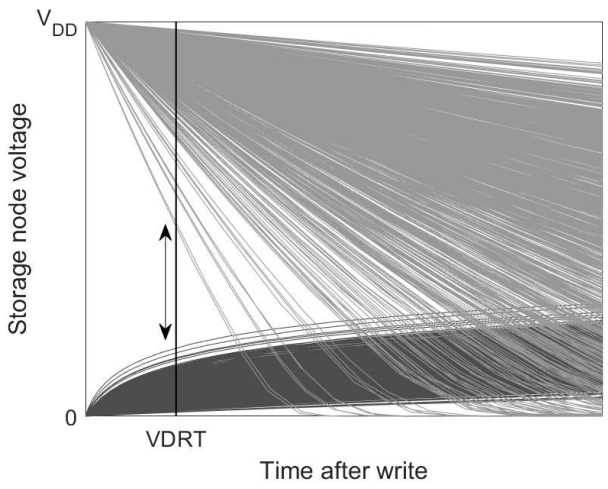


Fig. 2. Voltage-based DRT (VDRT) measurement illustration [10].

B. Read Port Variation

The read port has a strong impact on the DRT through the read access delay, which is determined by the on-current of the read transistors. A 1T read port, depicted in Fig. 1(b), contains a single transistor which is controlled by the SN, and conditionally enables a current path between the RBL and the RWL to either discharge RBL when the cell holds a '1', or keep it charged at V_{DD} when it holds a '0'. The RBL evaluation delay is often the dominant portion of the total read latency of the GC array due to the large RBL capacitance and degraded SN voltage [20]. Moreover, process variations that affect the threshold voltage (V_{T}) of MR have a strong impact on the RBL delay, limiting the DRT when the RBL discharge pace is either too slow (for '1') or too fast (for '0') under the target latency constraints [13]. Finally, a 1T read port is also affected by unselected transistors sharing the same column, which cause the RBL voltage to saturate at a level close to V_{DD}-V_{T} when other cells in the column store a '1'. In order to avoid RBL saturation, a 2T read port is incorporated in the 3T GC, as shown in Fig. 1(a). In this more robust read port, the RWL is connected to the gate of MR, decoupling MS from the RBL when the cell is not selected. Fig. 3 illustrates these phenomena by plotting the RBL voltage distributions as a function of the SN voltage across process variations following a read access from a 1T and 2T read ports. The simulations were conducted under worst-case biasing conditions with all the unselected cells in the column storing a '1', causing the RBL voltage of the 1T read port to saturate at a level close to V_{DD}-V_{T} before completely discharging to GND. On the other hand, the RBL voltage of the 2T read port discharges close to GND when the SN voltage is high enough in order to completely discharge RBL within the 1 ns read access.

The main drawback of the voltage-based VDRT methodology is that it fails to capture the effect of the read port

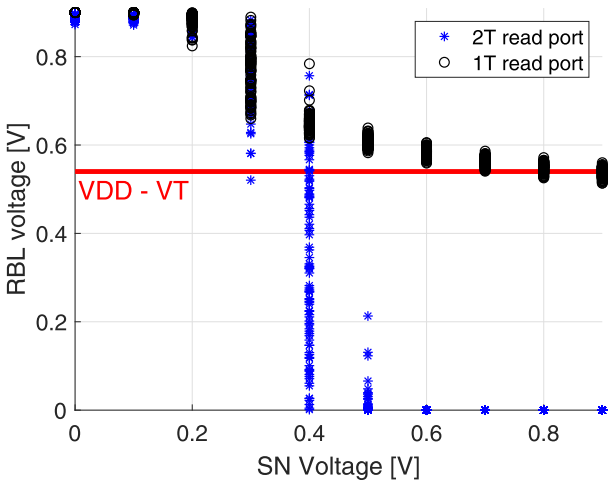


Fig. 3. RBL voltage distributions following 1 ns read access to a 1T read port across SN voltages.

variations on the DRT. In the next section, we present the current-based IDRT methodology, which provides much better accuracy than VDRT as it captures both the leakage and read port variations effects on the DRT.

III. Current-Based DRT (IDRT) Characterization Methodology

In this section, we describe the IDRT evaluation methodology, where the DRT is determined according to the maximum time interval following a write operation until the moment when the current through the active read port crosses a critical threshold at which the read operation for a specified read access latency would fail. The IDRT methodology consists of three phases that comprise:

1. A single EDRT simulation, comprising transferring a '1' or a '0' to SN of the GC, applying the opposite voltage level to the WBL and performing a read operation following a pre-determined idle time.

2. Reference current acquisition, comprising a replica of the GC read port, and applying a voltage source to the replica SN holding the estimated EDRT voltage, as extracted from the previous phase.

3. Evaluation phase, comprising a single GC and a replica GC, where the SN of the GC is connected using an ideal buffer to the replica SN. This test-bench allows to evaluate the IDRT using Monte-Carlo simulations by monitoring the current through the replica read port until it reaches the reference current, which is acquired in phase 2.

These phases are depicted in Fig. 4 for the example of a 3T NMOS-only GC as the device under test (DUT). The IDRT methodology is suitable for the evaluation of the DRT across LPVs as the initial trial-and-error EDRT simulations and the reference acquisition are performed *only once* at the given operating point, while the subsequent MC evaluation phase contains *only a single* transient simulation for each MC iteration. The three phases of the IDRT methodology are elaborately described in the following subsections.

A. EDRT Measurement

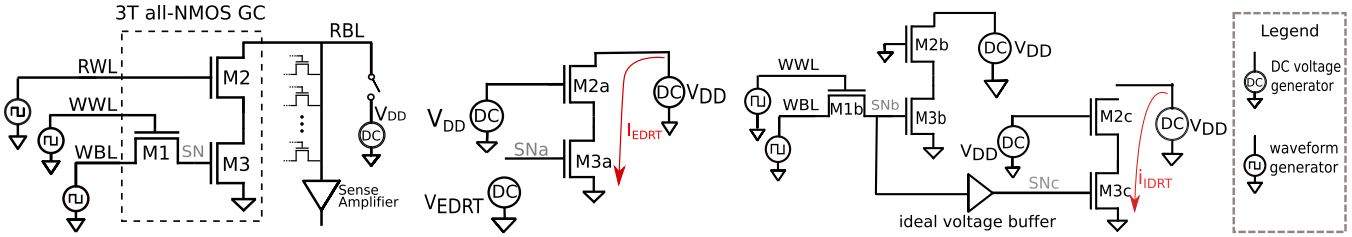
A single EDRT measurement is initially performed at the considered operating conditions, specified by the process corner, supply voltage, temperature, and operating frequency. The load of the GCs sharing the same column is added to the shared RBL and a sense amplifier is used to sense the RBL voltage, as depicted in Fig. 4(a). During the idle phase, when the GC is not accessed, the WBL is forced to the opposite voltage of the written value in order to apply worst-case biasing conditions in which the leakage through the SN is maximized, resulting in a faster deterioration of the stored value. Then, a trial read operation is performed following a determined idle phase and the output is compared with the written value. The simulation is repeated for increasingly longer idle periods until the moment when the output of the read operation is incorrect. The longest idle period which still ensures a correct read operation is used to sample the corresponding voltage of the SN (V_{EDRT}) to be used by the subsequent phase of reference acquisition.

