An Energy Efficient Non-Volatile Flip-Flop based on CoMET Technology

Robert Perricone∗, Zhaoxin Liang†, Meghna G. Mankalale†, Michael Niemier∗, Sachin S. Sapatnekar†, Jian-Ping Wang† and X. Sharon Hu∗

∗Department of Computer Science and Engineering, University of Notre Dame
Notre Dame, IN 46556, USA, Email: {rperricone,shu,mniemier}@nd.edu
†Department of Electrical and Computer Engineering, University of Minnesota
Minneapolis, MN 55455, USA, Email: {zliang,mankalale,sachin,jpwang}@umn.edu

Abstract—As we approach the limits of CMOS scaling, researchers are developing “beyond-CMOS” technologies to sustain the technological benefits associated with device scaling. Spintronic technologies have emerged as a promising beyond-CMOS technology due to their inherent benefits over CMOS such as high integration density, low leakage power, radiation hardness, and non-volatility. These benefits make spintronic devices an attractive successor to CMOS—especially for memory circuits. However, spintronic devices generally suffer from slower switching speeds and higher write energy, which limits their usability. In an effort to close the energy-delay gap between CMOS and spintronics, device concepts such as CoMET (Composite-Input Magnetoelectric-base Logic Technology) have been introduced, which collectively leverage material phenomena such as the spin-Hall effect and the magnetoelectric effect to enable fast, energy efficient device operation. In this work, we propose a non-volatile flip-flop (NVFF) based on CoMET technology that is capable of achieving up to two orders of magnitude less write energy than CMOS. This low write energy (≈2 aJ) makes our CoMET NVFF especially attractive to architectures that require frequent backup operations—e.g., for energy harvesting non-volatile processors.

I. INTRODUCTION

The persistent scaling of CMOS technology has allowed for computing performance to grow exponentially with each successive technology node generation. However, recent technology nodes have not enjoyed the same performance gain due in part to the growth of sub-threshold leakage current and minimal supply voltage scaling. To overcome the limits of CMOS technology, the semiconductor industry is exploring new technologies for the post-CMOS era [1]. These “beyond-CMOS” technologies represent a broad class of technologies that include charge-based transistor-like devices as well as spintronic devices [2].

Spintronic devices are of particular interest due to their inherent non-volatility, high integration density, radiation hardness, and negligible standby power [3]. To leverage the properties of spintronics, researchers have explored the full spectrum of on-chip integration from flip-flops (FFs), register files, and caches, up to a completely non-volatile architecture [3], [4], [5]. With the growth of the Internet of Things (IoT), these NV processors (NVPs) are gaining interest in a variety of application-spaces—e.g., in energy harvesting scenarios, for power state transitions, etc [3], [5], [6]. However, challenges stem from the fact that NV memories (NVMs) typically incur high write costs (i.e., in terms of energy) compared to their CMOS counterparts [7], [2]. Therefore, to make NV storage elements more suitable, processors must be judicious about when to perform a backup operation.

Using energy harvesting as an example, in a traditional non-volatile processor (NVP) one must harvest sufficient energy required by a backup operation before computation can be performed. Once computation has commenced, if ambient energy drops below a certain threshold, backup operations must ensue, and the aforementioned condition ensures that sufficient energy exists for said backup. However, in power constrained environments, the harvesting of sufficient energy becomes more challenging—especially to meet the energy costs of spintronic NVMs. As a result, computational progress is stifled. Similarly, for highly intermittent power supplies, more backup operations must be performed, which results in the harvested energy being primarily used for backups instead of computation. To address these concerns, complex backup policies have become ubiquitous in spintronic-based NVPs [5]. While backup policies can mitigate the high energy and performance overheads of backing up to a NVM, they do so at the expense of chip area and complexity.

To reduce performance overheads of writing to NVMs, previous works have investigated the use of non-volatile flip-flops (NVFFs) within registers that are placed in parallel with CMOS FFs to speedup backup and recovery operations [8], [7]. While NVFFs provide a means to overcome the poor switching speeds of NV devices, they do not reduce the overall energy cost of backup operations—i.e., the underlying NV technology remains a limitation.

