



HAL
open science

Benefits of design assist techniques on performances and reliability of a RRAM macro

B. Giraud, S. Ricavy, Y. Moursy, C. Laffond, I. Sever, V. Gherman, M. Pezzin, F. Lepin, M. Diallo, K. Zenati, et al.

► To cite this version:

B. Giraud, S. Ricavy, Y. Moursy, C. Laffond, I. Sever, et al.. Benefits of design assist techniques on performances and reliability of a RRAM macro. IMW 2023 - 2023 IEEE International Memory Workshop, May 2023, Monterey, United States. 10.1109/IMW56887.2023.10145984 . hal-04442641

HAL Id: hal-04442641

<https://hal.science/hal-04442641v1>

Submitted on 7 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Benefits of Design Assist Techniques on Performances and Reliability of a RRAM Macro

B. Giraud¹, S. Ricavy², Y. Moursy², C. Laffond¹, I. Sever³, V. Gherman¹, M. Pezzin¹, F. Lepin¹, M. Diallo¹, K. Zenati¹, S. Dumas², M. Vershkov³, A. Bricalli³, G. Piccolboni³, J.-P. Noel¹, A. Samir², G. Pillonnet², Y. Thonnart¹, G. Molas⁴

¹Univ. Grenoble Alpes, CEA, List, F-38000 Grenoble, France,

²Univ. Grenoble Alpes, CEA, Leti, F-38000 Grenoble, France,

³Weebit Nano Ltd

⁴Univ. Grenoble Alpes, CEA, Leti now Weebit Nano Ltd,
bastien.giraud@cea.fr

Abstract— This paper presents different design assist techniques and demonstrates their impact on enhancing the intrinsic RRAM performance. We show that the read-before-write, current-limitation and write-termination techniques reduce by -47%, -56% and -13% the power consumption during the writing process, respectively. Combined with write verification and error correction code, the overall improvements are 87% in energy saving and -55% on access time. Based on representative RRAM macro (130nm CMOS), statistic (128kb) and endurance (1M cycles) characterizations, this work contributes to accelerate RRAM industrial adoption by highlighting the design-technology co-optimization contribution.

Keywords—Non-volatile memory, ECC, smart algorithm, Adaptive and Reconfigurable Systems, Variant-tolerant

I. INTRODUCTION

Resistive RAM (RRAM) is a promising candidate to replace NOR eFlash in sub-40nm CMOS technologies [1-3]. It offers high-speed and low-voltage embedded non-volatile memories compatible with CMOS back-end process. However, the stochasticity of RRAM programming limits the intrinsic performance of memory arrays without the use of advanced design assist techniques [4-6]. The recent industrial adoption of RRAM in some major foundries has been done through a massive involvement of circuit design community to boost both the read-margin (RM) and endurance, while reducing the energy consumption [7-11]. Prior art proposed several programming techniques to guarantee RRAM functionality and to improve their performances. Nevertheless, due to the diversity of RRAM processes and design techniques in these previous works, the individual and mutual benefits on the tradeoff between RM, endurance, power consumption and speed are only partially reported. This paper highlights the additional performances provided by each main state-of-art custom programming strategies including current limiter (CL), write termination (WT), error correction code (ECC) driven by a flexible smart write algorithm (SWA). The overall performance improvements using all these design-aided techniques are demonstrated in our silicon measurements compared to the original performance of the RRAM macro.

II. TESTCHIP DESCRIPTION

Fig. 1 and Table I present the microphotograph, the cross-section view and the main characteristics of the RRAM macro to conduct our exploration of the benefits of design assist techniques. The macro includes a 128kb RRAM array

surrounded by on-chip programmable circuitries (CL, WT, ECC) controlled by the SWA. The RRAM bitcells are manufactured in the back end of line of 130nm CMOS process technology, between metal 4 and 5 layers (cross section view in Fig. 1). Two 88kb RRAM arrays include each 64k data-bits and 24k check-bits (ECC). The matrix size is enough to have a statistical representation of the cell-to-cell variation required for this study. This macro also embeds a RISC-V core for detailed RRAM characterization, an advanced FSM integrating the SWA as well as 64kb instruction and 64kb data memories. The analog and digital circuits are supplied by 4.6V and 1.2V, respectively.

