
An Energy-Efficient Design Paradigm for a Memory Cell Based on Novel Nanoelectromechanical Switches

Azam Seyedi, Vasileios Karakostas,
Stefan Cosemans, Adrian Cristal,
Mario Nemirovsky and Osman Unsal

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/105947>

Abstract

In this chapter, we explain NEMsCAM cell, a new content-addressable memory (CAM) cell, which is designed based on both CMOS technologies and nanoelectromechanical (NEM) switches. The memory part of NEMsCAM is designed with two complementary nonvolatile NEM switches and located on top of the CMOS-based comparison component. As a use case, we evaluate first-level instruction and data translation lookaside buffers (TLBs) with 16 nm CMOS technology at 2 GHz. The simulation results demonstrate that the NEMsCAM TLB reduces the energy consumption per search operation (by 27%), standby mode (by 53.9%), write operation (by 41.9%), and the area (by 40.5%) compared to a CMOS-only TLB with minimal performance overhead.

Keywords: nanoelectromechanical (NEM) switch, content-addressable memory (CAM) cell, translation lookaside buffer (TLB)

1. Introduction

Computing technology has witnessed an inimitable progress in the last decades, which is the result of CMOS technology scaling commensurate with Moore's law [1]. Transistor feature sizes have shrunk to half at each generation, and consequently, the number of transistors per chip has doubled every 2 years. However, CMOS scaling faces serious problems that occur due to the exponential increase of the leakage of current during technology scaling [2]. The subthreshold leakage is mainly affected by the subthreshold swing (S) of a device, which is defined as the amount of gate voltage reduction to reduce the subthreshold current by one decade ($S = dV_{gs}/d\log I_d$) [3]. For bulk CMOS, the subthreshold swing has

a substantially lower limit of 60 mV/decade, which leads to a large increase in the power density [4]. This limitation prevents manufacturers from fabricating smaller devices and forces them to look for alternative solutions targeting higher performance and efficiency. In order to maintain the scaling ability, a significant amount of research is ongoing to explore various non-CMOS technologies (emerging technologies) as a replacement for volatile and nonvolatile memories. This also motivates us, in this chapter, to exploit one of the most promising emerging technologies (nanoelectromechanical switches) to improve the processor performance.

Nanoelectromechanical (NEM) switches have been suggested as a promising candidate for replacing the CMOS technology [5]. NEM switches provide some unique characteristics that are not available in conventional MOS, such as near-zero leakage current and infinite sub-threshold slope (<0.1 mV/decade [6]). Such characteristics make them ideal for designing highly energy-efficient structures. However, NEMs have relatively long mechanical switching delay [5] compared to the intrinsic delay of CMOS devices, and to this date, they suffer from low endurance (10^{11} write cycles) [7]. Also, NEMs do not offer high turn-on current like CMOS transistors.

In spite of the large mechanical delay and limited number of reliable cycles, NEMs have been useful for a wide range of applications such as FPGAs (used as programmable routing switches) [8], adders [9], flip-flops [10], memories [4, 11], DACs [9], and ADCs [9], where the long switching time and limited number of hits are not important issues. Most of the mentioned circuits use the benefits of combining NEMs and CMOS technology in order to highlight the advantage points of each technology and alleviate the disadvantages to achieve low-power and high-performance operation for some critical components. Motivated by these observations, in this chapter, we describe a new content-addressable memory (CAM) cell design, NEMsCAM [12], based on both NEMs and CMOS technologies, to employ in processor structures where writes are relatively infrequent, for example, the translation lookaside buffers (TLB).

Content-addressable memories (CAM) have been widely adapted for applications that depend on fully associative and high-speed search operations, such as translation lookaside buffers (TLBs), network routers, and data compression [13]. Since the search operation requires fully parallel and fast comparisons, CAMs introduce high energy consumption and area constraints. Previous works have explored the design of CAMs with emerging memory technologies to mitigate these issues [14–18]. However, those CAMs suffer mainly from increased search latency due to the employed technology that prevents them from building performance critical structures such as TLBs.

The memory component of NEMsCAM cell is designed with two complementary nonvolatile NEM switches while the comparison circuits are designed with CMOS transistors to allow fast search operation. The novel structure of NEMsCAM along with unique characteristics of NEMs considerably reduces the layout area as well as the energy consumption. Also, in this design, both Out and OutB are available simultaneously, which is essential to design a CAM cell.

As a use case, we leverage the NEMsCAM cell to build fully associative TLBs. The TLB has been pointed out as a critical component of energy and performance in modern processors [19]. Translation lookaside buffer (TLB) is a cache that is employed to accelerate virtual-to-physical address translation [20]. The processor searches the TLB on every memory operation using the virtual page number. Hence, the TLB is a crucial component for the performance and power consumption of the computers [20]. We evaluate first-level data and instruction TLBs with 16 nm technology at 2 GHz frequency. Our analysis demonstrates that the proposed TLB reduces the energy per search and write operation and standby mode by 27%, 41.9% and 53.7%, respectively, and the area by 40.5% compared to a CMOS-only TLB. Also, both designs execute the search operation in one clock cycle. Furthermore, it is shown that the NEMs' increased write latency introduces minimal performance overhead (0.27% on average). The main contributions of this chapter are as follows: (1) Description of NEMsCAM cell design based on complementary nonvolatile NEM switches and CMOS transistors. (2) Explain the design of highly efficient first-level TLBs for data and instruction accesses based on the NEMsCAM cell. (3) Evaluate the proposed designs at both circuit and system level and compare to CMOS-only TLBs.

Section 2 provides background information whereas Section 3 and Section 4 describe the design of the NEMsCAM cell and the NEMsCAM TLB, respectively. In Section 5, we present our evaluation methodology and the obtained results. Finally, in Section Section 6, we summarize the chapter.

2. Background

This section provides background information related to this work. First, we briefly review recent available emerging technologies. Then, we describe NEM switches in detail and the prior art in their use of memory. Finally, we describe CMOS-only CAM cells and fully associative TLB structures.

2.1. Emerging technologies

In order to continue the trend of Moore's Law, many emerging technologies have been employed such as phase-change memory (PCRAM) [21], magnetoresistive RAM (MRAM) [22], spin-torque transfer magnetoresistive RAM (STT-RAM) [23], ferroelectric RAM (FeRAM) [24], memristor [14], and nanomechanical memory (NEM) [6]. Typical performance parameters of mentioned memory technologies are presented in **Table 1** [14, 25].

Ferroelectric RAM (FeRAM or FRAM) is one promising memory, which employs a ferroelectric gate cell instead of a poly-silicon cell and has been considered as a replacement for flash memory [24]. In spite of some disadvantages like lower density, higher cost, and poor scalability, it has some advantages over flash like faster programming time, lower power usage, and higher endurance.