In this work, we propose a NVFF based on CoMET technology [9] that is capable of achieving very low write energy (≈2 aJ). Furthermore, our CoMET NVFF removes the need to implement complex backup operations since backups are achieved implicitly without explicit write operations, which in turn allows for more harvested energy to be utilized to make computational progress. CoMET technology is also highly compatible with CMOS, which simplifies its integration. We describe how CoMET technology can be leveraged to create a low write energy NVFF, and present the design of a master-slave edge-triggered NVFF that requires just a single CoMET device. To evaluate the feasibility of our design, we performed micromagnetic simulations using simulation parameters that closely match realistic materials. Furthermore, our simulations include thermal noise at 300 K, which to the best of our knowledge is the first time CoMET has been considered in the context of thermal noise. Our results suggest that CoMET can
be leveraged to create a low write energy NVFF that utilizes up to two orders of magnitude less energy when compared to CMOS FFs, which allows for nearly free backup operations.

II. BACKGROUND

In Sec. II-A, we provide an overview of CoMET technology and a step-by-step summary of its operation. Then, in Sec. II-B, we present an overview of non-volatile processors (NVPs) and how they utilize NVFFs.

A. CoMET Overview

Composite input MagnetoElectric-based logic Technology (CoMET) [10] is a fast, low-energy spintronics-based logic device concept. Fig. 1 illustrates two cascaded CoMET inverter stages, and briefly illustrates the principle of operation of the first stage. A voltage applied on an input ferroelectric (FE) capacitor nucleates a domain wall (DW) via the magnetic field (ME) effect [11]. A composite ferromagnet (FM) structure with an in-plane magnetic anisotropy (IMA) layer is placed above the perpendicular magnetic anisotropy (PMA)–FM channel to enable fast, energy-efficient nucleation in the PMA–FM channel at the input end. The DW is propagated to the output end of the FM channel using a charge current applied to a layer of spin-Hall material placed under the PMA channel. The inverse–ME (IME) effect induces a voltage at the output node using the IME, and is transmitted to the next stage through the dual-rail inverter. The device concept has been shown through simulations to be viable when mapped to realistic material parameters [10].

In more detail, stage 1 (Fig. 2(a)) represents time \( t = 0 \) where a CoMET device is in steady-state with the IMA layer initialized to the +z direction, while the PMA layer is initialized to the −z direction. In stage 2 (Fig. 2(b)), a voltage (positive or negative) is applied to the input FE capacitor \( V_{FE} > 0 \), which charges the capacitor. This produces an effective magnetic field \( (B_{ME}) \) through the ME effect, which affects the composite structure. As a result, the magnetic direction of the IMA layer is flipped to −x and a DW is nucleated in the PMA layer with a down-up configuration (an initial state of −z would result in a up-down configuration).

For stage 3 (Fig. 2(c)), the DW is propagated through the PMA layer by setting the voltage \( V_{PROP} > 0 \), which sends a charge current through the SHM. Due to the combination of the SHE and the Dzyaloshinskii-Moriya interaction (DMI) [12], a magnetic torque is induced at the IMA-FM and SHM interface. As a result, the DW is propagated to the end of the PMA layer (+x direction). Furthermore, before the DW reaches the end of the PMA layer, the output FE capacitor is connected to ground and subsequently charged by setting the voltage \( V_{RST} > 0 \) (see Fig. 1) as a result of the electric field across it and the IME. Finally, in stage 4 (Fig. 2(d)), \( V_{RST} \) is set to 0 and the DW reaches the end to complete switching of the PMA layer. The magnetization of the PMA layer couples with the polarization of the output FE capacitor through the IME effect, which induces the output voltage \( (V_{OUT}) \).

In [10] a breakdown of the energy and delay of each stage in the CoMET operation is provided for technology feature sizes of 15 nm and 7 nm (unless otherwise noted, we assume 15 nm technology in our evaluations). Table I and Table II summarize the energy and delay per each CoMET operation for a 15 nm CoMET inverter from [10], respectively. In Table I, the energy required for charging the input FE \( (E_{FE}) \), turning transistors on \( (E_{TX}) \), Joule heating \( (E_{Joule}) \), and transistor leakage energy \( (E_{leak}) \) are presented for two different input voltages (i.e., 110 mV and 150 mV). Similarly, Table II presents the nucleation time \( (t_{nucleate}) \), DW propagation time \( (t_{propagate}) \), and the transfer time through the output inverter \( (t_{transfer}) \). This data will be used to evaluate lower energy, CoMET-based NVFFs.

