SuperVAULT: Superparamagnetic Volatile AUxiLiary Tamper-Proof Storage

Nikhil Rangarajan, *Member, IEEE*, Johann Knechtel, *Member, IEEE*, Dinesh Rajasekharan, *Member, IEEE*, and Ozgur Sinanoglu, *Senior Member, IEEE*

***Abstract*—Memory security has recently come into the spotlight as attackers have stepped up their efforts to gain illicit access to sensitive data or cause denial-of-memory-service via a variety of avenues like cold boot attacks, bus snooping, and physical probing or tampering. In this paper, we propose *SuperVAULT*, a novel secure storage solution for protecting secret data/keys by exploiting the superparamagnetic regime of nanomagnets. Through materials and dimensional engineering of the magnetic tunnel junction (MTJ) free layer, the energy barrier of spin-transfer torque magnetoresistive random access memory (STT-MRAM) cells can be designed to lie in the range of the thermal energy (k_BT). Such superparamagnetic MTJ (s-MTJ) cells, with an $\mathcal{O}$(10 ns) retention time, need to be refreshed frequently. In the absence of data refresh (under attack conditions), the data they hold is thermally corrupted to a random state after an arbitrary but short amount of time. We leverage this property to devise a secure memory primitive and showcase its potential against cold boot and Boolean satisfiability (SAT) attacks. Further, the overheads for s-MTJ-based STT-MRAMs is shown to be promising for on-chip implementations.**

***Index Terms*—Superparamagnet, Tamper-proof, Secure memory, Stochastic switching, Cold boot attack, SAT attack**

1. Introduction

Secure data storage has become paramount for cryptographic systems and hardware-supported enclaves for secure computation, especially in the face of advanced attacks leveraging bus snooping [1], cold memory boot [2], electro-optical probing [3], etc. However, conventional dynamic random-access memory (DRAM) modules remain vulnerable to these attack avenues, rendering security-critical applications susceptible to malicious data stealing and tampering. For example, the $\mathcal{O}$(s) data retention time of DRAM [4] can result in refresh times of several seconds, which leaves a long window of opportunity for an adversary to cool the memory module and perform unauthorized data retrieval attacks [2]. Further, optical contactless probing on the substrate backside, via electro-optical methods [5], can probe memory data or write/read operations to discern secret keys.

Prior solutions to securing memory systems against physical attacks include (i) encrypting the data before memory write [6], (ii) implementing physical shield structures and encapsulations [7], and (iii) locking the memory module such that any attempt to physically remove it results in data erasure [8]. However, most of these methods incur non-negligible overheads that can prohibit their ubiquitous large-scale implementation. Independently, emerging non-volatile memories (NVM) like spin-transfer torque magnetoresistive random-access memory (STT-MRAM), phase-change memory (PCM), or resistive random-access memory (RRAM) present an alternate paradigm in the quest for ultra-scaled and low-power storage systems. Nonetheless, these emerging memories are plagued by the same vulnerability of data retention after power down, perhaps to an even greater extent [9].

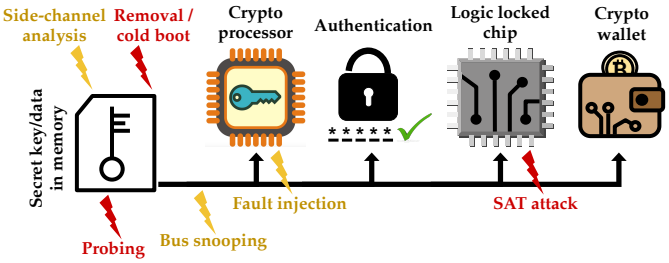

Fig. 1: Threat landscape for *SuperVAULT*. Here, the attacks highlighted in red are protected against, whereas those in yellow are not considered.

Scope of Work: In this paper, we propose an alternative to conventional DRAM-based secure storage solutions, in the form of *SuperVAULT*: A Superparamagnetic Volatile Auxiliary Tamper-Proof Memory.[1] *SuperVAULT* offers a contrasting departure from non-volatile auxiliary storage cells used to hold key bits, as well as from traditional ferromagnetic memories like the STT-MRAM. By leveraging the thermally-induced stochastic switching of nanomagnets in the superparamagnetic regime, *SuperVAULT* ensures optimal key degradation under attack conditions, wherein the attacker-recovered key bits are scrambled with a 50% probability, equating to random guessing of bits. Further, *SuperVAULT* incurs minimal overheads during normal operation.