Magnetoresistive RAM (MRAM), which has been designed with magnetic storage elements, is another emerging technology [22]. The elements are made of two ferromagnetic

	Traditional technologies				Emerging technologies				
	DRAM	SRAM	Improved flash		FeRAM	MRAM	PCRAM	Memristor	NEMS
			NOR	NAND					
Cell elements	1T1C	6T	1T		1T1C	1T1R	1T1R	1M	1T1N
Half pitch (F) (nm)	50	65	90	90	180	130	65	3–10	10
Smallest cell area (F ²)	6	140	10	5	22	45	16	4	36
Read time (ns)	<1	<0.3	<10	<50	<45	<20	<60	<50	0
Write/erase time (ns)	<0.5	<0.3	105	106	10	20	60	<250	1ns(140ps-5ns)
Retention time (years)	Seconds	N/A	>10	>10	>10	>10	>10	>10	>10
Write op. voltage (V)	2.5	1	12	15	0.9-3.3	1.5	3	<3	<1
Read op. voltage (V)	1.8	1	2	2	0.9-3.3	1.5	3	<3	<1
Write endurance	10 ¹⁶	10 ¹⁶	10 ⁵	10 ⁵	10 ¹⁴	10 ¹⁶	10 ⁹	10 ¹⁵	10 ¹¹
Write energy (fj/bit)	5	0.7	10	10	30	1.5 × 10 ⁵	6 × 10 ³	<50	<0.7
Density (Gbit/cm ²)	6.67	0.17	1.23	2.47	0.14	0.13	1.48	250	48
Voltage scaling	Fairly scalable					No	Poor	Promising	Promising
Highly scalable	Major technological barriers				Poor		Promising	Promising	Promising

Table 1. Traditional and emerging memory technologies characteristics [14, 25].

plates, each of them holds a magnetic field and an insulating layer separates them. One of the plates has a permanent magnetic field and the other has a variable one to be able to store data. MRAM is as fast as SRAM and as dense as DRAM. Also, it has a nonvolatility characteristic similar to flash and high endurance. However, it suffers from large cell size, high write current, and poor scalability, which greatly forbids being widely commercialized.

Another nonvolatile memory is phase change memory (PCRAM) [21]. Similar to optical storage devices, it stores data into a chalcogenide glass. The state of the glass is changed to crystalline or amorphous whenever an electric current passes through a heating element and

generates heat or quenches the glass. The main limitations of PCRAM are its high programming current and relatively long write/read time.

Memristor is considered to be one of the best candidates for future memory technologies. A memristor is a two-terminal nonvolatile memory and has been designed based on resistance switching effects [14]. Memristor has low write energy and high density due to multilayer crossbar architecture. Since the memristor crossbar-based architecture is highly scalable, it is predicted to be the selected candidate to use for future ultrahigh density memories. As no tunnel oxide is used in this device, memristor has higher endurance than flash memory. In spite of high density and endurance, read time is considerably high. Memristors may easily replace flash memories; however, they are not an appropriate option to employ in extremely fast system components.

Another promising candidate as a replacement for CMOS devices is the nanoelectromechanical switch (NEM) [6]. On/off state of NEM devices is determined by both electrical and mechanical forces between gate (a movable beam) and source terminals. Unique characteristics of NEM relays which are not available in conventional CMOS, such as zero leakage current and near infinite subthreshold slope, make them ideal for designing highly energy-efficient applications. In spite of low leakage current, NEM relays do not offer high turn-on current like CMOS transistors; moreover, they suffer from low endurance.

2.2. Nanoelectromechanical switches

Figure 1(a) and (b) shows the simplest NEM switch, 3T NEM that consists of three terminals: a cantilever beam (which is connected to the source terminal), a gate, and a drain. The voltage difference between the gate and the movable terminal (V_{GS}) controls the position of the beam and state of operation. When V_{GS} goes higher than a certain threshold value, called the pull-in voltage (V_{pi}), the electrostatic force exerted by the gate exceeds the elastic force of the beam and pulls the beam toward the drain until the beam collapses to the drain and forms a conductive channel from the beam (source) to the drain, thus closing the switch (on-state;

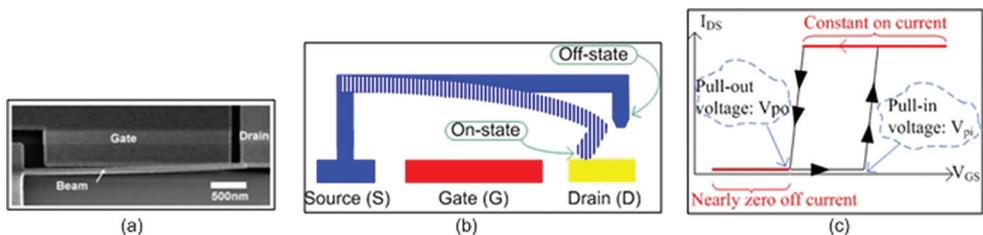


Figure 1. (a) SEM image of three-terminal NEM relay [28]. (b) Schematic of a 3T NEM relay consisting of source, gate, and drain; a cantilever beam connected to the source. The beam collapses to the drain when $V_{GS} \geq V_{pi}$ and is released when $V_{GS} < V_{po}$. (c) Typical I_{DS} - V_{GS} characteristics of 3T NEM array.