The benefit of design assist techniques has to be evaluated in light of area penalty. As the memory size is relatively limited here compared to an application purpose, the area overhead of the surrounding circuits is not negligible. By extrapolation, we estimate an overhead of 47% for 1Mb, which is comparable to previous works [10], and as low as 28% for 16Mb. Moreover, in particular application targets, the flexible programmability is no longer required, hence, reducing the footprint.

III. DESIGN TECHNIQUES AND RESULTING IMPROVEMENTS

Bitcell writing is performed per column thanks to a dedicated programming current and biasing voltage. Both of

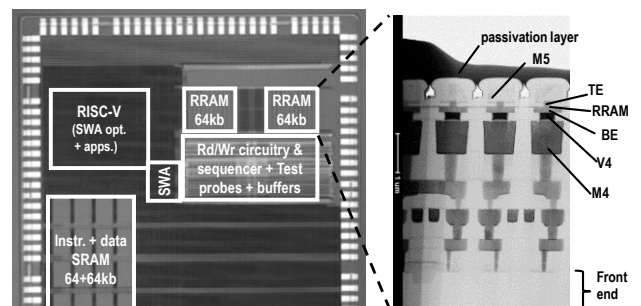


TABLE I. MAIN DEMONSTRATOR FEATURES

	This work
CMOS Technology	130nm
VDD	1.2/4.6 V
Cell Size	61F ²
Capacity	128kb (+ 48kb for ECC)
Array efficiency for 1Mb (array/total area)	53% ^a
SWA + ECC area w.r.t. RRAM macro area	1% + 2%

^a. Extrapolated

them can be widely and independently tuned for Set and Reset. This tunability is necessary to guarantee the functionality of the entire memory arrays, while taking into account cell-to-cell and filament growing process variabilities as well as variable IR drops due to high writing current. These variabilities reduce the read margin (RM) and limit achievable BER.

A. Embedded Design Techniques

In the presented system, several design assist techniques have been integrated in order to optimize intrinsic performances and energy of the proposed RRAM macro. These techniques are the following:

- **Read-Before-Write (RBW)** to save write energy of bits being already in the wanted state. Actually, RBW is adapted as the read energy is much lower than the write energy (by more than an order of magnitude),
- **Current limiter (CL)** that consists in limiting the write current during Set operation (from high to low resistive state) in order to prevent too high current from breaking down the oxide of the RRAM device,
- Abrupt or expendable **write termination (WT)** that stops a write operation once the filament is formed or broken in the case of Set or Reset (with or without a controlled time delay), respectively. WT aims at limiting the programming time and consumption, as well as ensuring that all the bitcells receive the same programming energy, leading to a narrower distribution after Set or Reset operations,
- **Write verify (WV)** that corresponds to reading the bits that have just been written. The WV principle consists in writing again until all the bits are successfully written or if the maximum number of WV is achieved,
- A binary linear block **ECC** with up to double-bit error correction capability. This requires 12 check-bits per code word with 32 data-bits. Partial 3-bit error detection was enabled without increasing the number of check-bits. This enables the detection of 78% of all possible 3-bit errors in a code word. Another ECC property is the inversion invariance, i.e., the inversion of all bits of a code word leads to another valid code word [12]. Moreover, all check-bits are also forced to either '0' or '1' if all data bits are at '0' or '1', respectively. Together with the ECC linearity, this feature guarantees the presence of *all-1* and *all-0* codes words which facilitates memory testing and forming.

In this study, we ran a detailed design of experiment leading to optimized programming conditions (voltage, current and time) that allowed to combine low BER and guarantee high reliability (retention and endurance). In the section below, the impacts of these previously listed design techniques will be assessed thanks to silicon data.

B. Improvements thanks to Design Techniques

The write verify feature reduces programming voltage by 10% and 20% in Set and Reset operations, respectively, as shown in Fig. 2. The average Set and Reset durations are thereby reduced down to 40ns. However, the SWA optimization priority favors reliability at the cost of slightly higher energy and durations.