B. Reference Current Acquisition

The reference acquisition phase is needed in order to evaluate the critical current through the read port (I_{EDRT}) that still ensures a successful read operation with the required read access latency at the considered operating point. The value of I_{EDRT} is then used during the evaluation phase as a critical threshold which represents the difference between an output of '1' and '0', depending on the polarity of the stored data and the type of transistors used to implement the GC read port. The value of I_{EDRT} is obtained using the test-bench shown in Fig. 4(b) where the GC read-port transistors (M2, M3 in Fig. 4(a)) are replicated (M2a, M3a). The previously evaluated V_{EDRT} is applied to the storage node (SNa), followed by the assertion of the RWL and the pre-charge of the RBL (e.g., to V_{DD} for an NMOS-only read port). Then, the I_{EDRT} is evaluated as the DC current which dis-charges the RBL for the considered operating point, representing the minimum/maximum read-port current that is necessary for a correct read operation.

C. Evaluation Phase

The evaluation phase is used to evaluate the DRT distribution under LPVs. The corresponding test-bench comprises a copy of the considered GC (M1b–M3b) in order to reproduce the data deterioration of the SN and the read port replica (M2c, M3c), whose current is monitored to evaluate the DRT, as depicted in Fig. 4(c). Similar to the VDRT methodology, the evaluation phase consists of a write operation to the GC (M1b–M3b), followed by an idle phase, during which the stored SNb voltage deteriorates over time as a result of the various leakage currents. The data deterioration is reproduced on the read port replica (M2c, M3c) using an ideal voltage buffer that applies the voltage of SNb to SNc, as presented in Fig. 4(c). Using this setup, the read port replica is constantly enabled and its drive current (i_{IDRT}) is continuously monitored. Therefore, the DRT is determined as



(a) Effective DRT (EDRT) measurement. (b) Reference current acquisition. (c) Evaluation phase.

Fig. 4. Phases of the current-based DRT (IDRT) characterization methodology.

TABLE I
ACCURACY OF THE IDRT METHODOLOGY

Technology	Bitcell	Temperature	EDRT	VDRT	VDRT Error	**IDRT**	**IDRT Error**
LP 28 nm FD-SOI	3T all-NMOS	0 °C	2.49 µs	3.15 µs	26.62 %	**2.49 µs**	**0.04 %**
LP 28 nm FD-SOI	3T all-NMOS	27 °C	673.60 ns	771.6 ns	14.55 %	**674.60 ns**	**0.15 %**
LP 28 nm FD-SOI	3T all-NMOS	85 °C	73.26 ns	70.2 ns	4.18 %	**75.23 ns**	**2.70 %**
LP 28 nm FD-SOI	4T all-NMOS	0 °C	788 µs	164 µs	79.19 %	**786.7 µs**	**0.17 %**
LP 28 nm FD-SOI	4T all-NMOS	27 °C	457 µs	112.2 µs	75.45 %	**459.1 µs**	**0.45 %**
LP 28 nm FD-SOI	4T all-NMOS	85 °C	81.6 µs	40.3 µs	50.62 %	**80.07 µs**	**1.88 %**
HP 28 nm Bulk	3T all-NMOS	0 °C	411.9 ns	451.9 ns	9.71 %	**416.0 ns**	**0.99 %**
HP 28 nm Bulk	3T all-NMOS	27 °C	128.4 ns	130.8 ns	1.91 %	**129.3 ns**	**0.66 %**
HP 28 nm Bulk	3T all-NMOS	85 °C	19.55 ns	18.13 ns	7.27 %	**20.27 ns**	**3.69 %**
HP 28 nm Bulk	4T all-NMOS	0 °C	44.42 µs	20.62 µs	53.59 %	**44.0 µs**	**0.94 %**
HP 28 nm Bulk	4T all-NMOS	27 °C	31.23 µs	10.53 µs	66.27 %	**30.78 µs**	**1.44 %**
HP 28 nm Bulk	4T all-NMOS	85 °C	18.34 µs	2.98 µs	83.76 %	**17.9 µs**	**2.39 %**

the time interval following the write operation at which the i_{IDRT} crosses the critical I_{EDRT} threshold as determined in the first and second phases.

Since the evaluation phase does not rely on any trial-and-error simulations, it can be performed across numerous MC runs in order to evaluate the impact of LPVs on the DRT at feasible run-times without the need of a large number of costly EDA software licenses. Furthermore, as opposed to the VDRT methodology, the effect of LPVs on the read port transistors is considered as well. As the test-bench relies on a replica of the GC read port, the level of variations applied to the read port of the considered GC (M2b, M3b) should identically affect the replicated transistors (M2c, M3c) for a consistent evaluation of the DRT across LPVs. This requirement is ensured by correlating the variations applied to the read-port transistors in the simulator settings (i.e., M2b with M2c and M3b with M3c for the considered GC).

IV. COMPARISON OF DRT CHARACTERIZATION METHODOLOGIES

In this section, the IDRT methodology is compared with the existing state-of-the-art EDRT and VDRT methodologies. First, a quantitative analysis of the accuracy achieved by the IDRT methodology is provided and compared to the other methodologies as simulated on a 3T and 4T all-NMOS GCs at 28 nm FD-SOI and 28 nm Bulk process technologies. Second, practical aspects of the three competing methodologies are compared.

A. Accuracy of the DRT Characterization Methodologies

The EDRT methodology provides the DRT value with the highest accuracy, and therefore it is considered the reference approach to quantify the accuracy of the IDRT and VDRT methodologies. For this analysis, both a 3T all-NMOS and a 4T all-NMOS GC are considered as DUTs for different temperatures and for both '0' and '1' data values. Further, the simulations were repeated in both low power (LP) 28 nm FD-SOI and high performance (HP) 28 nm Bulk CMOS technologies. Since the 3T GC features a full swing 2T read port, a conventional sense inverter was implemented in the EDRT testbench. On the other hand, the limited swing of the 1T read port featured in the 4T GC requires a differential sense-amplifier, which was biased with a reference voltage for comparison with the RBL voltage during read [14].