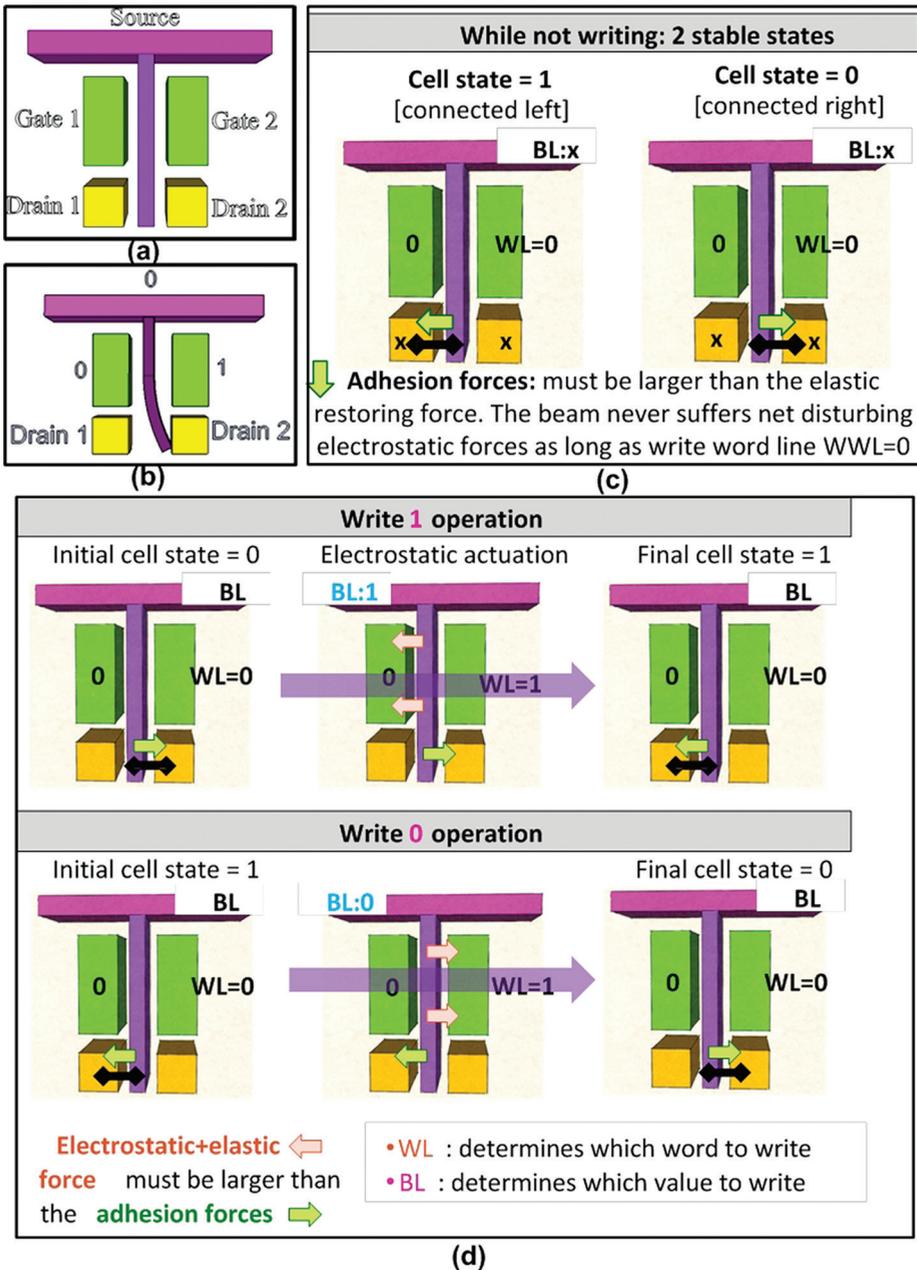


Figure 2. (a) A 5T NEM switch schematic. (b) The beam collapses to Drain2: VGS1 (VGate1-VSource) is 0 and VGS2 is 1. (c) The nonvolatile NEM considered here has two stable states. (d) Biasing scheme applied for writing the nonvolatile NEM switching. The BL assigns the value that will be written and chooses the cells for writing.

Figure 1c). In order to release the beam from the drain, VGS decreases to a voltage smaller than V_{pi} , called pull-out voltage (V_{po}), where the electrostatic force is not higher than the elastic force of the beam; at this moment the beam is disconnected from the drain (off-state; **Figure 1c).** Due to such sharp on/off transition, NEMS have zero off-state leakage as there is no path for current to flow. Moreover, because of the surface adhesion force between the two contacting regions, V_{po} is usually smaller than V_{pi} and the IV characteristics of NEMS exhibit such a hysteresis characteristic which enables NEM relays to be used as memory elements (**Figure 1c).**

There are several implementations of NEM switches [26, 27]. In this work, we consider 5-Terminal (5T) NEM switch which is illustrated in **Figure 2(a) and (b).** A suspended beam is anchored at the source. Gate terminal 1 and gate terminal 2 (Gate1 and Gate2) are located close to the beam. The beam is connected to Drain1 or Drain2 (two output nodes). The beam moves toward Gate1/Gate2 because of the electrostatic attraction and a voltage difference between Gate1 and Gate2 (**Figure 2b**), and then connects to Drain1/Drain2, creating a conductive path between this drain and the source. An advantage of employing two gates is that the electrostatic force can be utilized as both pull-in and pull-out voltages. Therefore, one does not have to rely on only the elastic restoring force of the beam. Hence, the scalability of the device and the operational margins are improved considerably. However, the write operations in NEM switches take multiple clock cycles [11], because the mechanical movement of the beam is fairly slow which is related to the device technology.

2.3. Nonvolatile nanoelectromechanical switches

In this work, we choose the NEM switches which exhibit nonvolatile characteristic: once they are connected to a drain, they remain in this location until the beam is pulled out by electrostatic forces from the opposite gate. We select the NEM switch described in [29]. **Figure 2(c)** demonstrates the two stable states of this NEM switch. As long as both gate1 and gate2 are at the same potential (wordline, $WL = 0$), the beam never suffers a net disturbing electrostatic force. **Figure 2(d)** shows the write operation of this switch.

2.4. Memory arrays based on nanoelectromechanical switches

Former studies have evaluated employing NEMs for memory usages [11, 30]. Some of them address the use of NEM switches to replace normal memory arrays, such as SRAM. Chong et al. [11] replace the two pull-down transistors in a 6T SRAM cell with NEM switches to reduce leakage and area. Some of these studies also discuss nonvolatile memory arrays [30]. The memory array structure disclosed in Ref. [31] is of particular interest to our proposal as we explain in Section 3.

2.5. Configuration elements based on nanoelectromechanical switches

Recently, NEMs have been used for configuration tasks. Dong et al. [8] employed 3T NEMs as configuration memory components in FPGAs, replacing a routing switch by one NEMs, or

an LUT cell by two NEMs. This design could be exploited for designing CAM cells; however, their proposed cell has many deficiencies. It relies only on the elastic restoring force for pull-out and suffers half-select conditions, it outputs only Out, not the complementary OutB, and it is volatile. The structure of the memory part of NEMsCAM which we describe in Section 3 depicts these mentioned shortcomings.

2.6. Content-addressable memory

A content-addressable memory (CAM) concurrently compares the search data with all of its stored data and returns the address of the matching location in a single clock cycle [27]. A typical CMOS-only CAM cell incorporates a SRAM cell to store the data bit and additional XOR circuits to compare the stored bit with the search data. CAMs propose a popular solution for a wide variety of applications that require high search speeds such as data compressions, network routers, and lookup tables [28]. However, the search operation in a CAM requires fully parallel comparison circuits to meet timing requirements. This results in high energy consumption and poses constraints on the number of entries affecting directly the effectiveness of the CAM.

2.7. Translation lookaside buffer

Virtual memory simplifies programming by abstracting and managing the available physical memory in pages. To accelerate virtual memory, processors employ the translation lookaside buffer (TLB) that holds recently used virtual-to-physical translations [20]. The processor searches the TLB on every memory operation using the virtual page number. In the case of a hit, the TLB returns the physical page number so that the memory operation can further proceed with accessing the memory hierarchy. However, in case of a miss, the memory operation will not complete until the address translation is retrieved from the memory (page walk) which might take up to hundreds of cycles. The TLB is hence a crucial component for the performance of the processor [20].

3. Design of NEMsCAM cell

In this section, we present the circuit details of our proposed NEMsCAM cell. We use the memory structure proposed in [31] to implement the storage part of NEMsCAM. That memory structure provides full-select behavior which is necessary to build a CAM; it also employs electrostatic pull-in and pull-out and does not need a cell selector component in the write path. The nonvolatile memory is designed based on the NEM switch proposed in [29], which can eliminate net-disturbing electrostatic force. **Figure 3(a)** depicts the configuration of the NEMsCAM cell. Out and OutB, the outputs, are connected to the transistors of the comparison part. We select CMOS for the comparison part to beware the long delay of the NEMs that happens because of the beams' mechanical movement, and that would slow down the search operation. **Figure 4(a)** represents the schematic of our NEM memory cell when it is programming to state "1." **Figure 4(b)** shows its switch model and **Figure 4(c)** shows a simplified NEM Verilog-A model between BL (source), Out (drain), and WL nodes, which we employ

