

# Comparing NVM Technologies through the Lens of Intermittent Computation

Tim Daulby  
University of Southampton  
Southampton, UK  
td2g14@soton.ac.uk

Alex S Weddell  
University of Southampton  
Southampton, UK  
asw@ecs.soton.ac.uk

Anand Savanth  
Arm Research  
Cambridge, UK  
anand.savanth@arm.com

Geoff V Merrett  
University of Southampton  
Southampton, UK  
gvm@ecs.soton.ac.uk

## ABSTRACT

Intermittent computing (IC) promises long lifetimes for IoT edge devices. Running directly from energy harvesting sources enables these devices to be deployed and left, potentially for decades. As the field of IC progresses from proof-of-concept to deployable devices, the research focus must shift from processor-centric schemes to consideration of the whole system. The non-volatile memory (NVM) technology, as well as the way it is used, will have a significant effect. Properties such as latency, read/write energy, and endurance can vary by orders of magnitude, and this may affect the viability of many schemes presented in the literature. This paper presents a review of the characteristics of both commercially-available and future NVM technologies, and recommends design considerations for IC systems which incorporate these.

## CCS CONCEPTS

• **Computer systems organization** → **Embedded and cyber-physical systems.**

## KEYWORDS

Intermittent Computing, Embedded Systems, IoT edge devices

### ACM Reference Format:

Tim Daulby, Anand Savanth, Alex S Weddell, and Geoff V Merrett. 2020. Comparing NVM Technologies through the Lens of Intermittent Computation. In *The 8th International Workshop on Energy Harvesting and Energy-Neutral Sensing Systems (ENSys '20)*, November 16–19, 2020, Virtual Event, Japan. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3417308.3430268>

## 1 INTRODUCTION

IoT edge devices are becoming increasingly pervasive, with many industries realising the value of a better connected physical world and cloud through live sensor data. However,

**Table 1: Comparison of NVM memory technologies for IoT Devices (\* signifies simulation data only)**

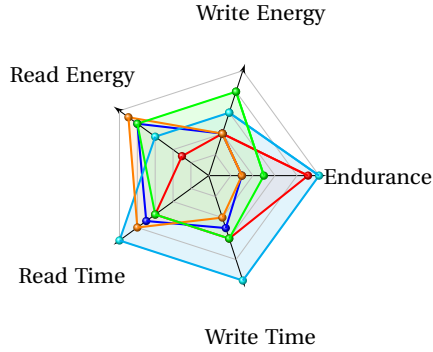
Technology	Write/Read Energy (per bit)	Write/Read Time (per bit)	COTS	Endurance (Write cycles)
NAND Flash[4]	470 pJ/46 pJ	200 $\mu$ s/25.2 $\mu$ s	Y	$10^5$
SRAM[1]	355 pJ/587 pJ	2.2 ns/2.1 ns	Y	Unlimited
FRAM[5]	1.4 nJ/1.4 nJ	120 ns/120 ns	Y	$10^{15}$
STT-MRAM[3]	2 nJ/34 pJ	250 ns/10 ns	Y	$10^5$
SOT-MRAM* [1]	334 pJ/247 pJ	1.4 ns/1.1 ns	N	$>10^{15}$
ReRAM[6]	1.1 nJ/525 fJ	10 $\mu$ s/5 ns	Y	$10^5$
PCM* [8]	13.5 pJ/2 pJ	150 ns/48 ns	N	$10^7$

the maintenance cost of battery powered systems, as well as their increased size, mass and cost, can be prohibitive. To overcome this, a new paradigm of battery-less devices is emerging, running intermittently on harvested energy. Intermittent computing (IC) works by saving state to non-volatile memory (NVM) before power failure. This state save can be periodic, at certain task boundaries, or triggered by a voltage interrupt. Most existing IC schemes have been implemented on commercial off-the-shelf (COTS) microcontrollers, e.g. MSP430FR5739, or in simulation. For systems to be deployed, greater consideration must be given to the impact of the whole system, including the available NVM technologies. Limited read/write speed, the high energy cost of accesses and poor endurance can all impact the system performance.

## 2 NVM TECHNOLOGY PROPERTIES

The emergence and subsequent proliferation of faster, more efficient byte-addressable NVM technologies has made IC viable, with some authors completely replacing RAM with NVM in a unified memory approach (i.e. saving all data usually separated to ROM and RAM in the same non-volatile memory space, greatly reducing checkpoint size)[7]. Table 1 and Figure 1 compare leading NVM technologies, with SRAM and Flash included for comparison. Whilst there are many other considerations for system designers, we consider these key criteria for IC systems.

**Energy consumption:** Perhaps the most obvious consideration for NVM in ultra-constrained IC devices. The energy consumption of embedded CPU cycles can be orders of magnitude lower than the NVM access cost. This motivates the use of IC schemes that minimise the number of accesses, such as managed-state checkpointing [9]. In addition, many



**Figure 1: Radar plot of NVM characteristics. Line colors correspond to the colors in Table 1.**

**Table 2: The number of clock cycles needed for NVM access at 24MHz CPU clock frequency.**

Technology	Write/Read Time (per bit)	Clock cycles (Write)	Clock cycles (Read)
NAND Flash	200 $\mu$ s/25.2 $\mu$ s	4800	605
SRAM	2.2ns/2.1ns	<<1	<<1
FRAM	120ns/120ns	2.9	2.9
STT-MRAM	250ns/10ns	6	0.24
SOT-MRAM	1.4ns/1.1ns	0.03	0.03
ReRAM	10 $\mu$ s/5ns	240	0.12
PCM	150ns/48ns	3.6	1.15

NVMs are asymmetric, i.e. writes are more expensive. This could motivate schemes that prefer re-execution over checkpointing, if clock cycles and reads are significantly cheaper.