**TABLE I: 15 NM COMET INVERTER ENERGY [10]**

<table>
<thead>
<tr>
<th>( V_{FE} ) (mV)</th>
<th>( E_{FE} ) (aJ)</th>
<th>( E_{TX} ) (aJ)</th>
<th>( E_{Joule} ) (aJ)</th>
<th>( E_{leak} ) (aJ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>0.8</td>
<td>24.2</td>
<td>1.6</td>
<td>16.3</td>
</tr>
<tr>
<td>150</td>
<td>1.3</td>
<td>30.6</td>
<td>1.3</td>
<td>22.8</td>
</tr>
</tbody>
</table>

**TABLE II: 15 NM COMET INVERTER DELAY [10]**

<table>
<thead>
<tr>
<th>( V_{FE} ) (mV)</th>
<th>( t_{nucleate} ) (ps)</th>
<th>( t_{propagate} ) (ps)</th>
<th>( t_{transfer} ) (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>35</td>
<td>38.7</td>
<td>8.3</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>38.7</td>
<td>8.3</td>
</tr>
</tbody>
</table>
B. NVP Overview

The advent of IoT has caused an increased interest in NVPs—especially for energy-harvesting applications [5], [13], [14]. Most NVPs can be classified as either explicit backup (EB-NVPs) or implicit backup (IB-NVPs) [14]. EB-NVPs are the most common, which involves volatile components for computation that must explicitly backup to NVMs in the event of a loss of power. IB-NVPs are NVPs where all memory components are comprised of NV technologies whereas logic can be volatile or NV. They are backed up implicitly due to the inherent non-volatility of the underlying memory components.

Most proposed NVPs are comprised of volatile components (i.e., CMOS) for computation and NV memory for backup. However, these EB-NVPs suffer in terms of computational progress due to the poor energy-delay of NV devices compared to their volatile counterparts. While this difference can be masked for NVPs with consistent power supplies, NVPs with intermittent power supplies suffer due to the high frequency and energy cost of backup/recover operations [5], [7].

To overcome the performance overheads associated backup/restore operations in NVPs, previous work has proposed integrating NV devices closer to the processor pipeline (i.e., register file, pipeline latches, etc. versus cache or off-chip). This granularity of integration is more preferable since the time required for a backup/recover operation can be reduced by performing backup/recover operations in parallel at the bit-level [7], [8]. However, the overall energy required for a backup/recovery operation remains unchanged. Ultimately, overcoming the energy-delay wall of NV devices requires a NV technology that has an energy-delay product (EDP) that is comparable or better than its CMOS counterpart.

Initial research suggests that CoMET is capable of achieving very low energy writes without significant performance overheads. As we will discuss in Sec. V-B, a CoMET NVFF achieves an EDP that is over an order of magnitude less than its CMOS counterpart. In the next section, we introduce our CoMET-based NVFF and discuss its operation.

III. CoMET-BASED NVFF DESIGN AND OPERATION

Given the basic operating procedure for CoMET in Sec. II-A, we now present a description of our CoMET-based NVFF. While the use of CoMET for combinational logic has been studied and shown to be promising [10], naively constructing a NVFF from combinational logic would only yield a 1.5× reduction in write energy compared to a low voltage CMOS equivalent whereas our design achieves a 42× reduction (discussed further in Sec. V-B). The key idea behind the design of CoMET-based NVFF is leveraging the nucleated DW as a NV storage element instead of the entire PMA layer. Thereby, a write operation only involves the nucleation of a DW, which per Table I, requires ≈1–2 aJ. Below, we present our CoMET NVFF and discuss how a single device can be used as a master-slave edge-triggered FF for use in IB-NVPs.

To transform CoMET into a low write energy NVFF, we divide the four stages discussed previously in Sec. II-A in half. The first half comprises stages 1 and 2, which will correspond to a write operation—i.e., DW nucleation. The second half comprises stages 3 and 4, which will correspond to an output/read operation—i.e., DW propagation and inducing an output voltage through the IME.