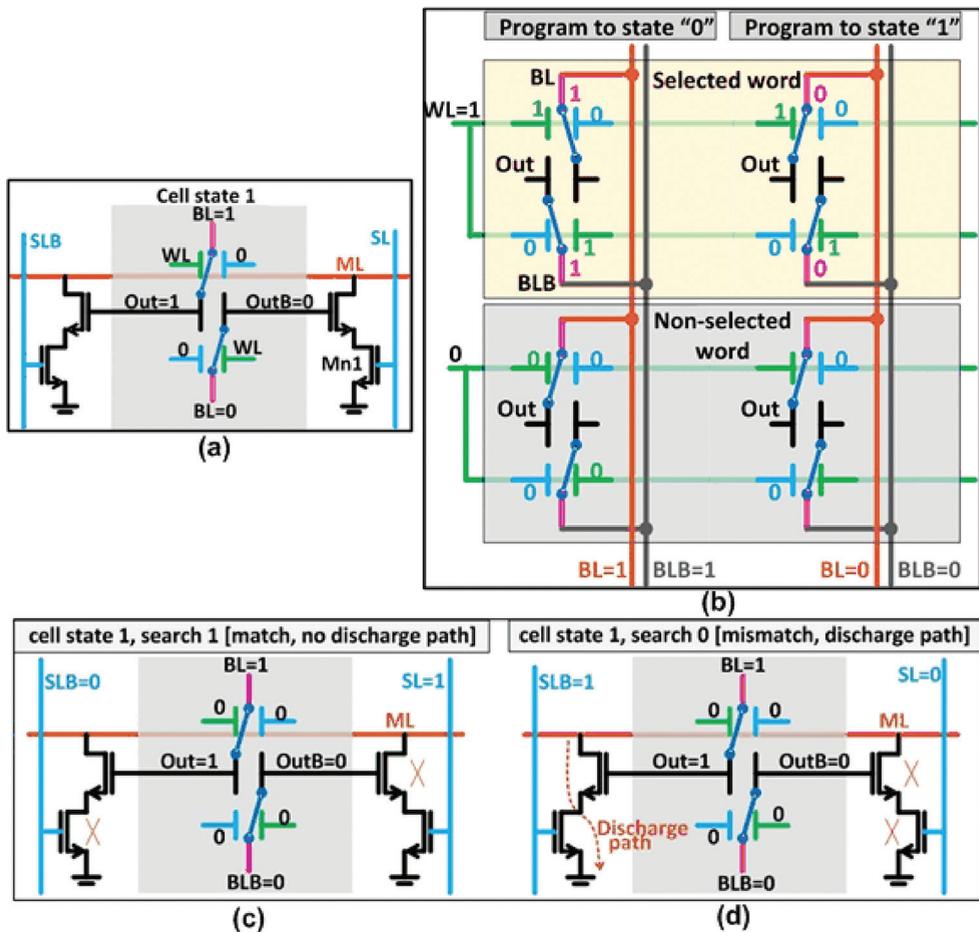


Figure 3. (a) Schematic of the proposed NEMsCAM cell. (b) NEMsCAM storage array organization and writing scheme. (c) When the NEMsCAM cell content matches the value being searched, there is no discharge path through the cell. (d) In the case of a mismatch, the cell tries to discharge the match line ML.

in our circuit analysis [11]. Other switches of NEM memory cell comply with similar switch modeling.

3.1. Circuit operations

3.1.1. Write biasing scheme

Figure 3(b) describes how the storage circuit of the NEMsCAM is written. When the WL (wordline) is activated, all beams on the row are sensitized. For the columns whose cells are to be programmed to 1, the bitlines are set to zero ($BL = BLB = 0$), and for the bitlines whose cells are to be programmed to 0, $BL = BLB = 1$ is applied.

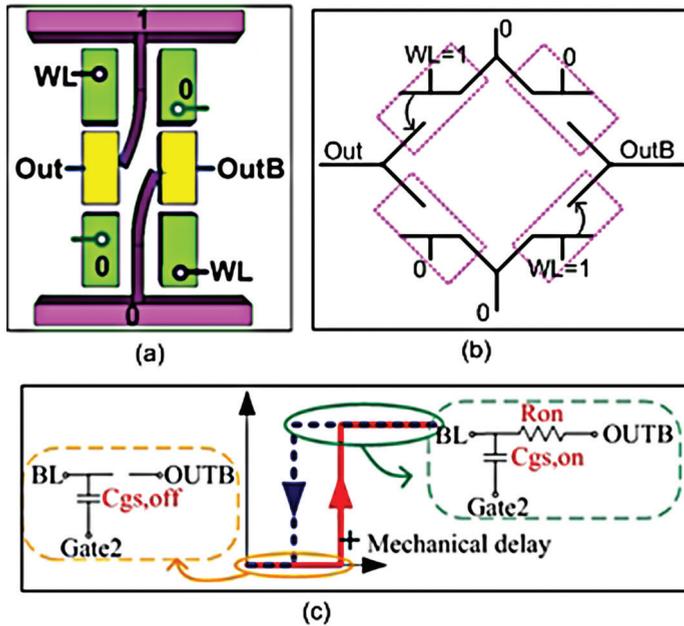


Figure 4. (a) Schematic of our NEM memory cell. (b) Simplified switch model of NEM memory cell. (c) A simple NEM Verilog-A model between BL (source), Out (drain), and WL nodes [11].

No cell suffers half-disturb situations, and since BLB and BL are always at the same potential during switching, there is no risk of short-circuit current running through the switches. This is critical because high currents through contact between the drain and the beam can be a source of failure. During typical operation, BLB is put at 0 and BL at 1. Cells whose beams are in state 0 hence have OutB = 1 and Out = 0. Keep in mind that there is no separate read operation in this memory design and there is no mechanical switch latency in the read path.

3.1.2. Search operation

Figure 3(c) and **(d)** shows a cell design that matches/mismatches the search data. If a cell(s) that is connected to a wordline entails a mismatch, the cell attempts to discharge the whole ML which is associated with that wordline, indicating an overall mismatch.

3.2. Cell architecture

Figure 5 presents the three-dimensional perspective of two neighbor NEMsCAM cells placed in the same column index of the array. As NEM switches have the possibility to be fully integrated with CMOS devices [32], they are located on top of the CMOS layer in this work and considerably decrease the layout space. The searchline (SL) wires are located parallel to the BL wires, whereas the matchlines (ML) and wordlines (WL) are located orthogonally to the BLs.

Using vertical NEM switches [27], the necessity of the long beam has a negligible effect on the layout space, since it is out of the plane. Two Gate1s are aligned and connected to their related WL, while the two Gate2s are connected to zero.

The drains are coupled from the opposite directions and build a cross configuration. The WL and BL wires can be combined with the real device terminals, resulting in a compressed layout. Eventually, the Vias connect OutB and Out to the CMOS layer which is placed under the NEMs layer. Because of this formation, our proposed NEMsCAM cell decreases the wire length, which considerably reduces the power consumption along with the near-zero leakage behavior of NEM switches.