Threat model: Figure 1 highlights the overarching scope of *SuperVAULT* and the associated threat landscape it caters to. Note that the current work safeguards against physical invasive attacks involving probing or removal, as well as SAT-based attacks. However, attacks based on side-channel analysis, fault injection, or bus snooping are not considered.

2. Superparamagnetic MTJs

Commercial STT-MRAMs have magnetic tunnel junctions (MTJ) with free layers (ferromagnets) engineered to retain

Nikhil Rangarajan, Johann Knechtel, and Ozgur Sinanoglu are with the Division of Engineering, New York University Abu Dhabi, Saadiyat Island 129188, UAE (e-mail: nikhil.rangarajan@nyu.edu, johann@nyu.edu, ozgursin@nyu.edu).

Dinesh Rajasekharan is with the Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, United States (email: dineshr@iitk.ac.in).

[1]Here, "auxiliary" is used to refer to the cells built solely for security purposes, as opposed to general-purpose memory used for all applications.

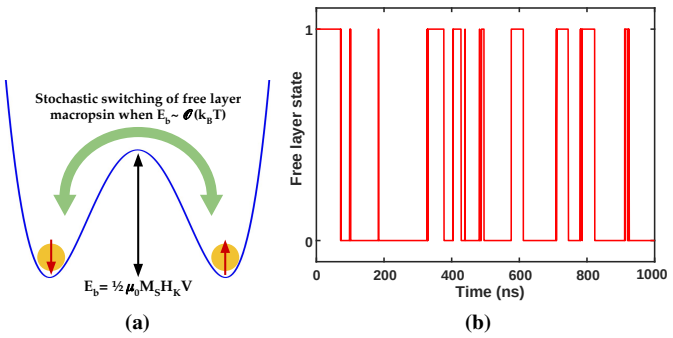

Fig. 2: (a) Two-well model of the free layer macrospin. When E_b is on the order of thermal energy $k_\mathrm{B}T$, the free layer switches stochastically between its two stable states, due to thermal fluctuations. (b) Stochastic telegraphic switching of a superparamagnetic free layer with $E_\mathrm{b} \sim 9\ k_\mathrm{B}T$.

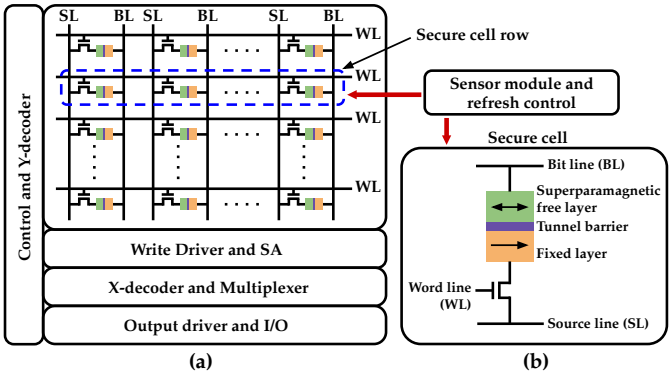

Fig. 3: (a) Architecture and organization of *SuperVAULT*, constructed using an STT-MRAM array [14] with secure cells. (b) Schematic of a secure cell with superparamagnetic free layer.

their states for >10 years [10]. The energy barrier for switching, given by $E_\mathrm{b} = 0.5 \times \mu_0 M_\mathrm{S} H_\mathrm{K} V$, can be controlled by manipulating the material properties and dimensions of the free layer. M_S is the saturation magnetization, H_K is the uniaxial anisotropy field, and V is the volume of the free layer.

Scaling down E_b below the thermal noise threshold brings the free layer into the superparamagnetic regime, where the ambient thermal fluctuations are enough to surmount the barrier and induce stochastic telegraphic switching. Under these conditions, the retention time of the free layer can be designed to lie within a few tens of nanoseconds. Hence, if such superparamagnetic MTJs (s-MTJs) are not refreshed before their retention time expires, the data bit they hold degrades and attains a random state.