Fig. 3: Proposed CoMET NVFF with write and read stages. A write involves DW nucleation while a read involves DW propagation and inducing an output voltage through the IME. The input voltage $V_{FE}$ represents the $D$-input to the flip-flop. Moreover, the read signal $V_{READ}$—which would be replaced by a clock signal if the CoMET NVFF is used for sequential logic—is placed at the $V_{PROP}$ input per Fig. 3. Correctly writing the CoMET-based NVFF requires that hold time ($t_{hold}$) of the input be at least as long as the nucleation time (i.e., $t_{hold} \geq t_{nucleate}$). For example, if writing a logic ‘1’ is represented by a voltage of 150 mV at the $D$-input, then $t_{hold} \geq 30$ ps. To output/read the CoMET-based NVFF, the $V_{READ}$ signal is raised, which causes the written DW to propagate and an output voltage to be induced through the IME effect—assuming $V_{RST}$ operates as described in Sec. II-A.

Unlike more traditional spintronic technologies that require at least two devices per master-slave edge-triggered NVFF [3], it is possible to utilize a single CoMET device. We show a negative edge-triggered NVFF in Fig. 3 by replacing “write” and “read” with “master” and “slave”, respectively. To complete the negative edge-triggered NVFF design, it is only necessary to replace the read signal by an active low clock signal $V_{CLK}$, which would propagate the DW during the low part of the clock cycle (CC). For correct operation, the CC time will need to be set according to Eq. 1 (assuming that writing and reading the NVFF is on the critical path).

$$CC \text{ Time} \geq 2 \times \max (t_{hold}, t_{propagate} + t_{transfer}) \quad (1)$$

For example, a 150 mV input from Table II would yield a clock cycle time of at least 94 ps. While the clock is high, a DW will be nucleated into the PMA layer as previously discussed. At the negative edge of the clock, the $D$-input goes low and the transistor connected to the clock signal will turn on (active low). This results in propagation of the DW followed by an induced voltage at the output of the inverter. This single device edge-triggered FF is made possible due to two key factors. First, is our use of a DW as a storage element and, second, is the signal timings of the basic CoMET operation.
Fig. 4: Y-slice view of the simulated IMA and PMA layers.

Fig. 5: Z-slice view of the simulated IMA and PMA layers.

specifically related to the charging of the input FE (i.e., for DW nucleation) followed by propagation of the DW.

For the low write energy NVFF to be functional, it is necessary to demonstrate DW stability—especially in the presence of thermal noise. In Sec. IV, we discuss our simulation infrastructure and setup used to assess DW stability.

IV. SIMULATION INFRASTRUCTURE

In this section we present the simulation infrastructure and the device input parameters we employ to evaluate our CoMET-based NVFF. Note that, the fundamental mechanisms that make up CoMET such as the ME effect [15], [16] and DW memories [17], [18] have all been experimentally demonstrated. Here, we leverage micromagnetic simulation to demonstrate the feasibility of utilizing the PMA-FM layer as a DW memory. We calibrated the simulation inputs to closely match existing materials.

To validate our design through simulation, we follow a similar simulation scheme presented in [19] and used to evaluate the original CoMET device in [10]. This approach involves utilizing both the LKh and Landau-Lifshitz-Gilbert (LLG) equations to capture the ME effect and magnetization dynamics, respectively [20]. Here, we map CoMET to realistic materials and perform LLG simulations to determine the required field strength from the ME effect necessary to nucleate a DW in the PMA layer. Once the field strength has been established, the LKh equation is solved to determine the input voltage and resulting nucleation energy.

To simulate the CoMET structure, we leverage the MuMax3 GPU-accelerated micromagnetic simulator [21]. The simulated device structure is illustrated in Fig. 4 (y-slice) and Fig. 5 (z-slice)—i.e., both the IMA and PMA layers are simulated. The IMA and PMA layers are assumed to be 32 nm × 16 nm × 1 nm and 64 nm × 16 nm × 1 nm (L × W × H), respectively. The total grid size is 128 nm × 32 nm × 64 nm with a mesh size of 1 nm × 1 nm × 1 nm. The magnetization direction of the IMA and PMA layers is initialized to +z and +z, respectively. The simulation environment will also include thermal noise set to 300 K with a time step of 8 × 10^{-17} s.