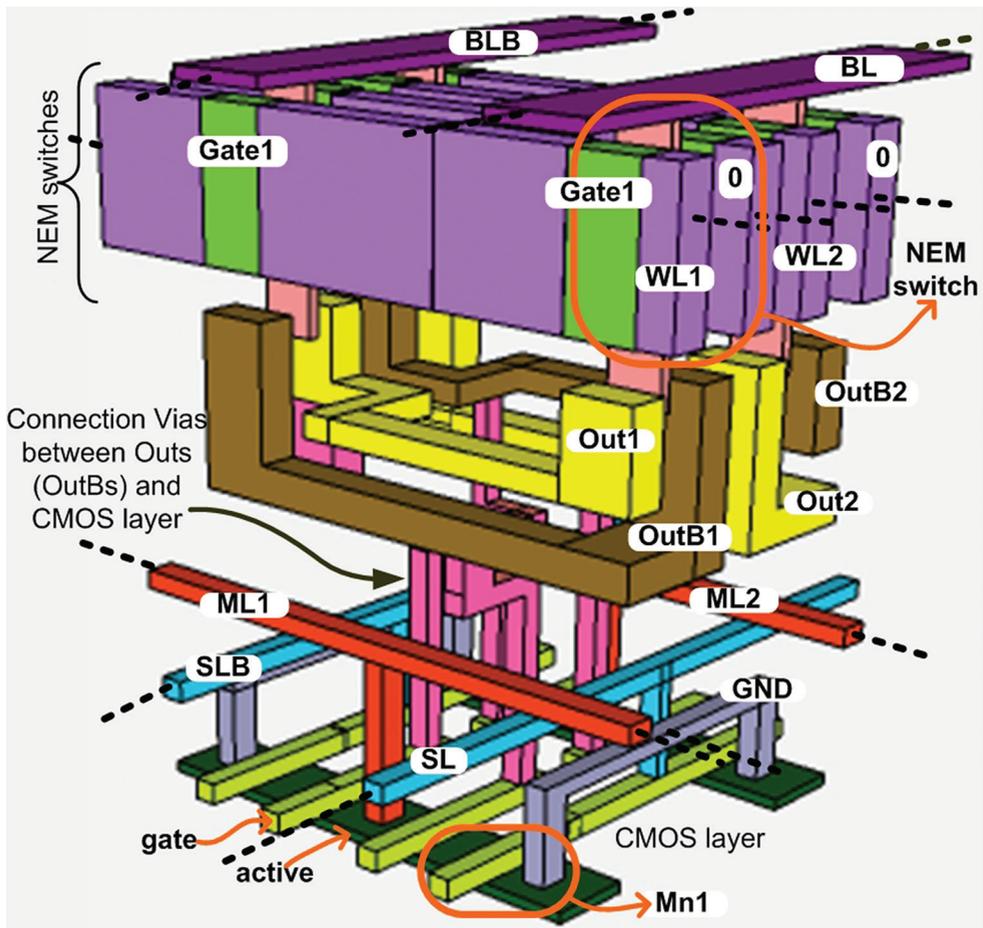


Figure 5. Three-dimensional view of two adjacent NEMsCAM cells in a CAM array.

4. A use case for NEMsCAM: TLB

As mentioned before, we leverage the design of the proposed NEMsCAM cell to build a fully associative translation lookaside buffer (TLB), called NEMsCAM TLB. In this section, we first elaborate on the motivation behind it and then we describe the design details and the circuit operations.

4.1. Motivation

Because of the importance of the TLB in the system's efficiency, processor designers have utilized a two-level TLB structure [33]. The first-level TLB is fully associative, small and provides a very fast search operation, while the second-level TLB is large and holds as many translations as possible. In order to achieve further system's performance, processors prepare separate TLBs for instructions and data [33].

The TLB hierarchy has been accounted for a substantial percentage of the power consumed in the chip [34, 35]. Intel recently informed that 13% of the total core energy comes from the TLBs designed for memory-intensive workloads [19]. Based on our evaluation base (Section 5), we discover that the TLB power consumption is overwhelmingly dominated by the first-level TLBs in terms of accesses across the TLB hierarchy (**Figure 6**). Moreover, by breaking down the power in the first-level TLBs, we detect that the CAM component contributes by 94%. In order to decrease this source of power consumption without diminishing the performance, we leverage our proposed NEMsCAM cell to design energy-efficient first-level TLBs.

4.2. Design

We design the NEMsCAM TLB with our proposed CAM cell and with typical SRAM memory circuits (**Figure 7**). The CAM part (**Figure 7a**) consists of the NEMsCAM cells and the neces-

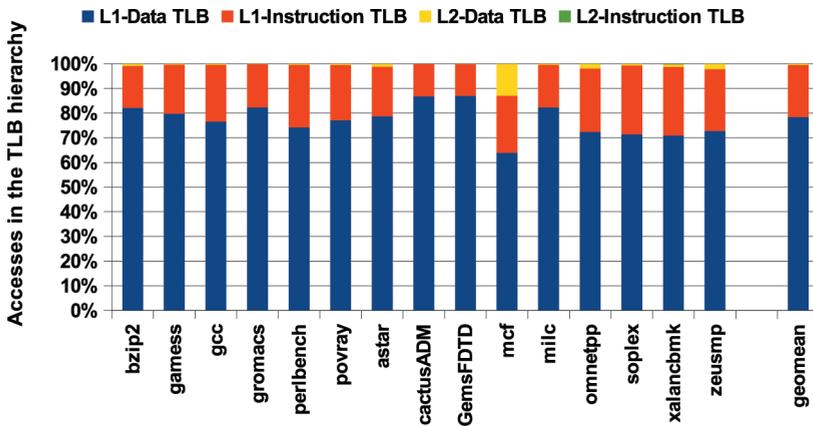


Figure 6. Breakdown of accesses in the TLB hierarchy.

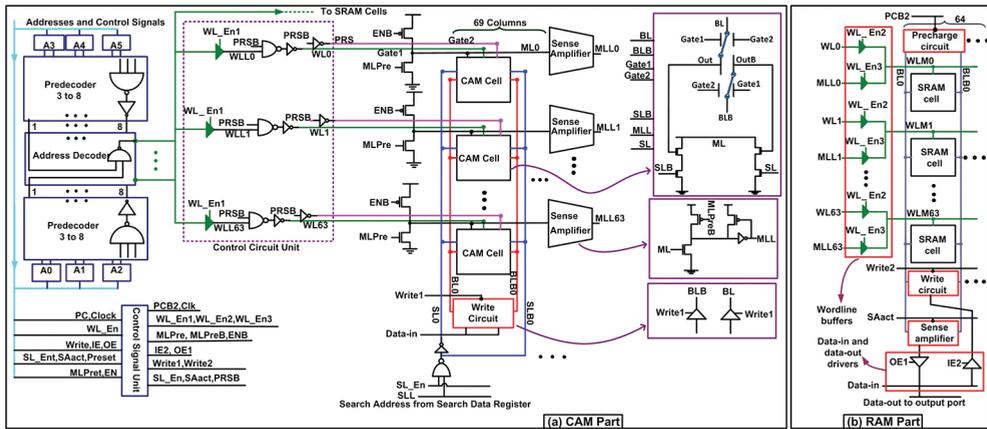


Figure 7. The circuit detail of (a) the proposed NEM-CMOS CAM and (b) a typical SRAM architecture in the proposed TLB structure.

sary peripheral circuitry optimized for both search and write operations. Similarly, the SRAM cells (**Figure 7b**) and the associated circuits are designed with CMOS technology. The control signal unit consists of the necessary inverter chains that generate the signals to control the TLB circuits so that the search and the write operations are performed correctly.