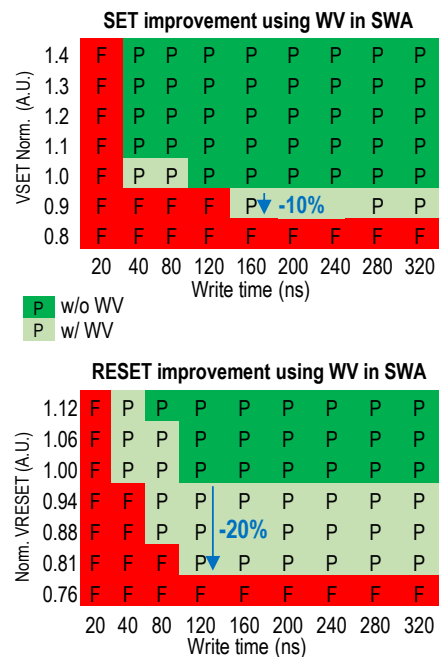


Fig. 2. Schmoos plots for write operations w/ & w/o WV (write verify).

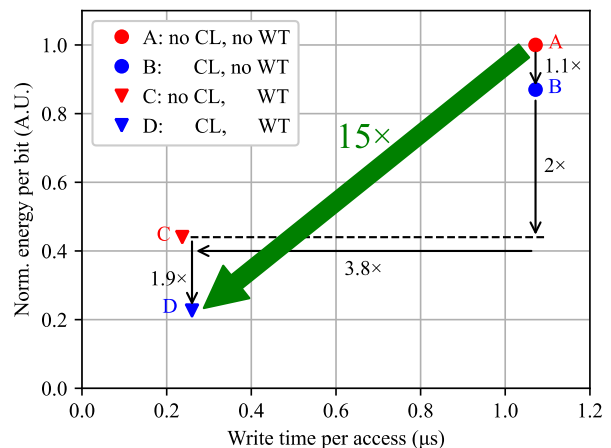


Fig. 3. Write time and energy improvements thanks to WT and CL.

The energy consumed during a Set operation is depicted in different configurations, improved by 10% when optimal CL value is chosen, by an extra 2x when enabling only WT and by an additional reduction of 1.9x when combining both WT and CL (Fig. 3). While CL enhances significant energy reduction, WT allows both energy and write time reduction. The CL/WT combination leads to a 15x Energy Delay Product saving. Besides, the CL prevents each bitcell from receiving too many energy that can conduct to a programming stuck, while WT aims at guaranteeing that all the bitcells receive a more homogeneous quantity of energy, leading to a much higher reliability.

Table II reports all the measured gains or drawbacks of each design technique independently from the other ones on energy, access time, RM and BER. For the sake of space, endurance is not analyzed here and will be displayed only in section IV. Extensive measures on every bitcell of this 128kb macro lead to reliable analyses.

RBW improves energy by -47% at the cost of +4% write access time. Actually, the read energy is 3% compared to write operation but the RBW prevents half of the bits in average to

TABLE II. MAIN BENEFITS OF DESIGN ASSIST TECHNIQUES

Design techniques	Energy	Access time	Additional benefits
RBW	-47%	+4% WR	NA
WT	-56%	-74% WR	NA
CL	-13%	<1% WR	NA
WV	+5%	+6% WR	RM x2
ECC	+0.01%	+18% RD	BER/100

be over-written. In addition, the read operation lasts 40ns as compared to a 1 μ s write time. It corresponds to 20ns required for RISC-V request cycle and 20ns of RRAM macro access time in the same order as [8-9] designed in 22nm technology.

WT permits to save energy by 56% and write access time by 74%, as depicted in vertical and horizontal transitions, respectively, from points A to C in Fig. 3.

CL enables 13% of energy saving without modifying the write access time, as represented by transition from A to B points. One CL per IO slice (16-bitcell column) offers less cell-to-cell variabilities compared to the configurations where each bitcell selector limits the current.

WV degrades energy and write access time by 5% and 6%, respectively, in favor of a x2 enhanced RM. 5% of extra energy correspond to the fact that all bits are verified once (3% of write energy for a read) and 2% of bits need to be rewritten after this verification phase. Similarly, 6% access time degradation is due to 4% of read time plus 2% of bits to be rewritten.