Figure 2(a) illustrates the two-well model for stochastic free layer switching. Generally, ferromagnets with $E_\mathrm{b} \geq 40\times$ the thermal energy ($k_\mathrm{B}T$) are used for stable memory applications [10]. However, $E_\mathrm{b} \sim 10\ k_\mathrm{B}T$ or less can be obtained by (i) using a material with small M_S, H_K, or by (ii) reducing the volume of the magnet. Thermally-induced stochastic switching for a representative superparamagnet is shown in Fig. 2(b).

We note that, although s-MTJs have been fabricated in prior works [11], their experimental maturity is still nascent. In this work, we aim to study the security implications and promises of this emerging device for secure storage applications.

3. SuperVAULT Construction and Working

The basic premise of *SuperVAULT* involves replacing the requisite number of MTJs in an STT-MRAM array with s-MTJs, such that only those cells are used for storing critical data or keys. For instance, consider a 128×128 STT-MRAM array employed for storing the key bits of an AES-128 circuit. Here, one particular row (128 cells) of the STT-MRAM can be implemented with s-MTJs, to store the 128-bit key. These s-MTJ-based cells are the *secure cells* of the memory. The *secure cells* are designed with a small retention time (10-20 ns), and thus require frequent refresh operations to maintain data integrity. In general, the whole memory is partitioned into dedicated *general-purpose* cells and "auxiliary" *secure cells*, designed solely for secure data storage. This is because converting all the MTJs into s-MTJs would incur large refresh costs and might cause reliability issues.

Figure 3 shows the implementation of *SuperVAULT*. Note that the *secure cells* need not be adjacent and can be placed anywhere on the memory array. However, for simplicity of routing and addressing, we consider them to be contiguous in Fig. 3. In this work, we leverage the existing STT-MRAM crossbar-array architecture and tailor it for secure data storage.

We consider in-plane CoFeB-based s-MTJs similar to [11], with 8 $\Omega\mu\mathrm{m}^2$ RA product and 140% TMR. However, we simulate cuboidal free layers of dimensions 60×45×2 nm^3, with uniaxial energy densities in the range of ∼7000-12000 $\mathrm{J/m}^3$, to achieve the various energy barriers desired. Regarding the memory access time of *SuperVAULT*, we argue that it will be comparable to state-of-the-art STT-MRAM implementations (∼1.3 ns [12]). Note that the refresh operation in the *secure* cells might contribute to additional delays in memory access. Therefore, the estimated memory access time for *SuperVAULT* would be ∼2 ns, including a refresh time of ∼750 ps [13].

The *secure cells* in *SuperVAULT* are to be provisioned with sensory mechanisms, to detect incoming probing or removal attacks. Such sensing schemes can be realized at reasonable overheads using, e.g., cryptographically secure shields [7] or resistance/capacitance sensor implementations [15], [16]. An active shield [7] placed over the terminals of the MTJ crossbar-array can be used to thwart the attacker from querying the memory. As soon as the attacker attempts to probe the memory, the monitoring data sequence passing through the mesh wires will get altered, thus detecting the incursion. The power overheads for implementing an active shield can be limited to <10%, by controlling its monitoring frequency.

TABLE I: Overheads for *secure cell* implementation, at various energy barriers.

	9 $k_\mathrm{B}T$	**12 $k_\mathrm{B}T$**	**15 $k_\mathrm{B}T$**
Refresh power	3.79 pW	13.3 pW	19.2 pW
Area	Nil	Nil	Nil

Operating temperature is 300 K. Refresh power is calculated at iso-performance for 128 *secure cells* in a 128×128 STT-MRAM array. A refresh time of 1 ns is considered. The refresh current for the s-MTJs is in the range of $\mathcal{O}(\mu\mathrm{A})$ owing to the extremely low energy barriers. Area overheads are negligible since there is no difference between a regular MTJ and an s-MTJ at the layout-level, including peripherals.