The results of this comparison are summarized in Table I, which reports the minimum DRTs (among '1' and '0') obtained for the EDRT, VDRT, and IDRT simulations. As expected, the VDRT methodology shows significantly larger errors compared to the IDRT methodology. This is because the VDRT does not account for the LPVs of the read port, which results in deviations up-to 83.76% from EDRT simulations. On the other hand, the IDRT methodology provides a maximum

TABLE II
COMPARISON OF DRT CHARACTERIZATION METHODOLOGIES

	EDRT	VDRT	IDRT
Accuracy	Highest	Low	High
Complexity	Low	Lowest	Medium
Nominal Simulation Time	High	Low	Highest
MC Simulation Time	Highest	Low	Low
License Number for MC Runs	Highest	Low	Low

TABLE III
RUNTIME COMPARISON OF DRT CHARACTERIZATION METHODOLOGIES

	EDRT	VDRT	IDRT
Nominal Reference Simulation	44 sec	2 sec	46 sec
1k Monte-Carlo Runs	1.25 hrs	125 sec	136 sec
10k Monte-Carlo Runs	∼13 hrs	1588 sec	1647 sec
100k Monte-Carlo Runs	∼5 days	∼4.2 hrs	∼4.4 hrs

deviation of only 3.69% due to its effective LPVs tracking of both the write and read ports.

B. Comparison of Practical Aspects in DRT Characterization Methodologies

A qualitative comparison based on four relevant practical aspects is provided in Table II for the EDRT, VDRT, and IDRT methodologies. The considered metrics are the accuracy, the complexity of the measurement, the nominal measurement time (i.e., the required time to obtain the DRT under typical operating conditions), the MC measurement time, and the number of EDA software licences required to execute a large number of MC runs for LPV analysis.

Section IV-A has shown that the EDRT methodology provides the highest accuracy and the proposed IDRT method is comparable, while the accuracy of the VDRT methodology is limited as it does not include the impact of LPVs on the GC read port. While both the EDRT and the VDRT methodologies can be performed with a single and simple testbench, the IDRT methodology is slightly more complex as it relies on three phases. The EDRT methodology relies on the execution of several simulations as the time duration of the idle phase is adjusted to find where the GC starts to fail. Thus, the nominal EDRT methodology takes longer than the VDRT methodology, where the SN voltage is monitored during a single simulation. Among the three approaches, the IDRT method has the longest nominal simulation time since it comprises the EDRT simulation as well as two additional simulations (i.e., reference acquisition and evaluation phase). However, due to the fact that both VDRT and IDRT methodologies do not require trial-and-error transient simulations when executed in the context of MC simulations, they results in a significantly shorter MC runtime and they do not require a large number of costly EDA software licenses, as opposed to the EDRT method. Table III depicts the achieved run-times of the three DRT characterization methodologies for a nominal simulation and a varying number of Monte Carlo runs. The simulations were initiated on a hyper-threaded 8-core Intel Core i7–3770 processor with 32 GB of RAM. The EDRT evaluation was made using the bisection method applied with Cadence's Ultrasim simulator, while the Monte-Carlo analysis was conducted using Cadence's Spectre simulator utilizing 8 Spectre licenses. The nominal simulation time of the IDRT methodology was found to be 46 seconds, which is slightly longer than the EDRT run due to the additional, second phase current acquisition. However, the evaluation phase of the IDRT methodology, which was conducted using 1 k, 10 k, and 100 k Monte-Carlo runs, was timed at 136 seconds, 1647 seconds, and approximately 4.4 hours, respectively, which are comparable to the Monte-Carlo run-times of the VDRT methodology. The slightly longer run-times of IDRT compared to VDRT are due to the additional components in the evaluation test-bench which are used to replicate the read port of the GC. As expected, the Monte-Carlo run-time of the EDRT methodology is over an order of magnitude longer, reaching over 5 days for 100 k runs.

To summarize, the VDRT methodology is suitable for a fast DRT evaluation in the early stage of the design phase. For a rigorous design-space exploration as well as for corner analysis (e.g., GPVs and temperature points), the EDRT method should be preferred as it provides the highest measurement accuracy. An accurate evaluation of the DRT across LPVs is enabled by the IDRT approach without the need of a large number of EDA software licenses.

V. DRT MEASUREMENT OF A 3T GC AND A 4T GC IN 28 nm FD-SOI

In this section, the DRTs of a 3T all-NMOS and a 4T all-NMOS GCs designed in 28 nm FD-SOI process technology are measured for different design choices as well as for global process and temperature variations using the EDRT methodology, while the impact of LPVs is evaluated with the IDRT approach. The following design and corner parameters are considered as a baseline: 0.9 V supply voltage, 1.3 V boosting voltage to drive the WWL, 0.7 V reference voltage for the 4T GC-eDRAM sense amplifier, minimum-sized transistors, a load of 127 GCs on the RBL to emulate a GC-eDRAM of 128 rows, body biasing (BB) is not applied, 1 ns period for single-cycle write and read operations, temperature at 27 °C, and typical (TT) process corner.

For the considered baseline operating point, the analyzed 3T and 4T GCs have nominal DRTs of 673 ns and 457 μs, respectively, as measured with the EDRT method with the WBL forced to the opposite value corresponding to the worst case.

A. Exploration of the Design Space

In addition to the choice of the GC topology, several design choices can improve the DRT of a given GC [10] and the impact of some of the most effective ones are evaluated in this subsection for the baseline 3T all-NMOS GC. The design space includes different values for the length of the write

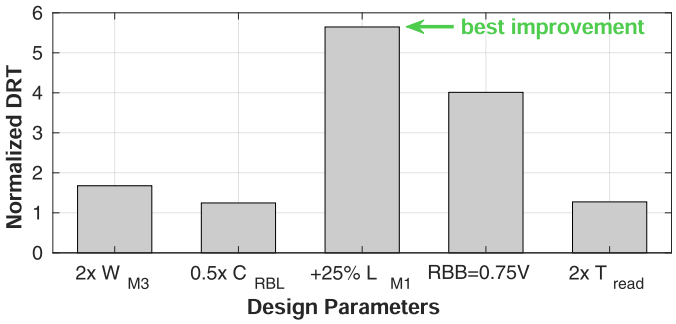


Fig. 5. Data retention time (DRT) improvement of a 3T GC for different design choices measured with the effective DRT (EDRT) method.