To make our simulations more realistic, we map our device input parameters to realistic materials. Based on the design of the IMA and PMA (FePt) layers.

TABLE III: IMA AND PMA MATERIAL PARAMETERS

<table>
<thead>
<tr>
<th>Material</th>
<th>Aex (J/m)</th>
<th>Msat (A/m)</th>
<th>KU (J/m²)</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoFeB [22]</td>
<td>10e-12</td>
<td>1.0e6</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>FePt [23], [24]</td>
<td>10e-12</td>
<td>0.42e6</td>
<td>0.46e6</td>
<td>0.2</td>
</tr>
</tbody>
</table>

To simulate the ME effect, an external field is applied to the IMA layer and to half of the PMA layer (the part directly underneath the IMA layer) as illustrated in Fig. 6. The field direction is completely in the −z direction (i.e., 270° from +x-axis). The upper bound of the applied field magnitude is 530 mT, which is the strength of the effective magnetic field from ME effect with a 188 mV input for \( V_{FE} \). The upper bound of the applied field magnitude was determined through repeated simulations to be the maximum required field to guarantee stable DW nucleation in all 100 of our simulations. To obtain the input supply voltage and nucleation energy, we follow the same approach as in [10], which involves solving the Landau-Khalatnikov (LKh) equation [25]. To ensure a domain wall is nucleated, the externally applied field is active for 300 ps—an order of magnitude greater than the expected nucleation time from [10]. Given the initial magnetization direction and applied field parameters, a DW is nucleated in the PMA layer with a down-up direction.

To establish the stability of the nucleated domain wall in the PMA layer, our simulations involve three parts: (i) a wait period where the IMA and PMA layers are allowed to couple without any external stimuli (Wait Time = 200 ps), (ii) an applied field period to simulate the ME effect (Applied Field Duration = 300 ps), and (iii) a final wait period where the DW stability of the DW will be determined (Final Wait Time = 500 ps). Collectively, the total time simulated is 1 ns, which requires 50 hours of computation time and precludes exploring longer simulation times.

V. RESULTS AND ANALYSIS

In Sec. V-A, we summarize the results of our CoMET simulations using the input data presented in Sec. IV. Next, in Sec. V-B, we compare our CoMET NVFF to previously proposed NVFFs from the literature.

A. Simulation Results

To capture how the DW changes in the PMA layer throughout the simulation, we first create a “snapshot” file for every
As discussed in Sec. IV, from simulation time $t_{sim} = 0$ ps until $t_{sim} = 200$ ps, the IMA and PMA layers are allowed to couple with no external stimuli. The shades of blue over this time period represent a more positive $z$ component of magnetization. Next, from $t_{sim} = 200$ until $t_{sim} = 500$ ps, an external field is applied to simulate the ME effect. When $t_{sim} = 230$ ps the yellow color indicates the formation of a negative magnetic domain and, hence, DW. However, the DW does not become well-formed until $t_{sim} = 300$ ps where a negative $z$-directional magnetic domain is denoted by the red/orange color. The external field is removed at $t_{sim} = 501$ ps and the DW remains fairly static throughout the remainder of the simulation. While the size of the DW oscillates, it is not essential that the DW maintain a certain size. It is only necessary that the leftmost side ($-x$ direction) of the PMA layer sustains the DW in order for the DW to be propagated to the rightmost side ($+x$ direction) during an operation. We repeat the above simulation process for 100 simulations with thermal noise at 300 K. We generated 100 unique thermal noise seeds to seed the random number generator of the MuMax3 simulator. The results from our simulations are presented in Table IV.