The address decoder, the write circuits, and the data-in drivers are used only for the write operation; however, the rest of the circuits are designed to be used during the search operation as well. BL and BLB are driven with predefined signals according to the operations. The control circuit unit is added to generate the necessary Gate1 and Gate2 signals during the search and write operations.

4.2.1. Search operation

Within the search operation, WLen3 becomes high and the WLM lines are connected to the ML lines. At the beginning of search operation, all ML lines are set provisionally in the precharged state as in a CMOS-only TLB. The search cycle begins when MLpre (the precharge signal) becomes high driving the ML to zero. At the same time, the SLs (search lines) are charged to their related data value; with this method, there is no need for a separate SL precharge phase. After this (completion of the precharge phase), the ENB signal becomes low and supplies the ML with the current source. During the evaluation phase, the stored bits of the CAM cells are compared against the data provided on the corresponding SLs.

In case of a match (TLB hit), the current source enabled by ENB pulls ML up and the ML voltage changes to high state. The state of each ML row is sensed and improved by an ML sense amplifier. Our used sense amplifier can be seen in this figure: an nMOS transistor, and also a half-latch circuit which stores the output data (**Figure 7a**). We choose the current-race scheme among various matchline-sensing techniques due to its simplicity and the average-low ML energy con-

sumption [13]. Alternatively, in case of a mismatch (TLB miss) the cell(s) that cause a mismatch counteract the current source and keep ML close to ground level. In the match case, matchline trips the half-latch circuit when it is charged to a voltage slightly higher than threshold voltage of Msense; whereas, it leaves the latch in its initial state in the mismatch case when remains at a much lower voltage [5, 27]. Finally, the ML sense amplifiers feed the wordline buffers mapping the match location to its corresponding encoded address as stored in the SRAM cells (Figure 7a). Figure 9 summarizes the signal behavior of the matching case for a cell of the NEMsCAM TLB.

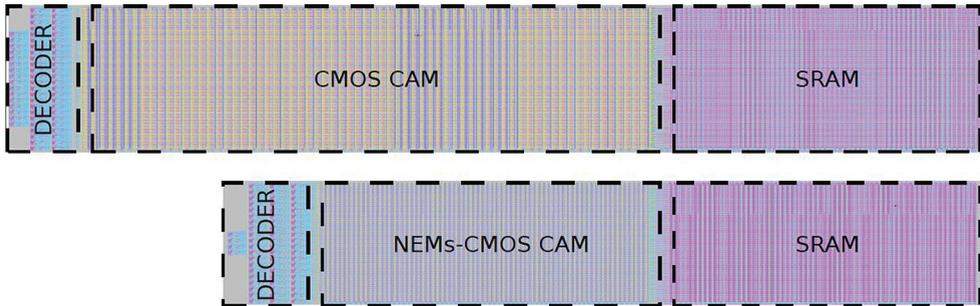


Figure 8. The layout of the DTLB implemented with CMOS-only CAM cells (top) and NEM-based CAM cells (bottom).

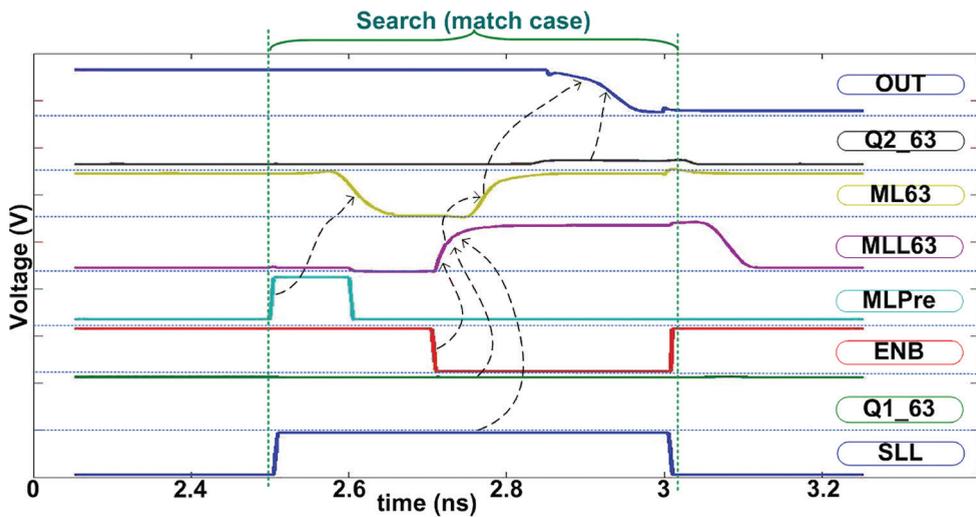


Figure 9. Simulation waveform of matching state for the first cell at the last row of the NEMsCAM DTLB.

4.2.2. Write operation

During the write operation, the WLen1 and WLen2 are high, the WL which is generated in address decoder is routed to the CAM and SRAM parts, and the data is written into the corresponding cells.

5. Experimental evaluation

In this section, we first describe our methodology to evaluate the NEMsCAM TLB, and then we present the results.

5.1. Methodology

We design NEMsCAM TLBs for DTLB (data) and ITLB (instruction) accesses based on [13] applying the TLB formation of a modern AMD server-oriented processor [33] (**Table 2**). For both CMOS-only and NEMsCAM TLB, we write the transistor-level netlists with all the tantamount resistance and capacitances of wires and essential circuitries. We simulate and optimize both TLB designs with Cadence Spectre exploiting 16 nm Predictive Technology Model [36] at $T = 25^\circ\text{C}$ and 2 GHz processor frequency. As mentioned before, for the NEM switches, we apply a naive Verilog-A pattern (**Figure 4**) with the following parameters: $V_{po} = 0.2\text{ V}$, $V_{pi} = 0.8\text{ V}$, $C_{gs-off} = 15\text{ aF}$, $C_{gs-on} = 20\text{ aF}$, $t_{mech} = 3\text{ ns}$ [11]. We optimize TLB circuits to minimize the energy consumption. We investigate that the search and write operations are performed correctly complying the timing necessities. We also assess the energy consumption per write and search operation and standby mode. Moreover, we plan the layouts [37], and span the wire lengths and optimize the wire capacitances in the netlists. Eventually, in order to evaluate the effect of the NEMsCAM TLBs at system's proficiency, we consume the Sniper simulator [38] with the configuration of **Table 2**, and run the TLB-intensive workloads from Spec2006 with the reference input set and execute for one billion instructions.

5.2. Results

5.2.1. Energy and area

Table 3 demonstrates the simulation outcomes for both CMOS-only and NEMsCAM TLBs. We perceive that the area decreases by 40.5% for the DTLB (**Figure 8**). The unique design of the NEMsCAM cell is the reason for this improvement. Furthermore, we discover that the energy consumption per write and search operation and standby mode decreases by 41.9%, 27%, and 53.7%, respectively, for the DTLB. This occurs due to the lower dimensions of the circuit leading to lower parasitic wire resistances and capacitances on the matchlines and the searchlines which in turn need fewer driving buffers. Also, the energy consumption further reduces due to the near-zero leakage current that NEMs prepare. Same results are achieved for the ITLB as well.

Per-core TLB organization		
Level 1	Data (DTLB)	64 entries, fully assoc.
	Instruction (ITLB)	48 entries, fully assoc.
Level 2	Data (L2-DTLB)	1024 entries, 4-way assoc.
	Instruction (L2-ITLB)	512 entries, 4-way assoc.