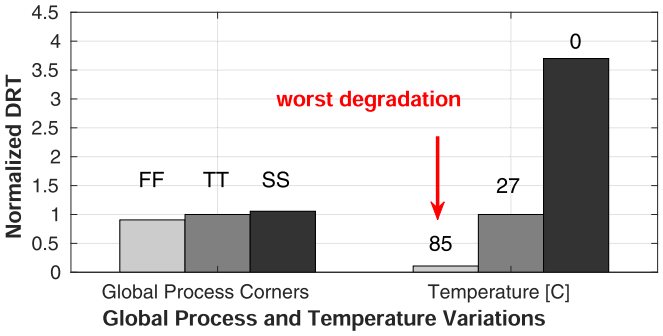
transistor (L_{M1}), the width of the read transistor (W_{M3}), the load on the RBL (C_{RBL}), the cycle time for the read operation (T_{read}), and the amount of reverse body biasing (RBB) applied on the write transistor. Starting from the baseline condition, the EDRT evaluation is performed for each design choice under nominal conditions.

The results of the design-space exploration are summarized in Fig. 5. Among the considered choices for DRT optimization, an increase of only 25% of L_{M1} provides a significannt improvement on the DRT that rises by 5× as the sub-threshold leakage through the write transistor is highly reduced. The leakage current can be further decreased by increasing the threshold voltage of the write transistor with the application of RBB. In particular, the DRT is improved by 4× when an RBB voltage of 0.75 V is applied to the write transistor. Further DRT improvements of up-to 2× can be achieved by doubling W_{M3} to ensure a successful read of a more degraded stored '1' (worst-case stored data for the considered GC), by a 50% reduction on the number of rows in the GC-eDRAMs which reduces C_{RBL} and diminishes the required drive current from the GC read port for a successful read, or by doubling the read-access time (T_{read}) to allow more time to discharge the RBL when a weaker '1' is stored in the GC. The considered design choices can also be applied together to further increase the overall DRT improvement.

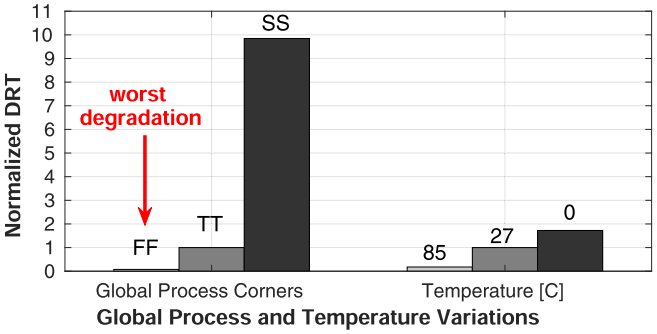
B. Global Process and Temperature Variations

Both manufacturing and environmental variations can degrade the DRT. In this subsection, the impact of global process and temperature corners on the DRT is evaluated with the use of the EDRT method. In particular, the slow (SS), typical (TT), and fast (FF) process corners have been considered together with 0°C, 27°C, and 85°C as temperature points.

The results of the EDRT analysis for global process and temperature variations are summarized in Fig. 6 for both a 3T GC and a 4T GC. For the case of a 3T GC (Fig. 6(a)), the worst degradation of the DRT is achieved at high temperatures, where the V_T of the write transistor is reduced and therefore its sub-threshold leakage current increases, resulting in a faster SN voltage deterioration. In particular, the DRT is reduced by a factor 9× at 85°C. Considering the different process corners, fast dies results in a DRT degradation of



(a) 3T GC.



(b) 4T GC.

Fig. 6. Data retention time (DRT) at global process and temperature variations measured with the Data retention time (DRT) method, normalized to the nominal (TT27°C) DRT.

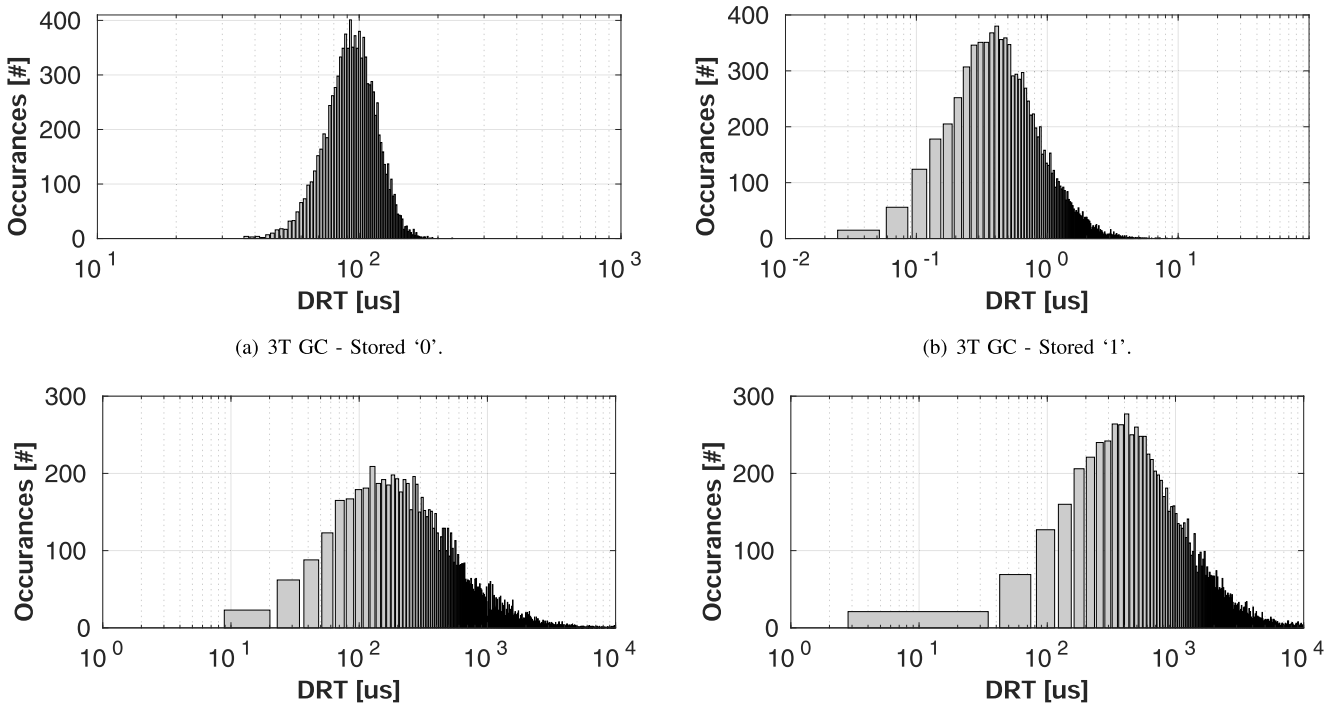
only 11% due to a relatively limited increase in sub-threshold conduction.