**Latency:** The access times of NVM technologies vary by orders of magnitude, but there are scenarios where this will not affect the performance of the system. Table 2 shows the number of cycles needed to access NVM at 24MHz, indicating how the latency of these technologies would negatively impact performance. Enabling non-blocking writes would mask some of this latency but, when saving/restoring state, NVM latency will have a direct impact due to constant accesses. For unified memory approaches, the latency of these technologies directly constrains the CPU clock frequency.

**Endurance:** Intermittent schemes often use NVM for checkpointing in ways that would exhaust some NVM cells within a matter of months or even days. In Hibernus++ [2], for example, the authors demonstrate their scheme with a Seiko watch power trace, interrupting every 0.4s. With FRAM, this scheme could checkpoint successfully for 6.34 million years (ignoring all other factors), but STT-MRAM and ReRAM with  $10^5$  endurance would last only 4.63 days.

For devices to be truly long-running, NVM technologies with greater write endurance must be used, or IC schemes must consider checkpoints as a finite resource that should be more carefully allocated. Schemes such as Hibernus++, which save the entire RAM contents to NVM, must be replaced by schemes such as allocated/managed state [9] which greatly

reduce the number of writes required. Additionally, methods that write minimal/small checkpoints must use wear-leveilling approaches to avoid over-using regions of NVM. Due to their greater access frequency, unified memory approaches remain unsuitable for many NVM technologies owing to their endurance limitations. Many task-based IC approaches depend on unified memory models, so this could present a significant challenge, unless endurance can be improved.

### 3 CONCLUSIONS

For IC devices to go beyond research, into deployment, it is necessary to consider the impact of the NVM on system performance. Some NVMs which may initially appear suitable, such as SOT-MRAM, have not yet been commercialised, typically due to fabrication challenges or their early state of development. The drawbacks of available technology must be factored in to IC design, e.g. schemes with frequent checkpoints must sacrifice performance to target greater endurance.

The ideal case for IC is NVM that can be used as unified memory, as it has high endurance and low latency similar to SRAM. Many schemes in the literature utilise unified memory approaches, however this paper has identified how the increased latency/energy cost over SRAM cannot be ignored, and that low endurance is prohibitive with many existing NVM technologies. This paper aims to promote awareness of these issues, so that future research into IC can consider the importance of NVM to the success of these schemes.

### ACKNOWLEDGEMENTS

This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P010164/1. Experimental data used in this article can be found at: <https://doi.org/10.5258/SOTON/D1594>.

### REFERENCES

- [1] Antaios. 2020. *Spin-Orbit Torque MRAM*. Technical Report. Antaios, 51 Avenue Jean Kuntzmann, 38830 Montbonnot – France. 3 pages. [https://www.antaios.fr/IMG/pdf/web\\_site\\_sot\\_whitepaper.pdf](https://www.antaios.fr/IMG/pdf/web_site_sot_whitepaper.pdf)
- [2] Domenico Balsamo et al. 2016. Hibernus++: A Self-Calibrating and Adaptive System for Transiently-Powered Embedded Devices. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 35, 12 (March 2016), 1968–1980. <https://doi.org/10.1109/TCAD.2016.2547919>
- [3] Yu-Der Chih et al. 2020. 13.3 A 22nm 32Mb Embedded STT-MRAM with 10ns Read Speed, 1M Cycle Write Endurance, 10 Years Retention at 150° C and High Immunity to Magnetic Field Interference. In *2020 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, 222–224.
- [4] Laura M Grupp et al. 2009. Characterizing flash memory: anomalies, observations, and applications. In *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*. 24–33.
- [5] Texas Instruments. 2009. FRAM – New Generation of Non-Volatile Memory. <https://www.ti.com/lit/ml/szzt014a/szzt014a.pdf>
- [6] Pulkit Jain et al. 2019. 13.2 A 3.6Mb 10.1Mb/mm 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. In *2019 IEEE Int. Solid-State Circ. Conf. - (ISSCC)*. IEEE, San Francisco, CA, USA, 212–214. <https://doi.org/10.1109/ISSCC.2019.8662393>
- [7] Hrishikesh Jayakumar et al. 2016. Energy-Aware Memory Mapping for Hybrid FRAM-SRAM MCUs in IoT Edge Devices. In *2016 29th Int. Conf. on VLSI Design (VLSID)*. IEEE, Kolkata, India, 264–269. <https://doi.org/10.1109/VLSID.2016.52>
- [8] Benjamin C Lee et al. 2009. Architecting phase change memory as a scalable dram alternative. In *Proc. 36th annual int. sym. on Comp. arch.* 2–13.
- [9] Sivert T Sliper et al. 2019. Efficient state retention through paged memory management for reactive transient computing. In *Proceedings of the 56th Annual Design Automation Conference 2019*. 1–6.