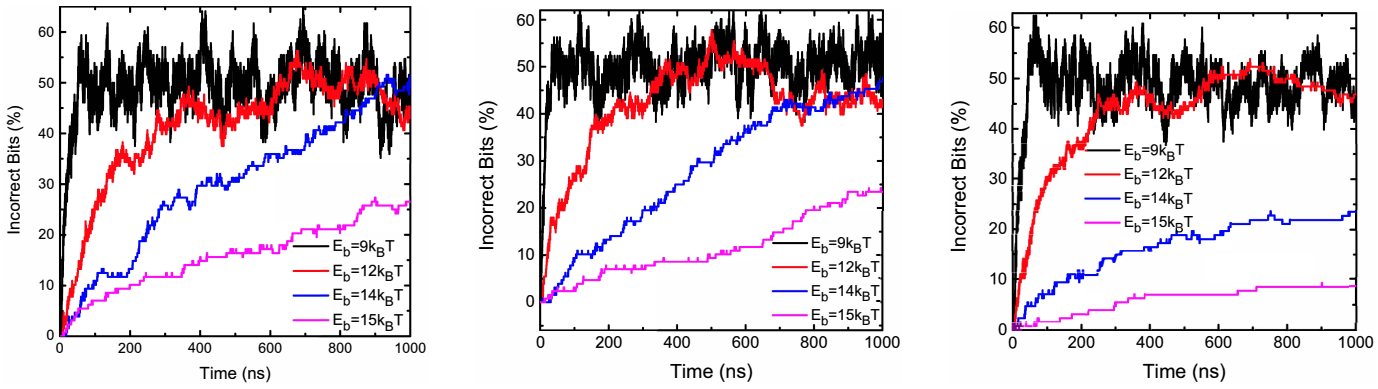

Fig. 4: Key degradation, as percentage of incorrect bits, under attack conditions, for various s-MTJ energy barriers. The degradation rate is highlighted for (a) nominal T = 300 K, (b) cooling rate of 0.01 K/ns, and (c) cooling rate of 0.10 K/ns.

Further, there is no impact on the memory performance, given that the authentication process can be completed well within the retention time of an s-MTJs, in each cycle. Based on such sensory modules, an *all-clear* signal is generated.

Under *normal conditions*, when the memory system is not under attack, the *all-clear* signal is high and the *secure cells* are refreshed before their retention time expires. The refresh costs of the considered 128×128 STT-MRAM array with 128 *secure cells*, at various E_b, are presented in Table I. Under *attack conditions*, the *all-clear* signal goes low, and the refresh operation of the *secure cells* is halted. Now, when the retention time expires, the original data stored in the *secure cells* becomes garbled. Such data loss in the volatile *secure cells* is not prohibitive since we assume that the sensitive data is either generated locally, using, e.g., physically unclonable functions (PUFs) or true random number generators (TRNGs), or obtained through a secure communication link. For logic locking, such techniques have been proposed in [17], [18].

We note that some sensory mechanism in conjunction with a conventional DRAM storage is also promising for secure storage applications. However, the larger retention time of DRAM could still enable an attacker to extract meaningful information after attack detection. The smaller retention time of s-MTJs and their rate of data degradation due to thermal noise play a big role in thwarting advanced cold boot attacks.

4. Attack Vectors and Results

In this section, we discuss the performance of *SuperVAULT* against two of the attack vectors highlighted in Fig. 1, viz. cold boot attacks and Boolean satisfiability (SAT)-based attacks.

4.1. Cold Boot Attacks

As mentioned, cold boot attacks exploit the latency between the logical switch-off of the target memory and the physical dissipation of its remnant state information. This latency is further exacerbated by cooling the memory cells, which reduces the entropy and halts data degradation.

SuperVAULT is able to circumvent such attacks based on temperature manipulation, by virtue of its extremely small retention time (∼10-20 ns). Figure 4 showcases the temporal key drift in *SuperVAULT* at various s-MTJ energy barriers. Here, a 128-bit key is stored in 128 *secure cells* of *SuperVAULT*, whose refresh operation is halted at $t = 0$ after the detection of the cold boot/removal attack. The initial operating temperature is assumed to be 300 K, which stays constant in the case of Fig. 4(a). For Figs. 4(b) and 4(c), a negative temperature gradient of 0.01 K/ns and 0.10 K/ns is applied, respectively. As seen in the figures, the percentage of incorrect bits rises sharply for $E_b = 9\ k_BT$, and hovers around 50% after a few ns. The rate of this key/data degradation is negligibly affected by the cooling processes in Figs. 4(b) and 4(c), for $E_b = 9$ and 12 k_BT. For free layers with larger E_b (14–15 k_BT), the rate of degradation is diminished when cooled at 0.10 K/ns.