On the other hand, for a 4T GC (Fig. 6(b)), the worst DRT is found at the FF corner, where the DRT degrades by 13× compared to its nominal value. This can be attributed to the faster RBL discharge for a weak '0' level, resulting in the RBL voltage decreasing beneath the reference voltage of the sense amplifier, hence resulting in a read failure. On the contrary, the SS corner resulted in a DRT improvement of almost 10× due to the longer retention of '0'. Compared to the 3T GC, temperature variation has a lower effect on the SN voltage, since the increased sub-threshold leakage from the WBL is compensated by the internal feedback. On the other hand, the effect of temperature inversion [25] at high temperatures (i.e., 85°C) implies a higher read current, resulting in an almost 6× lower DRT of the '0' level.

C. Local Process Variations

LPVs also affect the sub-threshold conduction and read current strength, and therefore, they significantly impact the DRT of GCs. In this subsection, the IDRT measurement is performed for the 3T and 4T GCs across LPVs with 10 k MC runs including device mismatch around a typical (TT) corner at room temperature.

The results of the LPV analysis are summarized in Fig. 7, where both a '0' and a '1' have been considered for the stored data. For the 3T GC, '1' is the worst-case data for the DRT where 99.99% of the collected DRT values are larger than 38 ns which corresponds to a 17× degradation as compared to



(a) 3T GC - Stored '0'.

(b) 3T GC - Stored '1'.

(c) 4T GC - Stored '0'.

(d) 4T GC - Stored '1'.

Fig. 7. Data retention time (DRT) of the 3T and 4T GCs measured across local process variations (LPVs) with the current-based DRT (IDRT) method for different data.

the nominal DRT measurement. The average DRT was found to be 680 ns, which deviates by 1% from the value found at the nominal IDRT simulation, as reported in Table I. For the 4T GC, the DRTs of '1' and '0' are comparable, with '0' yielding a slightly lower DRT and its average found at 460 μs, compared to 496 μs for '1'. The average DRT corresponds to the value reported in Table I as extracted from the nominal IDRT simulation, found at 459 μs. The minimum DRTs for '0' and '1' under LPVs are found at 14.5 μs and 18.7 μs, respectively, corresponding to over 32× and 25× degradations. Thus, for the considered case studies, LPVs show the strongest degradation on the DRT, and therefore, their evaluation with the proposed IDRT methodology is crucial for both a reliable design of GC-eDRAMs as well as for the choice of their refresh rate to ensure high yield without overly pessimistic margins.

VI. Gain-Cell Memory Macro Characterization Using the IDRT Methodology

The DRT has a significant impact on almost all GC-eDRAM characteristics, such as the refresh period, retention power, and memory availability. The refresh period is set according to the worst case DRT of a GC array, which becomes significantly lower due to LPVs as the on-chip memory capacity grows.

In this section, we demonstrate how the IDRT methodology can be used to extract GC-eDRAM characteristics of large embedded memories by evaluating the memory's DRT distribution. The accurate estimation of the DRT is critical as it enables to reduce unnecessary design guard bands on the refresh period, which impacts the retention power and memory availability.

A. DRT Affect on Memory Characteristics

1) Retention Power: An important aspect of the need for a periodic refresh is the additional power component associated with the refresh operation. Unlike SRAM, where the power consumption can be divided into dynamic power, which is consumed during read and write operations, and static power, which is consumed due to leakage during standby, GC-eDRAMs consume another power component due to refresh. In order to provide a fair comparison to SRAM, the static power component of GC-eDRAM is redefined as retention power, which includes both the leakage and refresh power components, as follows:

$$P_{\text{ret}} = P_{\text{leak}} + P_{\text{ref}} = P_{\text{leak}} + (E_{\text{write}} + E_{\text{read}})/T_{\text{ref}}, \quad (1)$$

where P_{ret} is the retention power, P_{leak} is the leakage power of the GC array, P_{ref} is the average refresh power, E_{write} and E_{read} are the total write and read energies required to refresh the entire GC array, and T_{ref} is the refresh period.

2) Memory Availability: An additional consequence of the refresh requirement is the limited memory availability to external accesses, since whenever the memory is occupied with refresh it is unavailable for conventional write and read operations. The memory availability (α) is directly dependent on the refresh period set by the DRT and it is defined as follows:

$$\alpha(\%) = (1 - (N_{\text{rows}} \times T_{\text{per}})/T_{\text{ref}})) \times 100, \quad (2)$$

where N_{rows} is the number of rows in the GC array, and T_{per} is the clock period. Note that write and read operations to the memory can be performed in parallel due to the two-ported

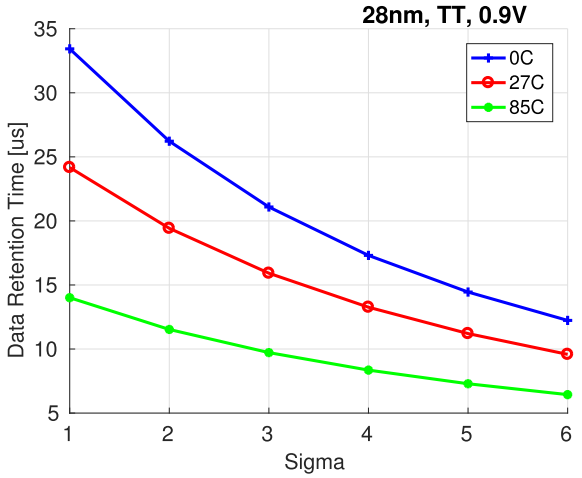


Fig. 8. DRT extrapolation of a 4T all-NMOS cell using the IDRT estimation between one to six sigma.

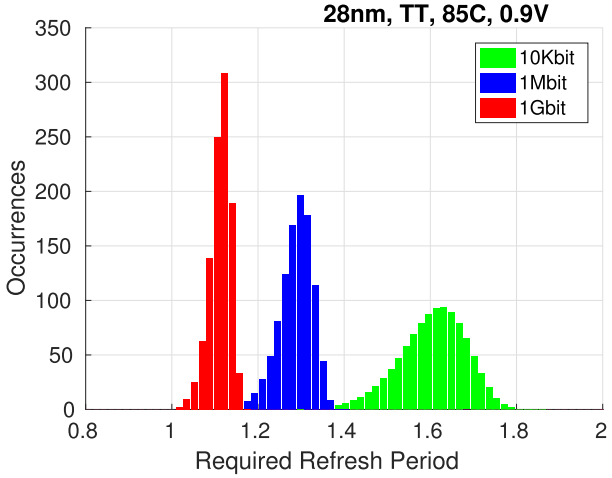


Fig. 10. Refresh period distributions for 10 kbit, 1 Mbit, and 1 Gbit GC-eDRAMs based on the 4T all-NMOS GC.