ECC leads to a negligible energy overhead and to 18% read time degradation while enhancing BER by more than 100x (from raw BER of 10^{-5} to $<3.10^{-8}$ demonstrated in endurance test). 18% extra read time corresponds to 20ns and 23.5ns RRAM macro access time without and with ECC decoding time, respectively.

IV. SMART WRITE ALGORITHM AND CUMULATIVE BENEFITS

A. Smart Write Algorithm Presentation

Fig. 4 describes the embedded SWA including a write part with ECC_ENC (encoding data), RBW, CL, WT and WV as well as a read part with SA (Sense Amplifier) and ECC_DEC (decoding data).

The writing parameters such as the Set and Reset voltages, the write pulse width, limit current and write termination current (respectively named V, T, I_{CL} and I_{WT} in the Fig. 4)

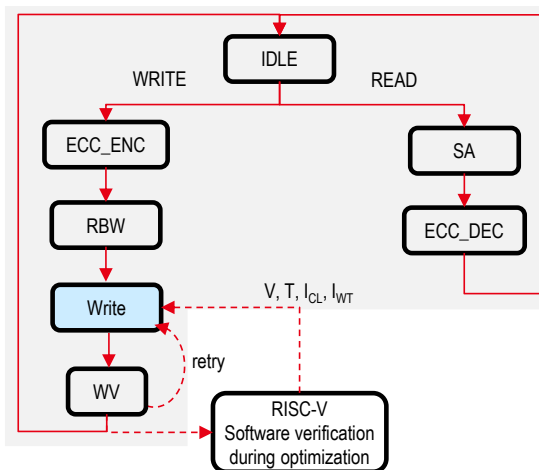


Fig. 4. SWA description.

can be widely and independently tuned. Set and Reset voltages can range from 1.2V to 4.6V thanks to a 5-bit voltage DAC, leading to 1024 possible combinations only for voltages. The pulse width can be varied from 5ns to 5 μ s (the small durations are only used for read operations). The limitation current and termination current can vary thanks to 2 4-bit current DACs from 50 to 425 μ A, and from 60 to 210 μ A, leading to 256 possible combinations only for current. Of course, only the combinations for which limit current is higher than termination current were considered. Actually, in the opposite case, the termination will never be valid. In the presented testchip, WV is managed thanks to a register enabling up to 32 write repetitions.

All these parameters are programmable through a RISC-V interface to adapt the SWA patterns to the target applications.

To cope with potential applications for IoT systems, the RISC-V core is in charge of optimizing SWA parameters with SW tuning loops during early life of the chip, the RRAM itself storing these parameters afterwards. The high flexibility of RISC-V has permitted to determine individually all the settings in order to maximize reliability.

B. Improvements of all Combined Design Techniques

After SWA optimizations, RBW, WT and ECC are enabled, while current limitation in Set has been fixed to 170 μ A and 3 WV iterations were allowed. Fig. 5 illustrates the boost of RM, thanks to SWA optimization, compared to the intrinsic one up to 26 μ A at 10^{-5} BER without cycling, better than [7, 10-11].

The endurance is then evaluated over 1M Set-Reset cycles in a 32-bit word, as shown in Fig. 6. Here, the resistance references for WV have been fixed differently for Set and Reset in order to maximize RM. While WV has no significant impact of LRS current, it clearly improves the high tail of HRS current distribution by pushing back all the measured currents below 20 μ A. RM becomes negligible without WV around 750k and 850k cycles (red stars of Fig. 6), but remains almost unchanged with WV over 1M cycles. WV efficiently mitigates the average HRS current degradation starting from 800k cycles, as shown inside the red box. Up to 2 outlier Set/Reset operations on random cells are excluded by taking into account the filtering action of the ECC: two cells per code word and Set/Reset operation may be filtered due the available double-bit error correction capability. Such filtering occurred