This observation shows that cold boot attacks would be unable to prevent key loss in *SuperVAULT* once the energy barrier of the free layer is sufficiently low. At most, the attacker can hope to recover a key with 40-60% correct bits. Note that advanced cryo-cooling rates are reported in the range of ∼0.001 K/ns [19]; our assumed temperature gradients are for even more aggressive attack scenarios. The conclusions drawn from Figure 4 can be generalized to other removal or probing attacks, as the key degradation after attack detection is inherent to the *secure cells* proposed using s-MTJs.

4.2. SAT Attacks

SAT attacks [20]–[22] are widely used to decipher the secret keys of logic-locked IPs. The secret keys are typically stored in a tamper-proof memory, to prevent key recovery attacks. Here, we demonstrate the potential of *SuperVAULT* for providing such tamper-proof storage against SAT attacks. For this experiment, we lock *c5315*, *c6288*, and *c7552* from the *ISCAS-85* benchmark suite [23], with 64, 128, and 256 randomly-placed key gates, respectively.

We run the conventional SAT attack [20] and the probabilistic SAT (PSAT) attack [24] on each of these locked circuits. The PSAT attack is chosen since it was shown to be more effective in resolving error-prone probabilistic circuits than the conventional SAT attack; thermally-induced key degradation is conceptually similar to errors observed for probabilistic circuit behavior. We observe that both the attacks always result

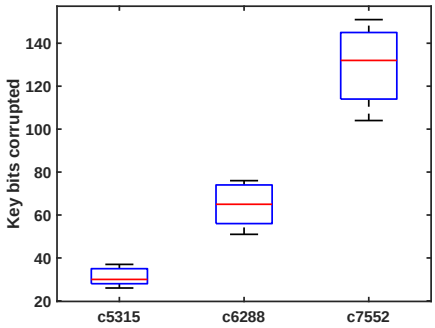

Fig. 5: Number of key bits corrupted, after 1,000 ns of thermal evolution, for *c5315*, *c6288*, and *c7552* circuits locked with 64, 128, and 256 key gates, respectively. The thermal evolution is repeated 10 times per key size.

in an *UNSAT* condition for over 1,000 attack runs for each benchmark, hence thwarting key recovery attacks altogether. This is because the significant and fast key degradation renders the input/output (I/O) behavior of the chip inconsistent the moment the s-MTJ refresh is halted. Such inconsistent behavior leads to an inconsistent SAT model for the SAT/PSAT attacks, as these models have to be build up iteratively over multiple I/O observations. With such inconsistent models, SAT/PSAT naturally cannot infer the correct key.

As seen from Fig. 5, about $\sim$50% of the key bits are corrupted due to the key drift. This is because, in the absence of key refresh, the thermally-induced degradation ensures that any key bit can be found in '0' or '1' state with equal probability, after arbitrary time. For the purpose of this simulation, we let the original key thermally evolve, and sample a degraded key after 1,000 ns. The telegraphic switching of a key bit, as shown in Fig. 2(b), can result in reverse flipping errors, wherein the initial key bit switches to an incorrect value and then back to its correct state. However, after thermal evolution for 1,000 ns, approximately 50% of the key bits in the ensemble will be in incorrect states, since each bit has two equiprobable states and all the bits evolve independently.

The onset of the SAT attack can be detected using scan chain access monitoring schemes [25]. As soon as an unauthorized scan access is detected, the authentication check fails and the refresh operation is halted, thus allowing the keys in the *secure* cells to drift. While prior works like [25] also flush the stored key bits after detecting scan-based attacks, the implementation approach is quite different. *SuperVAULT* does not require explicit erasure after attack detection, and halting the refresh operation is enough to cause key drift naturally. Further, the proposed *SuperVAULT* memory is a more encompassing device-level solution, capable of thwarting physical probing and removal attacks as well, whereas [25] focuses solely on scan-based attacks.

5. Conclusion

In this paper, we introduce a novel tamper-proof storage solution to thwart key stealing attacks on security-critical applications. The proposed *SuperVAULT* primitive utilizes the superparamagnetic phenomenon in small free layers of MTJs, to achieve $\mathcal{O}$(10 ns) retention times. *SuperVAULT* halts its key refresh operation under attack, whereafter the superparamagnetic MTJs evolve to a random state, thus rendering key recovery attempts futile. We demonstrate the efficacy of *SuperVAULT* against cold boot attacks at various cryo-cooling rates, verify its resilience against SAT attacks, and highlight its ultra-low overheads to show its potential for an efficient on-chip secure storage implementation.