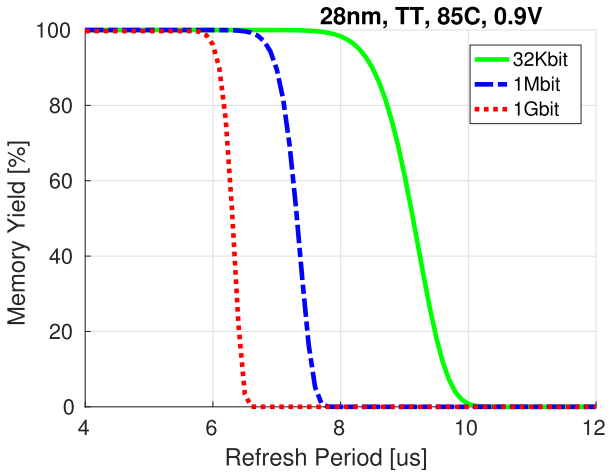


Fig. 9. Memory yield of 10 kbit, 1 Mbit, and 1 Gbit GC-eDRAMs based on the 4T all-NMOS GC for different refresh periods.

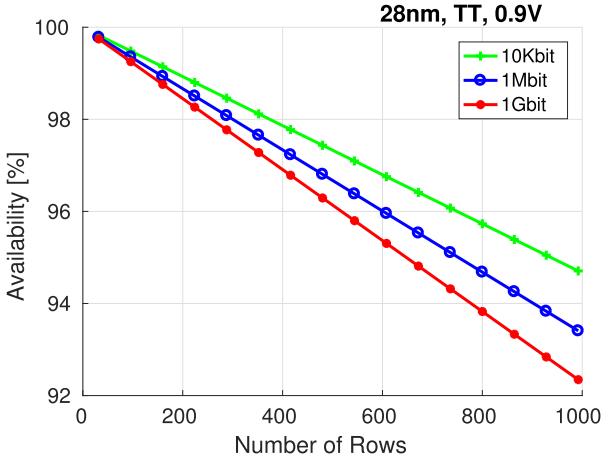


Fig. 11. Memory availability as a function of the number of GC array rows for a 10 kbit, 1 Mbit, and 1 Gbit GC-eDRAMs based on the 4T all-NMOS GC.

functionality of GC-eDRAM, hence the total number of cycles required for refresh is equivalent to N_{rows}.

B. Extraction of Key GC-eDRAM Characteristics

The DRT distributions that were presented in Fig. 7 can be used to evaluate the DRT of a small GC-eDRAM, as it is based on a limited number of MC runs (i.e., 10 k). However, in order to achieve high yield and support error-free operation of large GC-eDRAMs (i.e., 1 Mbit), a much higher number of MC runs is required. Unfortunately, running a brute-force MC analysis would require running over a million simulations, which is challenging and time consuming. Among the high-sigma evaluation techniques, importance sampling [26], [27] can provide an accurate DRT estimation by running the MC analysis around worst-case process corners instead of randomly choosing the process parameters. However, these evaluations would require trial-and-error runs in order to obtain the yield for each selected refresh period, hence they would still require an initial guess of the DRT margins and would result in long run-times. For an early design stage, high-sigma extrapolation techniques, such as Box-Cox power transformation, can be used to transform the DRT distribution into a normal shape and extract the high-sigma DRT values without the need for trial-and-error simulations [28], with an expected deviation of less than 2×. Fig. 8 presents the minimum DRT evaluation of the 4T GC across a six-sigma range at different temperatures, as extracted from a Box-Cox transformation, illustrating an over 2.5× reduction of the DRT from one to six sigma.

Using the high-sigma analysis, the DRT failure probability was obtained and used to evaluate the GC-eDRAM characteristics based on the minimum DRT evaluation, which sets the required refresh period. Fig. 9 depicts the expected memory yield for 10 kbit, 1 Mbit, and 1 Gbit 4T based GC-eDRAMs. As expected, the memory yield decreases as the memory size increases due to a higher probability for a DRT failure under a given refresh period. For example, the required refresh period for a target memory yield of 99% must be reduced from 7.9 μs to 5.8 μs as the GC-eDRAM capacity increases from 10 kbit to 1 Gbit.

Fig. 10 depicts the refresh period distributions obtained for 10 k TT chips containing 10 kbit, 1 Mbit, and 1 Gbit GC-eDRAMs under a 99% yield criterion. According to (1), the retention power is inversely proportional to the refresh period. Therefore, the reduction in the refresh period with

increasing memory size results in an almost 40% increase in the minimum retention power with the memory capacity increasing from 10 kbit to 1 Gbit.

Fig. 11 illustrates the average memory availability evaluated for 10 kbit, 1 Mbit, and 1 Gbit GC-eDRAMs with a varying ratio between rows and columns, assuming that all GC arrays are refreshed in parallel. As described in (2), the number of GC array rows has a linear effect on the memory availability. In addition, the decrease in average DRT with increasing memory size results in lower availability, going from 94.7% down to 92.3% between 10 kbit and 1 Gbit memories with 1000-row GC arrays.

VII. Conclusions

This paper presented a current-based data-retention-time (IDRT) characterization methodology. As the proposed methodology does not comprise trial-and-error transient simulations, it enables an accurate evaluation of the DRT in gain cell embedded DRAMs (GC-eDRAMs) across local process variations (LPVs) at a significantly shorter runtime compared to conventional trial-and-error transient simulations, which makes MC analysis feasible. When compared to the most accurate DRT methodology, the IDRT methodology shows an accuracy deviation of less than 3.7%. Together with the EDRT approach, the proposed methodology has been used to evaluate the DRT of GCs designed in 28 nm FD-SOI and 28 nm Bulk CMOS process technologies across a design and variations space, showing the LPVs to have the largest impact on the DRT among the considered variations. Moreover, the IDRT methodology was used to extract the high-sigma DRT and evaluate key GC-eDRAM characteristics, including retention power, memory availability, and maximum frequency, for different memory sizes. Thus, the proposed IDRT methodology is key for a reliable design of GC-eDRAMs as well as for the choice of their refresh rate without any undesirable design guard bands.