**TABLE IV: SUMMARY OF 15 NM CoMET SIMULATIONS**

<table>
<thead>
<tr>
<th>Applied Field (mT)</th>
<th>Stable DW %</th>
<th>$V_{FE}$ (mV)</th>
<th>$E_{FE}$ (aJ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>420</td>
<td>64%</td>
<td>150</td>
<td>1.5</td>
</tr>
<tr>
<td>530</td>
<td>100%</td>
<td>188</td>
<td>2.3</td>
</tr>
</tbody>
</table>

In Table IV, we provide the applied field magnitude, the percentage of simulations that achieved stable DWs, and the corresponding input voltage ($V_{FE}$) and nucleation energy ($E_{FE}$) associated with the applied magnitudes from solving the LKh equation. We first note that we performed 100 simulations with an applied field magnitude of 420 mT. However, a stable DW—similar to Fig. 8—was achieved in only 64% of the simulations. The ones that did not form a stable DW resulted in one of three possibilities: (i) no DW was nucleated, (2) the DW evaporated after a short period of time, or (3) the DW migrated across the PMA layer, but left the part of the PMA layer underneath the IMA in its initial magnetization state. For the 39% of simulations that did not achieve a stable DW, we increased the applied field magnitude in a binary search fashion until we found the smallest applied field magnitude necessary to achieve a stable DW, which we found to be 530 mT. This field magnitude represents a 26% increase from the upper bound (i.e., 420 mT) in [10]. We attribute this increase to our simulations utilizing thermal noise whereas the work in [10] simulates at 0 K.

**B. Benchmarking**

To benchmark the potential energy savings associated with our CoMET-based NVFF, we have gathered device data for both volatile and NV FFS. In Fig. 9, we present the energy versus delay of writing to a D flip-flop for various device technologies. Devices with blue markers represent BCBv3 [26] computed data for 15 nm devices. These ten data points represent high performance CMOS (CMOS HP), low voltage CMOS (CMOS LV), all-spin logic (ASL), charge-spin logic (CSL), spin torque oscillator logic (STOlogic), spin torque domain wall (STT/DW), nanomagnetic logic (NML), spin wave device (SWD), SpinFET, and spin majority gate (SMG). Devices with red markers represent either simulated or empirical data collected from previous works. These five data points represent 90 nm Spin-MTJ [27], 65 nm magnetic FF (MFF) [28], 180 nm resistive RAM (ReRAM) [29], spin-Hall effect NVFF (SHE-NVFF) [30], and 10 nm ferroelectric FET (FeFET) [7]. The green marker is our proposed CoMET NVFF (at 15 nm) from the 100 ps and 2.3 aJ DW nucleation time and energy found in this work.
Furthermore, our CoMET NVFF achieves reductions of an order of magnitude less than the selected spintronic devices, and is at least an order of magnitude more energy efficient than writing a CMOS HP and CMOS LV FF, respectively.

Fig. 9: Energy vs. delay of a D Flip-Flop for various devices.

Our CoMET NVFF achieves the lowest write energy among the selected devices, and is $360 \times$ and $42 \times$ more energy efficient than writing a CMOS HP and CMOS LV FF, respectively. Furthermore, our CoMET NVFF achieves reductions of $27 \times$ and $30 \times$ in terms of EDP compared to CMOS HP and CMOS LV. Compared to the other spintronic devices, CoMET is at least an order of magnitude more energy efficient than each technology with the exception of the SWD and SMG (both are $\approx 5 \text{aJ}$). In terms of EDP, our CoMET NVFF is at least an order of magnitude less than the selected spintronic devices.

VI. CONCLUSION

In this work, we presented a NVFF based on CoMET technology that is capable of achieving very low write energy. The low write energy is achieved by modifying the basic operation of CoMET so that data is written by nucleating a DW into the PMA layer. The ideal NVPs for this NVFF are ones that require frequent and/or large backup operations such as EB-NVPs. We have also discussed the design of a master-slave edge-triggered NVFF that can be formed from a single CoMET device. The master-slave NVFF would be useful for IB-NVPs and has reduced area and energy due to its single device design. To demonstrate the feasibility of our design, we performed 100 micromagnetic simulations with thermal noise at 300 K. Our results showed that a stable domain wall can be achieved under these conditions. For future work, we plan to quantify architectural-level energy efficiency.

ACKNOWLEDGMENT

This work was supported in part by ASCENT, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. This work was supported in part by C-SPIN, one of six centers of STARnet, a Semiconductor Research Corporation program, sponsored by MARCO and DARPA.

REFERENCES