References

[1] D. Lee et al. An off-chip attack on hardware enclaves via the memory bus. In *29th USENIX Security Symposium (USENIX Security 20)*, 2020.
[2] J. A. Halderman et al. Lest we remember: cold-boot attacks on encryption keys. *Communications of the ACM*, 52(5):91–98, 2009.
[3] M. T. Rahman et al. CONCEALING-Gate: Optical contactless probing resilient design. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 17(3):1–25, 2021.
[4] A. Rahmati et al. Refreshing thoughts on DRAM: Power saving vs. data integrity. In *Workshop on Approximate Computing Across the System Stack (WACAS)*, volume 44, 2014.
[5] N. Asadizanjani et al. Optical inspection and attacks. In *Physical Assurance*, pp. 133–153. Springer, 2021.
[6] D. McGrew and J. Viega. The Galois/counter mode of operation (GCM). *Submission to NIST Modes of Operation Process*, 20:0278–0070, 2004.
[7] J.-M. Cioranesco et al. Cryptographically secure shields. In *2014 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, pp. 25–31. IEEE, 2014.
[8] H. Seol et al. Amnesiac DRAM: A proactive defense mechanism against cold boot attacks. *IEEE Transactions on Computers*, 70(4):539–551, 2021.
[9] N. Rangarajan et al. SMART: A secure magnetoelectric antiferromagnet-based tamper-proof non-volatile memory. *IEEE Access*, 8:76130–76142, 2020.
[10] W. Zhao et al. Design considerations and strategies for high-reliable STT-MRAM. *Microelectronics Reliability*, 51(9):1454–1458, 2011.
[11] C. Safranski et al. Demonstration of nanosecond operation in stochastic magnetic tunnel junctions. *Nano Letters*, 21(5):2040–2045, 2021.
[12] T.-H. Yang et al. A 28nm 32Kb embedded 2T2MTJ STT-MRAM macro with 1.3 ns read-access time for fast and reliable read applications. In *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*, pp. 482–484. IEEE, 2018.
[13] G. Jan et al. Achieving sub-ns switching of STT-MRAM for future embedded LLC applications through improvement of nucleation and propagation switching mechanisms. In *2016 IEEE Symposium on VLSI Technology*, pp. 1–2. IEEE, 2016.
[14] L. Chang et al. Multi-port 1R1W transpose magnetic random access memory by hierarchical bit-line switching. *IEEE Access*, 7:110463–110471, 2019.
[15] P. Tuyls et al. Read-proof hardware from protective coatings. In *International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 369–383. Springer, 2006.
[16] D. J. Boday et al. Implementing carbon nanotube based sensors for cryptographic applications, 2014. US Patent 8,797,059.
[17] J. Rajendran et al. Fault analysis-based logic encryption. *IEEE Transactions on Computers*, 64(2):410–424, 2015.
[18] J. A. Roy et al. Ending piracy of integrated circuits. *Computer*, 43(10):30–38, 2010.
[19] B. Bikker. Measurement of cooling rate during plunge freezing of sample preparation in cryo electron microscopy. 2019.
[20] P. Subramanyan et al. Evaluating the security of logic encryption algorithms. In *2015 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, pp. 137–143. IEEE, 2015.
[21] D. Šišejković et al. Inter-lock: Logic encryption for processor cores beyond module boundaries. In *2019 IEEE European Test Symposium (ETS)*, pp. 1–6. IEEE, 2019.
[22] D. Šišejković et al. A secure hardware-software solution based on RISC-V, logic locking and microkernel. In *Proceedings of the 23th International Workshop on Software and Compilers for Embedded Systems*, pp. 62–65, 2020.
[23] M. C. Hansen et al. Unveiling the ISCAS-85 benchmarks: A case study in reverse engineering. *IEEE Design & Test of Computers*, 16(3):72–80, 1999.
[24] S. Patnaik et al. Spin-orbit torque devices for hardware security: From deterministic to probabilistic regime. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 39(8):1591–1606, 2019.
[25] N. Limaye et al. Thwarting all logic locking attacks: Dishonest oracle with truly random logic locking. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 40(9):1740–1753, 2021.