References

[1] F. Tu, W. Wu, S. Yin, L. Liu, and S. Wei, "RANA: Towards efficient neural acceleration with refresh-optimized embedded DRAM," in *Proc. ACM/IEEE 45th Annu. Int. Symp. Comput. Archit. (ISCA)*, Jun. 2018, pp. 340–352.

[2] C. Berry *et al.*, "IBM Z14: 14 nm microprocessor for the next-generation mainframe," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2018, pp. 36–38.

[3] M. Gautschi *et al.*, "Near-threshold RISC-V core with DSP extensions for scalable IoT endpoint devices," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2700–2713, Oct. 2017.

[4] Y.-H. Chen, T. Krishna, J. S. Emer, and V. Sze, "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 127–138, Jan. 2017.

[5] M. Tikekar, V. Sze, and A. Chandrakasan, "A fully-integrated energy-efficient H.265/HEVC decoder with eDRAM for wearable devices," in *Proc. Symp. VLSI Circuits*, Jun. 2017, pp. C230–C231.

[6] T. Luo *et al.*, "DaDianNao: A neural network supercomputer," *IEEE Trans. Comput.*, vol. 66, no. 1, pp. 73–88, Jan. 2017.

[7] M. Tikekar, V. Sze, and A. Chandrakasan, "An energy-efficient, fully integrated 1920×1080 H.265/HEVC decoder with edram," *Res. Abstr.*, vol. 49, no. 1, p. 1, 2018.

[8] C. Li and P. Ampadu, "A compact low-power eDRAM-based NoC buffer," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Rome, Italy, Jul. 2015, pp. 116–121. [Online]. Available: http://ieeexplore.ieee.org/document/7273500/

[9] N. Jing *et al.*, "An energy-efficient and scalable eDRAM-based register file architecture for GPGPU," *SIGARCH Comput. Archit. News*, vol. 41, no. 3, pp. 344–355, Jul. 2013.

[10] P. Meinerzhagen, A. Teman, R. Giterman, N. Edri, A. Burg, and A. Fish, *Gain-Cell Embedded DRAMs for Low-Power VLSI Systems-on-Chip*. Cham, Switzerland: Springer, 2017.

[11] R. Giterman, R. Golman, and A. Teman, "Improving energy-efficiency in dynamic memories through retention failure detection," *IEEE Access*, vol. 7, pp. 27641–27649, 2019.

[12] A. Bonetti, R. Golman, R. Giterman, A. Teman, and A. Burg, "Gain-cell embedded DRAMs: Modeling and design space," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, to be published.

[13] R. Giterman, A. Bonetti, A. Burg, and A. Teman, "GC-eDRAM with body-bias compensated readout and error detection in 28-nm FD-SOI," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 12, pp. 2042–2046, Dec. 2019.

[14] R. Giterman, A. Fish, N. Geuli, E. Mentovich, A. Burg, and A. Teman, "An 800-MHz mixed-V_T 4T IFGC embedded DRAM in 28-nm CMOS bulk process for approximate storage applications," *IEEE J. Solid-State Circuits*, vol. 53, no. 7, pp. 2136–2148, Jul. 2018.

[15] E. Amat, R. Canal, and A. Rubio, "Modem gain-cell memories in advanced technologies," in *Proc. IEEE 24th Int. Symp. On-Line Test. Robust Syst. Design (IOLTS)*, Jul. 2018, pp. 65–68.

[16] R. Giterman, A. Fish, N. Geuli, E. Mentovich, A. Burg, and A. Teman, "An 800 Mhz mixed-VT 4T gain-cell embedded DRAM in 28 nm CMOS bulk process for approximate computing applications," in *Proc. 43rd IEEE Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2017, pp. 308–311.

[17] W. Choi, G. Kang, and J. Park, "A refresh-less eDRAM macro with embedded voltage reference and selective read for an area and power efficient Viterbi decoder," *IEEE J. Solid-State Circuits*, vol. 50, no. 10, pp. 2451–2462, Oct. 2015.

[18] B. Frankel, R. Herman, and S. Wimer, "Queuing-based eDRAM refreshing for ultra-low power processors," *IEEE Trans. Comput.*, vol. 67, no. 9, pp. 1331–1340, Sep. 2018.

[19] A. Kazimirsky and S. Wimer, "Opportunistic refreshing algorithm for eDRAM memories," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 11, pp. 1921–1932, Nov. 2016.

[20] R. Giterman, A. Fish, A. Burg, and A. Teman, "A 4-transistor nMOS-only logic-compatible gain-cell embedded DRAM with over 1.6-ms retention time at 700 mV in 28-nm FD-SOI," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 4, pp. 1245–1256, Apr. 2018.

[21] R. Giterman, A. Teman, P. Meinerzhagen, L. Atias, A. Burg, and A. Fish, "Single-supply 3T gain-cell for low-voltage low-power applications," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 1, pp. 358–362, Jan. 2016.

[22] W. Zhang, K. C. Chun, and C. H. Kim, "A write-back-free 2T1D embedded DRAM with local voltage sensing and a dual-row-access low power mode," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 8, pp. 2030–2038, Aug. 2013.

[23] K. C. Chun, W. Zhang, P. Jain, and C. H. Kim, "A 2T1C embedded DRAM macro with no boosted supplies featuring a 7T SRAM based repair and a cell storage monitor," *IEEE J. Solid-State Circuits*, vol. 47, no. 10, pp. 2517–2526, Oct. 2012.

[24] R. Giterman, Y. Weizman, and A. Teman, "Gain-cell embedded DRAM-based physical unclonable function," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 12, pp. 4208–4218, Dec. 2018.

[25] Y. Zu, W. Huang, I. Paul, and V. J. Reddi, "Ti-states: Processor power management in the temperature inversion region," in *Proc. 49th Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO)*, Oct. 2016, pp. 1–13.

[26] R. Kanj, R. Joshi, and S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," in *Proc. 43rd Annu. Conf. Design Autom. (DAC)*, 2006, pp. 69–72.

[27] T. Doorn, E. Ter Maten, J. Croon, A. Di Bucchianico, and O. Wittich, "Importance sampling Monte Carlo simulations for accurate estimation of SRAM yield," in *Proc. 34th Eur. Solid-State Circuits Conf. (ESSCIRC)*, Sep. 2008, pp. 230–233.

[28] H.-H. Wang, Y.-L. Chen, C.-C. Yang, C.-K. Lin, and M.-C. Jeng, "Compact modeling for application-specific high-sigma worst case," in *Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD)*, Sep. 2013, pp. 61–64.