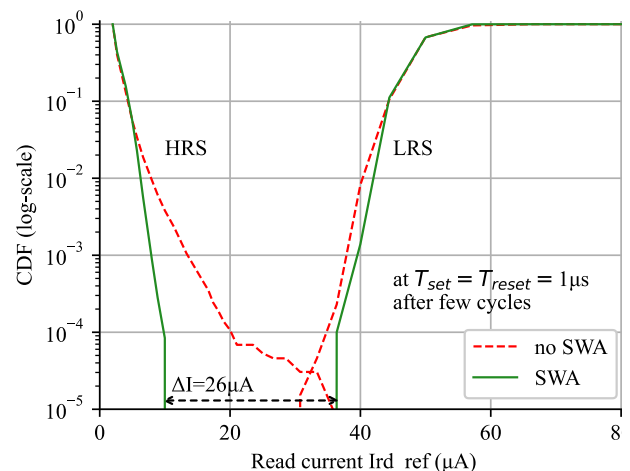


Fig. 5. Memory windows improvements achieving 26 μ A margin thanks to optimal SWA parameters.

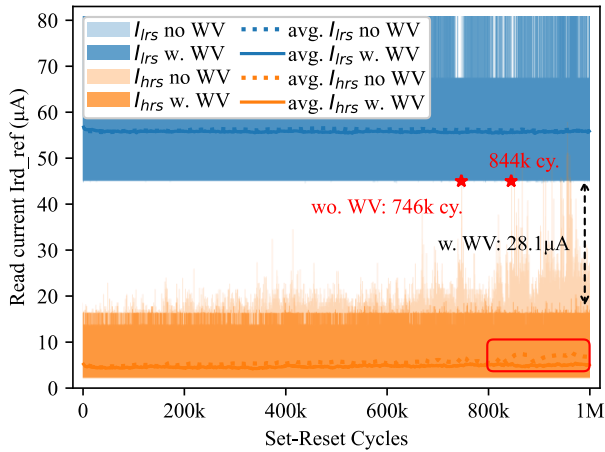


Fig. 6. Read current over 1M Set-Reset cycles.

in less than 2% of the Set/Reset operations.

Fig. 7 shows the average on 50k cycles of the read margin RM (read current difference) over 1M Set-Reset cycles. The resulting RM without WV starts at $28\mu\text{A}$ and is maintained above $14\mu\text{A}$ all over the cycling sequence, which is thus degraded by around 50%. Actually, this high degradation has been fitted with $-5.7\mu\text{A/Mcy}$ slope up to 800k cycles and a much higher slope of $-49.9\mu\text{A/Mcy}$ over the last 200k cycles. With WV enabled, the RM ranges from $29.9\mu\text{A}$ to $28.1\mu\text{A}$ all over the cycling sequence which is thus degraded by only 6%. The WV allows to maintain the RM degradation below $-2\mu\text{A/Mcy}$ over 1 Mcy, leading to 2x RM enhancement at 1 Mcy. No cycle presented 3 weak bits or more over 1M cycles on the 32 bits. Actually, RRAM endurance failure is due to a collapse of the resistance window: resistance can drop at an intermediate resistance between HRS and LRS [13], or be stuck at the LRS due to dielectric breakdown [14]. In our case, HRS current increase measured in Fig. 6 confirms that we are limited by gradual dielectric degradation.

Data with WV showed in Fig. 7 confirms that optimized programming schemes and controlled programming energy allow to reduce the degradation acceleration factor.

Table III lists the improvements when activating all the design assist techniques embedded into the SWA. The measurements point out that the energy is improved by -87%, the write access time by -55%, the memory windows by 2x, at the cost of 18% degradation in read time access due to ECC

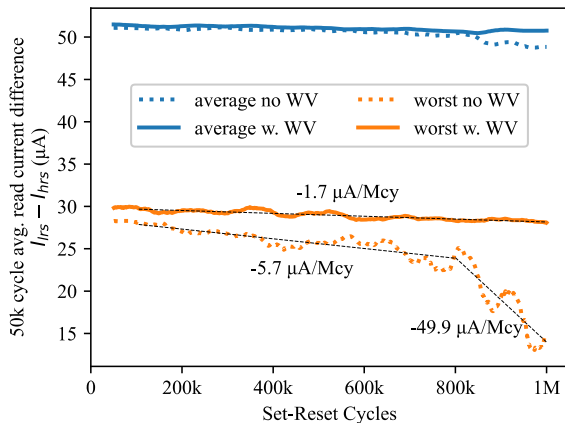


Fig. 7. Read Margin: current difference averaging on 50k cycles over 1M Set-Reset cycles.