Table 2. TLB organization of a modern processor [33].

Structure	Metric	CMOS	NEMs	Benefit (%)
DTLB 64 entries	Search operation (pJ)	4.529	3.308	27.0
	Write operation (pJ)	0.148	0.086	41.9
	Standby mode (pJ)	0.141	0.065	53.7
	Area (normalized)	1.000	0.595	40.5
ITLB 48 entries	Search operation (pJ)	3.658	2.805	23.3
	Write operation (pJ)	0.187	0.107	42.8
	Standby mode (pJ)	0.106	0.046	55.9
	Area (normalized)	1.000	0.661	33.9

Table 3. Energy consumption for search, write operations, and standby mode and normalized area footprint.

5.2.2. Latency

Figure 9 demonstrates the simulation waveform of the matching case for one cell of the NEMsCAM DTLB (data TLB) during the search operation. The waveform considers that the design supplies the purpose time requirement of one clock cycle per search operation. On the other side, the write operation takes six cycles in the NEMsCAM TLB (based on Ref. [11]), whereas it takes two cycles in the CMOS-only TLB. This slowdown is because of the long mechanical delay of the NEM switches. However, this latency barely affects the processor performance as shown next.

5.2.3. System

Figure 10 displays the energy reduction in the first-level TLBs due to the NEMsCAM utilization for several workloads. We observe that the search operation overcomes in the energy breakdown for both ITLB and DTLB and that the NEMsCAM TLBs reduce the energy spent by 28.7% on average. Taking into account that 13% of the total core energy comes from the TLBs [19], the NEMsCAM cell can considerably assist in reducing the total chip’s energy performance. Figure 11 demonstrates the evaluated execution overhead due to the utilization of the NEMsCAM TLBs. This overhead occurs because of the increased latency of the write operation in NEMsCAM. Albeit, the write operation: (a) occurs only after TLB misses which take place scarcely compared to TLB hits, and (b) adds latency to an already slow operation,

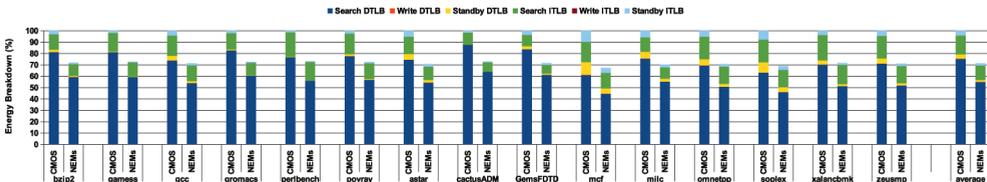


Figure 10. Dynamic energy consumption of the CMOS-only and the NEMsCAM DTLB and ITLB.

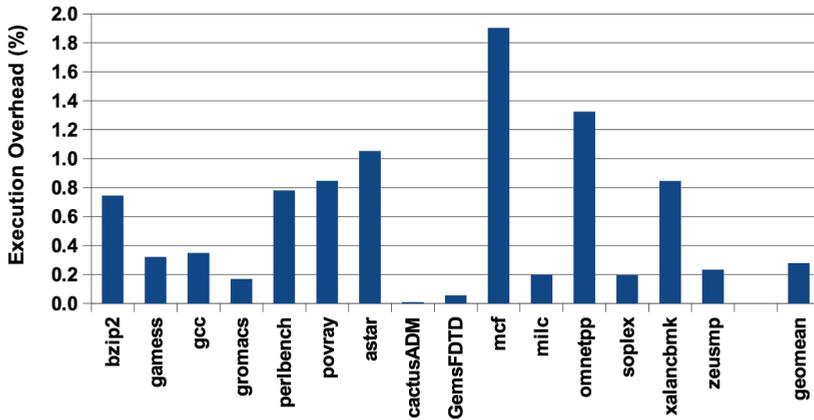


Figure 11. Execution time overhead due to NEMsCAM TLBs.

i.e., L2-TLB access (~7 cycles [39]), including potentially the penalty of L2-TLB miss (~100 s cycles [20]). Therefore, the NEMsCAMs TLB have an insignificant effect on the execution time for most applications (0.27% on average) while decreasing outstandingly the energy spent in the TLB hierarchy.

6. Summary

In this chapter, we describe the NEMsCAM cell design that combines both NEMs and CMOS to design low power and highly efficient processor structures such as TLBs. Our analysis shows that the NEMsCAM TLB exhibits significant benefits over the CMOS-only TLB in terms of energy consumption and area. However, the limited write endurance of current NEMs may delay their adoption until the technology improves.

Author details

Azam Seyedi^{1,2*}, Vasileios Karakostas^{1,2}, Stefan Cosemans³, Adrian Cristal^{1,2}, Mario Nemirovsky^{1,4} and Osman Unsal¹

*Address all correspondence to: azam.seyedi@bsc.es

1 Barcelona Supercomputing Center, Barcelona, Spain

2 Polytechnic University of Catalonia, Barcelona, Spain

3 IMEC International, INSITE, Leuven, Belgium

4 Catalan Institution for Research and Advanced Studies (CREA), Barcelona, Spain

References

- [1] Moore G. E., Cramming More Components onto Integrated Circuits. *Electronics*. 1965; 38:114–7.
- [2] International Technology Roadmap for Semiconductors, “Emerging Research Devices,2013 Edition.” available on http://www.semiconductors.org/clientuploads/Research_Technology/ITRS/2013/2013PIDS.pdf
- [3] Taur Y, Ning TH. *Fundamentals of Modern VLSI Devices*. 2nd ed. New York, NY, USA: Cambridge University Press; 2009.
- [4] Dadgour HF, Banerjee K. Design and Analysis of Hybrid NEMS-CMOS Circuits for Ultra Low-Power Applications. In: 44th ACM/IEEE Design Automation Conference (DAC '07); 4–8 June 2007. p. 306–11.
- [5] Akarvardar K, Elata D, Parsa R, Wan GC, Yoo K, Provine J, Peumans P, Howe R, Wong HSP. Design Considerations for Complementary Nanoelectromechanical Logic Gates. In: *Electron Devices Meeting, 2007 IEDM 2007 IEEE International*. 2007. p. 299–302.
- [6] Nathanael R, Pott V, Kam H, Jeon J, Liu T-JK. 4-Terminal Relay Technology for Complementary Logic. In: *IEEE International Electron Devices Meeting (IEDM)*. 2009. p. 223–6.
- [7] Gaddi R, Schepens C, Smith C, Zambelli C, Chimenton A, Olivo P. Reliability and Performance Characterization of a Mems-Based Non-Volatile Switch. In: *Reliability Physics Symposium (IRPS), 2011 IEEE International*. 2011. p. 2G.2.1–2G.2.6.
- [8] Dong C, Chen C, Mitra S, Chen D. Architecture and Performance Evaluation of 3D CMOS-NEM FPGA. In: *Proceedings of the System Level Interconnect Prediction Workshop*. Piscataway, NJ, USA: IEEE Press; 2011. p. 2:1–2:8. (SLIP '11).
- [9] Chen F, Kam H, Markovic D, Liu TJK, Stojanovic V, Alon E. Integrated Circuit Design with NEM Relays. In: *Computer-Aided Design, 2008 ICCAD 2008 IEEE/ACM International Conference on*. 2008. p. 750–7.
- [10] Han J-W, Ahn J-H, Kim M-W, Yoon J-B, Choi Y-K. Monolithic Integration of NEMS-CMOS with a Fin Flip-flop Actuated Channel Transistor (FinFACT). In: *Electron Devices Meeting (IEDM), 2009 IEEE International*. 2009. p. 1–4.
- [11] Chong S, Akarvardar K, Parsa R, Yoon JB, Howe RT, Mitra S, et al. Nanoelectromechanical (NEM) relays integrated with CMOS SRAM for improved stability and low leakage. In: *Computer-Aided Design - Digest of Technical Papers, 2009 ICCAD 2009 IEEE/ACM International Conference on*. 2009. p. 478–84.
- [12] Seyedi A, Karakostas V, Cosemans S, Cristal A, Nemirovsky M, Unsal O. NEMsCAM: A Novel CAM Cell Based on Nano-Electro-Mechanical Switch and CMOS for Energy Efficient TLBs. In: *2015 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*. 2015. p. 51–6.