Robert Giterman received the B.Sc. degree in electrical engineering from Ben-Gurion University, Be'er Sheva, Israel, in 2013, the M.Sc. degree from Ben-Gurion University in 2014 as part of a fast track program for outstanding students, and the Ph.D. degree, under the supervision of Prof. Alex Fish and Dr. Adam Teman, from the Emerging Nanoscaled Intergrated Circuits and Systems (EnICS) Lab, Bar Ilan University, Ramat-Gan, Israel, in 2018. Since November 2018, he has been a Post-Doctoral Researcher with the Telecommunications Circuits Laboratory, EPFL. As part of his research, he led several full test chip integrations and tape outs. He has authored or coauthored more than 30 journal articles and international conference articles and seven patent applications. He has presented his research at a number of international conferences. He has coauthored the book *Gain-Cell Embedded DRAMs for Low-Power VLSI Systems-on-Chip*. His research interests include embedded DRAM design and optimization for low-power and high-performance operation, error-correction and fault-tolerant circuits, development of hardware-security oriented embedded memories for use in low-power applications and high-end processors, and memory accelerators for machine learning.

Andrea Bonetti received the bachelor's degree in biomedical engineering and the master's degree in electronic engineering from the Politecnico di Milano, Milan, Italy, in 2009 and 2012, respectively, and the Ph.D. degree from the Telecommunications Circuits Laboratory (TCL), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2019.

In 2011, he visited the Bio Engineering Laboratory (BEL), Swiss Federal Institute of Technology (ETH) Zürich, Zürich, Switzerland. From 2012 to 2014, he worked as a Design Engineer with ams AG, Rapperswil, Switzerland, for the development of interfaces in MEMS sensors. In 2016, he visited the Circuits Research Laboratory, Intel Labs, Hillsboro, OR, USA. His current research activity is focused on error-tolerance techniques and memory design for low-power computing.

Ester Vicario Bravo received the B.Sc. degree in telecommunications engineering from the Technical University of Madrid in 2018. In 2018, she studied at the Ecole Polytechnique Federale de Lausanne (EPFL), where she collaborated with the Telecommunications Circuits Laboratory (TCL). She is currently pursuing the M.Sc. degree in embedded systems with TU Delft University with a specialization in computer architecture. Her current research is focused on dynamic memories and low-power circuits design.

Tzachi Noy received the B.Sc. and M.Sc. degrees in electrical engineering from Bar-Ilan University, Ramat-Gan, Israel, in 2007 and 2019, respectively. He is currently pursuing the Ph.D. degree in electrical engineering, under the supervision of Dr. Adam Teman, from the Emerging Nanoscaled Intergrated Circuits and Systems (EnICS) Lab, Bar-Ilan University. His research interests include EDA optimization heuristics, hardware implementations for low-power machine learning applications, design of very large-scale integration circuits, and high-performance computer architectures.

Adam Teman received the B.Sc. degree in electrical engineering and the M.Sc. degree from Ben-Gurion University (BGU), Be'er Sheva, Israel, in 2006 and 2011, respectively. He received the Ph.D. degree, under the supervision of Prof. Alexander Fish, from the VLSI Systems Center, BGU, in 2014.

He worked as a Design Engineer with Marvell Semiconductors from 2006 to 2007. From 2014 to 2015, he was a Post-Doctoral Researcher with the Telecommunications Circuits Lab (TCL), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, under a Swiss Government Excellence Scholarship. In 2015, he joined the Faculty of Engineering, Bar-Ilan University (BIU), as a Tenure Track Senior Lecturer with the Department of Electrical Engineering and as the Co-Director of the Emerging Nanoscaled Integrated Circuits and Systems (EnICS) Labs Research Center. He has coauthored the recently published book *Gain-Cell Embedded DRAMs for Low-Power VLSI Systems-on-Chip* (Springer). He has authored more than 80 scientific articles and eight patent applications, and has participated in more than 15 IC tapeouts. His research interests include embedded memories, energy efficient circuit design, hardware for artificial intelligence, hardware acceleration, and methodologies for physical implementation.

Dr. Teman is a member of the technical and review boards of several conferences and journals. His Ph.D. studies were conducted under a Kreitman Foundation Fellowship. He has been honored with Teaching Excellence recognitions at both BGU and BIU and has received multiple awards for outstanding research, including the Yizhak Ben-Ya'akov HaCohen Prize, the BGU Rector's Prize, the Wolf Foundation Scholarship, and the Intel Prize. He is an Associate Editor of the *Microelectronics Journal*.

Andreas Burg (Member, IEEE) was born in Munich, Germany, in 1975. He received the Dipl.Ing. degree from the Swiss Federal Institute of Technology (ETH) Zürich, Zürich, Switzerland, in 2000, and the Dr.Sc.Techn. degree from the Integrated Systems Laboratory, ETH Zürich, in 2006.

In 1998, he was with Siemens Semiconductors, San Jose, CA, USA. During his Ph.D. studies, he was with the Bell Labs Wireless Research for a total of 1 year. From 2006 to 2007, he was a Post-Doctoral Researcher with the Integrated Systems Laboratory and with the Communication Theory Group, ETH Zürich. In 2007, he co-founded Celestrius, an ETH-spinoff in the field of MIMO wireless communication, where he was responsible for the ASIC development as Director for VLSI. In 2009, he joined ETH Zürich as an SNF Assistant Professor and as the Head of the Signal Processing Circuits and Systems Group, Integrated Systems Laboratory. Since 2011, he has been a Tenure Track Assistant Professor with the École Polytechnique Fédérale de Lausanne (EPFL), where he is currently leading the Telecommunications Circuits Laboratory.

He is also a member of the EURASIP SAT SPCN and of the IEEE TC-DISPS. In 2000, he received the Willi Studer Award and the ETH Medal for his diploma and his diploma thesis, respectively. He was also awarded an ETH Medal for his Ph.D. dissertation in 2006. In 2008, he received a 4-years grant from the Swiss National Science Foundation (SNF) for an SNF Assistant Professorship. With his students, he received the Best Paper Award from the *EURASIP Journal on Image and Video Processing* in 2013 and best demo/paper awards at ISCAS 2013, ICECS 2013, and at ACSSC 2007. He has served on the TPC of various conferences on signal processing, communications, and VLSI. He was a TPC Co-Chair for VLSI-SoC 2012 and the TCP Co-Chair for ESSCIRC 2016 and SiPS 2017. He served as an Editor of the IEEE TRANSACTIONS OF CIRCUITS AND SYSTEMS in 2013 and on the Editorial Board of the *Microelectronics Journal* (Springer). He is an Editor of the *Journal of Signal Processing Systems* (Springer), and the *Journal of Low Power Electronics and Applications* (MDPI).