TABLE III. MAIN RESULTS WITH ALL TECHNIQUES

Design techniques	Energy	Access time	RM	BER	Endurance
All enabled	-87%	-55% WR +18% RD	x2	$<10^{-7}$	$>10^6$

decoding. BER and endurance achieve less than 10^{-7} and more than 1M cycles, respectively.

V. CONCLUSION

The paper explores the individual and cumulative benefits from most common programming techniques deployed in RRAM macro. The testchip including CL, WT, RBW, WV and ECC functionalities mastered from a flexible smart programming algorithm SWA embedded in on-chip RISC-V counteracts efficiently the inherent variability of 128kb RRAM. As described, each design assist technique acts differently on energy consumption, access time, BER and read margin. All activated, the endurance is 1M cycles while maintaining $28.1\mu\text{A}$ read margin without any read error ($\text{BER} < 10^{-7}$), consolidating the recent industrial RRAM adoption.

REFERENCES

- [1] C. Peters, F. Adler, K. Hofmann, J. Otterstedt, "Reliability of 28nm embedded RRAM for consumer and industrial products," IMW 2022.
- [2] EENews Europe, "Infineon's microcontrollers to go RRAM with TSMC", eewnews europe.com. <https://www.eewnews europe.com/en/infineons-microcontrollers-to-go-rram-with-tsmc/> (accessed Feb. 5, 2023).
- [3] Design & Reuse, "eMemory and UMC Bring New ReRAM Intellectual Property to Market", design-reuse.com. <https://www.design-reuse.com/news/50886/ememory-umc-rram-intellectual-property.html> (accessed Feb. 5, 2023).
- [4] A. Grossi et al., "Cell-to-Cell Fundamental Variability Limits Investigation in OxRRAM Arrays," in Electron Device Letters, 2018.
- [5] G. Molas and E. Nowak, "Advances in Emerging Memory Technologies: From Data Storage to Artificial Intelligence," Appl. Sci. 2021, 11, 11254.
- [6] L. Grenouillet et al., "16kbit 1T1R OxRAM arrays embedded in 28nm FDSOI technology demonstrating low BER, high endurance, and compatibility with core logic transistors," IMW 2021.
- [7] C-C. Chou et al., "An N40 256Kx 44 embedded RRAM macro with SL-precharge SA and low-voltage current limiter to improve read and write performance," ISSCC, pp. 478-480, 2018.
- [8] P. Jain et al., "A 3.6 Mb 10.1Mb/mm2 Embedded Non-Volatile ReRAM Macro in 22nm FinFET Technology with Adaptive Forming/Set/Reset Schemes Yielding Down to 0.5V with Sensing Time of 5ns at 0.7V," ISSCC, pp. 212-213, 2019.
- [9] C-C. Chou et al., "A 22nm 96KX144 RRAM Macro with a Self-Tracking Reference and a Low Ripple Charge Pump to Achieve a Configurable Read Window and a Wide Operating Voltage Range," Symp. VLSI Circuits, 2020.
- [10] J. Yang et al., "A 28nm 1.5Mb Embedded 1T2R RRAM with 14.8 Mb/mm2 using Sneaking Current Suppression and Compensation Techniques," Symp. VLSI Circuits, 2020.
- [11] J. Yang et al., "A 14nm-FinFET 1Mb Embedded 1T1R RRAM with a 0.022µm2 Cell Size Using Self-Adaptive Delayed Termination and Multi-Cell Reference," ISSCC, pp. 336-338, 2021.
- [12] V. Gherman, S. Evain and B. Giraud, "Binary Linear ECCs Optimized for Bit Inversion in Memories with Asymmetric Error Probabilities," Design, Automation and Test in Europe (DATE), 2020.
- [13] S. Balatti et al., "Voltage-Controlled Cycling Endurance of HfOx-Based Resistive-Switching Memory," IEEE Transactions on Electron Devices, vol. 62, no. 10, pp. 3365-3372, Oct. 2015.
- [14] D. Alfaro Robayo et al., "Endurance Statistical Behavior of Resistive Memories Based on Experimental and Theoretical Investigation," IEEE Transactions on Electron Devices, vol. 66, no. 8, pp. 3318-3325, Aug. 2019.