- [13] Pagiamtzis K, Sheikholeslami A. Content-Addressable Memory (CAM) Circuits and Architectures: A Tutorial and Survey. *IEEE J Solid-State Circuits*. 2006;41(3):712–27.
- [14] Eshraghian K, Cho KR, Kavehei O, Kang SK, Abbott D, Kang SMS. Memristor MOS Content Addressable Memory (MCAM): Hybrid Architecture for Future High Performance Search Engines. *IEEE Trans Very Large Scale Integr Syst*. 2011;19(8):1407–17.
- [15] Matsunaga S, Hanyu T. Quaternary 1T-2MTJ Cell Circuit for a High-Density and a High-Throughput Nonvolatile Bit-Serial CAM. In: 2012 42nd IEEE International Symposium on Multiple-Valued Logic (ISMVL). 2012. p. 98–103.
- [16] Nebashi R, Sakimura N, Tsuji Y, Fukami S, Honjo H, Saito S, et al. A Content Addressable Memory Using Magnetic Domain Wall Motion Cells. In: 2011 Symposium on VLSI Circuits (VLSIC). 2011. p. 300–1.
- [17] Rajendran B, Cheek RW, Lastras LA, Franceschini MM, Breitwisch MJ, Schrott AG, et al. Demonstration of CAM and TCAM Using Phase Change Devices. In: Memory Workshop (IMW), 2011 3rd IEEE International. 2011. p. 1–4.
- [18] Wang W. Magnetic Content Addressable Memory Design for Wide Array Structure. *Magn IEEE Trans*. 2011 Oct;47(10):3864–7.
- [19] A. S. Race to Exascale: Opportunities and Challenges. In: MICRO Keynote. 2011.
- [20] Jacob B, Mudge T. Virtual Memory: Issues of Implementation. *Computer (Long Beach Calif) [Internet]*. Los Alamitos, CA, USA: IEEE Computer Society Press; 1998;31(6):33–43. Available from: <http://dx.doi.org/10.1109/2.683005>
- [21] Lee B, Ipek E, Mutlu O, Burger D. Architecting Phase Change Memory as a Scalable DRAM Alternative. In: International Symposium on Computer Architecture (ISCA) [Internet]. 2009. Available from: <http://research.microsoft.com/apps/pubs/default.aspx?id=79150>
- [22] Zhu JG. Magnetoresistive Random Access Memory: The Path to Competitiveness and Scalability. *Proc IEEE*. 2008 Nov;96(11):1786–98.
- [23] Heidel DF, Marshall PW, Pellish JA, Rodbell KP, LaBel KA, Schwank JR, et al. Single-Event Upsets and Multiple-Bit Upsets on a 45 nm SOI SRAM. *IEEE Trans Nucl Sci*. 2009;56(6):3499–504.
- [24] Burr GW, Kurdi BN, Scott JC, Lam CH, Gopalakrishnan K, Shenoy RS. Overview of Candidate Device Technologies for Storage-Class Memory. *IBM J Res Dev*. IBM. 2008;52(4.5):449–64.
- [25] International Technology Roadmap for Semiconductors, “Process Integration, Devices, and Structures, 2011 Edition.” available on http://www.maltiel-consulting.com/ITRS_2011-Process-Integration-Devices-Structures.pdf
- [26] Ionescu AM, Pott V, Fritschi R, Banerjee K, Declercq MJ, Renaud P, et al. Modeling and Design of a Low-Voltage SOI Suspended-Gate MOSFET (SG-MOSFET) With a Metal-

- Over-Gate Architecture. In: Quality Electronic Design, 2002 Proceedings International Symposium on. 2002. p. 496–501.
- [27] Parsa R, Lee WS, Shavezipur M, Provine J, Maboudian R, Mitra S, et al. Laterally Actuated Platinum-Coated Polysilicon NEM Relays. *J Microelectromech. Syst.* 2013;22(3):768–78.
- [28] Vaddi R, Pott V, Chua GL, Lin JTM, Kim TT. Design and Scalability of a Memory Array Utilizing Anchor-Free Nanoelectromechanical Nonvolatile Memory Device. *Electron Device Lett IEEE.* 2012 Sep;33(9):1315–7.
- [29] Cosemans S. Data Storage Cell and Memory Arrangement. Google Patents; 2015.
- [30] Kim M-S, Jang WW, Lee J-M, Kim S-M, Yun E-J, Cho K-H, et al. NEMS Switch With 30 nm Thick Beam and 20 nm High Air Gap for High Density Non-Volatile Memory Applications. In: *Semiconductor Device Research Symposium, 2007 International.* 2007. p. 1–2.
- [31] Van Kampen RP. Four-Terminal Multiple-Time Programmable Memory Bitcell and Array Architecture [Internet]. Google Patents; 2009. Available from: <http://www.google.com/patents/US20090273962>
- [32] Chen C, Parsa R, Patil N, Chong S, Akarvardar K, Provine J, et al. Efficient FPGAs Using Nanoelectromechanical Relays. In: *Proceedings of the 18th Annual ACM/SIGDA International Symposium on Field Programmable Gate Arrays [Internet].* New York, NY, USA: ACM; 2010. p. 273–82. (FPGA '10). Available from: <http://doi.acm.org/10.1145/1723112.1723158>
- [33] Advance Micro Devices. Software Optimization Guide for AMD Family 15h Processors. Number 47414. In 2014.
- [34] Intel Strongarm Processor. www.intel.com/design/pca/applicationsprocessors/1110_brf.htm. In.
- [35] Sh-3 RISC Processor family. http://www.hitachi-eu.com/hel/ecg/products/micro/32bit/sh_3.html. In.
- [36] “Predictive Technology Model,” Available on <http://ptm.asu.edu/>.
- [37] “The Electric VLSI Design System,” Available on <http://www.staticfreesoft.com>.
- [38] Carlson TE, Heirman W, Eeckhout L. Sniper: Exploring the Level of Abstraction for Scalable and Accurate Parallel Multi-Core Simulation. In: *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis.* 2011. p. 52.
- [39] Intel Corporation, IntelR 64 and IA-32 Architectures Optimization Reference Manual. Number 248966–026. In.